



**HAL**  
open science

# Impact de l'organisation tridimensionnelle du génome sur la réplication de l'ADN chez la levure

Jade Pellet

► **To cite this version:**

Jade Pellet. Impact de l'organisation tridimensionnelle du génome sur la réplication de l'ADN chez la levure. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université Paris Sciences & Lettres, 2022. Français. NNT: . tel-04310679v1

**HAL Id: tel-04310679**

**<https://theses.hal.science/tel-04310679v1>**

Submitted on 27 Nov 2023 (v1), last revised 6 Sep 2024 (v2)

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

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

**THÈSE DE DOCTORAT**

**DE L'UNIVERSITÉ PSL**

Préparée à l'Ecole Normale Supérieure

**Impact de l'organisation tridimensionnelle du génome  
sur la réplication de l'ADN chez la levure**

Soutenue par

**Jade PELLET**

Le 08 décembre 2022

Ecole doctorale n° 515

**Complexité du vivant**

Spécialité

**Génétique et Génomique**

**Composition du jury :**

Chun-Long CHEN Directeur de recherche, Institut Curie	<i>Président</i>
Jean-Charles CADORET Maître de conférences, Institut Jacques Monod	<i>Rapporteur</i>
Armelle LENGRONNE Directrice de recherche, Institut de Génétique Humaine	<i>Rapporteur</i>
Sarah LAMBERT Directrice de recherche, Institut Curie	<i>Examineur</i>
Nataliya PETRYK Chargée de recherche, Institut Gustave Roussy	<i>Examineur</i>
Olivier, HYRIEN Directeur de recherche, Ecole Normale Supérieure	<i>Directeur de thèse</i>



## Remerciements

Tout d'abord, je remercie mon directeur de thèse Olivier H, pour m'avoir accueillie dans son laboratoire et accompagnée au cours de ces trois années. Au sein du laboratoire je remercie pour leur accompagnement Laurent, Florence et Benoît. Florence, qui a répondu à mes nombreuses questions et a toujours été là pour m'aider à mener à bien mes expériences. Laurent, pour toute l'aide qu'il m'a apporté dans l'analyse des résultats et dans la prise en main de R. Benoît, qui m'a lancé avec la SY14 pour m'apprendre à travailler avec les levures. Finalement ce qui devait prendre quelque mois a pris plus d'ampleur jusqu'à devenir mon projet de thèse. Merci pour m'avoir poussée à relever mon niveau d'exigence, pour toutes les discussions sur les résultats et les analyses, je n'aurais définitivement pas pu réussir sans ton aide. Je remercie aussi Etienne, grâce à tes codes j'ai réussi à me familiariser avec R. Je remercie Bertrand, Diletta et Victoria pour leur discussion sur la science mais pas seulement, leur aide et leur soutien.

Au sein de la section génomique fonctionnelle je remercie Isabelle, Morgane, Olivier B, Lucía, Nadia, Marine qui ont toujours été là pour répondre à mes questions et m'aider dans la conduite des expériences. Je remercie aussi Martine qui a fait en sorte que je dispose toujours de matériel propre quand j'en avais besoin.

Je remercie Alexandra, Manon et Maël de la cellule *bioclust* pour leur bienveillance et leur aide au cours de l'utilisation du cluster, ainsi que, Pierre, Catherine et Bilel du service informatique qui m'ont été d'une grande aide.

Je remercie les personnes qui m'ont conduite à faire une thèse, Christophe, Gwenola et Céline mes maîtres de stages, qui m'ont poussé à candidater pour ce sujet de thèse et ont accompagné mes premiers pas dans le monde de la recherche.

Je remercie mes amis, Julie, Chloé, Gaël, Olivier E, Adrien, Robin, Loïc, François et Thibault pour leur soutien, ainsi que les moments passés ensemble qui m'ont permis de me détendre et de relâcher la pression accumulée. Julie qui a toujours été là pour me remonter le moral quand j'en avais besoin. Thibault pour avoir consacré du temps à la relecture du manuscrit pour corriger les fautes de français. François pour avoir essayé de comprendre mon sujet. Je remercie spécialement ceux qui m'ont écouté répéter les séminaires de section, pour m'entraîner à parler en public. Sans vous, ces trois années auraient été beaucoup moins amusantes.

Je remercie aussi mon chat Wicket, qui m'a observé rédiger ma thèse, m'a écouté répéter chacune de mes présentations et a participé activement à mes réflexions sur la présentation des résultats. Pour finir, je remercie Samuel, pour m'avoir soutenue et encouragée non pas uniquement au cours de ces trois années, mais durant les sept dernières. Sans toi à mes côtés, je n'aurais certainement pas entrepris une aventure aussi compliquée.



# Table des matières

Remerciements .....	2
Liste des abréviations .....	8
Table des Tableaux .....	10
Table des figures.....	11
I. Introduction.....	15
I.1 La réplication de l'ADN chez la levure <i>Saccharomyces cerevisiae</i> .....	15
I.1.1 Les origines de réplication.....	15
I.1.2 Chargement des complexes de pré-réplication (pré-RC), phase G1 du cycle cellulaire.....	16
I.1.3 Activation des pré-RC .....	16
I.1.4 Élongation.....	19
I.1.5 Terminaison.....	21
I.2 Organisation temporelle de la réplication .....	22
I.2.1 Moment de réplication ( <i>timing</i> ).....	22
I.2.2 Efficacité des origines.....	23
I.2.3 Les facteurs limitants.....	23
I.3 Organisation spatiale de la réplication.....	25
I.3.1 L'organisation nucléaire orchestre la réplication de l'ADN.....	25
I.3.2 Régulation du moment de réplication par des protéines spécifiques .....	28
I.3.2.1 Réplication tardive des télomères.....	28
I.3.2.2 Réplication précoce des centromères.....	29
I.3.2.3 Régulation du moment de réplication par des protéines spécifiques ni télomériques ni centromériques .....	29
I.3.2.3.1 Fkh1 Fkh2 .....	29
I.3.2.3.2 Rpd3, Sir2, Sir3 et acétylation des histones .....	30
I.3.2.3.3 Relation entre efficacité des origines et abondance des MCM .....	32
I.3.3 Interdépendance entre organisation nucléaire et activation des origines par des protéines .....	32
I.4 Déterminer l'implication de l'organisation du génome dans différents processus cellulaires : fusion des chromosomes chez la levure. ....	35
II. Présentation du projet de thèse .....	37
III. Méthodes expérimentales .....	39
III.1 Transformation de la BY4742 et de la SY14 pour l'incorporation de BrdU dans l'ADN .....	39
III.1.1 Culture de levures .....	40
III.1.2 Amplification du locus URA3 .....	40
III.1.3 Modification du plasmide pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> (Theulot et al, 2022).....	41

III.1.4	Transformation des levures.....	42
III.1.5	Vérification de l'intégration des plasmides et de l'insertion du gène URA3.....	43
III.1.6	Mesure du nombre de copies de plasmide intégré par qPCR.....	43
III.1.7	Mesure de l'expression des gènes d'incorporation du BrdU par RT-qPCR.....	44
III.1.8	Test de croissance .....	45
III.2	Analyse de la réplication de la BY4742 et de la SY14 par séquençage Nanopore .....	46
III.2.1	Extraction d'ADN .....	46
III.2.2	Séquençage Nanopore .....	46
III.2.3	Analyse de la réplication par NFS (Theulot et al, 2022) .....	46
III.2.3.1	Marquage des fourches de réplication.....	46
III.2.3.2	Profil de directionnalité des fourches de réplication (RFD) .....	46
III.2.3.3	Analyse des évènements individuels d'initiation et de terminaison.....	48
III.2.3.4	Analyse de la vitesse des fourches de réplication .....	51
III.2.4	Analyse du moment de réplication ou <i>timing</i> par séquençage Nanopore : nanotiming 52	
III.2.4.1	Marquage d'un cycle cellulaire.....	52
III.2.4.2	Obtention des profils de Nanotiming.....	52
III.2.5	Comparer des profils de réplication provenant de deux génomes différents .....	53
III.2.5.1	Alignement des génomes .....	53
III.2.5.2	Conversion des informations.....	53
III.2.5.2.1	Fourches .....	54
III.2.5.2.2	Initiation et terminaison.....	55
III.2.5.2.3	Nanotiming.....	55
III.2.5.3	Annotations des génomes .....	55
IV.	Méthodes d'analyse .....	57
IV.1	Détection des fourches de réplication .....	57
IV.1.1	Fork-seq.....	58
IV.1.2	NFS Nanoforkspeed (Theulot et al; 2022) .....	59
IV.1.3	Forkfit .....	61
IV.1.4	Fork-ML.....	61
IV.1.5	Comparaison de Fork-ML et de NFS.....	64
IV.2	Analyse des profils de réplication .....	66
IV.2.1	RFD.....	66
IV.2.1.1	Test de Student.....	67
IV.2.1.2	Test du Chi <sup>2</sup> .....	69
IV.2.1.3	Comparaison du test de Student et du test de Chi <sup>2</sup> .....	73
IV.2.1.4	Analyse des profils de réplication à l'aide des tests de Chi <sup>2</sup> .....	75

IV.2.2	Timing .....	78
IV.2.3	Initiation Probability (IP).....	80
IV.2.4	Termination Probability (TP) .....	80
V.	Résultats .....	81
V.1	Incorporation de BrdU chez la BY4742 et la SY14.....	81
V.2	Analyse du <i>timing</i> de réplication par séquençage Nanopore .....	87
V.2.1	Mise au point de la méthode Nanotiming par Bertrand Theulot.....	87
V.2.2	Adaptation de la méthode dans la SY14 et la BY4742 .....	88
V.3	Analyses des profils de réplication .....	90
V.3.1	Comparaison des différentes analyses.....	90
V.3.2	La majorité des changements sont centromériques ou télomériques .....	95
V.3.2.1	Modifications des origines à proximité des centromères.....	97
V.3.2.2	Modification des télomères .....	101
V.3.3	Régulation des origines à une distance éloignée ou moyenne.....	109
V.4	Analyse de la vitesse de progression des fourches de réplication.....	114
VI.	Discussion .....	117
VI.1	Amélioration de l'analyse des profils de réplication .....	117
VI.1.1	Amélioration de la détection des fourches de réplication .....	117
VI.1.2	Amélioration des analyses statistiques .....	118
VI.1.2.1	Test de Student pour comparer les RFD.....	118
VI.1.2.2	Test de Chi <sup>2</sup> pour comparer les RFD .....	119
VI.1.2.2.1	Synthèse des comparaisons .....	119
VI.1.2.2.2	Exclusion des régions montrant des différences significatives entre les réplicats 123	
VI.1.2.2.3	Analyses visuelles.....	125
VI.1.2.3	<i>Timing</i> de réplication.....	127
VI.1.2.4	Conclusion sur les analyses statistiques .....	129
VI.1.3	Analyses complémentaires.....	129
VI.1.3.1	Études de la modification des zones de terminaison dans la SY14 et la BY4742	129
VI.1.3.2	Analyse de la vitesse de progression des fourches dans la SY14 et la BY4742 ...	132
VI.1.3.3	Analyser des modifications de vitesse localement.....	132
VI.1.3.4	Réaliser une carte des vitesses dans la SY14 et la BY4742.....	132
VI.2	La réplication de l'ADN chez la levure est faiblement affectée par l'organisation tridimensionnelle .....	132
VI.2.1	Les origines à proximité des centromères.....	133
VI.2.2	Les origines à proximité des télomères.....	133
VI.2.3	Diminution de l'activité des origines éloignées des centromères et des télomères ..	134

VI.2.3.1	Candidats potentiels pour l'inhibition des origines dans la SY14.....	135
VI.2.3.1.1	Rif1 .....	135
VI.2.3.1.2	Les histone désacétylases Sir2 et Sir3.....	136
VI.2.3.2	Modification des origines et organisation nucléaire.....	136
VI.2.3.2.1	Fkh1/2 .....	136
VI.2.3.2.2	Les TADs .....	137
VI.2.3.2.3	Organisation des origines au sein du noyau .....	138
VI.2.4	Conclusion et perspectives.....	139
VII.	Table supplémentaire.....	140
VIII.	Figures supplémentaires .....	151
IX.	Annexe : Genome-wide mapping of individual replication fork velocities using nanopore sequencing .....	208
X.	Bibliographie.....	209

## Liste des abréviations

ACS : *ARS Consensus Sequence*

ADN : acide désoxyribonucléique

ADNr : ADN ribosomal

ADP : Adénosine-DiPhosphate

APC/C : *Anaphase Promoting Complex/Cyclosome*

ARN : acide ribonucléique

ARNm : ARN messenger

ARN-t : ARN de transfert

ARS : *Autonomously replicating sequence*

ATP : Adénosine-TriPhosphate

BET : Bromure d'ETHidium

BRCT : BRCA1 C-terminal

BrdU : BromodésoxyUridine

BrdUMP : BromodésoxyUridine MonoPhosphate

BrdUTP : BromodésoxyUridine TriPhosphate

CDK : *cyclin-dependent kinases*

Cdt1 : *CDC10-dependent transcript 1*

CMG : CDC45-MCM-GINS

$C_T$  : *threshold cycle, cycle seuil*

DDK : *Dbf4-dependent kinase*

DMSO : diméthylsulfoxyde

dNTPs : désoxyriboNucléotides TriPhosphates

dTMP : désoxythymidine MonoPhosphate

dTTP : désoxythymidine TriPhosphate

Fkh : *Forkhead*

GINS : Go-Ichi-Ni-San

hENT1 : *human equilibrative nucleoside transporter 1*

Hi-C : *High-throughput Conformation Capture*

hsvTK : thymidine kinase du virus *Herpes simplex*

IP : *Initiation Probability*

LB : *lysogeny broth*

MCM2-7 : *MiniChromosome Maintenance 1 à 7*

MCM DH : complexe MCM double hexamère

MMS : Methyl MethaneSulfonate

NFS : NanoForkSpeed

ORC : *Origin Recognition Complex*

pb : paire de bases

PCNA : *Proliferation Cell Nuclear Antigen*

pré-RC : complexes de préréplication

Pol : polymérase

PP1 : protéine phosphatase 1

qPCR : PCR quantitative

RFC : facteur de réplication C (RFC)

RFD : *replication fork directionality*

Rif1 : *Rap1 interacting factor 1*

RPA : protéine de réplication A

RT-qPCR : Reverse transcription suivie par une PCR quantitative

TAD : Topologically Associated domains /domaines associées topologiquement

TP : *Termination Probability*

u.DO : unité de DO<sub>600nm</sub>

YPD : *Yeast Peptone Extract*

## Table des Tableaux

Tableau III-1 : Amorces pour amplifier le locus URA3 et vérifier l'intégration du gène dans le génome de la levure <i>S. cerevisiae</i> . .....	40
Tableau III-2 : Amorce pour la vérification de l'intégration des plasmides pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> . .....	42
Tableau III-3 : Amorces pour la qPCR et RT-qPCR. ....	44
Tableau V-1 : Modifications des origines à proximité des centromères. ....	97
Tableau V-2 : Tableau récapitulatif du nombre d'origines activées et non modifiées par Ctf19 dans la catégorie « proches » à proximité d'un centromère. ....	101
Tableau V-3 : Modification des origines situées à des télomères. ....	101
Tableau V-4 : Récapitulatif du nombre d'origines modifiées et non modifiées dans les classes de distances « moyennes » et « éloignées » des centromères et des télomères. ....	109
Tableau VI-1 : Evolution du nombre de comparaisons significatives pour les origines détectées en RFD en fonction de leur détection en IP et en <i>timing</i> . ....	121
Tableau VI-2 : Nombre d'origines détectées comme modifiées en <i>timing</i> de réplication et en RFD et/ou IP en fonction du % de tests significatifs sur le <i>timing</i> retenu comme seuil. ....	127
Tableau supplémentaire VII-1 : Liste des échantillons biologiques utilisés. ....	140
Tableau supplémentaire VII-2 : Table de conversion des données. ....	142
Tableau supplémentaire VII-3 : Liste des origines significativement modifiées dans la SY14. ....	144
Tableau supplémentaire VII-4 : Régulation des origines modifiées proches des centromères par Ctf19. ....	146
Tableau supplémentaire VII-5 : Régulation des origines augmentées à proximité des télomères. ....	147
Tableau supplémentaire VII-6 : Modifications des origines à une distance moyenne et éloignée des centromères et des télomères par Rpd3 et Fkh1/2. ....	148
Tableau supplémentaire VII-7 : Modifications des origines à une distance moyenne et éloignée des centromères et des télomères par Rif1 et Rif1-RBM. ....	149
Tableau supplémentaire VII-8 : Liste des souches utilisées dans ce manuscrit. ....	150

## Table des figures

Figure I-1 : Représentation schématique de l'activation des MCM en phase S.....	19
Figure I-2 : Représentation schématique d'une fourche de réplication. ....	20
Figure I-3 : Élimination des amorces d'ARN sur le brin discontinu. ....	21
Figure I-4 : Modèle de l'impact de l'organisation du noyau sur la réplication. ....	35
Figure II-1 : Schéma du séquençage nanopore. ....	37
Figure III-1 : Stratégie pour reconstituer le locus URA3 dans la BY4742 et la SY14.....	39
Figure III-2 : Exemple de profil RFD. ....	48
Figure III-3 : Exemple de profils IP et TP.....	51
Figure III-4 : Exemple de profil de <i>timing</i> . ....	53
Figure III-5 : Stratégie de coupure des fourches. ....	55
Figure IV-1 : Incorporation de BrdU lors du déplacement d'une fourche de réplication. ....	57
Figure IV-2 : Exemple de lectures nanopore <i>basecaller</i> par RepNano ou <i>Megalodon-based BrdU basecaller</i> .....	59
Figure IV-3 : Exemple de segmentation linéaire d'une fourche droite. ....	60
Figure IV-4 : Exemple de différentes stratégies de lissage du signal. ....	62
Figure IV-5 : Exemple d'annotation des lectures pour l'entraînement de fork-ML.....	64
Figure IV-6 : Fonction de distribution empirique de la distance entre les évènements d'initiation détectés par NFS ou Fork-ML. ....	65
Figure IV-7 : schéma représentant la stratégie de comptage des fourches de réplication. ....	67
Figure IV-8 : Test de Student avec et sans correction pour le chromosome IX. ....	69
Figure IV-9 : Test de Chi <sup>2</sup> sans correction du FDR pour le chromosome IX.....	70
Figure IV-10 : Test de Chi <sup>2</sup> après correction du FDR pour le chromosome IX.....	70
Figure IV-11 : Test de Student sans correction du FDR et test de Chi <sup>2</sup> après correction du FDR pour le chromosome IX.....	75
Figure IV-12 : Représentation des 7 régions présentant des différences significatives entre souches mais aussi entre réplcats. ....	77
Figure IV-13 : Exemple d'annotation de terminaison. ....	78
Figure IV-14 : Représentation du nombre de tests de Student significatifs sur le <i>timing</i> de réplication des origines actives. ....	80
Figure V-1 : Incorporation de BrdU chez la levure. ....	81
Figure V-2 : Croissance des différentes souches BT1, BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3.....	83

Figure V-3 : Nombre de copies du plasmide intégré dans les souches BT1, BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3. ....	84
Figure V-4 : Expression des gènes <i>hsvTK</i> et <i>hENT1</i> pour la BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3.....	85
Figure V-5 : Signal moyen des souches BT1 pour un <i>pulse</i> de 2min à 100µM et BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3 pour un <i>pulse</i> de 2min à 100µM ou 1000µM. ....	86
Figure V-6 : Profil de Nanotiming et Sort-seq dans la souche BT1 pour le chromosome VII.....	88
Figure V-7 : Profils Nanotiming à 10, 25, 50 et 100µM de BrdU et profil RFD pour le chromosome VII dans la SY14 et la BY4742.....	88
Figure V-8 : Profils de réplication du chromosome V dans la BY4742 et la SY14.....	91
Figure V-9 : Modification significative du TEL16L.....	92
Figure V-10 : Résultats des différentes analyses statistiques. ....	95
Figure V-11 : Distance entre les origines modifiées, les centromères et les télomères. ....	96
Figure V-12 : Exemple de modifications au niveau des centromères. ....	99
Figure V-13 : Exemple de modifications de <i>timing</i> observées dans la souche <i>Ctf19Δ</i> (Natsume et al., 2013).....	100
Figure V-14 : Modifications des origines télomériques.....	103
Figure V-15 : Modifications des origines télomériques par Rif1, $\gamma$ KU70 et Rif1-RBM. ....	105
Figure V-16 : Modification des origines à proximité des télomères et modification de l'expression des gènes. ....	106
Figure V-17 : Modifications des origines à proximité des télomères par Fkh1/2. ....	107
Figure V-18 : Représentation des origines comprise entre 25kb et 35kb de l'extrémité d'un chromosome dans la BY4742. ....	109
Figure V-19 : Modification des origines à des distances « moyennes » et « éloignées » des centromères et des télomères par Fkh1/2. ....	111
Figure V-20 : Modification des origines à une distance « moyennes » et « éloignées » des centromères et des télomères par Rpd3. ....	112
Figure V-21 : Modification des origines à des distances « moyennes » et « éloignées » des centromères et des télomères en fonction du signal MCM-ChIP-seq (Dukaj and Rhind 2021). ....	114
Figure V-22 : Vitesse de progression des fourches dans le génome de la BY4742 et le génome de la SY14.....	115
Figure VI-1 : Comparaison des résultats du test de Student pour différents seuil $\alpha$ . ....	119
Figure VI-2 : Exemple d'origines détectées par différents nombres de comparaison significatives. .	123
Figure VI-3 : Modifications significatives entre les réplicats.....	125
Figure VI-4 : Exemple de régions où l'interprétation du RFD est difficile. ....	126

Figure VI-5 : Origines significativement modifiées en <i>timing</i> avec un % de modification significative compris entre 25 % et 50%.....	129
Figure VI-6 : Analyses statistiques des TP par DEseq2. ....	130
Figure VI-7 : <i>Timing</i> de réplication dans la BY4742 pour les origines modifiées dans la SY14 en fonction de la proximité des origines avec les télomères et les centromères. ....	135
Figure VI-8 : TADs chez la levures et origines modifiées dans la SY14.....	137
Figure supplémentaire VIII-1 : Comparaison des profils de réplication entre la BY4742 et la SY14, chromosomes entiers.....	151
Figure supplémentaire VIII-2 : Origines détectées comme significativement modifiées par le RFD, l'IP et le <i>timing</i> . ....	160
Figure supplémentaire VIII-3 : Origines détectées comme significativement modifiées uniquement sur le RFD dont l'origine voisine possède un <i>timing</i> de réplication modifié significativement. ....	162
Figure supplémentaire VIII-4 : Origine détectée comme significativement modifiée uniquement sur le RFD, avec une modification visible de l'IP dont l'origine voisine possède un <i>timing</i> de réplication modifié significativement.....	163
Figure supplémentaire VIII-5 : Origines détectées comme significativement modifiées par l'IP et le RFD possédant une origines voisine modifié significativement en <i>timing</i> de réplication.....	164
Figure supplémentaire VIII-6 : Origines détectées comme significativement modifiées par le <i>timing</i> de réplication impactant une zone de terminaison sur le RFD. ....	165
Figure supplémentaire VIII-7 : Origines détectées comme significativement modifiées par l'IP et le RFD dans une région tardive.....	166
Figure supplémentaire VIII-8 : Origines détectées comme significativement modifiées uniquement sur le RFD étant dans une région tardive et possédant une modification visible mais non significative de l'IP.....	167
Figure supplémentaire VIII-9 : Origines détectées comme significativement modifiées uniquement sur le RFD étant dans une région tardive et ne possédant pas de modification visible de l'IP. ....	170
Figure supplémentaire VIII-10 : Origines détectées comme significativement modifiées uniquement sur le <i>timing</i> de réplication mais possédant une modification visible sur l'IP et/ou RFD.....	174
Figure supplémentaire VIII-11 : Origines détectées comme significativement modifiées par l'IP et le RFD et possédant une modification visible en <i>timing</i> mais non significative. ....	175
Figure supplémentaire VIII-12 : Origines détectées comme significativement modifiées par le <i>timing</i> et le RFD possédant une modification visible de l'IP mais non significative.....	176
Figure supplémentaire VIII-13 : Origines détectées comme significativement modifiées uniquement sur l'IP mais possédant une modification visible sur le RFD. ....	179

Figure supplémentaire VIII-14 : Origines détectées comme significativement modifiées par le <i>timing</i> et le RFD ne possédant pas de modification visible de l'IP. ....	180
Figure supplémentaire VIII-15 : Origines détectées comme significativement modifiées uniquement sur le RFD.....	181
Figure supplémentaire VIII-16 : Origines détectées comme significativement modifiées uniquement sur le <i>timing</i> de réplication. ....	183
Figure supplémentaire VIII-17 : Comparaison des profils de réplication entre la BY4742 et la SY14 pour les régions péricentromériques.....	184
Figure supplémentaire VIII-18 : Modifications du <i>timing</i> de réplication en absence de Ctf19. ....	189
Figure supplémentaire VIII-19 : Comparaison des profils de réplication entre la BY4742 et la SY14 pour les fusions chromosomiques. ....	191
Figure supplémentaire VIII-20 : Fixation de Rif1 à proximité des origines « proches » des télomères. ....	196
Figure supplémentaire VIII-21 : Profil de réplication de la BY4742. ....	201

# I. Introduction

## I.1 La réplication de l'ADN chez la levure *Saccharomyces cerevisiae*

Pour réaliser la duplication de l'ADN, les cellules ont besoin de deux éléments : une hélicase qui ouvre l'ADN et une polymérase qui synthétise la nouvelle séquence d'ADN. Ce processus doit être parfaitement régulé afin de dupliquer une fois et une seule l'intégralité du génome durant la phase de synthèse de l'ADN du cycle cellulaire ou phase S.

### I.1.1 Les origines de réplication

La réplication de l'ADN chez *S.cerevisiae* démarre en quelques centaines de sites nommés origines de réplication (pour revue : Costa and Diffley, 2022). Les origines de réplication ont été identifiées pour la première fois par un test génétique comme des séquences capables de maintenir un plasmide après plusieurs cycles cellulaires. Le maintien d'un plasmide repose sur sa capacité à se dupliquer afin qu'une copie puisse être transmise à chaque cellule fille lors de la division cellulaire (Stinchcomb et al., 1979). Ces séquences ont été nommées ARS pour *Autonomously Replicating Sequence*. Les ARS sont généralement contenues dans des régions de 100-200 paires de bases (pb). Elles se composent principalement de deux éléments, A et B. L'élément A contient l'ACS (*ARS Consensus Sequence*) : (A/T)TTTAT(A/G)TTT(A/T), qui fournit un site de liaison pour un complexe hétérohexamérique nommé ORC (*Origin Recognition Complex*; voir Chargement des complexes de pré-réplication (pré-RC), phase G1 du cycle cellulaire), essentiel au fonctionnement des origines chez tous les eucaryotes (Bell et Stillman, 1992). La mutagenèse systématique par substitution d'oligonucléotides de l'ARS1 a permis de montrer qu'elle se compose d'un élément A essentiel, qui recouvre l'ACS, et de trois sous-éléments B1, B2 et B3, dont la mutation individuelle ne réduit que partiellement l'activité de l'origine, dans l'ordre B1>B2>B3, dont l'effet est additif, et dont la triple mutation abolit le fonctionnement de l'origine (Marahrens and Stillman, 1992). L'élément B1, adjacent à l'ACS, contribue à l'affinité de ORC et peut donc être considéré comme une extension de son site de liaison (Rao and Stillman, 1995; Wilmes and Bell, 2002). L'élément B2 de ARS1 contient une copie presque parfaite de l'ACS en orientation inverse à celle de l'élément A. Des études fonctionnelles récentes suggèrent que le rôle de B2 est de fournir un site de liaison de ORC en orientation opposée à celui fourni par A/B1 (Coster and Diffley, 2017). L'élément B3 contient un site de fixation pour le facteur de transcription ABF1 mais peut être remplacé par un site de fixation pour un autre facteur de transcription, son rôle est de positionner les nucléosomes adjacents à l'origine (Marahrens and Stillman, 1992; pour revue : Newlon and Theis, 1993). Les éléments B possèdent une distribution asymétrique des adénines et thymine sur les deux brins de l'ADN qui empêche la formation de nucléosomes (Eaton et al., 2010) et facilite l'ouverture de la double hélice (Umek and Kowalski, 1988).

### I.1.2 Chargement des complexes de pré-réplication (pré-RC), phase G1 du cycle cellulaire

En 1992 Bell et Stillman ont isolé d'extraits nucléaires de levures un complexe hétérohexamérique capable de se lier à l'ADN et de protéger les éléments A et B1 de l'ARS1 d'une dégradation par la DNase I. Ce complexe est nommé *Origin Recognition Complex* (ORC). Des mutations qui inactivent l'ACS réduisent la liaison de ORC à l'ADN, ce qui montre que cette protéine reconnaît spécifiquement l'origine de répllication fonctionnelle. Les gènes codant les six sous-unités Orc1 à 6 sont essentiels chez *S. cerevisiae*, où leurs mutations hypomorphes produisent des défauts d'initiation de la répllication. Ces gènes sont conservés chez tous les eucaryotes.

L'étude de la protection des ARS dans des lysats cellulaires après traitement par la DNase I (DNase I *footprinting*) dans des cellules asynchrones, a montré que l'élément A est totalement protégé, suggérant que ORC est lié à l'ADN tout au long du cycle cellulaire et donc que la liaison d'autres partenaires est nécessaire pour démarrer la répllication. L'étude des profils de coupure par la DNase I (*footprint*) dans des cellules synchronisées a montré que les origines existent sous forme de deux états chromatinien au cours du cycle cellulaire. Dans l'état postrépllicatif, le *footprint* ressemble à celui obtenu *in vitro* avec ORC et ABF1 seuls. Ceci indique que la liaison de ces seules protéines n'est pas suffisante pour démarrer la répllication. Dans l'état pré-répllicatif, on observe une région supplémentaire de protection adjacente à celle couverte par ORC seul. Cet état est observé de la fin de la mitose à la fin de la phase G1. Après entrée en phase S, le profil de protection redevient semblable au *footprint* postrépllicatif (Diffley et al., 1994). L'étude de la formation de l'état ou complexe pré-répllicatif (pré-RC) dans des mutants conditionnels (Cocker et al., 1996) puis à l'aide de protéines purifiées (Evrin et al., 2009; Remus et al., 2009) a montré que celle-ci nécessitait, outre ORC, les protéines Cdc6, Cdt1 (*CDC10-dependent transcript 1*) et MCM2-7 (*MiniChromosome Maintenance 1 à 7*). ORC coopère avec Cdc6 et Cdt1 pour charger sur l'ADN le complexe MCM sous la forme d'un double hexamère (MCM DH) tête-contre-tête, qui entoure l'ADN double brin et est capable de glisser dessus (Evrin et al., 2009; Remus et al., 2009; pour revue : Labib, 2010). Ce chargement peut prendre place de la fin de la mitose à la fin de la phase G1. MCM2-7 est le moteur de l'hélicase répllicative, mais sous la forme de MCM DH il est inactif et le reste jusqu'à l'entrée en phase S.

### I.1.3 Activation des pré-RC

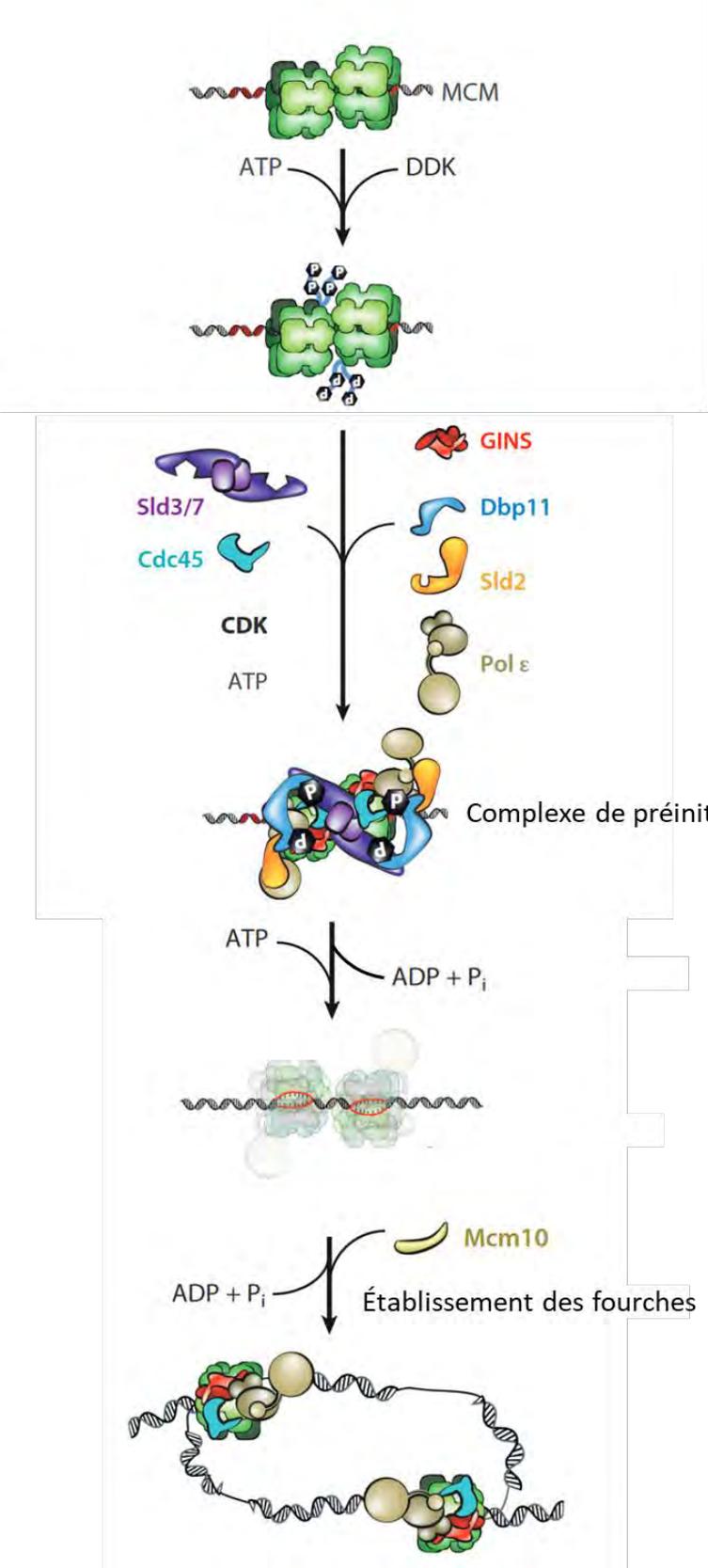
Les protéines impliquées dans l'activation des pré-RC ont été en majorité identifiées en étudiant l'impact de leur mutation sur le cycle cellulaire. Une mutation qui empêche l'entrée des cellules en phase S, impacte une protéine impliquée dans l'initiation de la répllication, tandis qu'une mutation bloquant les cellules en phase S impacte une protéine impliquée dans la progression de la répllication. Parmi les protéines identifiées, on retrouve deux *cyclin-dependent kinases* (CDK)

constituées de Cdc28 associées aux cyclines Clb5 ou Clb6. La suppression de Clb5 allonge fortement la durée de la phase S, tandis que la suppression de Clb6 seule a peu ou pas d'impact. Le double mutant *clb5 clb6* montre un retard de l'entrée en phase S (Schwob and Nasmyth, 1993), car celle-ci dépend maintenant de l'activation plus tardive des CDK normalement responsables de l'entrée en mitose (Cdc28 associé à Clb3 et Clb4). Une autre kinase qui impacte la progression de la phase S est Cdc7-Dbf4 ou *Dbf4-dependent kinase* (DDK). Cdc7 est présente tout au long du cycle cellulaire, néanmoins son activité est la plus forte lors de la transition entre la phase G1 et la phase S, et cette activité est abolie en absence de sa sous-unité régulatrice Dbf4 (Jackson et al., 1993). CDK et DDK jouent des rôles directs dans le recrutement de facteurs essentiels à l'activation du MCM DH. L'hélicase active se compose de MCM2-7, Cdc45 (Tercero et al., 2000), et du complexe GINS composé des protéines Sld5, Psf1, Psf2 et Psf3. GINS signifie Go-Ichi-Ni-San, 5, 1, 2, 3 en Japonais (Takayama et al., 2003). Ces trois facteurs forment l'hélicase active CMG (CDC45-MCM-GINS).

L'activation de l'hélicase commence par l'assemblage d'un complexe de pré-initiation (pré-IC) qui, outre les MCM, comprend la polymérase  $\epsilon$  (Pol  $\epsilon$ ), Sld3, Sld7, Sld2, Dpb11, GINS, et Mcm10. CDK et DDK sont inactives en G1 car leurs sous-unités régulatrices sont dégradées par l'*anaphase promoting complex/cyclosome* (APC/C). L'inactivation de l'APC/C à la fin de la phase G1 permet l'accumulation de Clb5 et 6 et de Dbf4 et l'activation de CDK et DDK. DDK phosphoryle MCM DH au niveau de Mcm4 et Mcm6, ce qui crée un site de liaison pour Sld3 qui, en complexe avec Sld7, recrute alors Cdc45 sur MCM DH (Sheu and Stillman, 2006). Plusieurs mutants de MCM, notamment phosphomimétiques, permettent de démarrer la réplication en l'absence de DDK, ce qui indique que MCM est le seul substrat critique de DDK (Hardy et al., 1997; Sheu and Stillman, 2010). CDK phosphoryle Sld2 et Sld3, ce qui crée des sites de liaison pour deux paires de motifs BRCT (BRCA1 C-terminal) répétés, situés sur Dpb11. Si l'on rend ces interactions constitutives à l'aide d'un mutant phosphomimétique de Sld2 et d'une fusion Dpb11-Sld3, CDK n'est plus nécessaire au démarrage de la réplication (Tanaka et al., 2007; Zegerman and Diffley, 2007). Ainsi, la phosphorylation de MCM par DDK recrute Sld3/7 et Cdc45 ; la phosphorylation de Sld2 et Sld3 par CDK recrute Sld2 sur le complexe MCM via Dpb11. Sld2 recrute à son tour Pol  $\epsilon$  et GINS (, Heller et al., 2011; Muramatsu et al., 2010).

La formation du pré-IC permet le relargage par MCM DH de l'ADP généré lors de son chargement et resté depuis associé. La liaison d'ATP frais déclenche la séparation du DH en hexamères individuels, l'association de Cdc45 et GINS pour former une paire de complexes CMG stables, et une première ouverture de l'ADN sur 6-7 pb, encore insuffisante pour évincer l'un des deux brins du canal formé par MCM. A l'étape finale, l'hydrolyse de l'ATP et l'association de Mcm10 convertissent ce CMG inactif en hélicase active, ayant évincé la matrice du brin discontinu et encerclant désormais la matrice

du brin continu sur laquelle CMG se déplace dans le sens 3'-5' (Figure I-1, pour revue : Costa and Diffley, 2022).



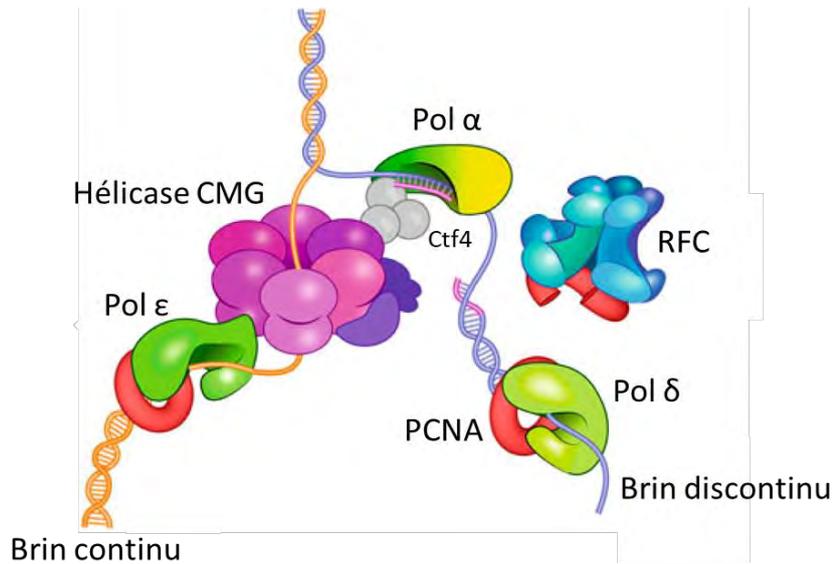
### Figure I-1 : Représentation schématique de l'activation des MCM en phase S.

Un double hexamère de MCM (MCM DH) chargé préalablement en phase G1 est phosphorylé par DDK, ce qui permet le recrutement de Sl3/Sl37 et Cdc45. La phosphorylation de Sl2 et Sl3 par CDK recrute Sl2 et Sl3/Sl37 sur le MCM DH via Dpb11. Sl2 recrute à son tour Pol  $\epsilon$  et GINS. Le remplacement de l'ADP par l'ATP permet une première ouverture de l'ADN, puis l'ajout de MCM10 et le relargage de l'ADP convertissent le MCM DH en deux hélicases CMG actives qui entament leur déplacement sur l'ADN simple brin. Adapté de Costa and Diffley (2022).

#### I.1.4 Élongation

La synthèse d'ADN nécessite la synthèse d'amorces puis l'addition de nucléotides dans le sens 5'→ 3' à partir de ces amorces. Étant donné la nature antiparallèle des deux brins de l'ADN, et le fait que les polymérases ne peuvent allonger une chaîne que dans le sens 5'-3', seul l'un des brins émanant de l'hélicase peut être recopié de manière continue dans le sens de progression de la fourche. L'autre brin est synthétisé de façon discontinue sous forme de petits fragments, les fragments d'Okazaki, qui sont allongés dans le sens opposé à la progression de l'hélicase avant d'être recollés entre eux. Trois enzymes coopèrent pour la synthèse d'ADN : (i) la polymérase  $\alpha$ -primase (Pol  $\alpha$ ) qui possède une activité primase et polymérase, ce qui lui permet de synthétiser quelques paires de bases d'ARN avec son activité primase avant de passer à l'activité polymérase pour allonger le fragment. Elle possède une processivité faible et n'est pas capable de corriger ses erreurs. (ii) La polymérase  $\epsilon$  (Pol  $\epsilon$ ), qui est responsable de la synthèse continue, dans le sens de progression de la fourche. (iii) La polymérase  $\delta$  (Pol  $\delta$ ), qui est responsable de la synthèse des fragments d'Okazaki. Toutefois, des données récentes suggèrent que la polymérase  $\delta$  intervient aussi dans la synthèse du brin continu au niveau de l'origine et lors de la terminaison (pour revue : Guillian et Yeeles, 2020).

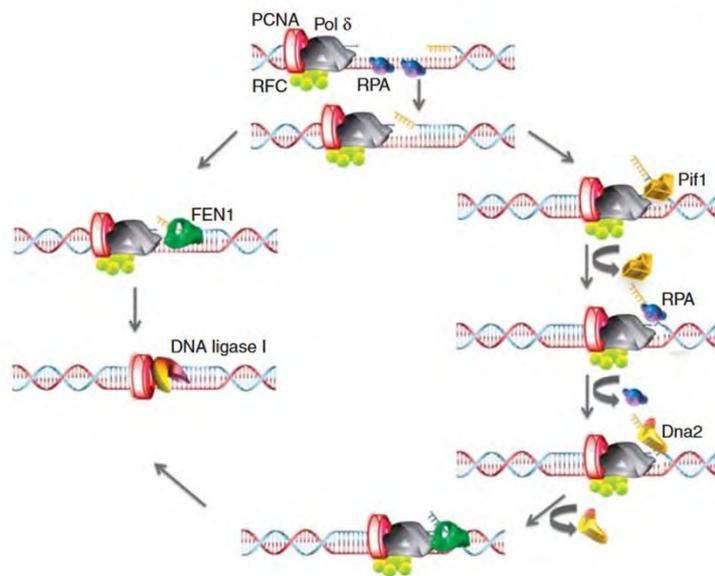
L'avancée de l'hélicase libre de l'ADN simple brin qui est recouvert par les protéines de réplication A (RPA), ce qui le protège des attaques par les nucléases et empêche la formation de structures secondaires. RPA sert aussi de plateforme d'assemblage de différents facteurs de signalisation ou de protection du génome. Le facteur de réplication C (RFC) a pour rôle de reconnaître les amorces synthétisées par la Pol  $\alpha$ -primase et d'y déposer un anneau trimérique : *Proliferative Cell Nuclear Antigen* (PCNA). PCNA offre une surface d'interaction pour l'arrière de Pol  $\delta$  et coulisse sur l'ADN double brin qui en émerge, à la manière d'une pince coulissante (*sliding clamp*), ce qui augmente fortement la processivité de Pol  $\delta$ . Pol  $\epsilon$  est intrinsèquement plus processive que Pol  $\delta$ , car elle possède un domaine qui encercle l'ADN double brin, mais elle peut aussi s'associer à PCNA pour atteindre sa processivité maximale (Figure I-2, pour revue : Balakrishnan et Bambara ; Guillian et Yeeles 2020)



**Figure I-2 : Représentation schématique d'une fourche de réplication.**

Pol  $\alpha$  synthétise des amorces d'ARN avant de les allonger. Elle est liée à l'hélicase par un trimer de Ctf4. Pol  $\epsilon$  synthétise le brin continu tandis que Pol  $\delta$  synthétise le brin discontinu par synthèse successive de fragments d'Okazaki. La processivité des deux Pol  $\epsilon$  et  $\delta$  sera augmentée par leur liaison à PCNA chargé par RFC. Adapté de Yao and O'Donnell, 2019.

Chaque fragment d'Okazaki comporte une amorce d'ARN synthétisée par Pol  $\alpha$ , qui est éliminée avant la ligation des fragments. Lorsque Pol  $\delta$  rencontre le fragment d'Okazaki précédent, elle déplace son extrémité 5', générant un rabat (*flap*) qui sera ensuite excisé par FEN1. L'intégralité de l'amorce sera retirée par plusieurs cycles successifs d'action de FEN1. Parfois, la Pol  $\delta$  déplace plusieurs nucléotides avant l'intervention de FEN1, ou encore, l'hélicase Pif1 allonge le rabat, ce qui recrute RPA. Dna2 coupe alors une partie de ce long rabat qui redevient suffisamment petit pour être éliminé par FEN1. Pour finir, les fragments d'Okazaki sont liés par la DNA ligase I (Figure I-3, pour revue : Balakrishnan and Bambara, 2013).



**Figure I-3 : Élimination des amorces d'ARN sur le brin discontinu.**

Lorsque la Pol  $\delta$  rencontre le fragment d'Okazaki précédent, elle déplace son extrémité 5' et forme un rabat (flap), indiqué en jaune. Ce rabat peut directement être clivé par FEN1, puis les deux extrémités 5'P et 3'OH voisines peuvent être liées par la DNA ligase I. Le rabat peut aussi être allongé par Pif1, recouvert par RPA qui recrute Dna2, raccourci par Dna2, et ensuite excisé par FEN1.

### I.1.5 Terminaison

Lorsque deux réplisomes se rencontrent, un événement de terminaison a lieu. Ce processus se déroule sans ralentissement ni pause des fourches de réplication (Dewar et al., 2015), en plusieurs étapes : (i) ouverture de l'ADN jusqu'à la rencontre des deux réplisomes. Lorsqu'il n'y a plus suffisamment de place entre les fourches pour la formation de supertours et l'action des topoisomérases, l'excès d'enlacements entre les brins parentaux va entraîner la rotation des fourches sur leur axe et la formation d'enlacements entre chromatides sœurs (précáténanes, convertis en caténanes après terminaison de la réplication). Les Topoisomérases II et III semblent être responsables de la résolution des caténanes aux événements de terminaison (pour revue : Dewar et Walter, 2017). (ii) La ligation des fragments provenant des deux réplisomes. Au point de rencontre des deux fourches, le brin discontinu d'une fourche est lié au brin continu de l'autre fourche après ligation des derniers fragments d'Okazaki. (iii) Le désassemblage du réplisome. C'est une étape cruciale pour éviter la re-réplication ou encore l'interférence avec d'autres processus ayant lieu sur l'ADN comme la transcription (pour revue : Dewar et Walter, 2017). L'étape-clé semble être l'ubiquitylation de MCM7 réalisée par SCFDia2, suivie d'un désassemblage de l'hélicase médié par Cdc48 (Maric et al., 2014). La régulation de l'ubiquitylation de MCM7 est cruciale car cette sous-unité ne doit pas être reconnue et

dissociée par Cdc48 lors de l'assemblage des pré-RC en phase G1, ni en phase S, tant que deux réplisomes ne se sont pas rencontrés. C'est pourquoi elle n'a lieu que lorsque MCM7 est accompagné de GINS et CDC45, seul CMG est donc sujet à dégradation (Maric et al., 2014). De plus, cette ubiquitination est inhibée tant que CMG reste associé à de l'ADN simple brin. Lorsque CMG rencontre l'extrémité 5' de la fourche opposée, il avance sur l'ADN double brin. La perte de contact avec l'ADN simple brin démasque alors le site de liaison de SCFDia2, ce qui déclenche l'extraction du réplisome (Deegan et al., 2020; Dewar et al., 2015; pour revue: Dewar and Walter, 2017).

## 1.2 Organisation temporelle de la réplication

### 1.2.1 Moment de réplication (*timing*)

McCarroll et Fangman (1988) ont analysé le moment de réplication de différentes séquences du génome en cultivant des levures dans un milieu contenant des isotopes lourds ( $^{14}\text{C}$  et  $^{15}\text{N}$ ) pendant plusieurs générations, ce qui permet de marquer l'intégralité de l'ADN. Les cultures sont ensuite arrêtées en phase G1, puis relâchées en phase S dans un milieu classique ( $^{12}\text{C}$  et  $^{14}\text{N}$ ) ; la réplication produit alors des ADN hybrides lourd/léger. La récupération des levures à différents moments après l'entrée en phase S permet de suivre la conversion de l'ADN lourd/lourd non répliqué en ADN lourd/léger répliqué. L'analyse de 12 fragments d'ADN a permis de montrer que chaque fragment possédait un moment ou *timing* de réplication différent. Par la suite, l'obtention à l'aide de puces à ADN de profils de réplication pour tous les chromosomes a permis de montrer que chaque chromosome présentait une alternance de régions répliquées précocement ou tardivement (Raghuraman et al., 2001).

L'étude de la réplication du chromosome VI par étalement d'ADN marqué au BrdU sur lame de verre (*DNA combing*) a permis d'analyser le profil de réplication de chromosomes VI individuels provenant de différentes cellules, à différents moments de la phase S, révélant ainsi la variabilité intercellulaire (Czajkowsky et al., 2008). Chaque cellule initie la réplication à des endroits différents, à différents moments de la phase S. Il n'existe donc pas d'ordre strict de déclenchement des origines. Les profils populationnels ne reflètent que la tendance moyenne de la population, où chaque cellule se réplique d'une façon différente des autres.

Afin de déterminer le moment de mise en place du *timing* de réplication, Raghuraman et al., (1997) ont excisé à différents moments du cycle cellulaire une cassette contenant l'ARS501, une origine tardive située à proximité d'un télomère. Lorsque la cassette est excisée en phase G1, l'ARS501 extrachromosomique se réplique tardivement, tandis que lorsque la cassette est excisée en phase G2/M, elle se réplique précocement. De plus, lorsque l'origine est excisée en phase G1, mais que les

cellules sont cultivées pendant plusieurs cycles cellulaires, l'origine se réplique précocement. Ces résultats montrent que le moment de répllication est déterminé en phase G1 de chaque cycle cellulaire.

### 1.2.2 Efficacité des origines

Malgré leur capacité à maintenir un plasmide, toutes les ARS de la levure ne sont pas actives dans leur contexte chromosomique. En effet, l'analyse de l'activité de différentes ARS en gel-2D a montré que les origines possédaient un signal de bulle de répllication plus ou moins fort, parfois indétectable, qui traduit la proportion de cellules dans lesquelles cette origine s'est activée (Friedman et al., 1997; Yamashita et al., 1997). La faible activité de certaines origines dans leur contexte chromosomique s'explique parfois par leur répllication passive par une fourche émise depuis une origine voisine. Par exemple, les origines de répllication qui entourent le locus HML sont inactives, mais lorsque l'activité de l'origine en amont est inhibée, ces origines peuvent initier la répllication car elles ne sont plus répliquées de manière passive (Vujcic et al., 1999). L'étude de l'activité des origines en gel 2D au cours de la phase S a montré que, le plus souvent, les origines les plus efficaces formaient une bulle de répllication de manière plus précoce que les origines les moins efficaces, ce qui confirme que les origines ne s'activent pas toutes au même moment de la phase S et que leur efficacité est corrélée à leur moment d'activation (Yamashita et al., 1997). Les fourches émises par les origines les plus précoces tendront à passiver plus fréquemment les origines voisines plus tardives. Néanmoins les origines tardives ne sont pas pour autant les moins efficaces, dès lors qu'elles sont suffisamment éloignées des autres origines pour ne pas subir de répllication passive (Friedman et al., 1997). Il convient donc de distinguer l'efficacité intrinsèque de l'origine, c'est-à-dire la capacité qu'a l'origine d'initier la répllication indépendamment de ses voisines, de l'efficacité observée de l'origine, qui dépend de l'activité des origines voisines. Comme il est difficile de déterminer l'efficacité intrinsèque d'une origine dans son contexte cellulaire, le terme d'efficacité est généralement utilisé pour parler de l'efficacité observée. Indépendamment du phénomène de passivation, les origines d'efficacité intrinsèque plus élevée se déclenchent en moyenne plus rapidement que les moins efficaces. Ainsi, la corrélation entre moment d'activation et efficacité des origines est plus marquée pour l'efficacité intrinsèque que pour l'efficacité observée (Yang et al., 2010).

### 1.2.3 Les facteurs limitants

Aparicio et al., (1999) ont analysé l'association de MCM, Cdc45 et des polymérases  $\alpha$  et  $\epsilon$  au cours du cycle cellulaire aux origines répliquées précocement et tardivement. Le complexe MCM s'associe en même temps aux origines précoces et tardives en phase G1. Cdc45 s'associe en phase G1 aux origines précoces et en phase S aux origines tardives. Les polymérases  $\alpha$  et  $\epsilon$  sont ensuite recrutées en début de phase S aux origines précoces. Lorsque le signal pour Cdc45 et les deux polymérases

diminue aux origines précoces, il augmente pour les origines tardives. Ces résultats montrent que Cdc45 n'est pas capable de se lier simultanément aux origines précoces et tardives et suggère un recyclage de ce facteur après terminaison de la réplication des régions précoces aux origines tardives. Un des éléments pouvant expliquer ce phénomène est que Cdc45 est un des facteurs limitants pour le déclenchement des origines et reste associé aux MCM et aux fourches pendant l'élongation. Le dosage des protéines impliquées dans le complexe de réplication a montré que Sld3, Sld7, Sld2, Cdc45, Dpb11 et Dbf4 étaient présents en quantité moindre que ORC et que le nombre total d'origines (Mantiero et al., 2011; Tanaka et al., 2011). Ceci suggère que toutes les origines ne peuvent pas avoir accès à ces facteurs simultanément. La surexpression de Sld3, Sld7 et Cdc45 entraîne l'activation précoce d'origines tardives et d'un plus grand nombre total d'origines, ce qui confirme que le niveau de ces protéines est normalement limitant pour le déclenchement des origines (Tanaka et al., 2011). La surexpression de Dbf4, Sld3, Sld2 et Dpb11 permet aussi aux origines tardives d'être activées plus précocement. Une combinaison de seulement trois de ces facteurs montre un effet plus faible sur le *timing* de réplication, et la surexpression de Dbf4 seul ne permet pas d'avancer le moment d'activation des origines (Mantiero et al., 2011). La surexpression de Cdc45, Dbf4, Sld2, Sld3, Sld7, et Dpb11, avance le moment de réplication et augmente l'efficacité de presque toutes les origines (Mantiero et al., 2011; McGuffee et al., 2013). L'existence de quelques origines réfractaires suggère la présence d'un autre mécanisme de régulation que la disponibilité en facteurs limitants. La suppression de l'histone désacétylase Rpd3, combinée à la surexpression des facteurs précédents, permet l'activation de toutes les origines de réplication. Ceci montre que le contexte chromatinien est important dans l'activation des origines (Mantiero et al., 2011, Rpd3, Sir2, Sir3 et acétylation des histones). De plus, la déplétion de Sld2 ou Sld3 affecte d'autant plus le *timing* de réplication des origines que les origines sont tardives (Lynch et al., 2019) suggérant que les origines précoces ont une plus grande capacité à fixer les facteurs limitants, ce qui leur permet de conserver leur moment de réplication lorsque la quantité de ces facteurs diminue. La surexpression de Dbf4, Sld3, Sld2 et Dpb11 induit un stress réplcatif provenant d'un manque de dNTPs (Mantiero et al., 2011), ce qui montre l'importance de la régulation temporelle de la réplication. Plusieurs éléments peuvent expliquer la nécessité pour la cellule de réguler la réplication de manière temporelle : (i) le *timing* de réplication organise la transcription et la compaction de l'ADN sous forme de chromatine. (ii) Le programme temporel de réplication permet de limiter les conflits entre réplication et transcription, c'est-à-dire éviter que les polymérases impliquées dans ces deux mécanismes ne se rencontrent. (iii) La régulation temporelle est simplement importante pour ne pas répliquer l'intégralité du génome en même temps. Cela permet d'éviter à la cellule de consommer rapidement les dNTPs et autres facteurs limitants, de gérer trop d'évènements de réparation de l'ADN en même temps, et de limiter le stress topologique induit par la formation de supertours lors de l'avancée des fourches de réplication. Les expériences générant une

reprogrammation du *timing* de réplication n'ont pas montré d'effet important sur le métabolisme cellulaire, suggérant que le programme de réplication a peu d'importance fondamentale à cet égard, et que son rôle essentiel est d'étaler la réplication du génome au cours du temps (pour revue : Rhind, 2022).

Le *timing* de réplication et l'efficacité des origines sont liés à la disponibilité en facteurs limitants. Les télomères, malgré leurs origines de réplication très efficaces, se répliquent en fin de phase S. Comme ils sont attachés à la périphérie nucléaire, ceci suggère un rôle possible du positionnement intranucléaire dans le programme temporel de réplication (McCarroll and Fangman 1998, Heun et al., 2001; Raghuraman et al., 2001).

### I.3 Organisation spatiale de la réplication

#### I.3.1 L'organisation nucléaire orchestre la réplication de l'ADN

En 1992, Ferguson et Fangman ont tenté de déterminer si le *timing* de réplication était intrinsèque à l'origine, c'est-à-dire si une origine précise était toujours activée au même moment de la phase S quel que soit l'ADN qui l'entoure, ou au contraire si l'activité de l'origine dépendait de son contexte chromosomique. Ils ont dans un premier temps analysé la réplication de l'ARS1, qui se réplique précocement. Lorsqu'elle est déplacée à proximité d'une région répliquée tardivement, contenant l'ARS501 et située à proximité du télomère du chromosome V, l'ARS1 se réplique plus tardivement. Ceci signifie que la capacité de l'ARS1 à initier la réplication de manière précoce dépend du contexte chromosomique. Ils se sont ensuite intéressés à l'origine tardive ARS501. Lorsque cette origine est placée sur un plasmide circulaire, elle se réplique de manière précoce. Encore une fois la séquence de cette ARS n'est donc pas capable de dicter à elle seule son *timing* de réplication. L'ajout de séquences télomériques à l'intérieur du plasmide ne suffit pas à induire une réplication tardive, mais la linéarisation du plasmide, suivie de l'ajout de séquences télomériques, le permet. L'ensemble de ces résultats indique que la proximité d'extrémités linéaires télomériques impose une réplication tardive. Pour éprouver cette hypothèse, Ferguson et Fangman (1992) ont supprimé 34kb contenant un domaine de réplication tardif adjacent à un télomère du chromosome V. Les régions désormais voisines du télomère se répliquent alors tardivement. L'ensemble de ces résultats montre que les télomères induisent une activation tardive des origines voisines.

Certaines origines comme l'ARS1412 et l'ARS1413 ne sont pas à proximité des télomères mais s'activent néanmoins tardivement. Lorsqu'on les place sur un plasmide circulaire avec leurs séquences adjacentes (environ 20kb), on observe une réplication tardive du plasmide alors que si on réduit la taille des régions adjacentes à environ 2kb le plasmide se réplique en début de phase S. Lorsque l'ARS1 est placée à côté de l'ARS1413 cette origine est activée tardivement (Friedman et al., 1996). Cela

signifie que le moment de réplication de ces origines est induit par leurs séquences flanquantes et est indépendant de la séquence de l'origine proprement dite. Heun et al. (2001) ont étudié la localisation nucléaire de différentes origines de réplication, notamment l'ARS1412 et l'ARS1413. Ils ont montré que les télomères, ainsi que l'ARS1412, sont situés à proximité de l'enveloppe nucléaire, tandis que l'ARS607, qui se réplique de manière précoce, est située au centre du noyau. De plus, un plasmide possédant l'ARS1412 et ses séquences flanquantes est aussi localisé à la périphérie nucléaire, tandis qu'un plasmide possédant l'ARS1412 mais pas ses séquences flanquantes est localisé de manière aléatoire au sein du noyau. Il existe donc un lien fort entre réplication de l'ADN et positionnement nucléaire.

Chez la levure l'ADN n'est pas organisé aléatoirement, mais comme chez d'autres organismes le génome est organisé selon la configuration dite "Rabl", du nom de son découvreur. Dans cette configuration les centromères sont regroupés et attachés au corpuscule polaire, tandis que les télomères sont attachés à l'enveloppe nucléaire. Cette organisation a été analysée plus en détail par quantification des interactions intra- et interchromosomiques par la méthode de Hi-C (*High-throughput Conformation Capture*, Duan et al., 2010). Les données valident la conformation Rabl. L'ADN ribosomique (ADNr) localisé au centre du chromosome XII est confiné à l'intérieur du nucléole, positionné diamétralement opposé aux centromères, et bloque toute interaction entre les deux extrémités du chromosome XII. L'attachement des centromères au corpuscule polaire, et des télomères à l'enveloppe nucléaire, contraint la position des bras des chromosomes et les pousse à interagir dans leur longueur. Les petits chromosomes interagissent plus fortement avec leurs voisins car ils sont en plus forte proximité. Les bras des longs chromosomes vont plus facilement s'écarter de l'axe centromère-télomère et établir ainsi moins de contacts avec les autres chromosomes. L'étude de l'organisation spatiale du génome a révélé une interaction entre les origines précoces (Duan et al., 2010; Knott et al., 2012; Lazar-Stefanita et al., 2017). L'analyse de profils de réplication, calculés à partir de l'abondance relative des séquences dans les données de conformation chromosomique, montre une vague de réplication, démarrant aux centromères pour se propager jusqu'aux télomères (Lazar-Stefanita et al., 2017), suggérant qu'une diffusion des facteurs limitants des centromères vers les télomères régule le moment d'activation des origines en fonction de leur position le long de cet axe.

Différentes modélisations de l'organisation tridimensionnelle du génome dans une population de cellules de levures ont permis de récapituler les observations expérimentales. Le premier modèle impose aux chromosomes l'attachement des centromères au corpuscule polaire et l'attachement des télomères à l'enveloppe nucléaire. Il postule une structure particulière à l'ADNr qui le pousse à occuper 1/3 du volume nucléaire et l'empêche d'interagir avec les autres chromosomes (Wong et al., 2012). Ce modèle permet de retrouver les observations expérimentales réalisées par Duan et al (2010). Le

second modèle impose lui aussi aux chromosomes les contraintes de l'attachement des centromères et des télomères mais il exclut explicitement l'ADNr dans une région opposée au corpuscule polaire (Tjong et al., 2012). Ce modèle permet lui aussi de récapituler les résultats expérimentaux obtenus par Duan et al (2010) et notamment de retrouver le regroupement des origines précoces dans la proximité des centromères. Ceci suggère que l'attachement des télomères et des centromères force certaines origines à résider à proximité des centromères, ce qui leur confère une réplication précoce, et d'autres origines à résider à la périphérie nucléaire, ce qui leur confère une réplication tardive.

Les domaines associés topologiquement (TAD) sont des régions d'ADN qui chez les métazoaires s'étendent sur des centaines de kilobases et au sein desquelles les interactions chromatinienne sont plus fréquentes qu'avec l'extérieur. Alors qu'on pensait initialement que les TADs n'existaient pas chez la levure, des recherches récentes ont mis en évidence chez cet organisme des TADs d'environ 200 kb, au sein desquels les origines sont plus synchrones que les origines situées dans des TADs différents, même à distance comparable (Eser et al., 2017). Ceci suggère encore un lien entre le moment de déclenchement des origines et l'environnement chromatinien local.

Cependant Lööke et al., (2013) ont approfondi les expériences de relocalisation des origines. Pour cela ils ont relocalisé différentes ARS au locus VPS13 du chromosome XII, puis examiné le moment auquel elles recrutent Cdc45. Le locus VPS13 est dépourvu d'origines sur 60 kb et normalement se réplique tard en phase S. Parmi les 7 ARS relocalisées 4 recrutent Cdc45, 60min après relargage en phase S, alors que dans leur contexte génomique elles recrutent Cdc45 à différents moments, avant 60 min pour trois ARS et après 60 min pour la dernière. Ceci suggère un impact de cette région sur le timing de réplication. Les 3 ARS restantes, une fois relocalisées, recrutent Cdc45 au même moment que dans leur contexte génomique, 30 min après relargage en phase S. Knott et al., (2012) venaient par ailleurs de montrer que la suppression des facteurs de transcription Forkhead (Fkh1 Fkh2) retardait de nombreuses origines précoces et avançait une partie des origines tardives (voir Fkh1 Fkh2). Lööke et al ont confirmé que les trois ARS insensibles à la relocalisation sont liées à Fkh1. De plus, lorsque l'origine précoce ARS607, qui est activée par Fkh1 et Fkh2, est positionnée à la périphérie nucléaire grâce à la fixation d'une protéine d'ancrage à proximité de l'ARS, son *timing* n'est pas ou très peu modifié, quelle que soit la protéine d'ancrage utilisée (Ebrahimi et al., 2010). L'ensemble de ces résultats montre que certaines origines sont sensibles à leur contexte nucléaire ou chromatinien, tandis que d'autres peuvent résister à ce contexte de façon dépendante de facteurs protéiques.

## I.3.2 Régulation du moment de réplication par des protéines spécifiques

### I.3.2.1 Réplication tardive des télomères

Diverses expériences de relocalisation des télomères ont été réalisées pour déterminer si leur positionnement périnucléaire est responsable de leur réplication tardive. Il existe plusieurs voies de positionnement des télomères à la périphérie nucléaire, une médiée par le complexe KU (*yku70-yku80*) et l'autre par les protéines Sir (Hediger et al., 2002). En absence d'*yku70*, les origines de réplication proches des télomères s'activent plus tôt, tandis que les origines situées à l'intérieur des chromosomes ne sont pas affectées (Cosgrove et al., 2002; Lian et al., 2011). En absence de Sir3, une réplication plus précoce des télomères est aussi observée (Stevenson and Gottschling, 1999). Ctf18-RLC est une autre protéine impliquée dans l'attachement des télomères à l'enveloppe nucléaire, nécessaire au positionnement télomérique effectué par l'un ou l'autre des complexes KU et Sir. Ainsi, un double mutant *Ctf18-yku70* ou *Ctf18-sir4* abolit complètement la localisation du télomère XIV-L à la périphérie tandis qu'un mutant *yku70* seul ou *sir4* seul affecte plus faiblement la localisation de ce télomère. Malgré l'implication de Ctf18 dans la localisation périnucléaire des télomères, sa mutation n'entraîne aucune modification du *timing* de réplication des origines télomériques (Hiraga et al., 2006). Ces résultats montrent que l'attachement des télomères à la périphérie nucléaire n'est pas nécessaire à leur réplication tardive.

Les levures mutantes pour *yku70* activent plus tôt leurs origines télomériques. Néanmoins, *yku70* ne se lie pas à ces origines, son effet n'est donc pas direct (Lian et al., 2011). La suppression de *yku70* entraîne par ailleurs un raccourcissement des télomères, et ce phénomène est connu pour avancer le moment de réplication des séquences subtélomériques (Bianchi and Shore, 2007). Lorsque la taille des télomères est rétablie dans le mutant *yku70*, le *timing* de réplication tardif des télomères est maintenu (Lian et al., 2011). En opposition à la mutation d'*yku70*, les levures mutantes pour Rif1 (*Rap1 interacting factor 1*) possèdent de longs télomères dont la réplication est avancée comme pour une souche *yku70Δ* (Lian et al., 2011). Rif1 est capable d'interagir avec la protéine phosphatase 1 (PP1), et lorsque le domaine d'interaction de Rif1 avec PP1 est muté, la réplication des télomères est avancée (Davé et al., 2014 ; Mattarocci et al., 2014). Cette mutation augmente la phosphorylation de MCM4 par DDK (Davé et al., 2014; Hiraga et al., 2014; Mattarocci et al., 2014). Rif1 retarde donc le déclenchement des origines télomériques par le recrutement de PP1 qui, en déphosphorylant les sites de MCM4 phosphorylés par DDK, s'oppose à l'activation des origines.

Rif1 ne se fixe pas seulement aux télomères (Hafner et al., 2018; Hiraga et al., 2018; Smith et al., 2016). Différentes analyses de la réplication en absence de Rif1 ont permis de mettre en évidence des origines non télomériques régulées par Rif1 (Hafner et al., 2018; Peace et al., 2014). Les premiers

résultats obtenus sur des cellules synchronisées avec de l' $\alpha$ -facteur montrent une disparité entre les origines activées en absence de Rif1 et les sites de fixation de cette protéine (Peace et al., 2014). La seconde analyse a été réalisée sur des cellules asynchrones. Les auteurs ont observé dans ce cas un moindre nombre d'origines modifiées et ont pu les corrélés à des sites de fixation de Rif1 (Hafner et al., 2019, 2018). Certaines origines moins actives en absence de Rif1 ne possèdent pas de site de fixation pour la protéine (Hafner et al., 2018 ; Peace et al., 2014). Ceci peut s'expliquer par une moindre disponibilité en facteur limitant. Lorsque Rif1 est active, la déphosphorylation de MCM4 empêche le recrutement du facteur limitant Cdc45. Lorsque Rif1 est absent, les origines qu'il inhibait deviennent capables de recruter Cdc45, qui devient moins disponible pour les autres origines, ce qui retarde leur déclenchement.

### 1.3.2.2 Réplication précoce des centromères

La relocalisation d'un centromère dans une région répliquée tardivement avance le moment de réplication de cette région et retarde celui de la région dépourvue de centromère (Pohl et al., 2012). L'avancée du moment de réplication induite par les centromères est médiée par le kinétochore car lorsque le site de liaison des protéines du kinétochores est supprimé le centromère n'est plus capable d'établir un moment de réplication précoce (Pohl et al., 2012). L'étude de la réplication dans une souche de levure mutante pour Ctf19, un composant du kinétochore, a montré qu'une partie des origines péri-centromériques étaient retardées (Natsume et al., 2013). L'effet du centromère sur le *timing* de réplication a une portée limitée puisque sa relocalisation n'affecte pas le *timing* de réplication de l'ensemble du chromosome, tout comme la suppression de Ctf19 (Pohl et al., 2012 ; Natsume et al., 2013). Lorsque Ctf19 est absent, le recrutement de Dbf4 au niveau des origines centromériques est réduit. Ainsi, la réplication précoce des origines centromériques est induite par le recrutement par Ctf19 du facteur Dbf4, qui est un des facteurs limitant pour l'activation des origines (Natsume et al., 2013).

### 1.3.2.3 Régulation du moment de réplication par des protéines spécifiques ni télomériques ni centromériques

#### 1.3.2.3.1 Fkh1 Fkh2

Knott et al (2012) ont étudié l'effet de la suppression de Fkh1 et Fkh2 sur l'incorporation de BrdU en début de phase S. Ils ont montré que 30% des origines étudiées, dont l'ARS607, incorporaient moins de BrdU dans le double mutant que dans le type sauvage, et que 23% en incorporaient davantage. Le mécanisme d'action de Fkh1/2 (Fkh1 et/ou Fkh2) sur les origines précoces semble direct puisque 42% des origines activées par Fkh1/2 possèdent un site de fixation pour Fkh1 détectable en CHIP-seq. Néanmoins, 11% des origines réprimées par Fkh1/2 possèdent un site de fixation pour Fkh1

en ChIP-seq (Ostrow et al., 2014). La surexpression de Fkh1/2 augmente la liaison de Fkh1/2 et l'incorporation de BrdU à des origines auparavant classées comme réprimées par Fkh1/2 ou non régulées par ces protéines (Peace et al., 2016). Ces résultats indiquent que Fkh1/2 ont une activité de stimulation des origines mais que leur dosage limite le nombre d'origines qu'ils peuvent stimuler.

Le mécanisme d'action de Fkh1/2 est lié à Dbf4. En effet Fkh1 se lie à la partie C-terminale de Dbf4 et lorsque que cette partie est supprimé (Dbf4ΔC), le phénotype réplcatif est le même que pour les levures mutantes pour Fkh1/2. De plus, Fkh1 reste lié à l'ADN même lorsqu'il ne peut pas s'associer à Dbf4, et Dbf4 seul n'est pas capable de se lier à l'ADN. Enfin, l'expression d'une protéine de fusion de Dbf4ΔC avec le domaine de liaison à l'ADN des protéines Fkh est capable de restaurer la réplication précoce des origines activées par Fkh1/2 en absence de Fkh1. Ceci s'accompagne d'une inactivation des origines centromériques, car cette fusion protéique perturbe le recrutement de Dbf4 par Ctf19 (Fang et al., 2017). Ceci illustre la différence de mécanisme de recrutement de DDK, via Ctf19 aux origines centromériques, et via Fkh1/2 à une grande partie des origines précoces non centromériques.

#### 1.3.2.3.2 Rpd3, Sir2, Sir3 et acétylation des histones

L'histone désacétylase Rpd3 est impliquée dans le contrôle d'une centaine d'origines de réplication (Aparicio et al., 2004 ; Knott et al., 2009 ; Vogelauer et al., 2002). En absence de Rpd3 le moment d'activation de ces origines est avancé et l'acétylation de leur chromatine augmente (Aparicio et al., 2004 ; Knott et al., 2009). De plus, Mantiero et al (2011) ont montré que la surexpression de Sld2, Sld3, Dbf4 et Dpb11, même renforcée par Cdc45 et Sld7, permettait l'activation précoce de beaucoup d'origines tardives mais pas toutes. Les origines inefficaces associées aux régions tardives, hétérochromatiques du locus HML (ARS301 and ARS302/303/320) ne sont pas affectées. Elles ne se déclenchent que lorsqu'on exprime les facteurs limitants dans un mutant *rpd3*. Ceci indique que c'est leur hypoacétylation par Rpd3 qui rend ces origines réfractaires à une surexpression des facteurs limitants. De plus, une étude très récente (He et al., 2022) suggère que Rpd3 agit en s'opposant à la liaison de Fkh1/2 aux origines, car en absence de cette histone désacétylase, la fixation de Fkh1/2 aux origines qu'ils régulent et aux origines inhibées par Rpd3 augmente. Fkh1/2 se révèlent nécessaires à la réactivation des origines inhibées par Rpd3 en l'absence de cette désacétylase. De plus, la suppression de Rpd3 permet la survie de levures Dbf4ΔC. Rpd3 semblerait donc agir en s'opposant au recrutement de Fkh1/2 et donc au recrutement de DDK.

Deux autres histones désacétylases, Sir2 et Sir3, régulent la réplication de l'ADN. Sir3 est impliqué dans la réplication tardive des régions péricentromériques (Stevenson and Gottschling, 1999). L'absence de Sir2 ou Sir3 permet de réduire le défaut de croissance induit par un mutant hypomorphe de Cdc6 (Cdc6-4 ; Pappas et al., 2004), ce qui suggère que ces deux protéines inhibent la formation du

pré-RC. De plus, le défaut de croissance induit par Cdc6-4 est aussi secouru lorsque le résidu 16 de l'histone H4 normalement désacétylé par Sir2 est remplacé par une glutamine. Ce résultat suggère que la désacétylation de l'histone H4 par Sir2 est responsable de son effet sur la formation du pré-RC (Crampton et al., 2008). Hoggard et al (2018) ont montré que dans les souches Cdc6-4 en absence de Sir2, le complexe MCM est recruté de manière plus importante aux origines de réplication précoce et que ce recrutement a un lien avec l'acétylation des histones H4K16 bordant les origines de réplication. Sir2 et Sir3 sont capables de désacétyler ce résidu et sont situés à proximité des origines de réplication. L'ensemble de ces résultats suggère que la désacétylation des histones par Sir2 et Sir3 décroît le recrutement du complexe MCM aux origines de réplication précoce permettant ainsi l'activation d'origines tardives (pour revue : Lengronne and Pasero, 2020).

L'ADNr est présent entre 100 et 200 copies dans le génome et ses origines représentent un tiers du total des origines de la levure. Cependant seulement 20% de ces origines sont activées à chaque phase S (Brewer and Fangman, 1988). Sir2 est responsable de l'inhibition de la réplication des origines de l'ADNr (Pasero et al., 2002). En absence de l'histone désacétylase Sir2, qui agit sur l'ADNr, l'efficacité des origines augmente dans l'ADNr mais diminue sur le reste du génome. A l'inverse, en absence de Rpd3, l'activité de ces origines augmente (Yoshida et al., 2014). En absence de Sir2 et Rpd3, le profil de réplication est similaire à celui d'une souche sauvage. L'inhibition d'origines au sein du génome en absence de Sir2 peut être contrecarrée par une surexpression des facteurs limitants. Ces résultats suggèrent que Sir3 et Rpd3 régulent l'activation des origines de l'ADNr mais aussi du reste du génome en modulant la compétition entre ces deux compartiments du génome pour l'accès aux facteurs limitants. Cependant l'étude très récente de He et al (2022) suggère que si Sir2 retarde bien le déclenchement des origines du ADNr, ce qui facilite le déclenchement d'autres origines dans le génome, la fonction de Rpd3 qui retarde le déclenchement de certaines origines hors du ADNr est indépendante de Sir2 et du nombre de copies du ADNr. Ces auteurs proposent que Rpd3 ne régule pas les origines tardives indirectement, par activation des origines du ADNr, mais directement, en s'opposant au recrutement de Fkh1/2 et donc de Dbf4.

Enfin, l'ancrage de l'histone acétylase GCN5 à l'origine tardive ARS1412 avance son *timing* de réplication (Vogelauer et al., 2002). L'ensemble de ces résultats illustre les multiples mécanismes de contrôle de l'initiation de la réplication par acétylation et désacétylation d'histones dans le voisinage ou à distance des origines de réplication.

### 1.3.2.3.3 Relation entre efficacité des origines et abondance des MCM

Toutes les origines ne sont pas régulées par Fkh1/2, Rif1 ou Rpd3. Il pourrait donc exister d'autres mécanismes régulateurs. Yang et collaborateurs (2010), ont pu modéliser des profils expérimentaux de *timing* de réplication en postulant que les origines sont marquées par une ou plusieurs molécules initiatrices ayant toutes la même probabilité de déclenchement par unité de temps, et que l'efficacité globale de chaque origine est déterminée par le nombre d'initiateurs qui s'y lient. Ainsi, les origines les plus efficaces, qui ont fixé le plus d'initiateur, sont aussi en moyenne les plus précoces. Les auteurs ont proposé le complexe MCM comme étant l'initiateur responsable du *timing* de réplication. L'analyse de profils MCM CHIP-seq et *timing* de réplication montre une bonne corrélation entre l'efficacité intrinsèque d'une origine calculée à partir du modèle et l'abondance des MCM (Das et al., 2015). Néanmoins cette tendance est loin d'être absolue. Ainsi les origines télomériques possèdent une forte quantité de MCM mais sont actives en fin de phase S, et les origines péri-centromériques activées par Ctf19 possèdent un signal moins fort qu'attendu, ce qui montre l'importance des autres mécanismes discutés plus haut (Das et al., 2015). De plus, lorsque la quantité de MCM intracellulaire est artificiellement réduite, le *timing* de réplication n'est pas perturbé uniformément. Les origines qui possédaient le signal le plus fort en CHIP-seq sont moins affectées que celle qui possédaient le signal le plus faible. A l'inverse, la surexpression des MCM ne perturbe pas le profil de *timing*, ce qui suggère que ces protéines sont normalement saturantes et que c'est la capacité intrinsèque de chaque origine à accepter les MCM qui détermine la quantité qui s'y lie, de sorte que des origines de capacité différente sont différenciellement affectées par un déficit en MCM (Dukaj and Rhind, 2021).

Rhind (2022) a proposé un modèle simple d'activation des origines,  $f=m*a$ , où  $f$  représente la probabilité d'une origine de s'activer indépendamment des mécanismes de passivation,  $m$  la quantité de MCM et  $a$  la capacité des MCM à être activés. Ce dernier point illustre le fait que les télomères, malgré leur forte quantité de MCM, ne s'activent qu'en fin de phase S car certaines protéines réduisent leur capacité à s'activer. Ce modèle est également compatible avec un effet de la structure tridimensionnelle, qui serait de rapprocher ou d'éloigner les MCM des facteurs qui limitent leur activation, et donc d'affecter leur capacité à être activés.

### 1.3.3 Interdépendance entre organisation nucléaire et activation des origines par des protéines

Pour expliquer l'action de Fkh1/2 sur la réplication, Knott et al (2012) ont proposé un mécanisme où Fkh1/2 regroupent les origines qu'ils activent à proximité des origines péri-centromériques. Cette supposition s'appuie sur des résultats de conformation chromosomique

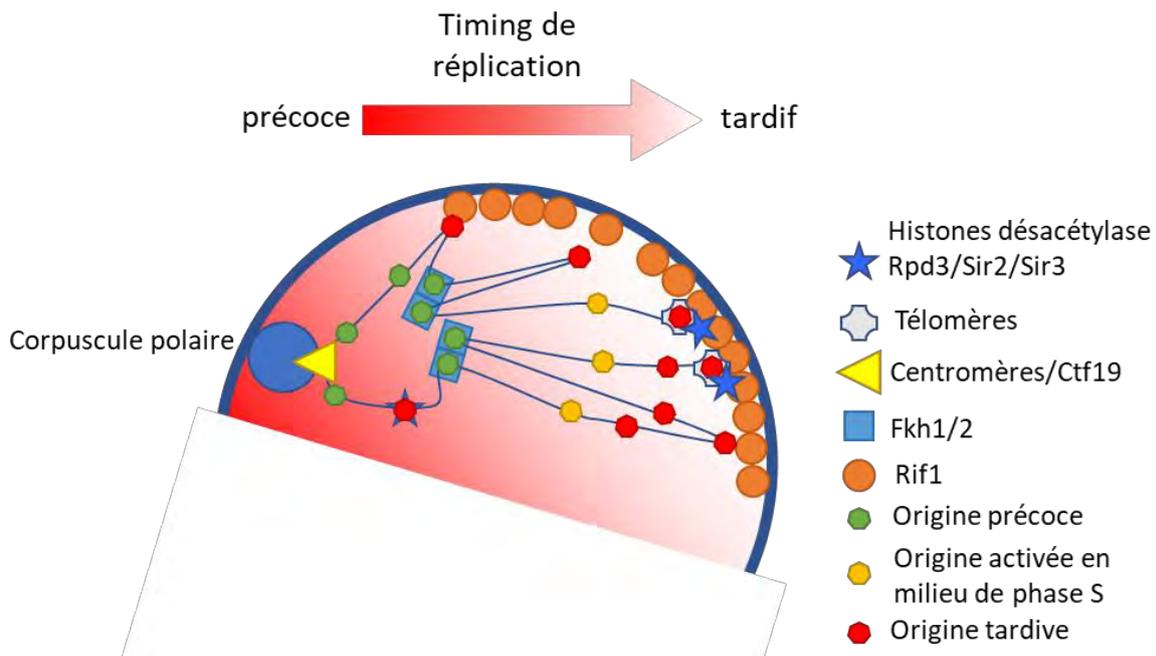
montrant un enrichissement des interactions entre les origines activées par Fkh1/2, les origines péri-centromériques et les origines précoces non régulées par Fkh1/2 (Knott et al., 2012). Les origines inhibées par Fkh1/2 seraient quant à elles refoulées dans un environnement moins favorable à la réplication, du fait du regroupement des origines les plus actives par ces protéines. Cette hypothèse est appuyée par la capacité de Fkh1 et Fkh2 à former des homodimères (Ostrow et al., 2017). Lorsqu'une souche de levure mutante pour Fkh1/2 est complétée avec des versions de Fkh1 sauvage ou mutée dans le domaine de dimérisation, le *timing* de réplication précoce est rétabli dans le premier cas mais pas dans le second. Cela montre la nécessité pour Fkh1 de former un dimère pour réguler la réplication. Ce processus de dimérisation n'est cependant pas impliqué dans la liaison de Fkh1 à l'ADN car Fkh1 muté dans son domaine de dimérisation se lie toujours aux origines qu'il régule (Ostrow et al., 2017). On pourrait donc supposer que la dimérisation de Fkh1 permet d'accrocher les origines régulées par Fkh1/2 ensemble et de créer un environnement favorable. Néanmoins Fkh1/2 ne sont pas responsables de la formation des TADs (Eser et al., 2017).

L'activation d'origines par Fkh1/2 induit leur relocalisation à l'intérieur de l'espace nucléaire en phase G1, après le chargement de Cdc45 et Sld3 dépendant de DDK, mais cette relocalisation ne nécessite pas d'activité CDK. La relocalisation est abolie lorsque l'action de DDK est perturbée en présence de PP1, ou d'un analogue de l'ATP qui inhibe la kinase Cdc7, ou lorsque la liaison entre Fkh1 et Dbf4 est abolie. De plus, la relocalisation n'a pas lieu lorsque l'ARS n'est pas fonctionnelle en l'absence de son ACS (Zhang et al., 2019). Il existe donc une relocalisation des origines activées par Fkh1/2 en phase G1 dépendante de DDK. Ce phénomène pourrait faciliter l'activation en début de phase S des origines relocalisées, en les concentrant dans un milieu nucléaire riche en facteurs limitants, et inhiber les autres en les laissant dans un contexte moins favorable. Mais cette relocalisation n'est pas forcément essentielle pour l'activation précoce des origines par Fkh1/2 car lorsque l'ARS607, activée par Fkh1/2, est placée à la périphérie nucléaire à l'aide d'une protéine d'ancrage, la relocalisation devient impossible mais le *timing* précoce est maintenu (Ebrahimi et al., 2010). La relocalisation dépendante de DDK pourrait tout aussi bien être une conséquence de l'activation des origines et non une cause. Ebrahimi et Donaldson (2008) ont étudié la relocalisation des télomères à l'intérieur du noyau au cours du cycle cellulaire et il se trouve que cette relocalisation est liée à la réplication puisqu'elle intervient après la réplication et qu'un délai de réplication des télomères par la suppression d'ARS dans leur proximité induit un délai de relocalisation.

Fkh1/2 ont tendance à imposer une position intranucléaire aux origines qu'ils régulent, tandis que Rif1 serait plutôt impliqué dans une localisation périphérique. Rif1 est un substrat de la palmitoyltransferase Pfa4, et la palmitoylation de Rif1 est liée à sa localisation à la périphérie nucléaire. Néanmoins la suppression de Pfa4 n'entraîne pas ou peu de diminution de la liaison de Rif1 aux

téломères (Park et al., 2011) et le profil de réplication *pfa4Δ* est similaire à celui d'une souche sauvage (Peace et al., 2014). Ces résultats montrent que la localisation à la périphérie nucléaire n'est pas essentielle pour l'action de Rif1.

L'ensemble de ces éléments permettent d'envisager un modèle adaptée de ceux décrits précédemment pour *Schizosaccharomyces pombe* (Nathanailidou et al., 2020; Pichugina et al., 2016) et *S. cerevisiae* (Hafner et al., 2018; Zhang et al., 2019). Dans ce modèle, Fkh1/2 et Ctf19 créent un endroit propice à la réplication dans une région située entre le corpuscule polaire et le centre du noyau, par recrutement de DDK via Dbf4. Rif1 est quant à lui localisé à la périphérie nucléaire où il crée une zone moins propice à la réplication par recrutement de la phosphatase PP1, qui antagonise l'action de DDK. Ainsi s'établit une vague de *timing* de réplication précoce au niveau du corpuscule polaire et du centre du noyau et tardive à la périphérie nucléaire. Les centromères, les télomères, le regroupement d'origines par Fkh1/2 ainsi que l'attachement de Rif1 à la périphérie nucléaire vont de plus contraindre la localisation des origines non directement régulées par ces protéines, de manière favorable ou non à leur activation. C'est pourquoi lorsque la quantité de facteurs limitant augmente, ce qui pourrait homogénéiser leur concentration intranucléaire, la majorité des origines s'activent de manière plus précoce (Mantiero et al., 2011). Sir2 et Sir3 inhibent certaines origines précoces en réduisant la fixation de MCM, ce qui facilite l'activation d'origines tardives placées dans un environnement moins favorable, tandis que Rpd3 antagonise l'action de Fkh1/2. Lorsque Fkh1/2 sont absents, le regroupement des centromères et l'action de Ctf19 maintiennent encore un contexte favorable à la réplication. En absence de Ctf19, seules les origines les plus proches des centromères sont affectées puisque les autres sont toujours dans un contexte favorable maintenu par Fkh1/2 (Figure I-4).



**Figure I-4 : Modèle de l'impact de l'organisation du noyau sur la réplication.**

Les origines à proximité des centromères bénéficient d'une réplication précoce grâce à l'action de Ctf19. Fkh1/2 regroupe un certain nombre d'origines à proximité des centromères, ce qui en refoule d'autres vers la périphérie nucléaire. Ctf19 et Fkh1/2 créent un environnement favorable à la réplication précoce. Rif1 est situé le long de l'enveloppe nucléaire où il est lié aux télomères et aux origines qu'il régule. Rif1 crée un environnement défavorable à la réplication en recrutant PP1 ce qui induit une réplication tardive des origines à proximité. Les origines situées entre la zone favorable et la zone défavorable se répliquent en milieu de phase S. Un niveau de contrôle supplémentaire est assuré par les histones désacétylases qui inhibent certaines origines non télomériques en réduisant la liaison des MCM ou la liaison de Fkh1/2.

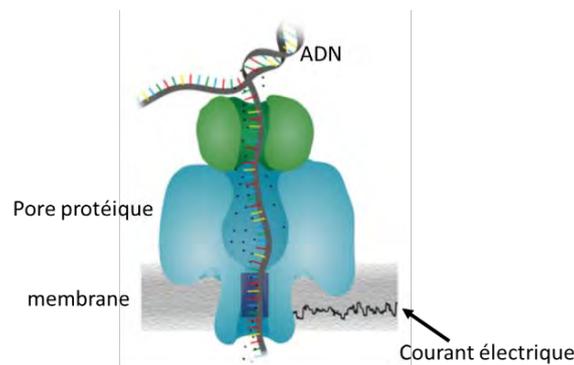
#### 1.4 Déterminer l'implication de l'organisation du génome dans différents processus cellulaires : fusion des chromosomes chez la levure.

L'information génétique chez les eucaryotes est organisée en différents chromosomes, suggérant que l'évolution a favorisé une partition de l'information génétique en différents fragments d'ADN. Deux laboratoires se sont demandés si cette partition était nécessaire à la viabilité cellulaire. Ils ont réalisé des séries de suppression des régions télomériques et centromériques à l'aide de CRISPR-cas9 suivies de fusions des extrémités créées. Ainsi Luo et al (2018) ont créé une souche de levure à 2 chromosomes et Shao et al (2018) une souche de levure à un chromosome. Ces souches présentent un léger défaut de croissance, suggérant un faible impact sur la viabilité cellulaire. Néanmoins, lorsque l'unique chromosome est circularisé, la croissance est davantage affectée, surtout en présence d'un agent génotoxique comme le méthyl méthanesulfonate (MMS) ou la camptothécine (Shao et al 2019). Le principal phénotype observé pour ces souches est une moindre capacité à sporuler.

Ces différentes souches représentent de parfaits modèles pour étudier l'impact de l'organisation de l'information génétique sur différents processus cellulaires. L'analyse de la conformation chromosomique au sein des souches possédant un caryotype modifié révèle une complète réorganisation tridimensionnelle de leur génome. La suppression des centromères entraîne une perte des interactions entre les régions centromériques et une interaction plus forte entre les centromères restants. La perte des télomères implique aussi une perte de contact entre les régions télomériques. La perte de la configuration Rab1 provoque une diminution des interactions interchromosomiques et une augmentation des interactions intrachromosomiques (Lazar-Stefanita et al., 2022; Shao et al., 2018). Dans un premier temps, Shao et al (2018) et Luo et al (2018) se sont intéressés à la modification de l'expression des gènes et ont trouvé seulement quelques gènes modifiés. La souche de levure à un chromosome montre 9 gènes inhibés et 19 gènes activés tandis que la souche de levures à 2 chromosomes montre 6 gènes inhibés et 32 gènes activés. La majorité des gènes inhibés dans les deux souches sont situés à proximité des télomères restants, suggérant un renforcement de l'inhibition de ces régions en absence des autres télomères. Ces résultats montrent que la réorganisation nucléaire a peu d'impact sur la transcription des gènes. Néanmoins, Di Stefano et al., (2020) ont examiné les changements d'expression des gènes dans différentes souches au caryotype réorganisé. Ils ont mis en évidence une relation entre position nucléaire et modification de l'expression des gènes. Plus un gène se retrouve éloigné de la périphérie nucléaire après fusion des chromosomes, plus son expression est augmentée. Ces résultats montrent que le positionnement nucléaire possède un effet modéré sur l'expression des gènes.

## II. Présentation du projet de thèse

Avant mon arrivée au laboratoire, Magali Hennion, post-doctorante, avait montré qu'il était possible d'utiliser le séquençage Nanopore (Figure II-1) pour détecter la bromodésoxyuridine (BrdU), un analogue de la thymidine. En effet, les courants électriques observés en Nanopore sont différents aux positions où le BrdU a remplacé la thymidine (Hennion et al ; 2020). De nombreuses méthodes d'étude de la réplication reposent sur l'incorporation d'analogue de nucléotides dans l'ADN lors de sa synthèse. La réalisation d'un marquage bref au BrdU (*pulse*) suivi d'une chasse par la thymidine permet de détecter et d'orienter les fourches de réplication sur des lectures de molécules individuelles d'ADN, dont on peut déterminer la position génomique. Ceci a permis à Magali Hennion de développer Fork-seq et d'analyser en détail la réplication du génome de la levure *S. cerevisiae*. Cette méthode permet non seulement d'assembler des profils populationnels de direction moyenne des fourches de réplication (*replication fork directionality*, RFD), mais aussi de recenser la localisation d'évènements individuels d'initiation et de terminaison signalés par la présence de deux fourches divergentes ou convergentes sur une même molécule. Nous avons donc à notre disposition un outil puissant pour questionner les mécanismes de réplication.



**Figure II-1 : Schéma du séquençage nanopore.**

Un pore protéique est inséré dans une membrane qui possède une différence de potentiel électrique de part et d'autre générant un courant électrique. Lorsqu'une molécule d'ADN passe à travers le pore protéique, les bases présentes au point de constriction, représenté en violet, modifient le courant électrique. Ces modifications du courant électrique sont ensuite converties en séquence d'ADN par des algorithmes complexes.

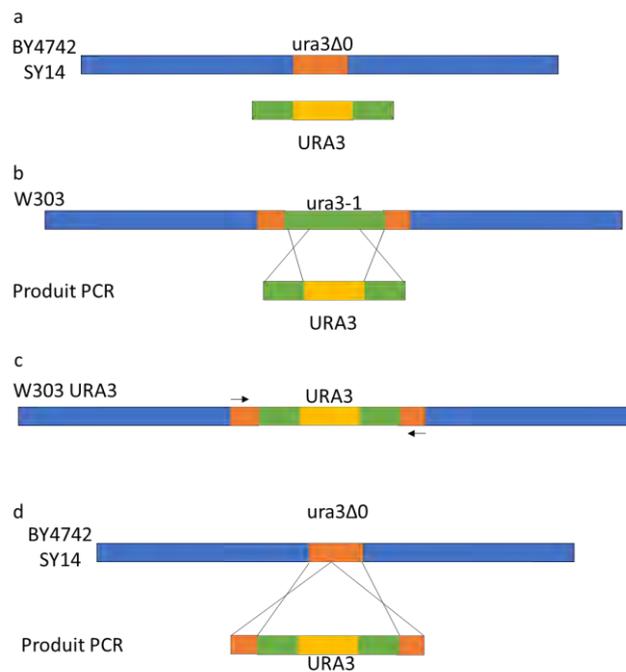
Les études réalisées jusqu'à présent pour déterminer l'impact du positionnement nucléaire sur la réplication ne pouvaient pas être transposées à grande échelle. Les études de relocalisation requièrent pour chaque ARS la modification du génome de la levure, puis l'analyse de la réplication et du positionnement de l'origine. 829 séquences ayant été identifiées comme des ARS (Siow et al., 2012), il est difficile d'envisager l'étude systématique de chaque origine. Il est donc nécessaire d'utiliser un

modèle permettant de perturber globalement l'organisation du génome. Les souches au caryotype réorganisé représentent des modèles de choix pour étudier l'impact de l'organisation du génome sur la réplication. Les protéines qui régulent la réplication étant toujours présentes, ces modèles devraient permettre de distinguer la contribution de l'organisation du génome. C'est pourquoi mon projet de thèse s'est porté sur la détermination des changements du programme de réplication du génome dans la souche de levure à un chromosome : SY14.

### III. Méthodes expérimentales

#### III.1 Transformation de la BY4742 et de la SY14 pour l'incorporation de BrdU dans l'ADN

Afin que la BY4742 et la SY14 puissent incorporer le BrdU présent dans leur milieu extérieur au sein de l'ADN, nous voulions insérer le plasmide pBL-hsvTK<sub>co</sub>-hENT1<sub>co</sub> (Theulot et al ; 2022) aux loci URA3 et AUR1. La BY4742 et la SY14 sont *ura3Δ0*, c'est-à-dire qu'elles ne possèdent plus le gène URA3. Dans un premier temps, il était nécessaire de reconstituer le locus URA3 (Figure III-1). Dans un second temps, les souche URA3 transformées avec le plasmide ne pouvant pas être sélectionnées sur milieu sans uracile, nous avons modifié le plasmide en lui insérant une cassette de résistance pour un antibiotique : la nouséothricine, puis nous avons sélectionné les clones ayant incorporé le plasmide sur un milieu contenant la nouséothricine.



**Figure III-1 : Stratégie pour reconstituer le locus URA3 dans la BY4742 et la SY14.**

a) La BY4742 et la SY14 sont *ura3Δ0*. Elles ne possèdent donc aucune séquence permettant la recombinaison homologue avec le gène URA3. b) La W303 est *ura3-1*, elle peut donc effectuer la recombinaison homologue avec un produit PCR contenant le gène URA3. c) Le gène URA3, avec ses séquences flanquantes, a été amplifié à partir de la W303-URA3. d) Un produit PCP contenant le gène URA3 avec ses séquences flanquantes peut être intégré par recombinaison homologue dans la SY14 et la BY4742.

### III.1.1 Culture de levures

Les souches BY4742 et SY14 ont été obtenues auprès du laboratoire de Zhongjun Qin. Les différentes souches sont stockées dans un mélange contenant 50% (volume/volume, v/v) de glycérol 40% (masse/volume, m/v) et 50% (v/v) de milieu *Yeast Peptone Extract* (YPD) à -80°C, avant d'être déposées sur une boîte YPD agar à 30°C pendant deux à trois jours. Les boîtes sont ensuite placées à 4°C pour une durée maximale de deux semaines.

En amont d'une expérience, une préculture est réalisée. Pour cela un clone isolé est récupéré sur boîte YPD agar et ensemencé dans un Erlenmeyer contenant 1/5 de son volume en YPD. Les cultures sont placées à 30°C, sous agitation à 180 rpm pendant une nuit. Le jour de l'expérience, les cellules sont diluées dans du milieu frais à une densité optique à 600nm ( $DO_{600nm}$ ) de 0,1, puis placé à 30°C sous agitation à 180 rpm. Les cultures sont ensuite utilisées en phase exponentielle de croissance après au moins deux doublements et tant que la  $DO_{600nm}$  est comprise entre 0,6 et 0,8.

### III.1.2 Amplification du locus URA3

Le locus URA3 a, dans un premier temps, été amplifié à partir d'un plasmide contenant ce locus (Tableau III-1). La PCR est réalisée avec la polymérase haute-fidélité Q5 (New England Biolabs®), 200µM de dNTPs, 0,5µM de chaque amorce et 1µg de plasmide. Le programme PCR comporte une première étape de dénaturation 3min à 95°C puis 35 cycles avec 30s de dénaturation à 95°C, 30s d'hybridation à 53°C et 1min30 d'amplification à 72°C. A l'issue de ces cycles, une étape d'élongation supplémentaire à 72°C pendant 10min est réalisée. Le produit PCR obtenu a été intégré dans la souche W303 par transformation (voir Transformation des levures). L'ADN de la souche transformée a été extrait (voir Extraction d'ADN) puis amplifié avec ses séquences flanquantes à l'aide d'amorces dans ces séquences (Tableau III-1), le mix PCR ainsi que le programme utilisé sont les mêmes que lors de la PCR précédente. Le produit PCR obtenu est utilisé pour transformer la BY4742 et la SY14 en souche URA3.

Région à amplifier	Amorce 1	Amorce 2
Gène URA3 à partir du plasmide	CATGTACGGGTTACAGCAGA	GTTCTTGATTTGTGCCCCGT
URA3 avec ses séquences flanquantes à partir de la W303	GCTACTGCGCCAATTGATGA	GTTCTTGATTTGTGCCCCGT
Vérification de l'intégration d'URA3	TGCACATGAACAAACACCAGA	ACACGTTGCTATGCTTCAA

**Tableau III-1 : Amorces pour amplifier le locus URA3 et vérifier l'intégration du gène dans le génome de la levure *S. cerevisiae*.**

### III.1.3 Modification du plasmide pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub> (Theulot et al, 2022)

Le plasmide pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub> a été modifié par des techniques de biologie moléculaire classiques. Brièvement, le plasmide pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub> et le plasmide pAG25 (Addgene) qui contient la cassette de résistance à l'antibiotique nouséothricine *nat1* ont été ouverts par l'enzyme de restriction *NotI*. La cassette *nat1* et le plasmide pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub> ont ensuite été ligués à l'aide de la T4 DNA ligase (New England Biolabs®). Le plasmide obtenu est appelé pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub>-NOUR.

Afin d'obtenir ce plasmide en grande quantité, il a été intégré au sein de bactéries *E.coli* compétentes DM5α (New England Biolabs®) par choc thermique. Une fois transformées, les bactéries sont étalées sur une boîte de sélection contenant 100µg/mL d'ampicilline et placées une nuit à 37°C. Le lendemain, les clones isolés sont testés par PCR à l'aide de deux amorces dirigées contre des séquences contenues dans le plasmide (Tableau III-2). Brièvement, les clones sont placés dans un tube PCR avec 50µL d'H<sub>2</sub>O pendant 5min à 95°C. 2µL de cette solution sont utilisés pour réaliser une PCR avec la Taq polymérase (New England Biolabs®), 0,2µM d'amorce et 200 µM de dNTPs. Le programme PCR contient une première étape de dénaturation pendant 5min à 95°C, puis 35 cycles d'amplification sont réalisés avec une étape de dénaturation 30s à 95°C, une étape d'hybridation 30s à 53°C, une étape d'élongation 2min30 à 68°C. A l'issue de ces cycles, une dernière étape d'élongation 10min à 68°C est réalisée. La taille des produits obtenues est vérifiée en déposant 4µL de produit PCR sur un gel d'agarose à 1% (m/v) avec 0,025% de Bromure Ethidium (BET, v/v). Deux clones sont ensuite placés dans 50mL de milieu *lysogeny broth* (LB) avec 100µg/mL d'ampicilline une nuit à 37°C. Le lendemain, le plasmide est extrait avec le kit NucleoBond® Xtra midi (Macherey-Nagel®) puis l'intégration de la cassette est vérifiée par séquençage de sanger réalisé par Eurofins.

Région à amplifier	Amorce 1	Amorce 2
Plasmide dans les bactéries	TAATACGACTCACTATAGGG	CTG CCA TTT TGG TTA AAG TAC
pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> -NOUR bordure gauche	TGAAAACCTCTGACACATGCAG	GTCATTATAGAAATCATTACGACCGAGA TTCCC
pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> -NOUR bordure droite	GGCTGTGGTTTCAGGGTCCATA AAGC	CTTGATTAGGGTGATGGTTCACG
pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> -NOUR multicopies	TGAAAACCTCTGACACATGCAG	CTTGATTAGGGTGATGGTTCACG
pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> bordure gauche	GGG CAT CCC TTT GCA TGT CT	CCTGGCCTTTTGCTGGCCTT
pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> bordure droite	AAGAAGTTGCATGCGCCTAT	CCC GGA AAA GTT TCA ACA AGA
pBL-hsvTK <sub>CO</sub> -hENT1 <sub>CO</sub> multicopies	AAGAAGTTGCATGCGCCTAT	CCTGGCCTTTTGCTGGCCTT

**Tableau III-2 : Amorce pour la vérification de l'intégration des plasmides pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub>.**

#### III.1.4 Transformation des levures

10 à 20 unités de DO<sub>600nm</sub> (u.DO) sont utilisées, une u.DO représente 1ml de culture à une DO<sub>600nm</sub>=1, soit environ 2x10<sup>7</sup> cellules. Les cellules sont dans un premier temps rendues compétentes pour l'incorporation de plasmide. Pour cela les cultures sont centrifugées 5min à 3100g, les culots sont lavés avec un volume d'eau équivalent au volume de culture. Après une nouvelle centrifugation 5min à 3100g, les cellules sont reprises dans 1mL de tampon TE-AcLi (10mM Tris-HCL pH 7,5 ; 1mM EDTA ; 100mM AcLi), puis centrifugées 5min à 3100g. Ces nouveaux culots sont repris dans 50µL de TE-AcLi avant d'être centrifugés 3 min à 2600g. Pour la transformation, les cellules sont ensuite reprises dans un volume final de 100µL de TE-AcLi en prenant en compte le volume du culot. 25µL d'ADN amplifié par PCR (gène URA3) ou 1µg de plasmide linéarisé par *StuI* (pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub> et pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub>-NOUR) sont ajoutés à l'échantillon avec 25µL d'ADN de sperme de saumon, et 600µL de PEG-AcLi (10mM Tris-HCL pH 7,5 ; 1mM EDTA ; 100mM AcLi ; PEG40%). Les cellules compétentes sont ensuite incubées 1h à 30°C, 30min à 42°C, 10min à 30°C. A l'issue de ces incubations, les cellules sont centrifugées 5min à 650rpm, reprises dans 1mL d'H<sub>2</sub>O puis de nouveau centrifugées 5min à 650rpm avant d'être reprises dans 50µL d'H<sub>2</sub>O et étalées sur boîte YPD pour les transformations avec les plasmides et sur milieu minimum sans uracile pour la transformation avec le gène URA3. Pour les transformations avec les plasmides, les boîtes sont répliquées le lendemain sur milieu YPD contenant l'antibiotique (Auréobenzidine 0,1µg/mL ou Nourseothricine 100µg/mL).

### III.1.5 Vérification de l'intégration des plasmides et de l'insertion du gène URA3

Une fois que des clones isolés ont poussé sur les boîtes de sélection, 6 clones sont étalés sur une nouvelle boîte de sélection afin de vérifier la bonne croissance sur milieu sélectif. Afin de vérifier l'intégration des plasmides et l'insertion du gène URA3, l'ADN des clones ayant poussé sur les boîtes de sélection est extrait par un protocole dérivé de celui décrit par Lööke et al., (2017). Pour cela, 5 à 6 colonies sont prélevées sur la boîte et re-suspendues dans 100µL de SDS-AcLi (200mM AcLi ; 1% SDS (m/v)) et placées à 70°C pendant 15min. Ensuite, 300µL d'éthanol 100% (v/v) sont ajoutés, puis les échantillons sont centrifugés 5min à 15000g et lavés dans 500µL d'Ethanol 70% (v/v). Après une nouvelle centrifugation 5min à 15000g, et après évaporation des restes d'éthanol 5min à 50°C, les culots sont repris dans 100µL d'H<sub>2</sub>O puis une nouvelle centrifugation est réalisée 1min à 15000g, l'ADN contenue dans le surnageant est récupéré. 4µl d'ADN sont utilisés pour réaliser la PCR, à l'aide de la Taq Polymérase (New England Biolabs®), 0.2µM d'amorce et 200µM de dNTPs. Le cycle PCR se compose de 30s de dénaturation à 95°C, suivi de 35 cycles avec 30s de dénaturation à 95°C, 30s de liaison des amorces à 53°C et 1min30 d'élongation à 68°C. A l'issue des 30 cycles, une dernière étape d'élongation à 68°C pendant 5min est réalisée. L'insertion du gène URA3 a été vérifiée par PCR à l'aide d'amorces dans ses séquences flanquantes (Tableau III-1). L'insertion des plasmides a été vérifiée selon la procédure décrite par Viggiani and Aparicio (2006), qui permet de vérifier l'intégration du plasmide à l'aide d'une amorce dans les séquences flanquantes du gène d'insertion et d'une amorce au sein du plasmide, les deux bordures d'insertion du plasmide peuvent être vérifiées par cette procédure. L'insertion du plasmide en une copie est déterminée par deux amorces au sein du plasmide. Le résultat de cette PCR indique si le plasmide est inséré en une ou plusieurs copies, mais ne nous indique par le nombre exact de copies du plasmide (Tableau III-2).

### III.1.6 Mesure du nombre de copies de plasmide intégré par qPCR

Le nombre de copies du plasmide intégré après transformation est déterminé par PCR quantitative (qPCR) à l'aide d'un couple d'amorces dans le gène hENT1<sub>CO</sub> et un couple d'amorces dans le gène hsvTK<sub>CO</sub> (Tableau III-2). Le nombre de copies est normalisé à l'aide d'un couple d'amorces dans le gène ALG9 et d'un couple d'amorce dans le gène UBC6. Ces gènes ont été choisis pour la stabilité de leur expression dans des conditions de croissance variées (Tableau III-3, Teste et al., 2009). La qPCR est réalisée uniquement sur des ADN (voir III.1.3) dont la qualité est vérifiée à l'aide des ratios  $\text{Absorbance}_{260\text{nm}}/\text{Absorbance}_{230\text{nm}}$  ( $A_{260\text{nm}}/A_{230\text{nm}}$ ) et  $\text{Absorbance}_{260\text{nm}}/\text{Absorbance}_{280\text{nm}}$  ( $A_{260\text{nm}}/A_{280\text{nm}}$ ). Seuls les ADN possédant le premier ratio supérieur à 1,8 et le second ratio supérieur à 2 sont conservés. La qPCR est réalisée en triplicats dans une plaque 384 puits avec 7,5 ng ou 15ng d'ADN par puits. Ensuite un mélange contenant 500nM de chaque amorce et 5µL de PowerUp SYBR® master mix (Thermo Fisher™) est ajouté dans chaque puits. La qPCR est réalisée sur l'appareil CFX384 Real time

system BioRad® avec le programme d'amplification suivant : une première étape de dénaturation 2min à 95°C, puis 40 cycles d'amplification avec 3s de dénaturation à 95°C, puis 30s d'élongation et d'hybridation à 65°C et une dernière étape d'élongation 5min à 65°C.

Les résultats obtenus à l'issue de la qPCR sont analysés avec le logiciel BioRad® CFX manager qui détermine le cycle seuil ou *threshold cycle* ( $C_T$ ), cycle pour lequel la quantité de produit amplifié atteint le seuil de fluorescence détectable. Les éventuelles différences d'efficacité d'amplification entre ALG9, UBC6, hENT1<sub>co</sub> et hsvTK<sub>co</sub> sont prises en compte par l'utilisation lors de la PCR d'un témoin possédant une unique copie du plasmide. Ainsi le nombre de copies du plasmide intégrées est déterminé par la formule suivante :

$$\Delta\text{échantillon} = \text{moyenne } C_T \text{ hENT1}_{co} \text{ échantillon} - \text{moyenne } C_T \text{ ALG9}_{\text{échantillon}}$$

$$\Delta\text{monocopie} = \text{moyenne } C_T \text{ hENT1}_{co} \text{ monocopie} - \text{moyenne } C_T \text{ ALG9}_{\text{monocopie}}$$

$$\text{nombre de copies} = 2^{-(\Delta\text{échantillon} - \Delta\text{monocopie})}$$

Le nombre de copies est déterminé pour les couples hENT1<sub>co</sub>-ALG9, hENT1<sub>co</sub>-UBC6, hsvTK<sub>co</sub>-ALG9, hsvTK<sub>co</sub>-UBC6 pour les deux concentrations d'ADN. On calcule ensuite la moyenne des huit déterminations du nombre de copies.

Régions à amplifier	Amorce 1	Amorce2
hENT1 <sub>co</sub>	GCT AGA CCT TGA CCA GAC ATG AT	AGG TTC TTT GTT TGG TTT GGC T
hsv-TK <sub>co</sub>	GTC ACC GTT TGG AGC CAA CA	AAC GCT GGT CCA AGA CCA CA
ALG9	TTA TTG CCT TCT GCC GTT GCC	ACC CAG TGG ACA GAT AGC GT
UBC6	CTA CAA AGC AGG CTC ACA AGA	GTT GGG GCG AGC AAG AAT AT
ADNr	ATGGTTTCAACGGGTAACGG	CTTCCTGGATGTGGTAGCC

**Tableau III-3 : Amorces pour la qPCR et RT-qPCR.**

### III.1.7 Mesure de l'expression des gènes d'incorporation du BrdU par RT-qPCR

L'ARN de 2 à 5 u.DO de culture de levures en phase exponentielle de croissance est extrait avec le kit Monarch® Total RNA Miniprep Kit (New England Biolabs®). A l'issue de l'extraction les ARN sont dosés par spectrophotométrie ( $A_{260nm}$ ) puis dilués à une concentration de 250ng/μL et stockés à -80°C. La qualité des ARN extraits est vérifiée au Fragment Analyzer™ avec le kit ARN. Seul les ARN montrant deux pics bien distincts d'ARN ribosomiques et un *RNA Quality Number* supérieur à 8 déterminé avec le logiciel ProSize sont conservés.

La *reverse transcription* est réalisée avec le kit superscriptIV (Invitrogen™) pour 1µg d'ARN, les amorces utilisées pour l'amplification de l'ADN complémentaire (ADNc) sont les *random hexamer primer* (Thermo Fisher™). Ensuite la qPCR est réalisée avec les même amorces que pour mesurer le nombre de copies, ainsi qu'un couple d'amorces supplémentaires dirigées contre l'ADNr afin d'avoir un témoin pour valider la qualité de la reverse transcription (Tableau III-3). Les ADNc sont déposés dans une plaque 384 puits dans les concentrations 1/1000 et 1/2000 en triplicats et la qPCR est réalisée avec le programme d'amplification précédent. Les puits possédant un C<sub>T</sub> inférieur à 35 sont exclus car le C<sub>T</sub> est considéré trop faible pour être fiable. De plus, lorsque la *reverse transcription* s'est bien déroulée, les C<sub>T</sub> correspondant à l'ADNr sont inférieurs à 17 pour la dilution 1/1000 et inférieurs à 19 pour la dilution 1/2000. Lorsque le C<sub>T</sub> de l'ADNr pour un échantillon est supérieur à ces valeurs l'échantillon est exclu de l'analyse car la *reverse transcription* n'a pas été assez efficace. L'expression est ensuite quantifiée à l'aide de la formule suivante :

$$\text{Expression } hENT1_{CO} = 2^{-(\text{moyenne } hENT1_{CO} - \text{moyenne } ALG9)}$$

L'expression pour chaque couple hENT1<sub>CO</sub>-ALG9, hENT1<sub>CO</sub>-UBC6, hsvTK<sub>CO</sub>-ALG9, hsvTK<sub>CO</sub>-UBC6 est déterminée pour chaque dilution et on calcule la moyenne des quatre mesures réalisées pour chaque gène.

### III.1.8 Test de croissance

Après une préculture d'une nuit, les cellules sont diluées à une DO<sub>600nm</sub> = 0,025. La DO<sub>600nm</sub> est ensuite mesurée toutes les deux heures pendant 8 à 10h. Les tests de croissance ont aussi été réalisés avec 10mM de BrdU ou un volume équivalent de diméthylsulfoxyde (DMSO), afin de suivre le ralentissement de croissance induit par l'incorporation de BrdU. Le nombre de doublements effectués par les cellules ainsi que leur temps de doublement sont estimés avec les formules suivantes :

$$\text{nombre de doublements} = \frac{\log(DO_{finale}) - \log(DO_{initiale})}{\log(2)}$$

$$\text{temps de doublement} = \frac{\text{temps de culture} \times \log(2)}{\log(DO_{finale}) - \log(DO_{initiale})}$$

## III.2 Analyse de la réplication de la BY4742 et de la SY14 par séquençage Nanopore

### III.2.1 Extraction d'ADN

L'extraction d'ADN est réalisée sur 20 à 100 u.DO avec le protocole décrit (Hennion et al., 2022). Pour le séquençage nanopore, l'ADN est repris dans du TE *low* EDTA (Tris 10mM pH 8,0 ; EDTA 0,1mM) et dosée au Qubit (Invitrogen™). Ce protocole a initialement été développé pour permettre l'isolation de grandes molécules d'ADN. J'utilise ce protocole pour isoler l'ADN pour la qPCR. Toutefois, la taille des fragments n'étant pas cruciale, les étapes de mélange sont réalisées au vortex afin d'augmenter le rendement. L'ADN est ensuite repris dans l'H<sub>2</sub>O et dosé par spectrophotométrie ( $A_{260nm}$ ).

### III.2.2 Séquençage Nanopore

Suivant la quantité de données nécessaire pour les analyses ultérieures, différents types de séquençages ont été réalisés. Le premier, permettant d'obtenir une faible quantité de données, a été effectué au laboratoire sur l'appareil MinION (Oxford Nanopore Technologies plc.), selon le protocole décrit par Hennion et al., (2022). Le second, permettant d'obtenir une grande quantité de données, a été effectué sur PromethION (Oxford Nanopore Technologies plc.), par nos collaborateurs au sein du Génoscope d'Evry (Tableau supplémentaire VIII-1)

### III.2.3 Analyse de la réplication par NFS (Theulot et al, 2022)

#### III.2.3.1 Marquage des fourches de réplication

Les fourches de réplication sont marquées sur une culture de levure asynchrone en phase exponentielle de croissance par l'ajout de 1mM de BrdU pendant 2min (*pulse*) suivi de l'ajout de 10mM de thymidine pendant 20min (chasse). Les cellules sont ensuite centrifugées 5min à 3100g et rincées avec de l'H<sub>2</sub>O. Cette étape est répétée une seconde fois avant l'extraction de l'ADN (voir Extraction d'ADN). Voir Détection des fourches de réplication pour une présentation détaillée des signaux obtenus après séquençage et de leur interprétation.

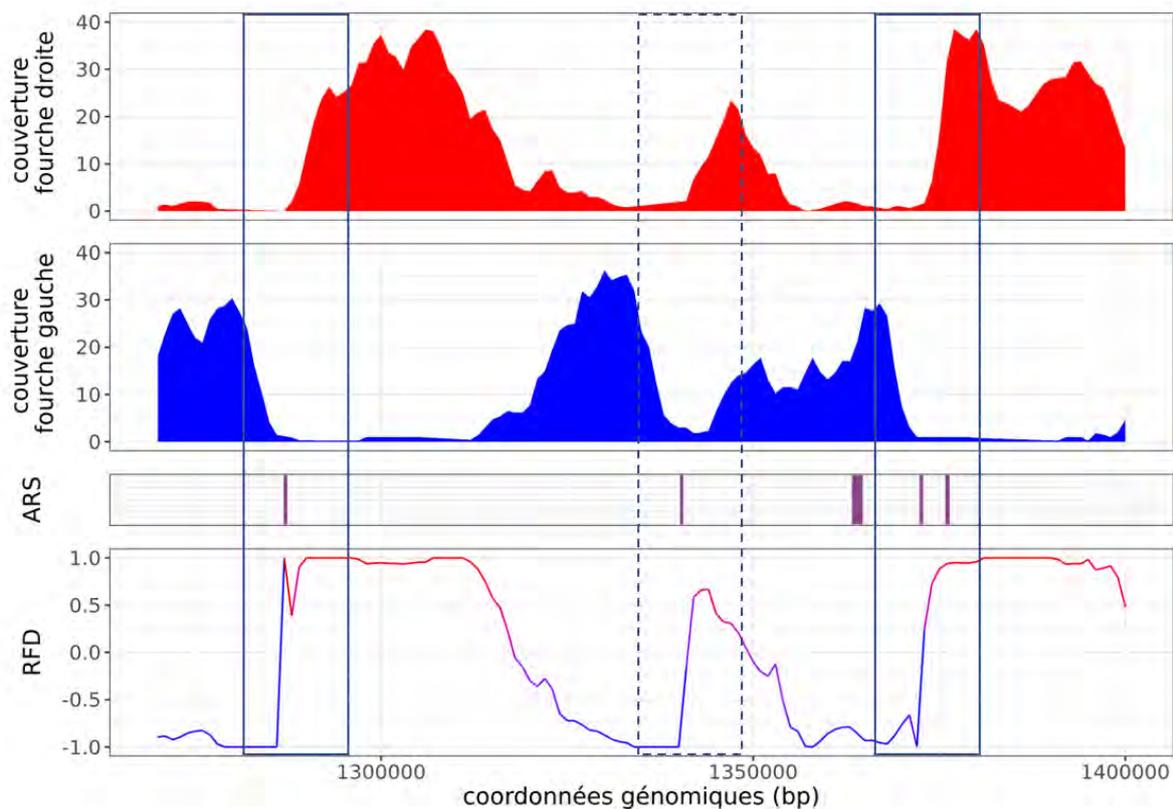
#### III.2.3.2 Profil de directionnalité des fourches de réplication (RFD)

La couverture en fourches gauches et droites n'est pas homogène sur le génome. Ainsi il existe des régions dépourvues de fourches droites, majoritairement répliquées vers la gauche, et des régions dépourvues de fourches gauches, majoritairement répliquées vers la droite (Figure III-2). Les changements de direction de réplication le long du génome indiquent la présence d'origines actives, autour desquelles les directions de réplication divergent, ou de sites de terminaison, autour desquels

elles convergent. Sur ce principe, il est possible de calculer la direction moyenne, ou directionnalité, des fourches de réplication (RFD) :

$$RFD = \frac{Droite - Gauche}{Droite + Gauche}$$

où "Droite" et "Gauche" indiquent le nombre de fois qu'un locus donné s'est répliqué vers la droite ou la gauche. Une RFD = +1 indique que la région est toujours répliquée vers la droite. Une RFD = -1 indique que la région est toujours répliquée vers la gauche. Les valeurs de RFD intermédiaires mesurent la prédominance relative des fourches droites ( $0 < RFD < 1$ ) ou gauches ( $-1 < RFD < 0$ ). Les zones d'initiation sont ainsi caractérisées par des pentes ascendantes du profil de RFD, qui correspondent à la localisation connue des origines chez la levure (Figure III-2). Les zones de terminaison, quant à elles, sont caractérisées par des pentes descendantes (Figure III-2). Le  $\Delta RFD$  ou différence entre la valeur de RFD au début de la pente et la valeur à la fin de la pente représente l'efficacité observée de la zone d'initiation ou de terminaison. Un  $\Delta RFD$  de -2 correspond à une origine qui dans 100% des cellules émet une fourche vers la droite (RFD=1) et dans 100% des cellules une fourche à gauche (RFD=-1) (Figure III-2, cadre bleu). Une valeur de  $\Delta RFD$  inférieure indique une moindre efficacité de l'origine (Figure III-2, cadre pointillé). Pour déterminer la RFD, la couverture en fourches droites et la couverture en fourches gauches ont été calculées au nucléotide près en utilisant la position du début (X0) et de la fin du *pulse* (X1, voir Détection des fourches de réplication, pour la définition des positions X0 et X1), puis la RFD a été calculé avec la formule  $\frac{\text{couverture droite} - \text{couverture gauche}}{\text{couverture totale}}$ . Pour diminuer les temps de calcul on utilise par la suite la moyenne de la RFD par fenêtre de 1kb.



**Figure III-2 : Exemple de profil RFD.**

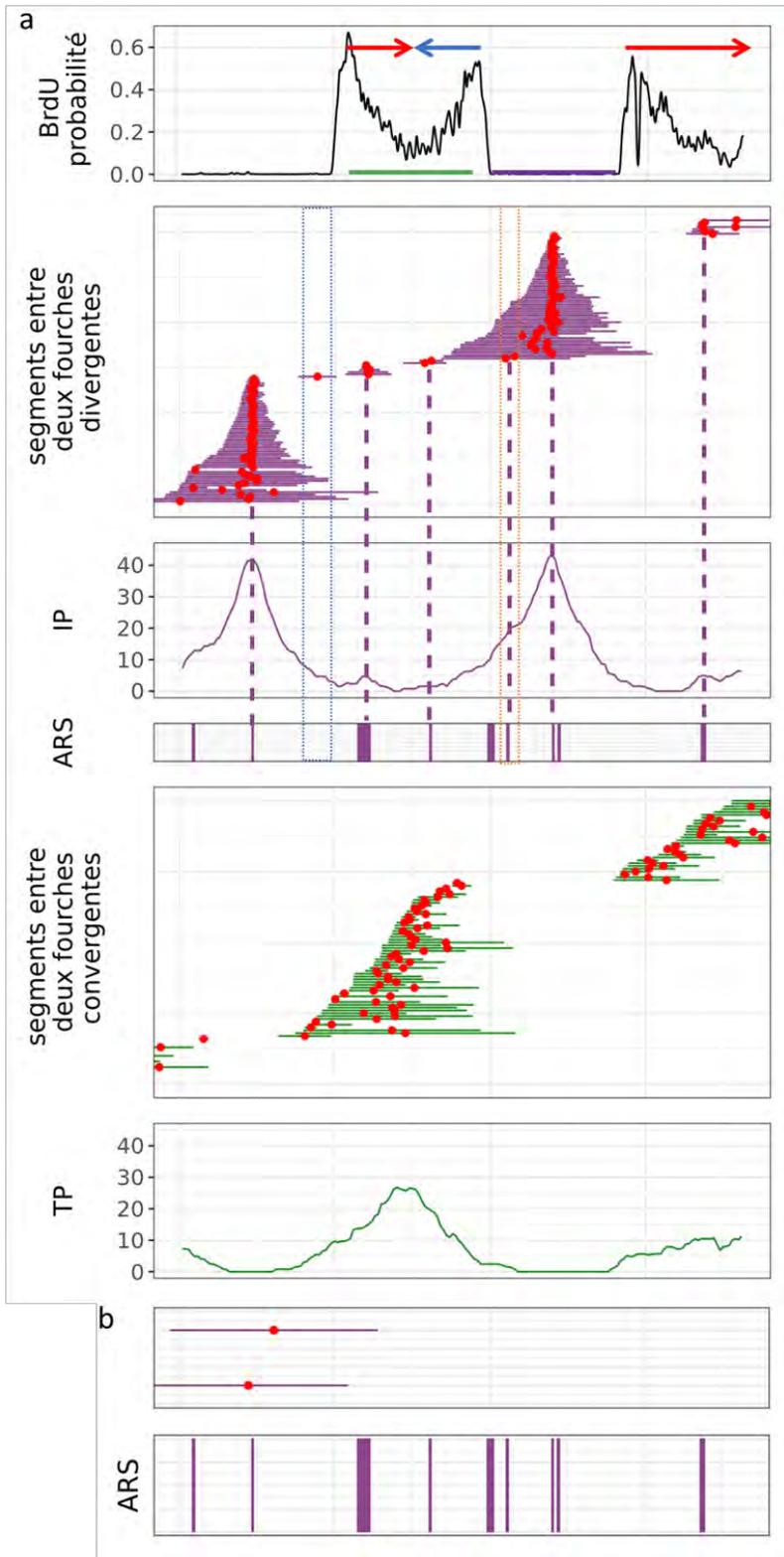
La couverture en fourche droite correspond à la couverture des segments X0-X1 des fourches droites. La couverture des fourches gauches correspond à la couverture des segments X0-X1 des fourches gauches. Les zones d'initiation sont encadrées par des cadres bleus, les deux zones d'initiation les plus efficaces montrent un  $\Delta$ RFD de -2. Entre ces deux zones d'initiation, on trouve une zone moins efficace encadrée en pointillés. L'analyse de la couverture révèle que cette zone d'initiation envoie de nombreuses fourches vers la gauche RFD=-1, mais elle envoie moins de fourches vers la droite RFD=0,5, sans doute parce que l'origine située à sa droite envoie plus souvent des fourches vers la gauche.

### III.2.3.3 Analyse des événements individuels d'initiation et de terminaison

Le séquençage nanopore permet de détecter des événements individuels d'initiation ou de terminaison lorsque la même lecture de séquence montre deux fourches divergentes ou convergentes (Figure III-3a). La position de l'évènement d'initiation peut alors être estimée comme le milieu du segment entre les X0 de deux fourches divergentes et la position de l'évènement de terminaison, comme le milieu entre les X1 de deux fourches convergentes, si l'on suppose que les deux fourches ont progressé à la même vitesse (voir Détection des fourches de réplication). Le travail de Hennion et al. (2020) démontre que les milieux des segments encadrés par deux fourches divergentes s'alignent précisément sur les origines de réplication annotées dans la base de données OriDB (Figure III-3a, point rouge, ligne pointillée violette). Parfois, il arrive que le milieu d'un segment ne corresponde pas à une

origine de réplication annotée. Hennion et al. (2020) ont montré que 9% des évènements d'initiation avaient lieu en dehors des localisations des origines connues (Figure III-3a, cadre pointillé bleu). Le milieu de certains segments qui chevauchent plusieurs ARS ne correspond pas à la localisation d'une ARS (Figure III-3b). Dans ce cas, il est possible que deux initiations aient pris place à l'intérieur de ces segments, et donc que les deux fourches qui ont formé le segment au moment du *pulse* proviennent de deux évènements d'initiation différents.

Une autre façon de représenter les données est de calculer la couverture entière des segments localisés entre fourches divergentes ou convergentes. Les pics de couverture représentent les localisations les plus probables des initiations et des terminaisons (Figure III-3a, IP, TP). Le nombre d'évènement d'initiation ou de terminaison détectés dans chaque expérience dépend de la taille des lectures et de leur nombre. Afin de pouvoir comparer les profils de couverture en segments compris entre deux fourches convergentes ou en segments compris entre deux fourches divergentes, nous avons normalisé ces données pour chaque expérience par le nombre total de segments détectés lors de l'expérience. Comme le nombre de copies de l'ADNr peut varier entre 100 et 200, nous avons exclu les segments qui chevauchaient l'ADNr. Nous définissons ainsi un effectif normalisé des segments entre fourches divergentes, que nous appelons IP pour *Initiation Probability*, car il refléterait la probabilité d'initiation à chaque locus si l'on ne faisait aucune hypothèse sur les vitesses relatives des fourches droites et gauches. A l'inverse, si l'on fait l'hypothèse que les deux fourches ont voyagé à la même vitesse et donc, que l'évènement d'initiation a pris place au centre du segment, la hauteur des maxima correspond à la probabilité d'initiation à cette position, mais les valeurs bordant chaque pic ne traduisent que la propagation des fourches à partir de ce pic. Néanmoins, une origine faiblement efficace située à proximité d'un pic fortement efficace n'apparaîtra que si l'on fait figurer les milieux des segments (Figure III-3a, cadre pointillé orange). À ces approximations près, la hauteur des pics des profils normalisés permet de comparer l'efficacité des initiations entre deux profils et au sein d'un profil. L'effectif normalisé des segments entre fourches divergentes, quant à lui, est appelé TP pour *Termination Probability*. Les mêmes limites d'interprétation que pour l'IP s'appliquent, d'autant plus que les évènements de terminaisons sont beaucoup plus dispersés que les évènements d'initiations.



### Figure III-3 : Exemple de profils IP et TP.

a) Lorsque deux fourches divergentes ou convergentes sont visibles sur une lecture, il est possible d'identifier respectivement un évènement unique d'initiation ou de terminaison. Les segments entre les X0 de deux fourches divergentes sont représentés par des lignes violettes. Les centres de ces segments sont indiqués par un point rouge. Ils s'alignent avec les ARS. Il est aussi possible d'observer un segment qui ne chevauche aucune ARS (cadre pointillé bleu). Ce segment représente une initiation dispersive Hennion et al. (2020) avaient identifié 246 initiations en dehors des ARS. Le profil IP permet de visualiser quatre initiations, deux fortement efficaces et deux faiblement efficaces, mais il ne permet pas de visualiser l'initiation dispersive (cadre pointillé bleu) et une initiation peu efficace à proximité d'une fortement efficace (cadre pointillé orange). Les segments entre les X0 de deux fourches convergentes sont représentés par des lignes vertes. La dispersion des segments et de leur centre est large ce qui montre que les terminaisons n'ont pas lieu en un site précis, contrairement aux initiations.

b) Exemple d'un segment qui contient plusieurs ARS et dont le centre ne s'aligne pas avec une ARS. Il est possible que les deux fourches divergentes de ce segment proviennent de deux origines de réplication différentes et qu'une terminaison ait eu lieu avant le *pulse*.

#### III.2.3.4 Analyse de la vitesse des fourches de réplication

La comparaison de la vitesse de progression des fourches sur le génome de la BY4742 et de la SY14 a été réalisée sur les vitesses des fourches dont le segment X0-X1 (voir Détection des fourches de réplication) ne contenait aucune partie d'aucun élément connu pour ralentir la vitesse de progression des fourches, c'est-à-dire l'ADNr, les ARN-t, les centromères et les télomères (Theulot et al., 2022). Après obtention des vitesses moyennes pour chaque réplicat, j'ai déterminé si la vitesse de progression des fourches était différente dans les deux souches par un test de contraste en modèle linéaire (Theulot et al., 2022).

Ensuite, j'ai comparé la vitesse de réplication de la BY4742 ou de la SY14 (sans les éléments précédents) à la vitesse de progression des fourches dont le segment X0-X1 contient tout ou partie d'un centromère ou d'un télomère, et pour la SY14 d'un centromère supprimé ou d'une fusion chromosomique par un test de contraste en modèle linéaire (Theulot et al., 2022). Les fourches qui traversaient l'ADNr ou un ARN-t ont été exclus de l'analyse. Selon la même procédure pour la SY14, la vitesse de progression des fourches traversant le centromère a été comparée à la vitesse de progression des fourches traversant les centromères supprimés, et la vitesse de progression des fourches croisant les télomères a été comparée à la vitesse de progression des fourches traversant les fusions chromosomiques.

### III.2.4 Analyse du moment de réplication ou *timing* par séquençage Nanopore : nanotiming

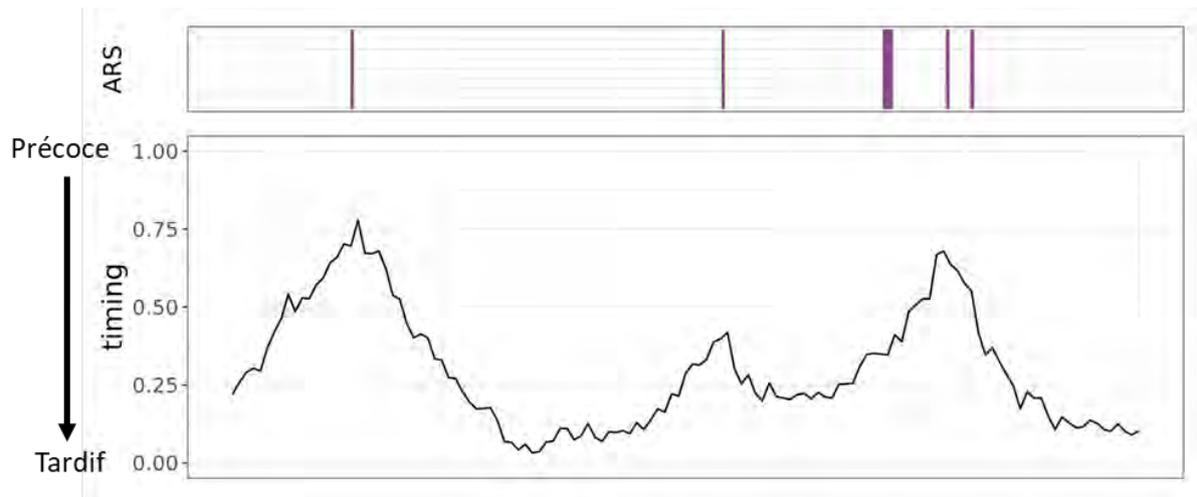
Le moment de réplication ou *timing* se caractérise par le moment d'activation d'une origine au cours de la phase S. Il est essentiel de connaître le *timing* de réplication pour comprendre les phénomènes de réplication passive d'origines. Une origine activée en début de phase S va répliquer passivement les origines voisines lorsqu'elles sont activées plus tardivement. Voir Mise au point de la méthode Nanotiming par Bertrand Theulot pour le principe de la méthode nanotiming.

#### III.2.4.1 Marquage d'un cycle cellulaire

Des cellules asynchrones en phase exponentielle de croissance sont marquées au BrdU pendant un cycle cellulaire par l'ajout d'une faible concentration 10 et 25 $\mu$ M pour la BY4742 et 25 $\mu$ M et 50 $\mu$ M pour la SY14. La  $DO_{600nm}$  est mesurée avant la réalisation du marquage, les cultures sont ensuite divisées en deux pour que les deux marquages soient effectués dans les mêmes conditions. Après division la  $DO_{600nm}$  est mesurée, puis après 1h30 toutes les 15min jusqu'à ce que la  $DO_{600nm}$  ait doublé. Les cultures sont alors immédiatement récupérées pour réaliser l'extraction d'ADN (voir Extraction d'ADN).

#### III.2.4.2 Obtention des profils de Nanotiming.

Les données de courants électriques issues du séquençage nanopore sont converties en séquence d'ADN (*basecaller*) à l'aide de *Megalodon-based BrdU basecaller*. La moyenne des données de probabilité de BrdU obtenues pour chaque thymidine, est calculée pour des fenêtres de 10pb dans un premier temps pour la conversion des données (voir Nanotiming). Les moyennes de BrdU sont ensuite calculées pour des fenêtres de 1kb. Les lectures possédant une médiane de BrdU inférieure à 0,02 sont exclues, car cette valeur correspond au niveau de bruit de fond dans un échantillon non marqué au BrdU (Theulot et al., 2022), ces lectures correspondent donc aux brins d'ADN ayant servi de matrice pour la réplication. De la même manière, les régions répliquées avant l'ajout de BrdU sont exclues en filtrant les fenêtres de 1kb possédant une valeur inférieure à 0,02. Le profil génomique est ensuite établi en réalisant la moyenne des différentes lectures par fenêtre de 1kb. Afin de faciliter l'analyse statistique des profils entre la BY4742 et la SY14, les données sont ensuite replacées sur une échelle comprise entre 0 et 1. Chaque origine de réplication correspond à un pic de *timing*. Plus le maximum du pic est proche de 1, plus l'origine se déclenche précocement (Figure III-4). Ensuite, les fourches de réplication émanant des origines vont répliquer les régions avoisinantes un peu plus tard, ainsi les zones de terminaisons correspondront à des vallées sur le profil de *timing* ( Figure III-4).



**Figure III-4 : Exemple de profil de *timing*.**

Les régions précoces possèdent un *timing* proche de 1, plus la valeur de *timing* décroît, plus la région est répliquée tardivement. Les pics de *timing* correspondent aux origines de réplication. La localisation de ces pics correspond à la localisation des origines connues chez la levure (ARS).

### III.2.5 Comparer des profils de réplication provenant de deux génomes différents

#### III.2.5.1 Alignement des génomes

La comparaison de la réplication entre la BY4742 et la SY14 ne peut se faire que sur des régions strictement similaires entre les deux souches. Afin de trouver ces régions, le génome de la SY14 (Shao et al., 2018) a été aligné sur le génome à l'aide de Burrows-Wheeler Aligner (BWA, Li and Durbin, 2010) avec les paramètres par défaut, selon la même procédure le génome de la BY4742 a été aligné sur le génome de la SY14. Cela a permis d'obtenir un premier jeu d'informations concernant les régions similaires présentes dans les deux génomes. Une comparaison visuelle des profils RFD pour les régions similaires a montré quelques légères différences sur certains chromosomes. Un second alignement a été réalisé à l'aide du logiciel Geneious et de l'algorithme d'alignement Minimap2 (Li, 2018). Ainsi, une seconde table de conversion a été créée. L'analyse visuelle des profils a montré une bonne concordance, excepté pour le chromosome XII au niveau du ADNr. Le ADNr possédant de nombreuses copies difficiles à aligner, les coordonnées du chromosome ont été ajustées à la main (Tableau supplémentaire VII-2).

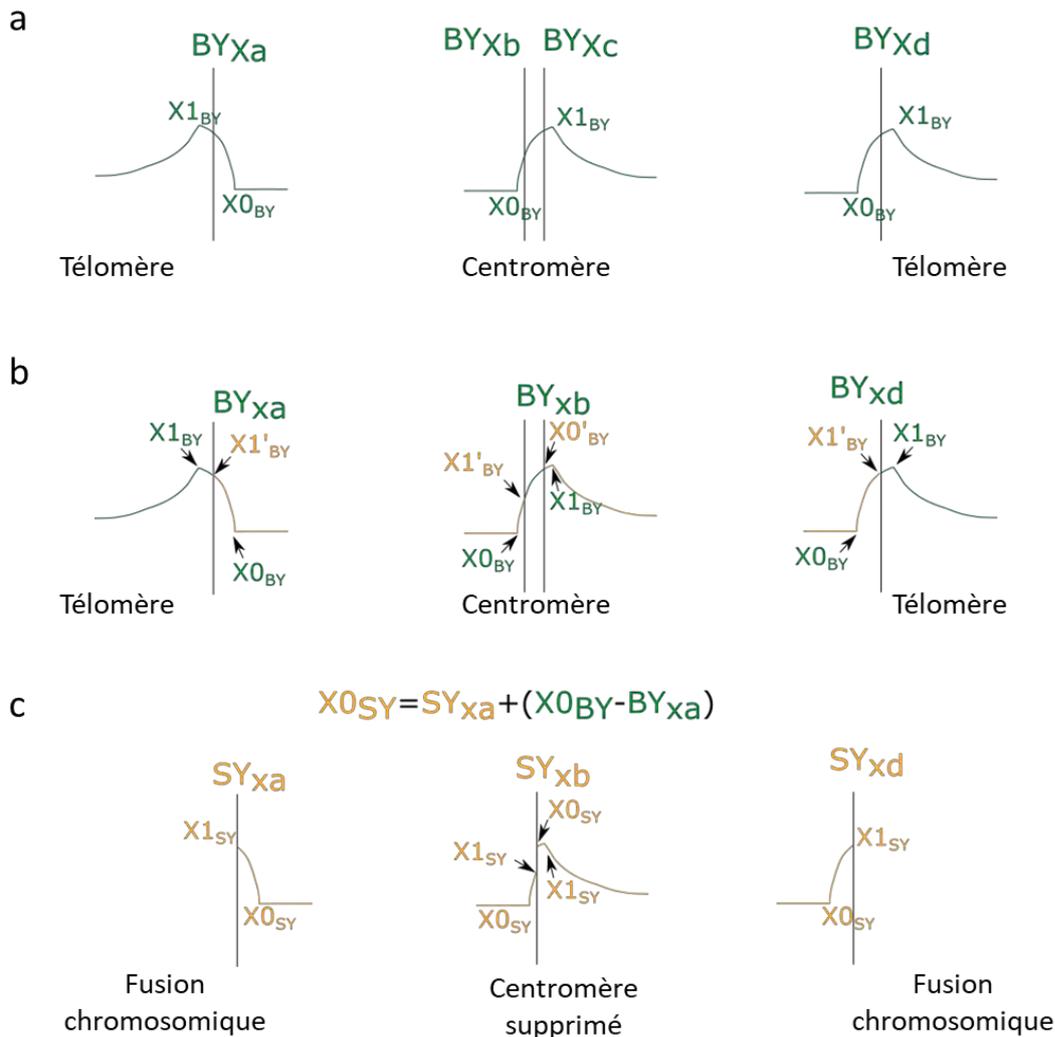
#### III.2.5.2 Conversion des informations

Après avoir obtenu les régions similaires entre les deux génomes, nous avons décidé, pour faciliter l'analyse, de convertir les coordonnées d'un génome vers l'autre. J'ai choisi de convertir les informations de la BY4742 vers la SY14 afin de visualiser plus facilement les fusions chromosomiques.

### III.2.5.2.1 Fourches

Pour convertir les coordonnées des fourches de répliation, elles sont dans un premier temps éventuellement tronquées, afin d'éliminer les paires de bases qui seraient absentes du génome de la SY14. Ainsi, un nouveau X0 ou X1 est attribué à la fourche qui correspond à la première base présente dans le génome de la SY14. Puis les coordonnées de X0 et X1 sont converties sur le génome de la SY14 en calculant la distance entre l'extrémité d'une région similaire dans la BY4742 et X0 ou X1, et en ajoutant cette distance à la coordonnée de l'extrémité de la région similaire dans la SY14 (Figure III-5). Le RFD est ensuite recalculée par la méthode décrite précédemment.

La comparaison de la vitesse de progression des fourches dans la SY14 et la BY4742 a été réalisée sans conversion des informations ni élimination des informations qui n'appartenaient pas aux deux génomes, c'est-à-dire en mesurant la vitesse de chaque fourche alignée sur son génome de référence.



### Figure III-5 : Stratégie de coupure des fourches.

a) Les fourches des échantillons BY4742 peuvent traverser les centromères ou les télomères qui sont supprimés dans la SY14. b) Seules les régions qui sont communes à la BY4742 et la SY14 sont conservées (régions en jaune). Cette étape induit la création d'un nouveau  $X0$ ,  $X0'_{BY}$  et/ou d'un nouveau  $X1$ ,  $X1'_{BY}$ . Les fourches qui traversent les centromères sont divisées en deux fourches, une avant le centromère et une après le centromère. c) Les coordonnées des fourches dans la BY4742 sont converties sur le génome de la SY14, grâce à la table de coordonnées obtenue après l'alignement des génomes. La table contient les positions des régions similaires  $BY_{Xn}$  et  $SY_{Xn}$ . La distance entre les coordonnées du génome de la BY et la position des régions similaires est calculée avant d'être ajoutée à la position de la région similaire dans la SY14, ainsi les fourches de la BY4742 possèdent de nouvelles coordonnées  $X1_{SY}$  et  $X0_{SY}$ .

#### III.2.5.2.2 Initiation et terminaison

Les segments entre deux fourches divergentes et les segments entre deux fourches convergentes subissent la même procédure que les fourches de réplication. Après suppression des paires de bases uniques à l'un ou l'autre génome, les coordonnées des segments sont converties sur le génome de la SY14 et les IP et TP sont recalculées comme précédemment.

#### III.2.5.2.3 Nanotiming

La conversion des données de *timing* est effectuée après l'étape de filtration des lectures non marquées et des fenêtres répliquées avant l'ajout de BrdU. Pour garder un maximum d'informations, les données de probabilité de BrdU calculées pour des fenêtres de 10pb, présentes sur les deux génomes, sont conservées. Les coordonnées sont ensuite converties sur le génome de la SY14 selon la même procédure que pour les données des fourches ou des IP et TP. Pour chaque lecture, la moyenne de BrdU est calculée de nouveau sur les fenêtres de 1kb présentes dans les deux génomes. Une nouvelle étape de filtration est réalisée pour retirer les fenêtres avec une moyenne de BrdU inférieure à 0,02. Ensuite, le profil génomique est établi comme précédemment.

#### III.2.5.3 Annotations des génomes

Les génomes de référence de la SY14 et de la BY4742 ne sont pas annotés. Les coordonnées des centromères, des télomères, des ARN-t et des gènes dont l'expression est modifiée dans la SY14 sont donc manquantes. Pour définir les coordonnées de ces divers éléments, leur séquence a été téléchargée sur la plateforme *Saccharomyces Genome Database* (SGD). Les séquences des centromères et des télomères ont ensuite été alignées sur le génome de la BY4742 et de la SY14 (uniquement pour le centromère et les deux télomères restants) à l'aide de BWA avec les paramètres par défaut. Les positions des centromères supprimés dans la SY14 ont été déterminées en

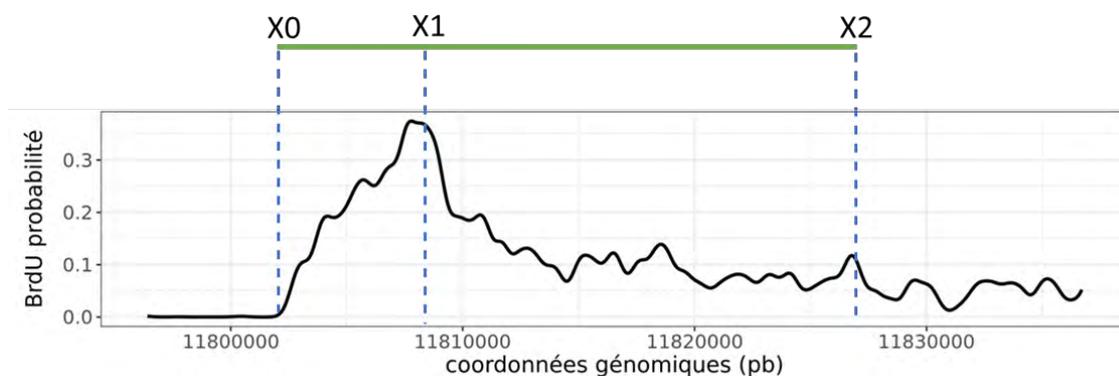
convertissant la coordonnée du début de chaque centromère dans le génome de la BY4742 sur le génome de la SY14. Les fusions chromosomiques sont définies par les coordonnées de l'extrémité des chromosomes de la BY4742 s'alignant sur le génome de la SY14. Les séquences des gènes modifiés ainsi que des ARN-t ont été alignées sur les deux génomes.

Les séquences des ARS ont été téléchargées depuis la plateforme OriDB (Siow et al., 2012) et ont été alignées sur le génome de la BY4742 à l'aide de BWA (paramètres -w 2000 -r 2.5). Les coordonnées des ARS pour la SY14 ont ensuite été obtenues par conversion des coordonnées sur le génome de la BY4742 vers le génome de la SY14. Cette procédure permet d'éviter l'alignement des ARS sur un chromosome qui n'est pas leur chromosome d'appartenance et de vérifier quelles ARS sont dans des régions supprimées dans la SY14. La suppression du CEN2 dans la SY14 a entraîné la suppression de l'ARS208, cette origine est donc exclue de la suite de l'analyse.

## IV. Méthodes d'analyse

### IV.1 Détection des fourches de réplication

Le défi principal de l'étude de la réplication par séquençage Nanopore est la détection des fourches de réplication. A l'issue du séquençage, nous obtenons des données de courant électrique qui sont ensuite converties en séquences d'ADN. Afin de déterminer la position des BrdU, un logiciel approprié (*Megalodon-based BrdU basecaller*) assigne à chaque site thymidine (T) une certaine probabilité que le site soit occupé par un BrdU (B). L'ajout de BrdU lors du *pulse* provoque une entrée de BrdU dans la cellule et son incorporation dans l'ADN pendant l'avancée de la fourche. Le taux de BrdU visible sur les lectures nanopore augmente progressivement, car l'équilibre des concentrations intra- et extracellulaires de BrdU n'est pas instantané. La chasse par la thymidine provoque ensuite une diminution plus lente du signal jusqu'à un niveau d'équilibre entre BrdU et thymidine. Ainsi nous pouvons définir la position de la fourche au début du pulse, X0, au début de la chasse, X1, et à l'équilibre, X2 (Figure IV-1). Notons que l'équilibre est atteint au bout d'un temps nettement inférieur à la durée totale de la chasse (20 min). La position de X2 est donc moins précisément définie que celle de X0 ou X1.



**Figure IV-1 : Incorporation de BrdU lors du déplacement d'une fourche de réplication.**

Le signal présenté ici correspond au signal visible après séquençage nanopore, basecalling et lissage. X0 est la position où la fourche se trouvait au début du pulse, X1 au début de la chasse, X2 à l'atteinte de l'équilibre des concentrations de thymidine et de BrdU. Les pentes des signaux reflètent la cinétique selon laquelle les concentrations intracellulaires de BrdU et de thymidine s'équilibrent avec leurs concentrations extracellulaires.

L'enjeu est de détecter le maximum de fourches avec la plus grande précision, mais l'importance de la précision sur les trois paramètres X0, X1, X2 dépend des informations que l'on souhaite en tirer. Pour déterminer la RFD, on peut utiliser l'information de direction de la fourche entre X0 et X1 ou entre X0 et X2. Un manque de précision sur la position de ces valeurs n'aura que peu d'impact sur la RFD

tant que la direction est déterminée avec exactitude. Pour déterminer la vitesse de progression des fourches, X0 et X1 doivent être déterminés avec précision. En effet, une erreur de 1kb sur la longueur du segment X0-X1, synthétisé en 2 min, entraînerait une erreur de vitesse de 500pb/min, soit 25% de la vitesse moyenne si on considère que les fourches voyagent à 2kb/min (Theulot et al, 2022). La position de X2 ne peut pas être exploitée pour une mesure de vitesse, car le temps auquel cette position est atteinte n'est pas connu. Pour déterminer la position d'événements uniques d'initiation ou de terminaison, on détermine le centre du segment s'étendant entre les X0 de deux fourches divergentes ou les X1 de deux fourches convergentes. Une erreur de 1kb sur un des X0 (ou X1) entraînera une erreur de 0,5kb sur la position de l'évènement d'initiation (ou de terminaison). Quatre méthodes de détection des fourches ont été développées au laboratoire :

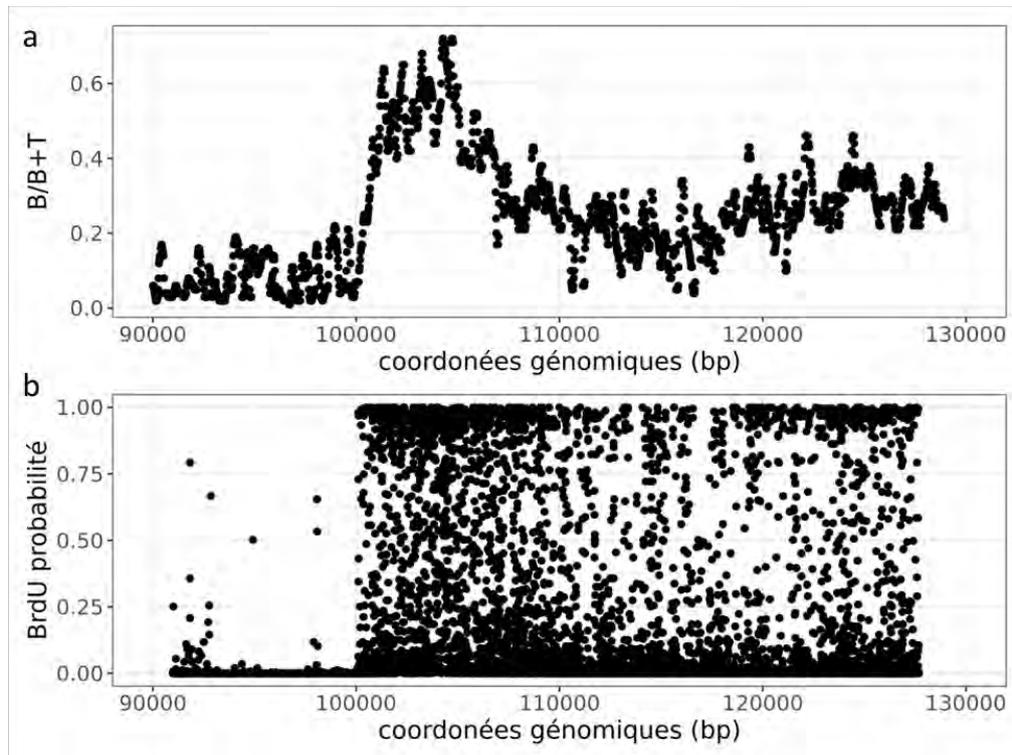
- Fork-seq (Hennion et al., 2020)
- Forkfit
- NFS (Theulot et al., 2022)
- Fork-ML

#### IV.1.1 Fork-seq

L'algorithme de détection des fourches de réplication contenue dans Fork-seq repose sur la segmentation en morceaux linéaires du signal BrdU par un algorithme Ramer-Dougllass-Pucker. Ce processus a été développé par Magali Hennion post-doctorante au laboratoire, pour l'algorithme de *basecalling* RepNano (Hennion et al., 2020). Par la suite, un nouvel algorithme a été développé : *Megalodon-based BrdU basecaller* (Theulot et al., 2022). Il possède plusieurs avantages : (i) il est plus facile et rapide d'utilisation car les données de courant électrique obtenues après séquençage nanopore peuvent être utilisées seules tandis que RepNano requiert les données de courant électrique, mais aussi les données de séquences d'ADN obtenues après une première étape de *basecalling*. (ii) *Megalodon-based BrdU basecaller* permet une détection plus précise du BrdU (Theulot et al, 2022).

Les données à la sortie des deux *basecallers* sont différentes. RepNano indique un ratio  $\frac{BrdU}{BrdU+Thymidine}$  en fonction des coordonnées génomiques tandis que *Megalodon-based BrdU basecaller* indique la probabilité que la thymidine séquencée soit un BrdU (Figure IV-2). Pour utiliser l'algorithme de détection des fourches de Fork-seq les données de *Megalodon-based BrdU basecaller* doivent donc être adaptées afin que le signal ressemble à celui généré par RepNano (Figure IV-2), ou alors l'algorithme de détection des fourches doit être adapté pour détecter des fourches sur ce

nouveau signal. Malheureusement, cet algorithme était rédigé en python et après le départ de Magali Hennion, personne n'était à l'aise avec ce langage de programmation, Laurent Lacroix a donc décidé de créer un algorithme similaire avec R.



**Figure IV-2 : Exemple de lectures nanopore *basecaller* par RepNano ou *Megalodon-based BrdU basecaller*.**

a) RepNano indique une proportion de BrdU par fenêtre de 10 nucléotides, ici la moyenne de cette proportion est représentée par fenêtres de 20 nucléotides. b) *Megalodon-based BrdU basecaller* indique la probabilité que la thymidine séquencée soit un BrdU.

#### IV.1.2 NFS Nanoforkspeek (Theulot et al; 2022)

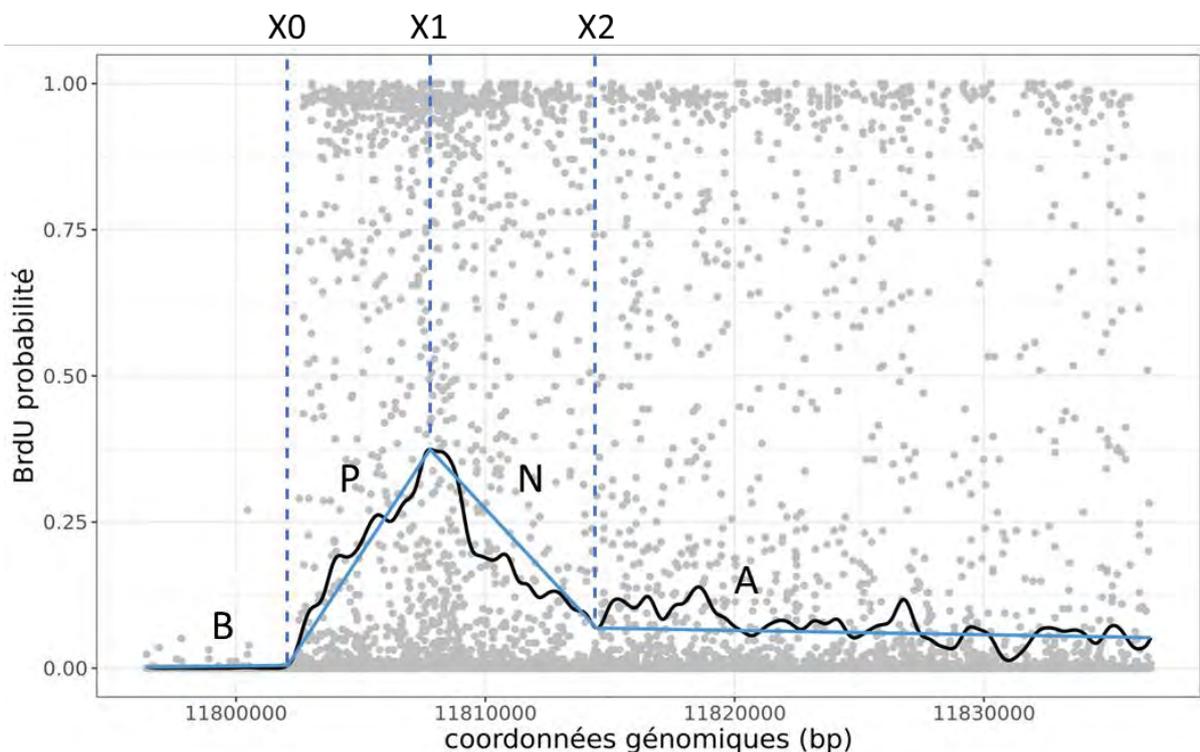
Laurent Lacroix a développé NanoForkSpeed (NFS) sur la base de Fork-seq, c'est-à-dire en utilisant un algorithme de Ramer-Dougllass-Pucker afin de réaliser la segmentation en morceaux linéaires du signal *après basecalling*. Cette méthode repose sur une première étape de lissage qui consiste à réaliser la moyenne de BrdU sur une fenêtre glissante de 100pb, puis une moyenne avec un poids gaussien sur 2,5kb (Figure IV-3, ligne noire). Ce lissage permet d'obtenir un signal analysable par l'algorithme (Figure IV-3, ligne bleue). A l'issue de la segmentation linéaire du signal, les segments sont classés en 4 catégories :

- B : segment plat avec un niveau de BrdU similaire au bruit de fond
- A : segment plat avec un niveau de BrdU supérieur au bruit de fond

- P : segment avec une pente ascendante
- N : segment avec une pente descendante

Les fourches sont ensuite détectées par la recherche d'expression régulière "BP(P|A)\*N+" pour les fourches droites, en excluant les fourches où le segment suivant est un B. Pour les fourches gauches, l'expression est "P+(N|A)\*NB" et les fourches précédées par un segment B sont exclues (Figure IV-3). L'exclusion des segments B évite la détection de fourches qui se chevauchent, ainsi que d'initiations et de terminaisons durant le *pulse*. Pour la fourche représentée sur la figure IV-3, les positions de X0 et X1 sont déduites de la position du segment P. Connaissant la distance entre X0 et X1 et la durée du marquage, la vitesse de progression des fourches est donnée par la formule suivante :

$$vitesse = \frac{X1 - X0}{durée\ du\ pulse\ (2min)}$$



**Figure IV-3 : Exemple de segmentation linéaire d'une fourche droite.**

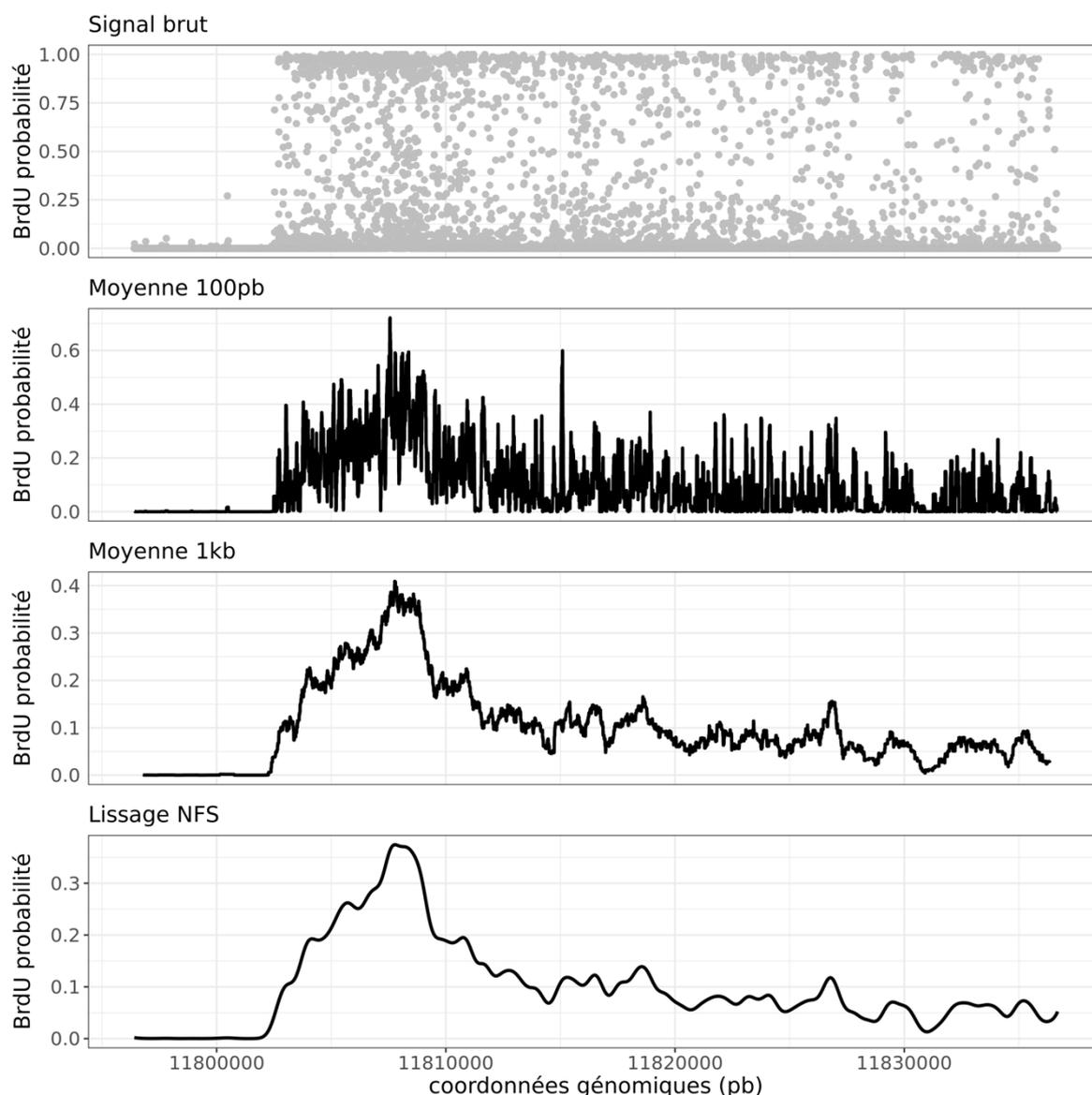
Les points gris représentent les données de probabilité de BrdU obtenues à l'issue du *basecalling*. Le signal noir correspond au même signal après lissage, et le signal bleu, à la segmentation du signal noir par l'algorithme de Ramer-Dougllass-Pucker. Les lettres indiquent le type de segment.

### IV.1.3 Forkfit

Étienne Jean, Ingénieur au Laboratoire, a développé Forkfit pour déterminer la vitesse de progression des fourches de réplication. Cette méthode est basée sur un modèle mathématique de l'incorporation de BrdU pendant le *pulse* et la chasse. Ainsi l'algorithme teste différents paramètres du modèle, dont la vitesse de la fourche, afin d'avoir le meilleur ajustement du modèle au signal expérimental. Ceci permet d'obtenir la vitesse de progression de la fourche, X0, X1 et X2. L'utilisation du modèle permet aussi d'extrapoler la position de ces points sur des signaux tronqués, ce qui permet d'utiliser davantage de signaux. Cependant, des expériences-contrôles ont montré que la vitesse des fourches déduite de Forkfit avait tendance à augmenter avec la dose de BrdU utilisée lors du *pulse*, phénomène qui n'était pas observé par NFS, ni lorsque la vitesse était déterminée en positionnant manuellement X0 et X1. Cette augmentation semblait donc refléter un biais du modèle Forkfit, qui nous a fait renoncer à son utilisation.

### IV.1.4 Fork-ML

Au cours du développement de Fork-fit, Étienne Jean a développé un algorithme d'apprentissage automatique pour la détection des fourches de réplication. Il est basé sur une segmentation sémantique du signal en trois catégories : fourche droite, fourche gauche, absence de fourche. Le réseau est entraîné à partir de lectures avec des fourches manuellement annotées. L'annotation manuelle permet d'entraîner le réseau à détecter des fourches qui ne sont pas situées entièrement sur la lecture (avec un X0, X1 ou X2 manquant), ou encore des initiations ou des terminaisons durant le *pulse*. Lorsque Étienne Jean a quitté le laboratoire le réseau avait été entraîné avec les données de RepNano (Hennion et al, 2019). Pour utiliser Fork-ML avec *Megalodon-based BrdU basecaller*, il était donc nécessaire de ré-entraîner le réseau. Dans un premier temps, il fallait déterminer comment utiliser l'information de *Megalodon-based BrdU basecaller*, car il est difficile de distinguer les fourches avec le nuage de points de la probabilité de BrdU. Nous avons testé deux options : (i) utiliser un signal moyen sur des fenêtres de 100pb ou 1kb, coulissant sur 10pb (ii) utiliser le même lissage que pour NFS (Figure IV-4). Le signal moyen sur des fenêtres coulissantes de 100pb possède trop de bruit pour bien distinguer à l'œil X0, X1 et X2. Le signal moyen sur des fenêtres de 1kb coulissantes sur 10pb ou lissé selon la même méthode que NFS est assez similaire. Nous avons choisi d'utiliser le même signal que NFS afin de pouvoir comparer les deux méthodes de détection.

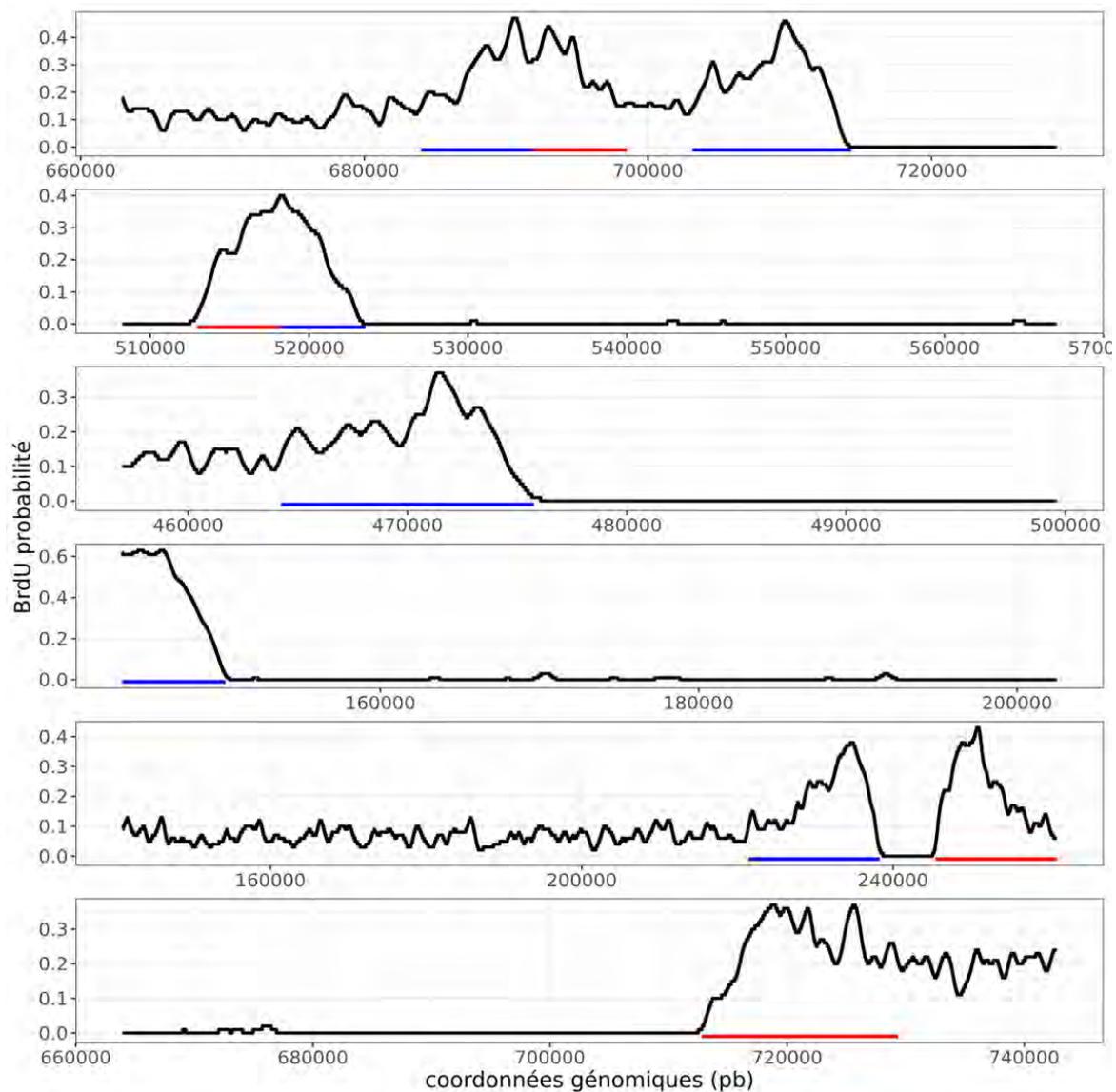


**Figure IV-4 : Exemple de différentes stratégies de lissage du signal.**

Le signal brut est obtenu à l'issue du *basecalling* par *Megalodon-based BrdU basecaller*. La moyenne 100pb a été réalisée en calculant la moyenne du signal brut sur des fenêtres de 100pb, coulissant d'un pas de 10pb. La moyenne 1kb a été réalisée de la même manière que la moyenne 100pb, en augmentant la taille de la fenêtre à 1kb. Le lissage NFS correspond à une moyenne sur une fenêtre glissante de 100pb, suivie d'une moyenne avec un poids gaussien sur 2,5kb.

Nous voulions utiliser Fork-ML plutôt que NFS pour réaliser les profils RFD et détecter les segments porteurs d'un évènement individuel d'initiation ou de terminaison, car cette méthode permet de détecter un plus grand nombre de fourches. Pour maximiser le nombre de fourches détectées, l'annotation manuelle de la direction de réplication a été effectuée même si X0 ou X1 ou X2 n'étaient pas visibles sur la lecture. Les évènements d'initiation ou de terminaison durant le *pulse*

ont aussi été annotés (Figure IV-5 panneau 4). Cette méthode permet d'exploiter un plus grand nombre de signaux, mais des biais d'annotation manuelle sont possibles. Par exemple, si la lecture s'arrête avant la fin de la descente du signal (Figure IV-5, panneau 3 et 6), plusieurs interprétations sont possibles : une fourche unique ou un évènement de terminaison dont la seconde fourche est incomplète. C'est pourquoi nous avons établi certaines règles pour l'annotation. Pour les initiations ou les terminaisons durant le *pulse*, le début et la fin des fourches sont respectivement placées au centre de l'évènement (Figure IV-5, panneau 1 et 2). Pour les initiations ou les terminaisons qui comportent un segment plat entre les deux fourches qui est différent du signal avant le *pulse*, ce segment est exclu des fourches (Figure IV-5, panneau 1). Pour l'annotation des fourches en bordure de lecture, lorsque le signal ne retombe pas au niveau du bruit de fond d'après *pulse*, la fourche s'arrête à mi-distance du début du signal et de l'extrémité de la lecture (Figure IV-5, panneau 3 et 6). Le système d'annotation peut parfois manquer de précision sur les longues lectures. Il peut donc y avoir un petit décalage entre le vrai début du *pulse* et le début de l'annotation. Parfois, pour des signaux tronqués dans leur partie ascendante, nous décidons de prolonger l'annotation jusqu'à la fin de la lecture (Figure IV-5, panneau 4). 12 000 lectures provenant de la SY14, la BY4742 et la BT1 ont été ainsi annotées pour l'entraînement du réseau. La détection effectuée par Fork-ML a ensuite été vérifiée visuellement sur 100 lectures. Sur ces 100 lectures, Fork-ML détecte la totalité des fourches dans 42 lectures. Ces 100 lectures représentent 216 fourches, dont 137 sont détectées par Fork-ML. Les autres fourches sont soit mal détectées (33), soit non détectées (30).



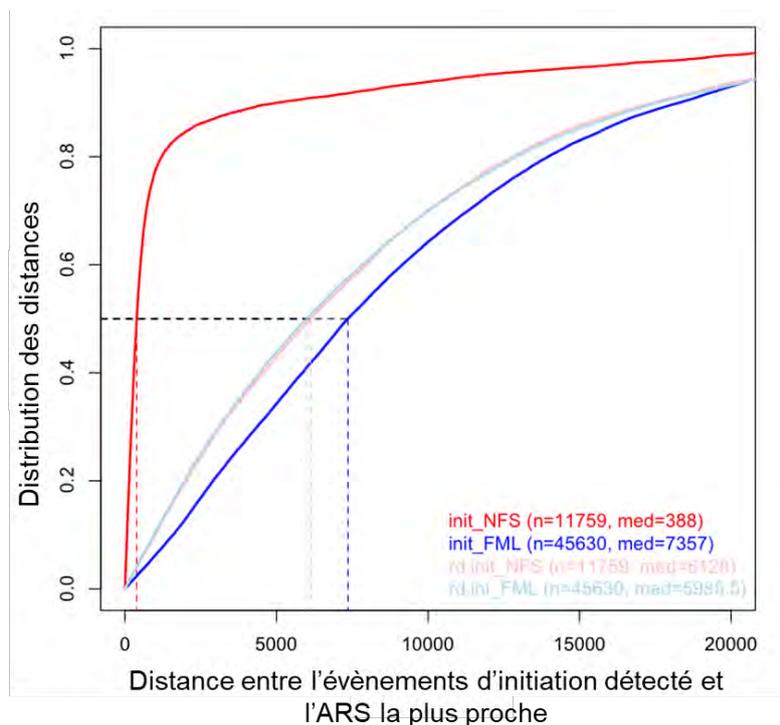
**Figure IV-5 : Exemple d'annotation des lectures pour l'entraînement de fork-ML.**

Les lignes rouges représentent les fourches droites et les lignes bleues les fourches gauches.

#### IV.1.5 Comparaison de Fork-ML et de NFS

Fork-seq et Forkfit ne pouvant être utilisés, nous avons à notre disposition Fork-ML et NFS pour détecter les fourches de réplication. La comparaison de la réplication entre la BY4742 et la SY14 requiert principalement la réalisation d'un profil RFD afin d'avoir une image globale de la réplication, nous avons donc besoin de détecter un maximum de fourches avec une bonne précision sur leur direction. Fork-ML semblait un bon candidat, puisqu'il est conçu pour détecter les fourches tronquées, ainsi que les initiations et les terminaisons durant le *pulse*. Ainsi, sur un séquençage promethION, Fork-ML détecte 105 953 fourches contre 50 260 pour NFS. Nous avons alors comparé la détection Fork-ML et NFS sur les 100 lectures ayant servi à évaluer la qualité de la détection par Fork-ML. NFS détecte

117 fourches lorsque Fork-ML en détecte 137, mais commet seulement 3 erreurs de détection lorsque Fork-ML en commet 33. Afin de déterminer l'impact des erreurs réalisées par Fork-ML sur la qualité globale de la détection, nous avons analysé la distance moyenne entre les événements unique d'initiation visible sur les lectures nanopore et les coordonnées des ARS. La distance médiane aux ARS pour NFS est de 338bp, alors qu'elle est de 7357pb pour Fork-ML, ce qui signifie que les erreurs réalisées par Fork-ML déterminent fortement la qualité de la détection (Figure IV-6). La majorité des erreurs de Fork-ML proviennent d'une sur-détection des fourches : une fourche est détectée à plusieurs reprises en lui assignant des directions différentes. Nous avons donc choisi NFS pour détecter les fourches de réplication dans la BY4742 et la SY14.



**Figure IV-6 : Fonction de distribution empirique de la distance entre les évènements d'initiation détectés par NFS ou Fork-ML.**

Les distances entre les initiations détectées par NFS (init\_NFS) ou Fork-ML (init\_FML) et l'ARS la plus proche ont été calculées. Pour contrôle, les initiations détectées par ces deux méthodes ont été redistribuées aléatoirement sur le génome (rd.init\_NFS et rd.init\_FML) et les distances entre l'ARS la plus proche et les origines redistribuées ont été calculées.

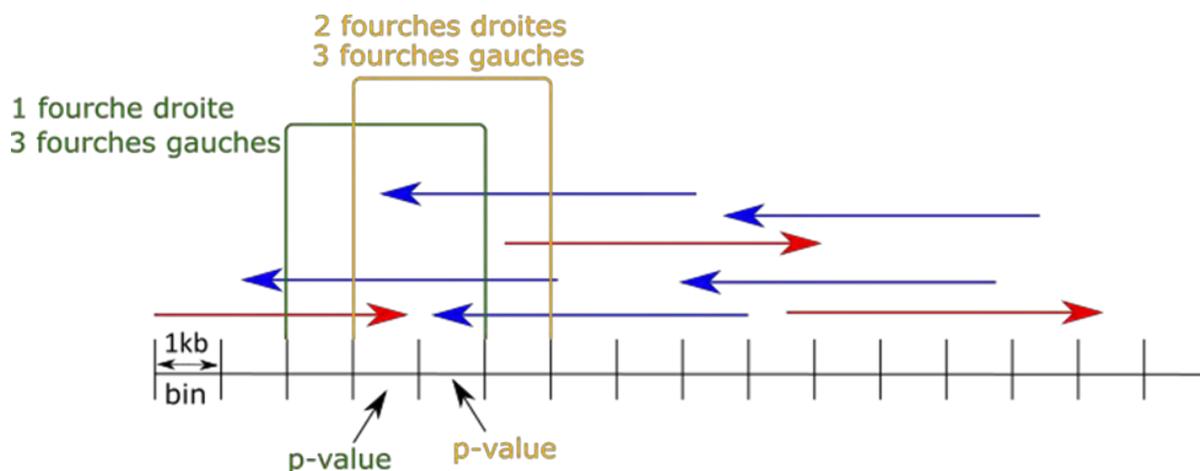
## IV.2 Analyse des profils de réplication

### IV.2.1 RFD

Une première analyse visuelle des RFD a montré des modifications importantes au niveau des centromères et des fusions chromosomiques. Il s'agit d'un résultat attendu, car ces régions ont été modifiées lors de la construction de la SY14. En revanche, il est aussi possible de distinguer des modifications d'efficacité des origines dans d'autres régions, quoique l'amplitude des effets soit moindre. Nous souhaitons donc mettre en place une stratégie d'analyse robuste pour nous permettre de distinguer les différences de RFD entre la BY4742 et la SY14 qui sont biologiquement significatives et celles qui sont liée à des fluctuations expérimentales. Détecter les différences entre les profils RFD de la BY4742 et de la SY14 revient à déterminer si, dans une région définie, les proportions de fourches droites et gauches sont significativement différentes.

Les individus d'un test statistique doivent être choisis de manière aléatoire et indépendante. Deux fourches provenant de deux lectures différentes issues d'ADN d'une même culture proviennent, sauf hasard difficilement plausible, de deux cellules différentes, et sont donc considérées comme indépendantes. En revanche, deux fourches présentes sur la même lecture proviennent soit de la même origine, auquel cas elles ne sont pas indépendantes, soit de deux origines voisines qui se déclenchent peut-être indépendamment l'une de l'autre, mais proviennent néanmoins de la même cellule. Toutefois, le nombre de lectures possédant plusieurs fourches étant faible, nous avons décidé de considérer toutes les fourches comme indépendantes.

Les origines étant espacées de 40 kb en moyenne, on s'attend à ce qu'un changement d'efficacité d'une origine se traduise par un changement de RFD sur une distance comparable. Afin de distinguer de petites différences de RFD ponctuelles qui se maintiennent à grande distance, nous avons effectué les tests statistiques sur des fenêtres de 3kb, 5kb, 11kb, 21kb et 41kb. Pour éviter un biais dû au choix de la position des fenêtres, nous avons utilisé des fenêtres glissantes chevauchantes, avec un pas de 1kb. La valeur de *p-value* du test est ensuite assignée à la fenêtre de 1kb (*bin*) située au centre de la fenêtre plus grande. Pour réaliser les tests, nous avons dénombré les segments X0-X1 gauches et droits présents dans chaque fenêtre, sachant qu'un segment est compté dans une fenêtre dès qu'il la chevauche sur plus de 500pb (Figure IV-7).



**Figure IV-7 : schéma représentant la stratégie de comptage des fourches de réplication.**

Les flèches rouges représentent les fourches droites, les flèches bleues représentent les fourches gauches. Le cadre vert représente la première fenêtre de comptage de 3 kb qui contient une fourche droite, comptée dans deux bins, et trois fourches gauches : une comptée dans 3 bins, une autre dans deux bins et une dernière dans 1 bin. Chacune de ces fourches sera comptée une fois et une seule dans la fenêtre de 3 kb. La *p-value* en vert du test effectué sur cette première fenêtre est assignée à la bin centrale. La seconde fenêtre de comptage est représentée en jaune, elle est déplacée de 1 bin vers la droite et contient deux fourches droites chacune présente sur une bin. Elle contient aussi trois fourches gauches, deux présentes sur 3 bins et une présente sur 1 bin. Encore une fois, la *p-value* en jaune du test effectué sur cette fenêtre est assignée à la bin centrale.

Nous avons essayé différentes approches statistiques : (i) le test de Student, (ii) le test du Chi<sup>2</sup>. Usuellement, le test permettant de comparer des proportions est le test du Chi<sup>2</sup>. Dans notre cas, il ne pouvait être réalisé qu'en comparant les échantillons deux à deux. Cela ne nous permettait pas de tenir compte de la variabilité entre les réplicats, ce qui ne semblait pas optimale. Le test de Student, quant à lui, permet de comparer les 6 échantillons en un seul test, et tient compte de la variabilité entre les réplicats, mais il est habituellement utilisé pour comparer des moyennes.

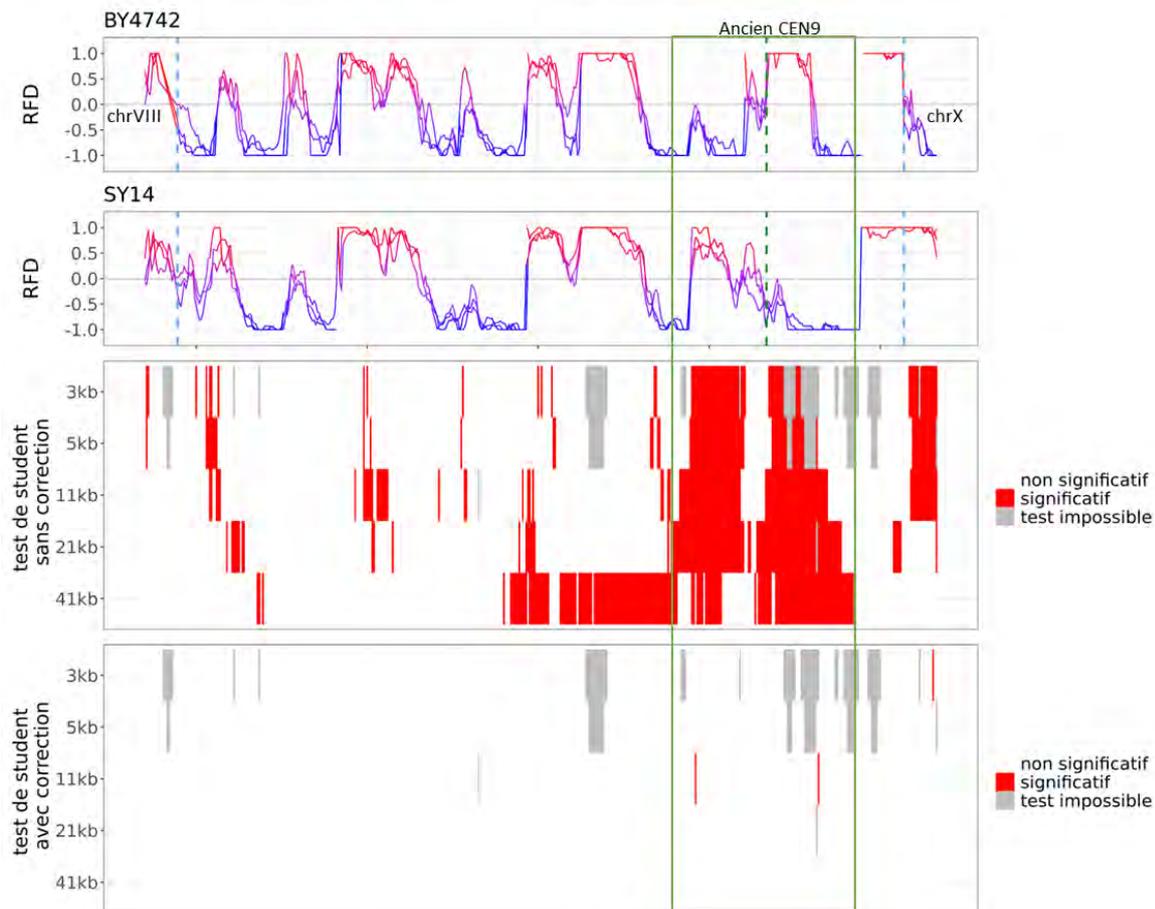
#### IV.2.1.1 Test de Student

Le test est réalisé sur la proportion de segment X0-X1 droit  $\left(\frac{\text{nombre de segment droit}}{\text{nombre de segment totale}}\right)$  pour chaque fenêtre coulissante avec la correction de Welsh, qui permet de réaliser le test lorsque la variance des deux populations est différente. Dans notre cas, nous avons deux populations : une population de proportion de segment X0-X1 droit dans la BY4742, dont nous étudions un échantillon avec trois représentants, et une population de proportion de segments droits dans la SY14, dont nous étudions aussi un échantillon avec trois représentants. Comme nous ne connaissons pas les populations, nous ne pouvons pas déterminer si la variance des deux populations est égale. De plus, le test ne peut pas être réalisé lorsque la variance de l'échantillon est nulle, c'est-à-dire lorsque la proportion de fourches

droites est la même dans tous les réplicats d'une souche. Ce phénomène se produit essentiellement lorsque la couverture est faible, mais est aussi d'autant plus fréquent que la proportion est proche de 0% ou de 100%.

Le résultat d'un test statistique s'interprète grâce à la *p-value* et en ayant choisi un seuil de significativité  $\alpha$ . Lorsque la *p-value* est supérieure au seuil, l'hypothèse  $H_0$  est acceptée avec un risque de se tromper égal à  $\alpha$ . Pour notre test, cela signifie que la proportion de fourche droite dans la SY14 est similaire à la proportion de fourche droite dans la BY4742. Lorsque la *p-value* est inférieure ou égale à  $\alpha$ , l'hypothèse  $H_1$  est acceptée avec un risque de se tromper égal à  $\alpha$ . Pour notre test, cela signifie que la proportion de fourche droite dans la SY14 est différente de la proportion de fourche droite dans la BY4742. J'ai choisi pour seuil  $\alpha=0,01$  pour tous les tests suivants, ce qui signifie que le risque de déterminer qu'un test est positif alors qu'il n'y a pas de différence biologique est de 1%.

Le génome de la SY14 possède environ 11 800 fenêtres de 1kb. Comme les tests sont réalisés sur cinq échelles différentes, le génome est analysé par la réalisation de 59 000 tests. Avec un seuil  $\alpha=0,01$  on s'attend à détecter 590 tests positifs même s'il n'y a aucune différence entre les deux profils. Il est donc nécessaire d'appliquer une correction pour tenir compte de la multiplicité des tests. Nous avons choisi d'utiliser une correction permettant de contrôler le *False Discovery Rate* (FDR), qui est le taux de faux positifs que l'on s'attend à obtenir. La méthode la plus couramment utilisée pour contrôler le FDR est la méthode de Benjamini et Hochberg, néanmoins cette méthode est applicable uniquement lorsque les tests sont indépendants. Nos tests ne sont pas indépendants, car une même fourche peut être comptée dans des fenêtres d'échelles différentes et à une même échelle dans plusieurs fenêtres. Nous avons donc utilisé, pour réduire le FDR, la méthode de Benjamini et Yekutieli, qui reste utilisable lorsqu'il existe une dépendance entre les tests. Cependant, après correction pour tests multiples, le nombre de tests positifs est excessivement réduit. Même les régions de fortes différences parfaitement convaincantes à l'œil, par exemple au voisinage des centromères supprimés, n'apparaissent plus significatives (Figure IV-8, cadre vert). Nous aurions pu envisager d'augmenter le nombre d'échantillons pour la BY4742 et la SY14 afin que le test de Student soit plus puissant et moins sensible à la correction, mais le coût engendré était trop important. Nous avons donc choisi d'abandonner le test de Student au profit du test du Chi<sup>2</sup>.



**Figure IV-8 : Test de Student avec et sans correction pour le chromosome IX.**

Les bins significatives ( $p\text{-value} \leq 0,01$ ) sont représentées en rouge, les bins non significatives ( $p\text{-value} > 0,01$ ) en blanc et les bins où le test est impossible, car les variances sont nulles, en gris. Le cadre vert indique une région de forte modification qui n'est plus détectée lorsque les résultats des tests sont corrigés pour tenir compte de la multiplicité des tests.

#### IV.2.1.2 Test du $\chi^2$

Le test du  $\chi^2$  est réalisé à partir du nombre de fourches gauches et de fourches droites comptées dans la fenêtre de test. Les valeurs sont classées selon la table de contingence suivante :

	Nombre de fourches droites	Nombre de fourches gauches
Échantillon 1	N11	N12
Échantillon 2	N21	N22

Ensuite, le test de  $\chi^2$  ne peut être effectué que si l'effectif théorique de chaque classe est supérieur à 5, c'est pourquoi j'ai déterminé l'effectif théorique de chaque classe à partir des effectifs observés du tableau de contingence. L'effectif théorique de N11 est déterminé par la formule suivante :

$$N11_{théorique} = \frac{(N11_{observé} + N21_{observé}) \times (N11_{observé} + N12_{observé})}{N11_{observé} + N21_{observé} + N11_{observé} + N12_{observé}}$$

Lorsqu'une des valeurs théoriques était inférieure à 5, je n'ai pas réalisé le test. Le nombre de régions où ce critère n'est pas satisfait est assez important à petite échelle, mais diminue à grande échelle (Figure IV-9, régions en gris).

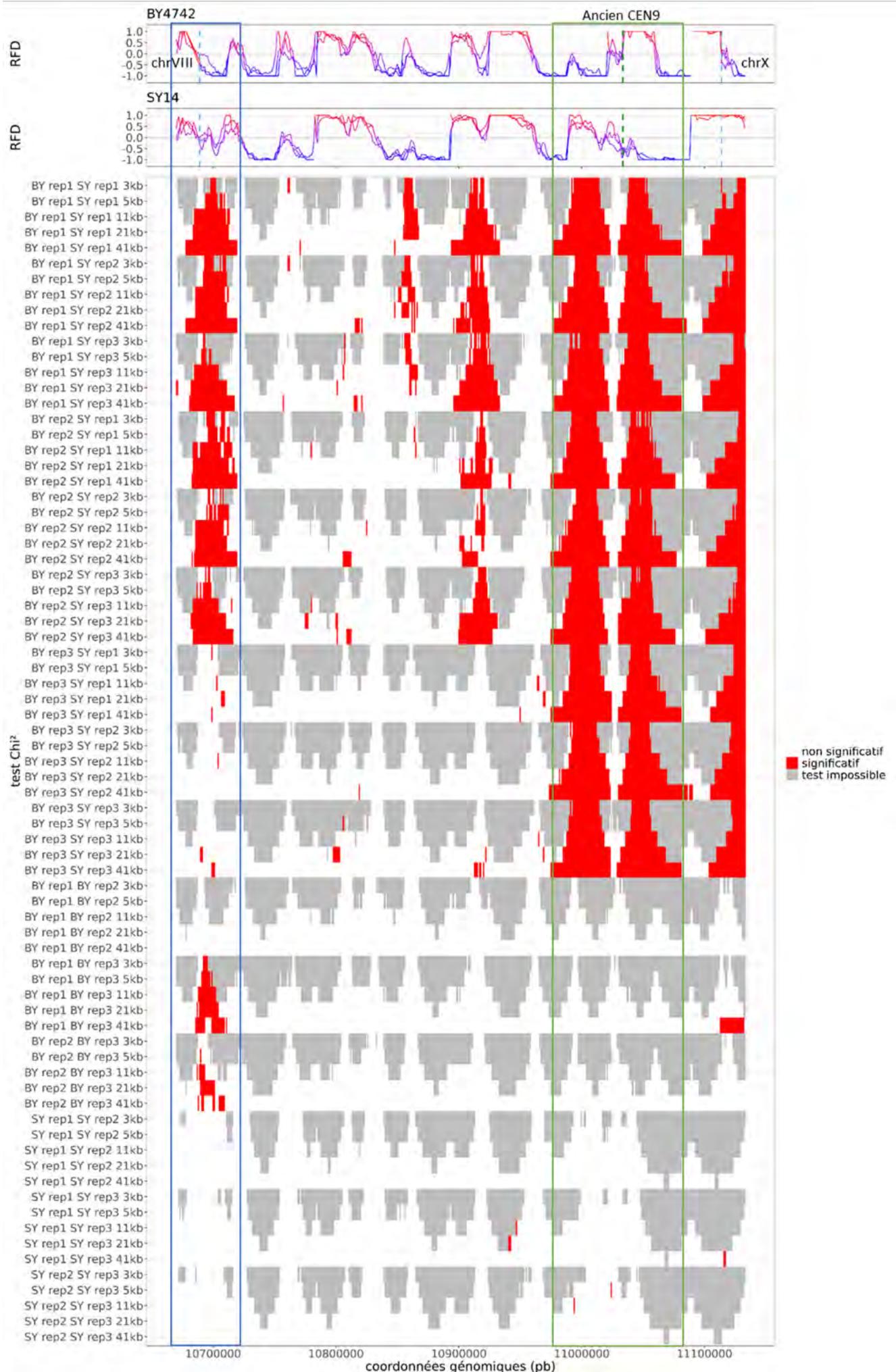
Le test de Chi<sup>2</sup> compare les échantillons deux à deux : 15 séries de test sont donc réalisées, 9 pour comparer la BY4742 et la SY14 et 6 pour comparer les réplicats de chaque souche entre eux. Les tests du Chi<sup>2</sup> permettent de bien visualiser les modifications importantes (Figure IV-9, cadre vert). Cependant, certaines régions apparaissent aussi significatives lorsque l'on compare les réplicats (Figure IV-9, cadre bleu). Le nombre de test réalisés, et donc le nombre de faux positifs, est plus important que pour le test de Student. Si l'on considère uniquement les régions du génome où les tests sont réalisés (c'est-à-dire les régions où l'effectif théorique de chaque classe est supérieur à 5), le nombre de tests réalisés sur le génome est de 399 017. Avec un risque  $\alpha=0,01$ , le nombre de faux positifs est de 3990. Afin de contrôler le FDR, les résultats des tests du Chi<sup>2</sup> ont été ajustés avec la méthode de Benjamini et Yekutieli. Contrairement au test de Student, après correction, les tests du Chi<sup>2</sup> pointent toujours vers les régions possédant d'importantes modifications, mais aussi des régions avec des modifications plus faibles (Figure IV-10, cadre vert = région fortement modifiée ; cadre bleu = région de faible modification). De plus, la correction permet d'éliminer la majorité des régions significatives dans les comparaisons entre réplicats.

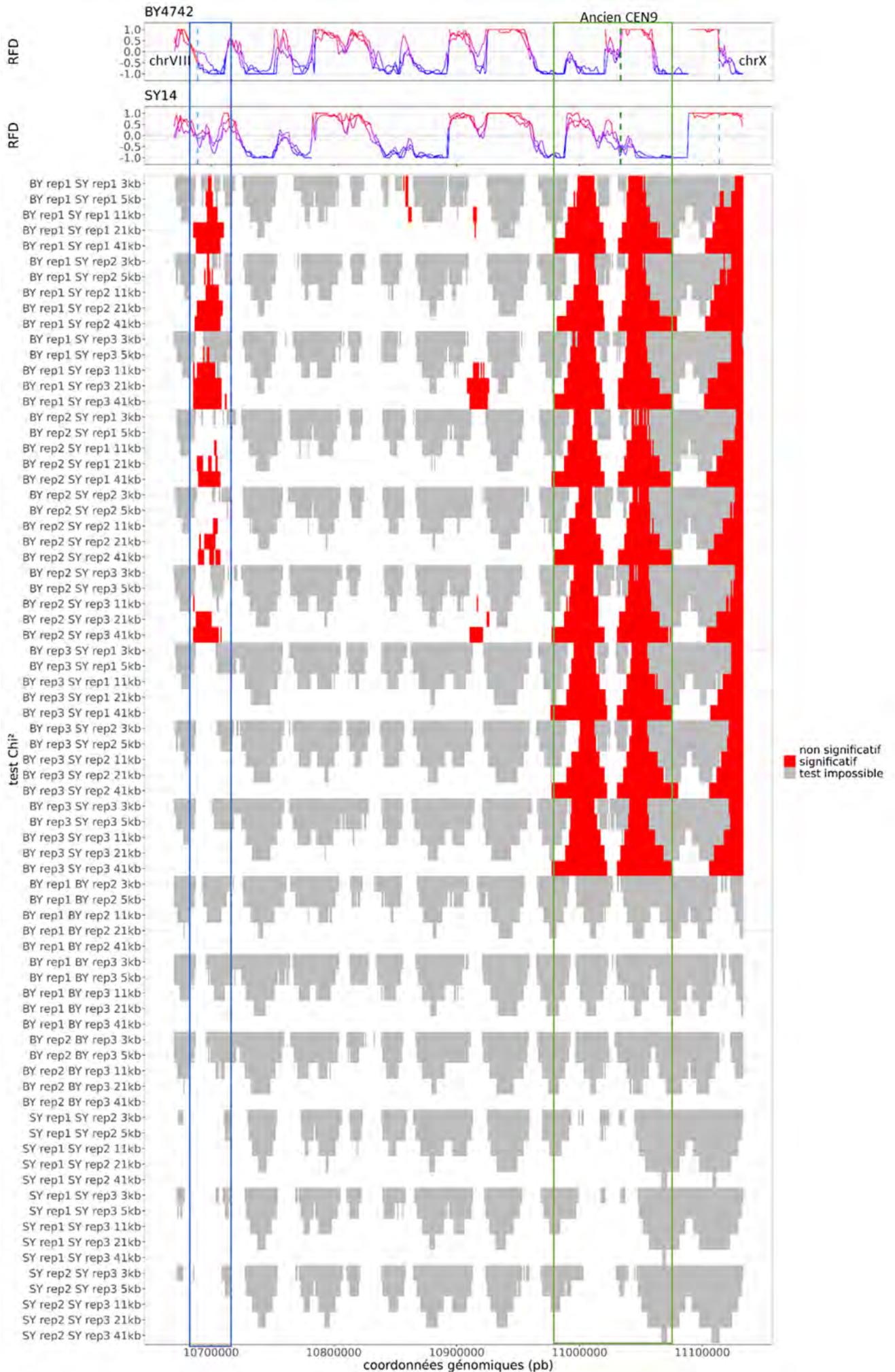
**Figure IV-9 : Test de Chi<sup>2</sup> sans correction du FDR pour le chromosome IX.**

Les bins significatives ( $p\text{-value} \leq 0,01$ ) sont représentées en rouge, les bins non significatives ( $p\text{-value} > 0,01$ ) en blanc et les bins où le test est impossible, car l'effectif théorique du nombre de fourches droites ou de gauches dans la BY4742 ou la SY14 est inférieur à 5, sont en gris. Le cadre bleu indique une région de modification où le test est aussi significatif entre certains réplicats. Le cadre vert indique une région de forte modification où l'ensemble des tests réalisés entre les souches sont significatifs.

**Figure IV-10 : Test de Chi<sup>2</sup> après correction du FDR pour le chromosome IX.**

Les bins significatives ( $p\text{-value} \leq 0,01$ ) sont représentées en rouge, les bins non significatives ( $p\text{-value} > 0,01$ ) en blanc et les bins où le test est impossible, car l'effectif théorique du nombre de fourches droites ou gauches dans la BY4742 ou la SY14 est inférieur à 5, sont en gris. Le cadre bleu indique une région de faibles modifications qui n'est détectée que dans un sous-ensemble de tests réalisés entre souches. Le cadre vert indique une région de forte modification où l'ensemble des tests réalisés entre les souches sont significatifs. Sur ce chromosome, plus aucun test n'est significatif entre les réplicats.

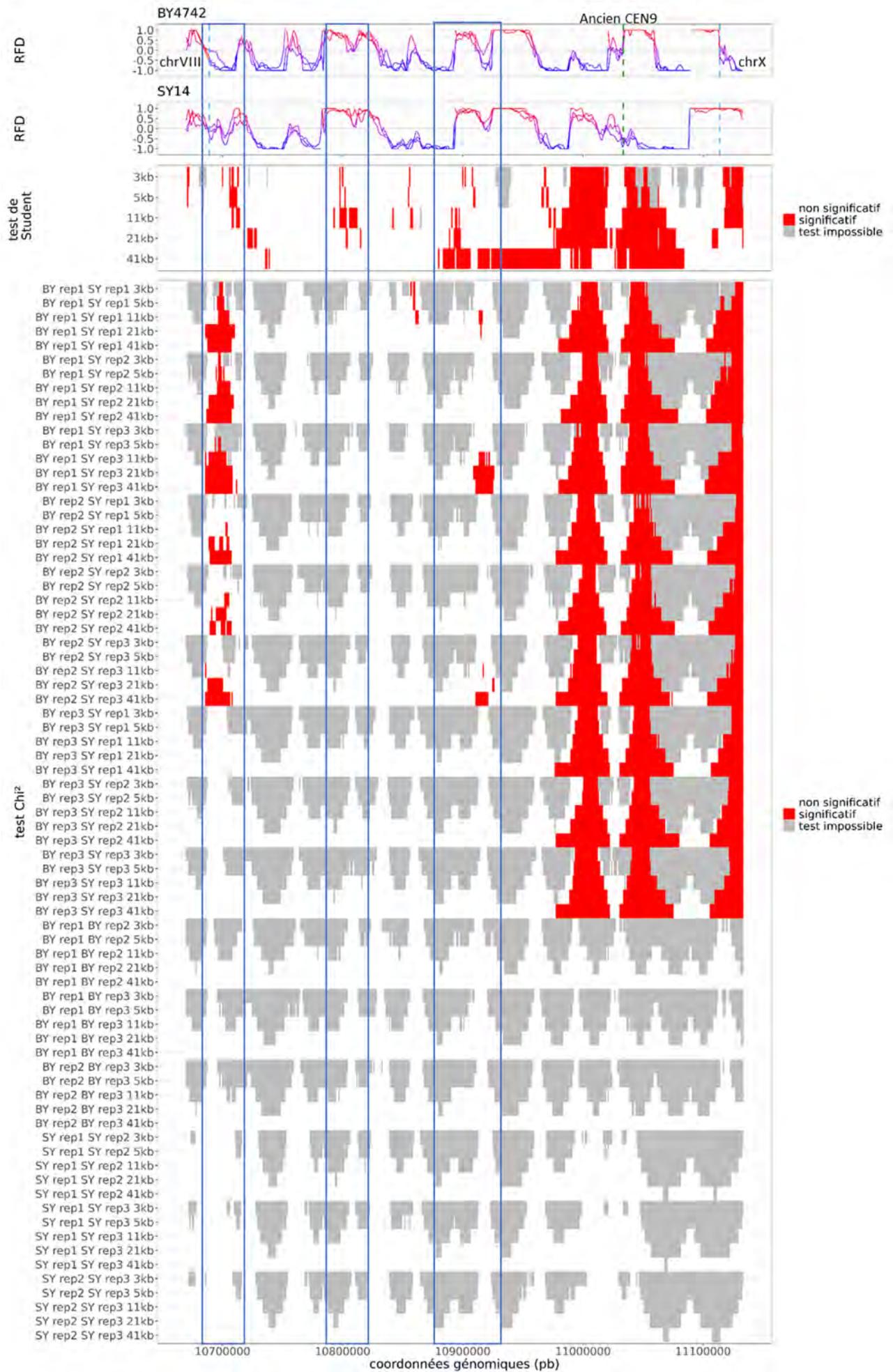




### IV.2.1.3 Comparaison du test de Student et du test de Chi<sup>2</sup>

Globalement, les résultats des tests du Chi<sup>2</sup> forment des pyramides qui pointent les zones de modification, tandis que les résultats du test de Student non-corrigé sont plus éparpillés. Sur la Figure IV-11, les cadres bleus représentent des exemples de régions où le test de Student et les tests du Chi<sup>2</sup> sont en désaccord. Certains de ces intervalles sont détectés dans toutes les comparaisons entre souches par les tests de Chi<sup>2</sup> (zone la plus à droite). D'autres intervalles ne sont détectés que par une fraction de ces comparaisons (coordonnée 1070000 et coordonnée 1090000). D'autres enfin ne sont pas détectés par les tests de Chi<sup>2</sup> (coordonnée 1080000). À la coordonnée 1070000, les tests de Chi<sup>2</sup> pointent l'activation d'une origine peu efficace dans la SY14 et, comme attendu, la diminution d'efficacité de l'origine suivante et l'apparition d'une zone de terminaison entre les deux. Le test de Student, quant à lui, ne pointe que ces deux derniers événements. À la coordonnée 1080000, le test de Student détecte une modification d'une zone d'initiation, mais seul un réplicat SY14 montre à l'œil une RFD différente de la BY4742. Tous les tests de Chi<sup>2</sup> sont négatifs sur cette même région. À la coordonnée 1090000, le test de Student pointe à gauche une origine dont la modification semble peu convaincante à l'œil, et à droite une origine dont la modification n'est détectée que pour l'échelle 41kb. Dans les tests de Chi<sup>2</sup>, l'origine de gauche n'est pas détectée, mais l'origine de droite est détectée dans 3 tests sur 9 aux échelles 11kb et 21 kb et dans 2 tests sur 9 à l'échelle 41 kb.

Dans l'ensemble, ces observations indiquent que le test de Student sans correction semble posséder de nombreux faux positifs que la synthèse des tests de Chi<sup>2</sup> ne possède pas. Il est donc préférable d'analyser les profils RFD avec la synthèse des tests de Chi<sup>2</sup>.



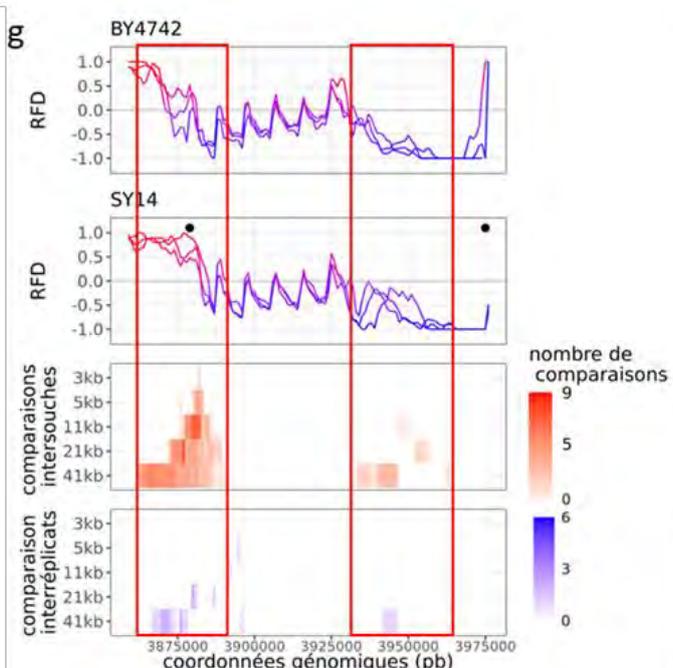
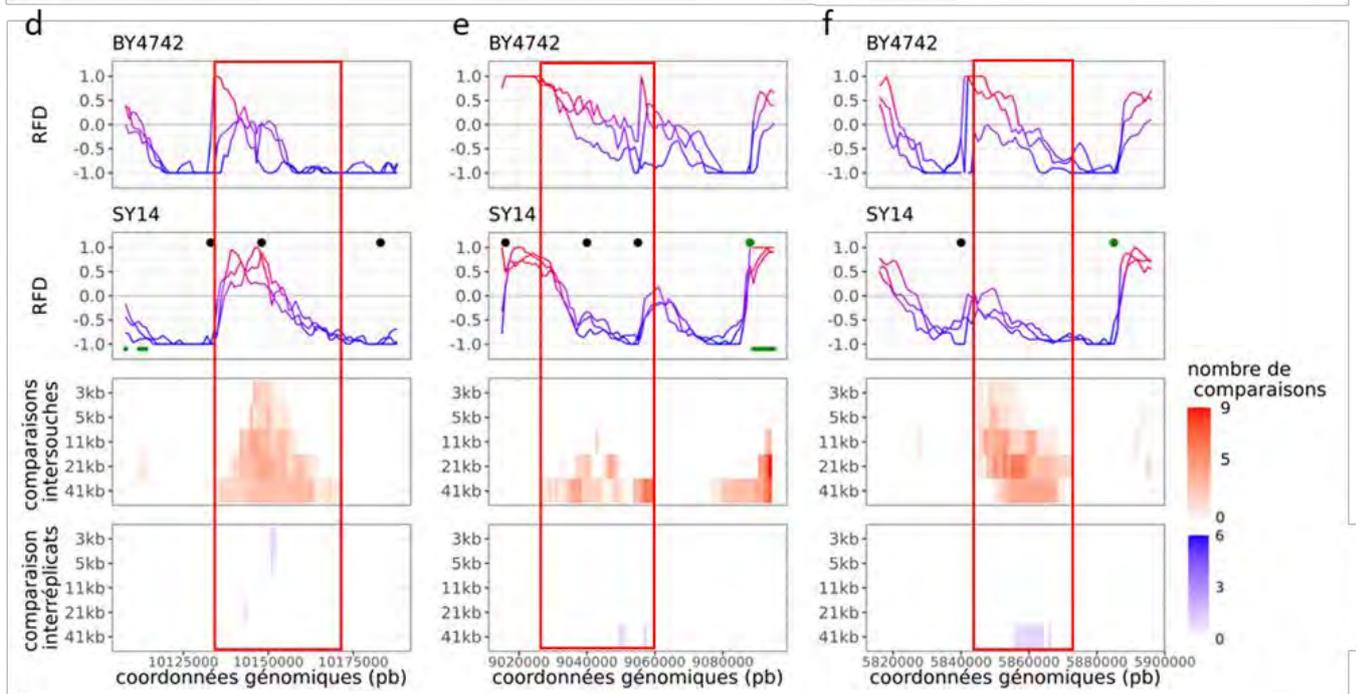
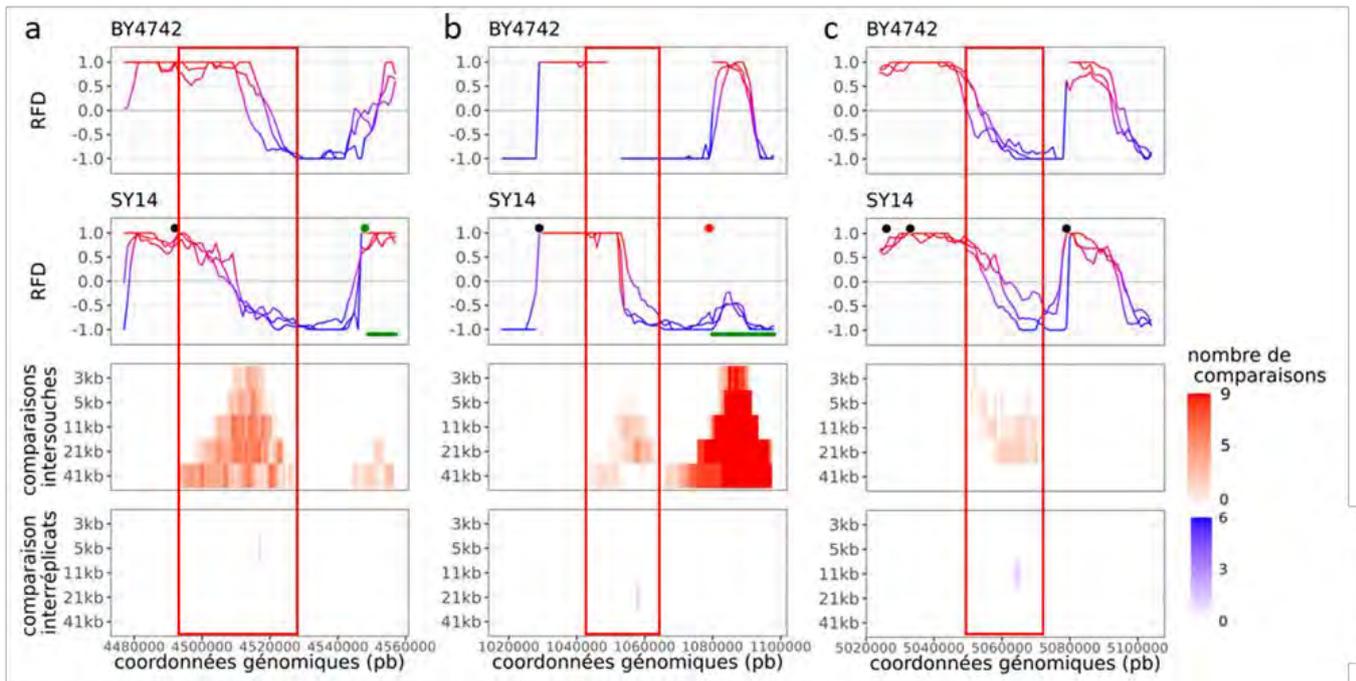
#### Figure IV-11 : Test de Student sans correction du FDR et test de Chi<sup>2</sup> après correction du FDR pour le chromosome IX.

Les bins significatives ( $p\text{-value}\leq 0,01$ ) sont représentées en rouge, les bins non significatives ( $p\text{-value}>0,01$ ) en blanc et les bins où le test est impossible, en gris. Le test de Student est impossible quand la variance d'un des deux échantillons est nulle. Le test de Chi<sup>2</sup> est impossible quand l'effectif théorique des fourches droites ou gauches dans la BY4742 ou la SY14 est inférieur à 5. Les cadres bleus servent à délimiter les régions discutées dans le texte.

##### IV.2.1.4 Analyse des profils de réplication à l'aide des tests de Chi<sup>2</sup>

Les résultats des tests statistiques pour les 9 comparaisons entre souches ne sont pas toujours similaires. Il convient donc de choisir un nombre minimum de tests positifs pour conclure qu'une région est significative ou non. Le nombre de comparaisons pour lesquelles une bin est significative est un indicateur de la force de la modification. Si l'écart entre les deux RFD est très important, tous les tests seront significatifs. Un écart plus faible ne sera éventuellement détecté que dans une fraction des tests. Lors de notre analyse, nous voulons nous intéresser aux zones de forte modification, mais aussi à des modifications plus faibles. C'est pourquoi nous avons choisi de considérer qu'une région est significativement modifiée dès qu'une comparaison entre les réplicats est significative sur au moins deux fenêtres adjacentes ou pour deux tailles de fenêtres différentes.

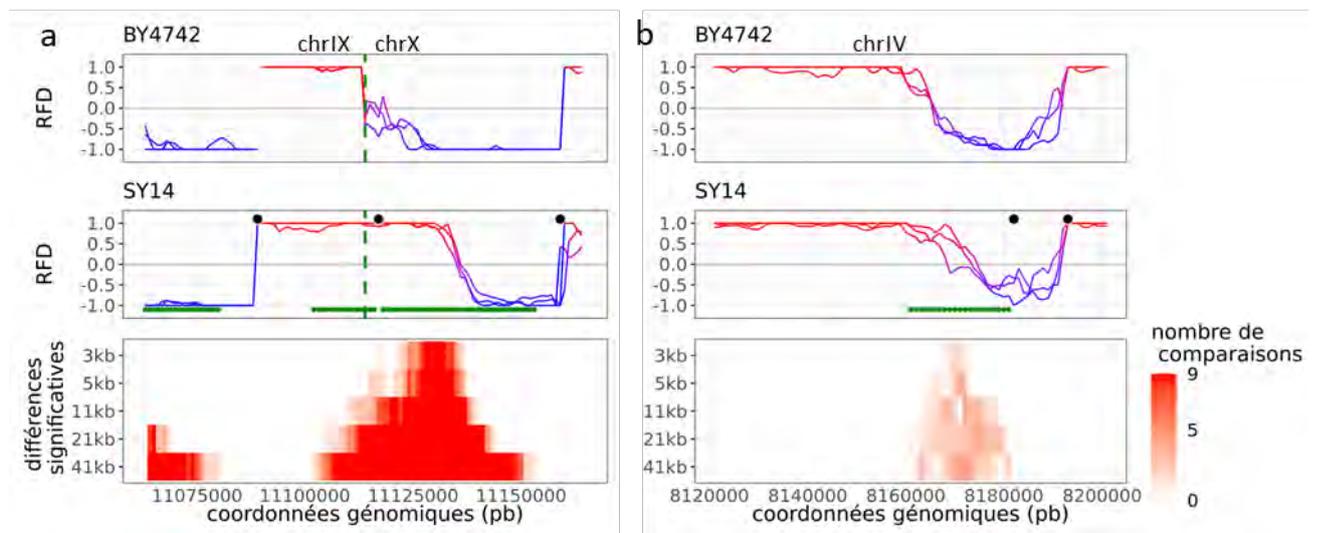
Sept régions génomiques significativement différentes entre la SY14 et BY4742 sont également significativement différentes entre réplicats (Figure IV-12). On peut distinguer deux cas : (i) un des réplicats d'une souche se comporte comme l'autre souche (Figure IV-12 d, e, f) ou s'en rapproche (Figure IV-12) ; (ii) les réplicats des deux souches sont tous semblables sauf un (Figure IV-12b, c). L'ensemble de ces modifications provient d'un réplicat qui possède un comportement aberrant. Garder ces régions pour l'analyse revient à considérer que le comportement aberrant se produit rarement dans la population. C'est-à-dire que si on réalise une centaine de profils RFD pour la SY14 et la BY4742, il y aura très peu de profils présentant ce comportement. En revanche, exclure ces régions revient à considérer que le comportement aberrant est fréquent dans la population, c'est-à-dire que sur une centaine de profils RFD pour la SY14 et la BY4742, ce comportement sera fréquent. La meilleure manière de vérifier si ce comportement est fréquent ou non dans la population serait de réaliser de nouveaux échantillons BY4742 et SY14, ce qui engendrerait un coût important. J'ai donc choisi d'être prudente et de considérer que la valeur aberrante pouvait être trouvée de manière régulière dans la population. C'est pourquoi, j'ai choisi d'exclure ces régions de l'analyse. L'exclusion s'applique à l'ensemble des modifications ayant lieu dans la zone formée par les tests significatifs entre souches dès lors qu'une bin de la zone est aussi significative entre les réplicats (Figure IV-12, cadre rouge).



**Figure IV-12 : Représentation des 7 régions présentant des différences significatives entre souches mais aussi entre réplicats.**

Les régions significatives entre les souches sont représentées en rouge, tandis que les régions significatives entre réplicats sont représentées en bleu. L'intensité de la couleur correspond au nombre de comparaisons pour lesquelles la région est significativement modifiée. Les origines actives sont indiquées par des cercles noirs, les origines significativement diminuées par des cercle rouges, les origines significativement augmentées par des cercles verts. Les zones de terminaisons modifiées sont indiquées par des lignes vertes.

Une fois les régions significativement différentes retenues pour l'analyse définie, j'ai examiné visuellement leurs profils afin de déterminer la nature des changements détectés. Dans un premier temps, je souhaitais obtenir la position des origines actives dans la SY14 afin de pouvoir comparer les origines modifiées et non modifiées. J'ai déterminé les origines actives dans la SY14 ou la BY4742, en les définissant par une pente ascendante en RFD, un pic IP ou un pic de *timing* (Figure IV-12, cercle noir). Ensuite, j'ai déterminé le type de modification des origines dans les régions significatives. Pour cela, j'ai comparé à l'œil l'amplitude du saut de RFD aux origines et déterminé si elles avaient une efficacité diminuée ou augmentée dans la SY14 (Figure IV-12a,b,e,f cercle rouge et vert). Lorsque les zones significatives correspondaient à des déplacements d'une zone de terminaison, j'ai annoté la terminaison à partir de l'origine modifiée lorsque celle-ci était dans la zone de tests significatifs ou au bord de cette zone. J'ai terminé l'annotation de la terminaison à la fin de la zone significative ou jusqu'à l'origine suivante (Figure IV-13a,b ligne verte).



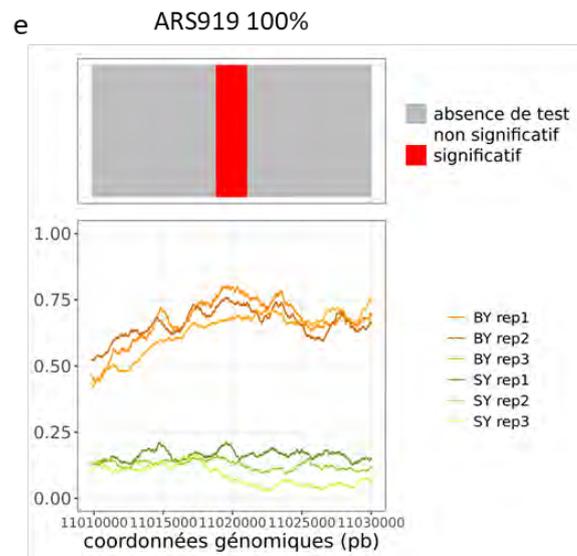
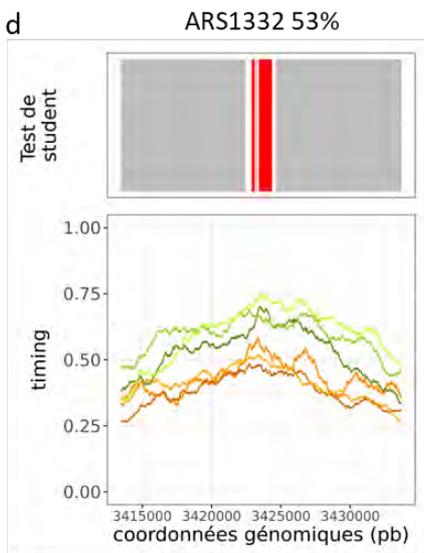
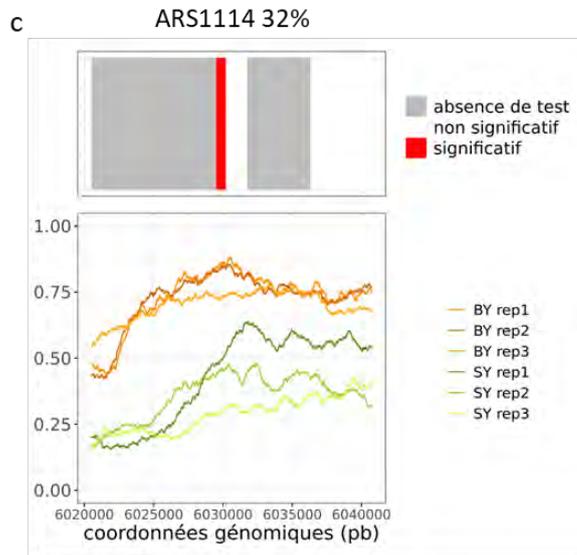
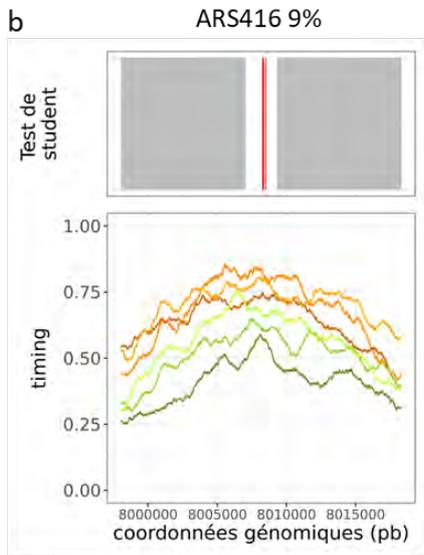
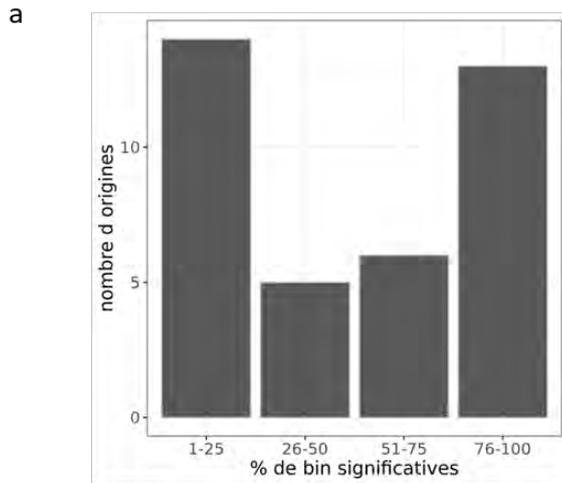
### Figure IV-13 : Exemple d'annotation de terminaison.

a-b) Les terminaisons annotées sont indiquées par une ligne verte et les origines actives par un cercle noir.  
a) Fusion entre les chromosome IX et X représentés par les pointillés verts : la zone de terminaison est annotée sur toute la longueur de la zone de test significatif. b) Zone de terminaison modifiée sur le chromosome IV : la zone de terminaison est annotée sur toute la longueur de la zone de test significatif.

#### IV.2.2 Timing

Les données qui permettent de calculer le *timing*, c'est-à-dire la probabilité moyenne de BrdU sur une fenêtre (voir Obtention des profils de Nanotiming.), sont dans un premier temps obtenues par fenêtres de 10pb. Gaël Millot, Biostatisticien à l'institut Pasteur, a suggéré de réaliser les tests à partir d'une probabilité moyenne de BrdU calculée par fenêtres de 1kb coulissantes avec un pas de 10 pb. L'analyse de l'intégralité du génome reviendrait à effectuer un test tous les 10 nucléotides soit 1 180 000 tests. Le nombre de tests significatifs par chance (pour un seuil alpha 0,01) serait alors de 11 800, il est donc nécessaire de corriger les tests. Néanmoins, vu le nombre de test réalisés, il serait peu probable de trouver des régions significatives après correction. C'est pourquoi nous avons réalisé le test uniquement sur les origines actives définies précédemment (voir Analyse des profils de réplication à l'aide des tests de Chi<sup>2</sup>). Les moyennes de probabilité de BrdU ont été calculées par fenêtres de 1kb coulissantes tous les 10pb sur l'intégralité du génome, puis un test de Student a été réalisé sur les fenêtres comprises entre +/- 1,5kb et l'ARS. J'ai choisi d'utiliser une fenêtre supérieure à la taille de l'ARS pour tenir compte de la possibilité que le pic de *timing* ne soit pas exactement centré sur l'ARS. Cette procédure réduit le nombre de test à 136 263, le nombre de faux positifs détectés est donc désormais de 1 362. Néanmoins, la modification de chaque origine est testée à plusieurs reprises. Nous avons testé 305 origines, donc si le nombre de faux positifs est réparti de manière équitable entre les origines, alors on a 4 faux positifs par origine. Néanmoins, la répartition du nombre de faux positifs par origine est aléatoire. Le choix d'un seuil du nombre de bin significatives détectées par origine devrait donc permettre d'éviter de considérer une origine comme significativement modifiée alors qu'elle ne l'est pas et nous affranchir d'utiliser une méthode de correction des tests qui pourrait être trop stricte.

La répartition du nombre d'origines en fonction du pourcentage de bins significatives montre qu'il existe un grand nombre d'origines avec un pourcentage inférieur à 25%, que ce nombre diminue pour les origines possédant entre 26 % et 75% de bins significatives et qu'il remonte pour les origines possédant entre 76 et 100% de leur bins significatives (Figure IV-14a). Comme nous souhaitons nous intéresser à des petites modifications, nous avons jugé que prendre un seuil compris entre 76 % et 100% serait trop strict. Nous avons choisi de conserver les origines possédant au moins 25% de bins significatives. La Figure IV-14 illustre 4 origines et leur pourcentage de fenêtres positives. On constate que le nombre de fenêtres significatives reflète assez bien la force de la modification de l'origine.



**Figure IV-14 : Représentation du nombre de tests de Student significatifs sur le *timing* de réplication des origines actives.**

a) Nombre d'origines en fonction du pourcentage de bin significatives. b-c-d-e) Représentation de différentes origines avec les résultats des tests de Student. Les bins significatives ( $p\text{-value} \leq 0,01$ ) sont représentées en rouge, les bins non significatives ( $p\text{-value} > 0,01$ ) sont représentées en blanc et les bins où le test n'a pas été réalisé car elles sont situées à plus de 1,5kb du début ou de la fin de l'ARS sont représentées en gris.

#### IV.2.3 Initiation Probability (IP)

Afin d'appuyer les conclusions obtenues par l'analyse du *timing* et du RFD, nous avons aussi décidé de réaliser un test statistique sur les données d'IP. Pour cela, les données sont assimilées à des données génomiques d'expression des gènes. En effet, lors de l'analyse de données RNA-seq, on détermine pour chaque gène sa couverture en ARN messenger. Pour l'IP, on détermine la couverture en segments compris entre deux fourches divergentes autour des origines de réplication. Il existe de nombreux algorithmes R pour détecter des différences d'expression lors d'analyses de données de RNA-seq. Nous avons choisi la méthode LimmaTrend. Ainsi, la couverture normalisée des trois réplicats BY4742 au niveau de l'ARS est comparée à la couverture normalisée des trois réplicats SY14 pour cette même ARS. Ainsi, chaque origine active (voir Analyse des profils de réplication à l'aide des tests de  $\text{Chi}^2$ ) se voit assigner une *p-value* qui indique si sa couverture en segments compris entre deux fourches divergentes est significativement différente entre la BY4742 et la SY14. Cette méthode ajuste automatiquement les résultats pour tenir compte du FDR. Les origines ont été considérées comme significativement modifiées dès lors que leur *p-value* était inférieure ou égale à 0,01.

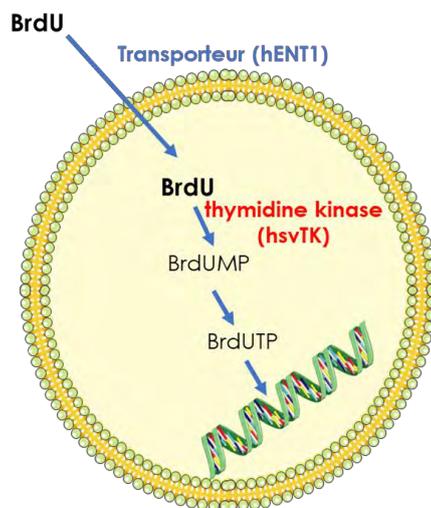
#### IV.2.4 Termination Probability (TP)

Dans un premier temps, notre analyse s'est focalisée sur la recherche de modifications au niveau des zones d'initiation, c'est pourquoi les profils TP ont simplement été analysés visuellement.

## V. Résultats

### V.1 Incorporation de BrdU chez la BY4742 et la SY14

Pour utiliser NFS et nanotiming chez la BY4742 et la SY14, il est nécessaire de modifier ces souches afin qu'elles incorporent le BrdU. En effet, la levure n'est pas naturellement capable d'incorporer dans l'ADN la thymidine ou ses analogues ajoutés au milieu extérieur, car elle ne possède pas d'enzyme capable de transformer la thymidine en thymidine monophosphate. Pour permettre l'incorporation de BrdU dans l'ADN, il est nécessaire de lui faire exprimer une thymidine kinase, qui va phosphoryler le BrdU en BrdU MonoPhosphate (BrdUMP) que les kinases intracellulaires peuvent convertir en BrdU TriPhosphate (BrdUTP, Figure V-1). La souche de référence utilisée pour développer nanotiming et NFS est la BT1 (Theulot et al., 2022) dont j'ai participé à la construction. Cette souche possède en plus de la thymidine kinase du virus Herpes simplex (hsv-TK), le transporteur passif hENT1 (*human equilibrative nucleoside transporter 1*) qui permet une meilleure pénétration du BrdU dans l'espace intracellulaire (Figure V-1). De plus, elle a été construite avec le plasmide pBL-hsvTK<sub>co</sub>-hENT1<sub>co</sub>, qui contient les séquences des gènes *hsvTK* et *hENT1* avec des codons optimisés pour la levure, afin de permettre une traduction plus efficace de l'ARN messager (ARNm) (Theulot et al., 2022).



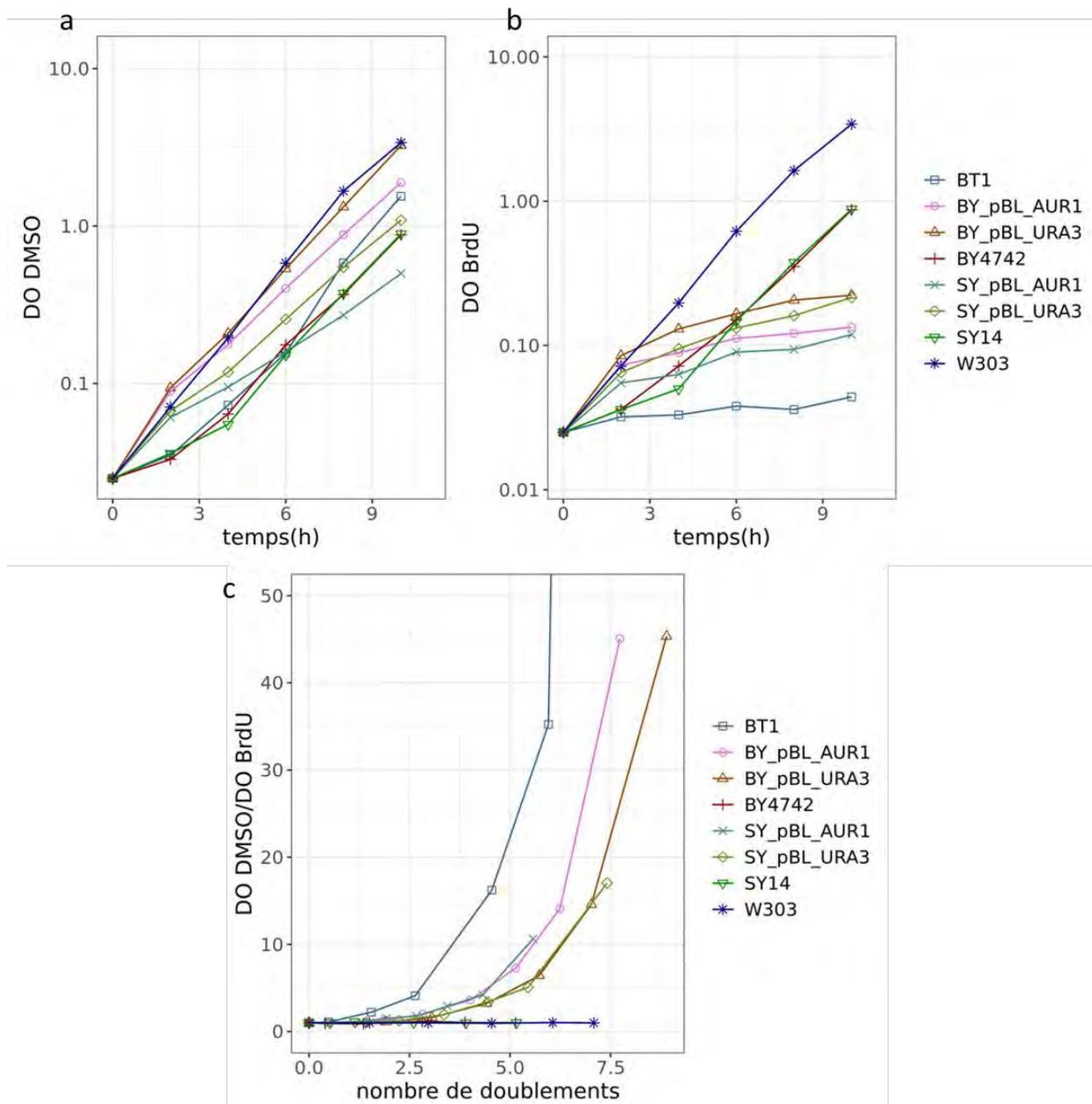
**Figure V-1 : Incorporation de BrdU chez la levure.**

Le BrdU du milieu extracellulaire peut diffuser de manière passive, ou son transport peut être facilité avec hENT1. Une fois le BrdU dans le milieu intracellulaire, une thymidine kinase est nécessaire pour transformer le BrdU en BrdUMP. Le BrdUMP peut alors être transformé en BrdUTP par les kinases intracellulaires et incorporé dans l'ADN au cours de la réplication.

La BT1 possède le plasmide pBL-hsvTK<sub>co</sub>-hENT1<sub>co</sub> intégré au locus *URA3*. Cependant, la SY14 et la BY4742 ne possèdent pas de séquences du gène *URA3* permettant d'insérer ce plasmide par recombinaison homologe. C'est pourquoi nous avons inséré le plasmide pBL-hsvTK<sub>co</sub>-hENT1<sub>co</sub> au

locus *AUR1* (BY-pBL-AUR1, et SY-pBL-AUR1). L'incorporation de BrdU a, dans un premier temps, été facilement mesurée, bien qu'indirectement, en comparant l'évolution de la densité optique à 260 nm ( $DO_{260}$ ) au cours du temps en présence de 10mM de BrdU ou en présence du diluant du BrdU, le DMSO. En présence de BrdU, les souches capables de l'incorporer voient leur croissance ralentie. Sur la Figure V-2b, seules la W303, la SY14 et la BY4742 montrent une croissance similaire en BrdU et en DMSO. Le ralentissement de croissance observé pour la BY-pBL-AUR1 et la SY-pBL-AUR1 est bien moindre que pour la BT1. Toutefois, la BY-pBL-AUR1 et la SY-pBL-AUR1 croissent moins vite que la BT1 en présence de DMSO seul (Figure V-2 a). On peut supposer que le ralentissement dépend du nombre de doublements effectués en présence de BrdU. Nous avons donc examiné l'évolution du ratio  $DO_{600nm} DMSO/DO_{600nm} BrdU$  en fonction du nombre de doublements en DMSO. Plus ce ratio est élevé, plus le ralentissement de croissance en présence de BrdU est fort (Figure V-2 c). Le ralentissement de croissance reste plus marqué pour la BT1 que pour les autres souches à nombre de doublements identique. Nous avons testé plusieurs clones ayant incorporé le plasmide sans en trouver un qui ait une croissance en BrdU similaire à la BT1.

Nous avons fait l'hypothèse que l'insertion du plasmide au locus *URA3* permettrait une meilleure incorporation de BrdU et donc établi une stratégie pour intégrer le plasmide au locus *URA3* dans la SY14 et la BY4742. Dans un premier temps, nous avons reconstitué le locus *URA3* dans nos deux souches en intégrant le gène *URA3* par recombinaison homologue. Ensuite, nous avons cherché à obtenir des souches possédant une mutation dans *URA3* (*ura3-1* obtenue dans la W303), afin de pouvoir sélectionner les clones ayant incorporé le plasmide par leur capacité à synthétiser l'uracile. Lors de l'intégration de ce gène, nous avons fait face à des difficultés expérimentales pour sélectionner les clones incapables de synthétiser l'uracile. L'acide 5-fluoroorotique est métabolisé en un composé toxique par les levures *URA3*. Nous avons essayé de sélectionner les levures *ura3-1* sur un milieu contenant ce composé, mais n'avons obtenu aucun clone. C'est pourquoi nous avons modifié le plasmide pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub> afin de lui ajouter une cassette de résistance à la nouséothricine. Ainsi, nous pouvions incorporer le plasmide au locus *URA3* dans les souches *URA3* et sélectionner les clones ayant incorporé le plasmide sur un milieu contenant de la nouséothricine (BY-pBL-URA3, et SY-pBL-URA3). Les ralentissements de croissance de ces nouvelles souches BY-pBL-URA3 et SY-pBL-URA3 sont moins importants que ceux des souches BY-pBL-AUR1 et SY-pBL-AUR1 (Figure V-2). Le locus *URA3* n'est donc pas responsable de la forte incorporation de BrdU dans la BT1. Lorsque les plasmides s'intègrent dans le génome de la levure, ils peuvent s'incorporer en plusieurs copies. Nous avons donc supposé que la BT1 avait intégré le plasmide en un plus grand nombre de copies que la BY-pBL-URA3, la SY-pBL-URA3, la BY-pBL-AUR1, et la SY-pBL-AUR1. Ceci devrait conférer à la BT1 une expression plus importante des ARNm *hsvTK* et *hENT1* et donc améliorer l'incorporation du BrdU.

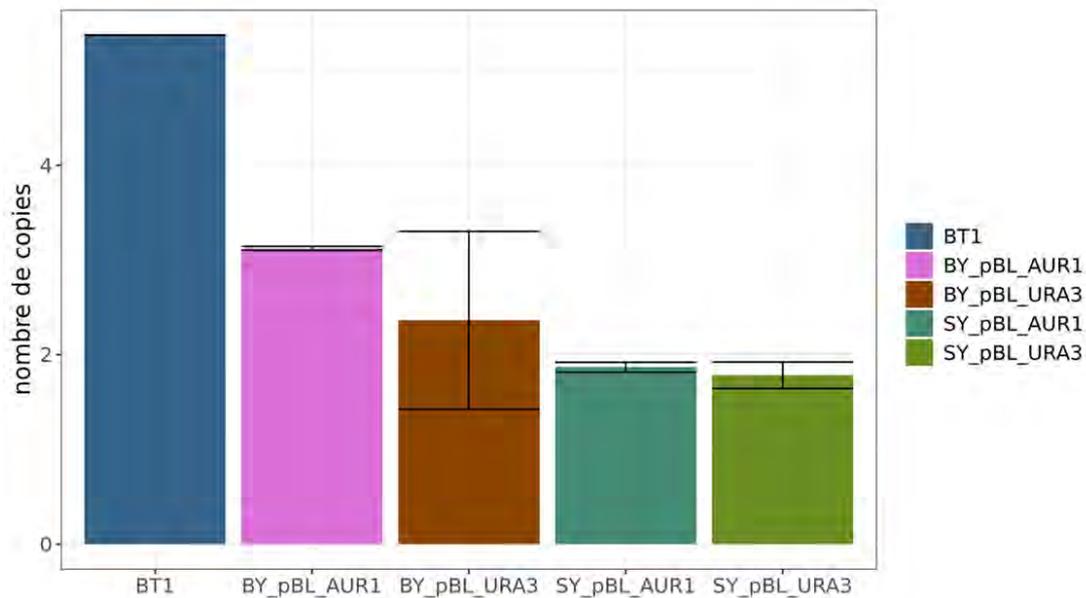


**Figure V-2 : Croissance des différentes souches BT1, BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3.**

a)  $DO_{600nm}$  en fonction du temps pour les souches cultivées avec du DMSO. b)  $DO_{600nm}$  en fonction du temps pour les souches cultivées avec 10mM de BrdU. c) Ratio  $DO_{600nm}$  DMSO/  $DO_{600nm}$  BrdU en fonction du nombre de doublements effectués par les cellules en présence de DMSO.

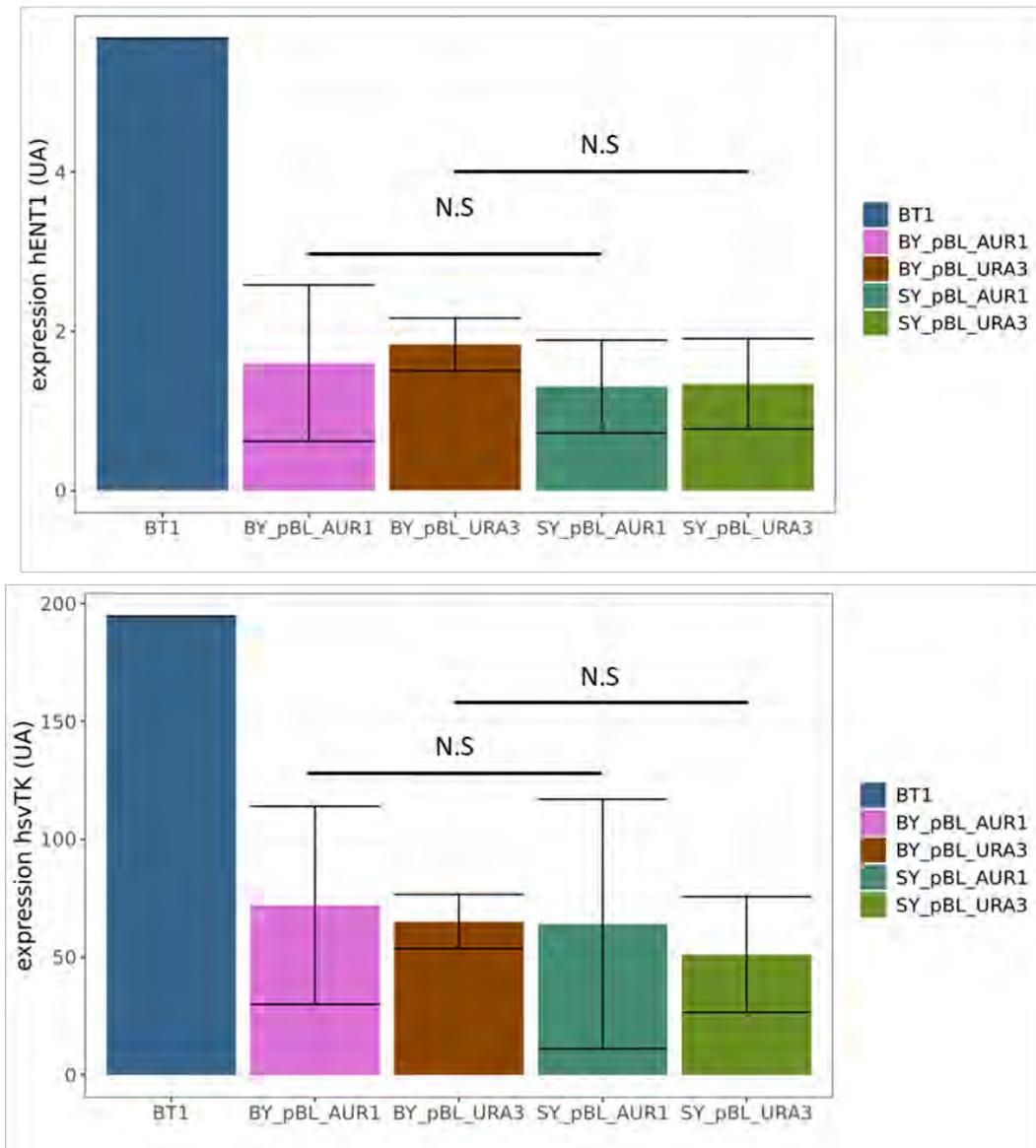
J'ai déterminé par qPCR le nombre de copies du plasmide intégrées pour chacune des souches. La BT1 possède 5 copies du plasmide tandis que BY-pBL-URA3, SY-pBL-URA3 et SY-pBL-AUR1 en possèdent 2, et BY-pBL-AUR1 3 (Figure V-3). Ainsi, la BT1 possède plus de copies des gènes *hsvTK* et *yENT1*, ce qui pourrait expliquer son plus fort ralentissement de croissance en BrdU. Nous avons

déterminé la quantité d'ARNm hENT1 et hsvTK par RT-qPCR. La quantité d'ARNm pour les deux gènes est environ trois fois plus élevée dans la BT1 que dans la BY-pBL-URA3, la SY-pBL-URA3, la BY-pBL-AUR1, et la SY-pBL-AUR1 (Figure V-4). La recherche de clones possédant un plus grand nombre de copies de plasmide intégrées au locus *URA3* ou *AUR1* a été infructueuse. Néanmoins, l'expression moyenne des ARNm hsvTK et hENT1 ne montre pas de différence significative entre la BY4742 et la SY14 pour les deux loci d'insertion du plasmide (Figure V-4). On peut donc espérer raisonnablement une incorporation de BrdU similaire entre la BY4742 et la SY14.



**Figure V-3 : Nombre de copies du plasmide intégré dans les souches BT1, BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3.**

La moyenne du nombre de copies est représentée avec l'écart-type. Réplicats biologiques : BT1 n=1, BY-pBL-AUR1 n=3, BY-pBL-URA3 n=3, SY-pBL-AUR1 n=3, SY-pBL-URA3 n=3.



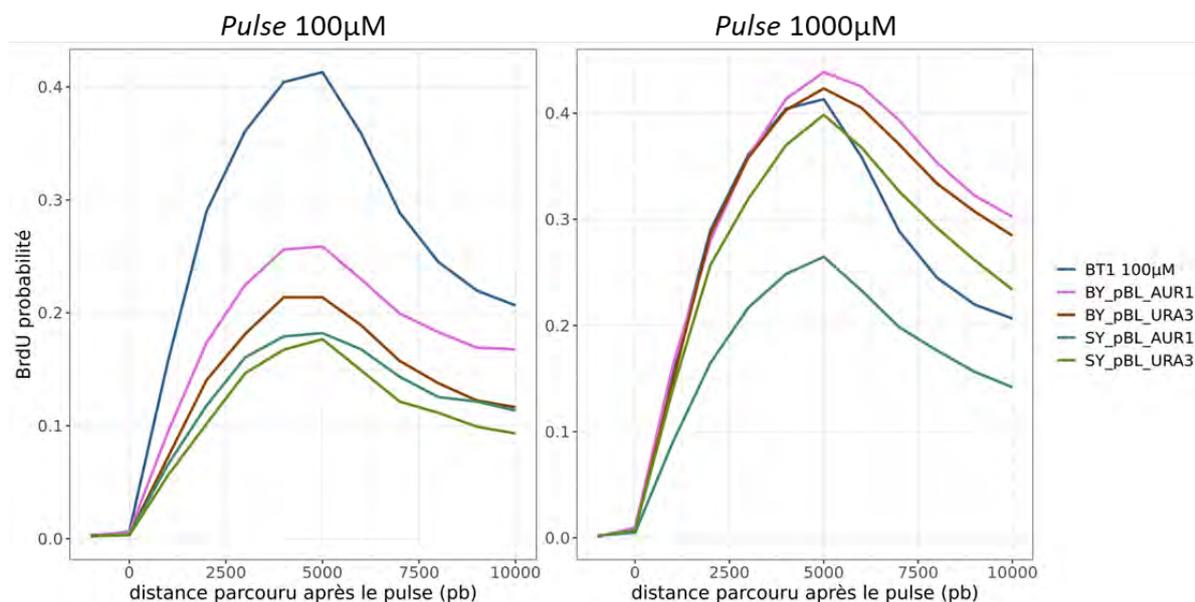
**Figure V-4 : Expression des gènes hsvTK et hENT1 pour la BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3.**

La moyenne de l'expression est représentée ainsi que l'écart-type. Réplicats biologique : BT1=1 BY-pBL-AUR1 n=4, BY-pBL-URA n=3, SY-pBL-AUR1 n=4, SY-pBL-URA3 n=4. Réplicats expérimentaux (RT-qPCR) : BT1 n=4, BY-pBL-AUR1 n=1, BY-pBL-URA n=1, SY-pBL-AUR1 n=1, SY-pBL-URA3 n=1. Un test de Student des moyennes d'expression a été réalisé entre les souches BY-pBL-AUR1 et SY-pBL-AUR1 et entre les souches BY-pBL-URA3 et SY-pBL-URA3. N.S = non significatif,  $p\text{-value} > 0.05$ .

Notre objectif était d'avoir un signal nanopore dans la BY4742 et la SY14 similaire à celui obtenu avec la BT1, notre souche de référence. L'incorporation de BrdU semblait toutefois moindre, compte tenu des résultats de la croissance en BrdU et de l'expression des ARNm. Nous avons réalisé un premier test par séquençage nanopore pour déterminer la concentration de BrdU à utiliser. La

concentration en BrdU utilisée en routine avec la BT1 est de 100µM. Nous avons donc réalisé deux *pulses* de BrdU de 2min à 100µM et 1000µM suivis respectivement d'une chasse thymidine à 1000µM et 10000µM.

Afin de comparer le signal Nanopore de la BT1 à celui de la BY-pBL-URA3, la SY-pBL-URA3, la BY-pBL-AUR1, et la SY-pBL-AUR1, les échantillons provenant des *pulses* avec les deux concentrations ont été analysés avec *Megalodon-based BrdU basecaller* et NFS (Theulot et al., 2022). Ainsi nous avons détecté les fourches de réplication de ces échantillons et comparé leur profil moyen à celui des échantillons de la BT1 obtenus précédemment (Theulot et al., 2022, Figure V-5). Le *pulse* à 100µM de BrdU donne pour la BY4742 et la SY14 un signal bien inférieur à celui observé pour la BT1, avec un pic d'amplitude à 0,25-0,35 au lieu de 0,5. La BY-pBL-URA3, BY-pBL-AUR1 et la SY-pBL-URA3 atteignent une amplitude similaire à la BT1 100µM lorsque le *pulse* est réalisé à 1000µM. La SY-pBL-AUR1 à 1000µM possède quant à elle une amplitude de 0,35, ce qui est inférieur à l'amplitude de la BY-pBL-AUR1, bien que les deux souches possèdent une expression similaire des deux ARNm. Cependant, la BY-pBL-AUR1 possède une copie supplémentaire du plasmide. Nous avons donc choisi d'écarter BY-pBL-AUR1 et la SY-pBL-AUR1, pour réaliser l'ensemble de nos expériences avec la BY-pBL-URA3 et la SY-pBL-URA3, car ces deux souches possèdent le même nombre de copies du plasmide, une expression similaire des ARNm *hsvTK* et *hENT1* et un signal nanopore similaire. Les *pulses* seront réalisés à 1000µM, car l'amplitude obtenue est similaire à celle de la BT1 100µM.



**Figure V-5 : Signal moyen des souches BT1 pour un *pulse* de 2min à 100µM et BY-pBL-AUR1, BY-pBL-URA3, SY-pBL-AUR1 et SY-pBL-URA3 pour un *pulse* de 2min à 100µM ou 1000µM.**

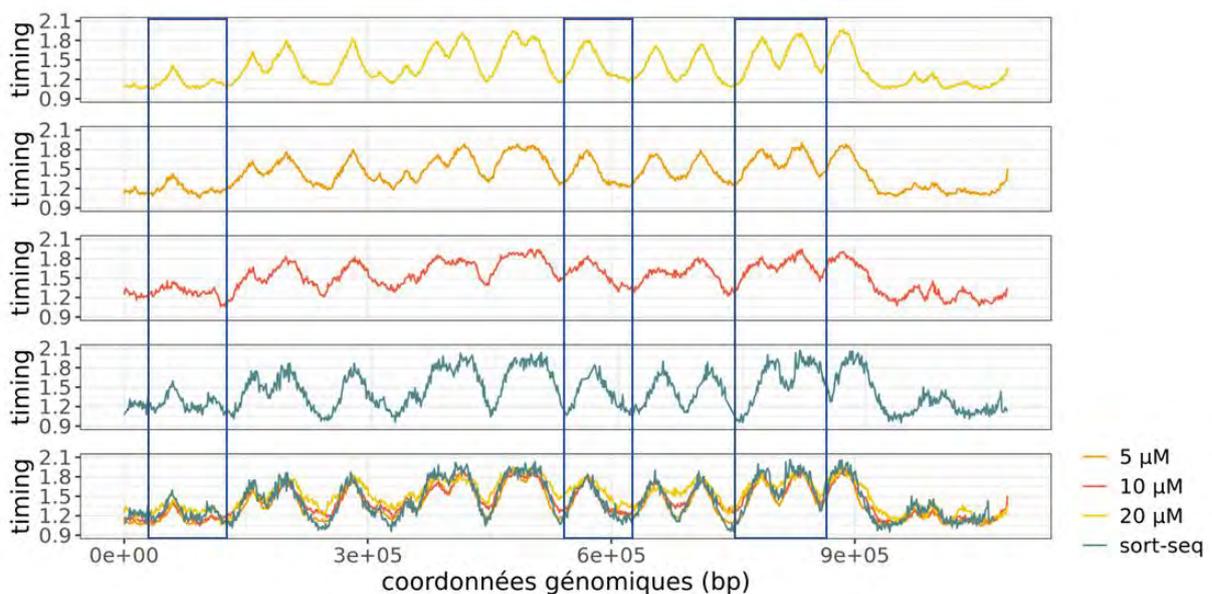
Le signal moyen représente la moyenne sur 1kb de tous les signaux des fourches détectées par NFS.

## V.2 Analyse du *timing* de réplication par séquençage Nanopore

### V.2.1 Mise au point de la méthode Nanotiming par Bertrand Theulot

Bertrand Theulot, doctorant au laboratoire, et Étienne Jean, ingénieur, ont développé une méthode d'étude du *timing* de réplication basée sur l'incorporation d'une faible quantité de BrdU dans des cellules asynchrones. Ils ont, dans un premier temps, réalisé trois marquages de BrdU dans trois cultures différentes avec 5 $\mu$ M, 10 $\mu$ M ou 20 $\mu$ M BrdU pendant un cycle cellulaire. Ils ont observé un gradient de concentration sur les lectures avec les pics de concentration les plus importants correspondant aux ARS.

Sachant que la quantité de thymidine intracellulaire augmente au cours de la phase S (Chabes et al., 2003; Koç et al., 2004, 2003), Bertrand Theulot a suggéré qu'en début de phase S, lorsque la concentration en thymidine est faible, le BrdU est incorporé dans l'ADN en plus grande proportion que par la suite, quand la quantité de thymidine intracellulaire augmente. La synthèse accrue de thymidine endogène crée donc une baisse progressive du rapport BrdUTP/dTTP (désoxythymidine TriPhosphat) au cours de la phase S. Bertrand Theulot a donc réalisé un profil moyen de la probabilité d'incorporation du BrdU en fonction des coordonnées génomiques pour chaque concentration et a comparé les résultats avec les données de Sort-seq obtenues dans la souche BT1 (Figure V-6). Les profils de *timing* obtenus pour les trois doses sont assez similaires à celui obtenu en Sort-seq. Néanmoins, le profil à 5 $\mu$ M possède des pics mieux définis, et une meilleure adéquation générale au profil Sort-seq, que les profils réalisés à 10 $\mu$ M et 20 $\mu$ M (Figure V-6, cadre bleu). Ainsi, la réalisation d'un marquage d'une culture de levures asynchrones au BrdU pendant un cycle cellulaire permet d'obtenir un profil génomique du moment de réplication moyen en phase S.



### Figure V-6 : Profil de Nanotiming et Sort-seq dans la souche BT1 pour le chromosome VII.

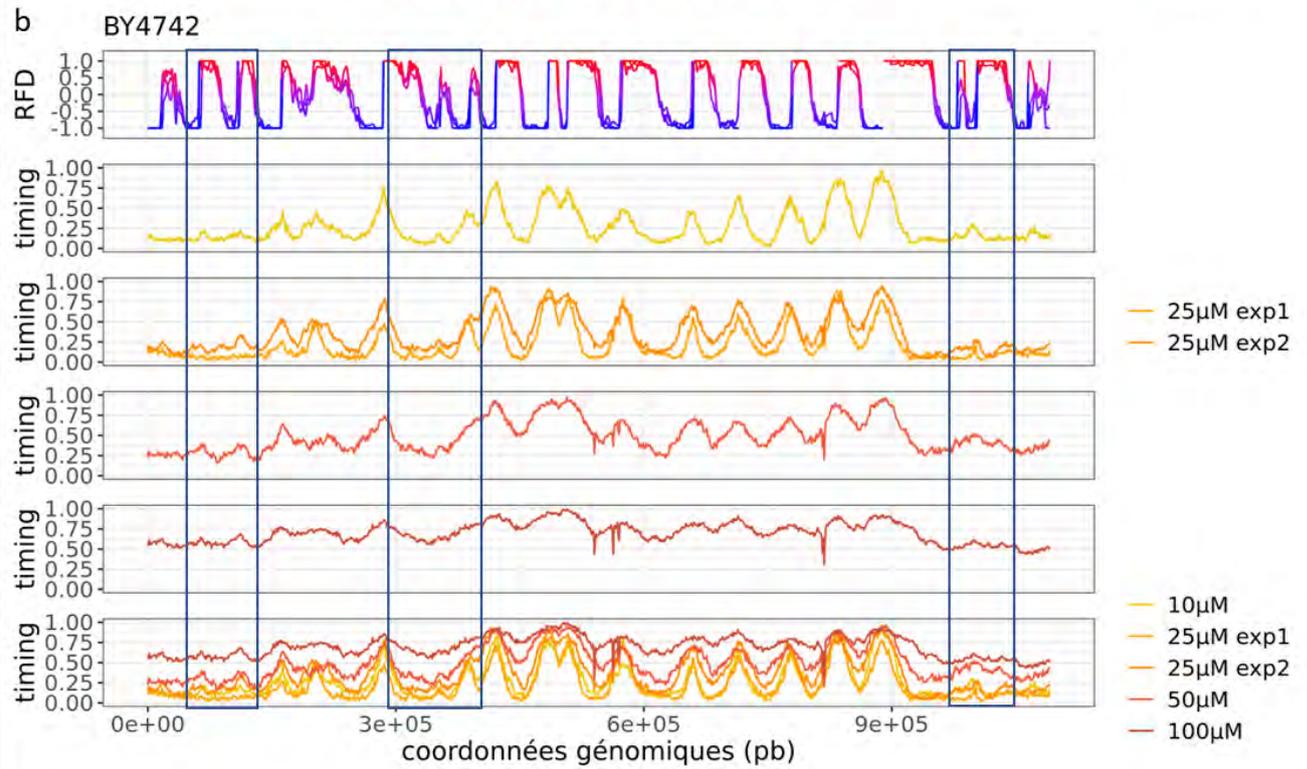
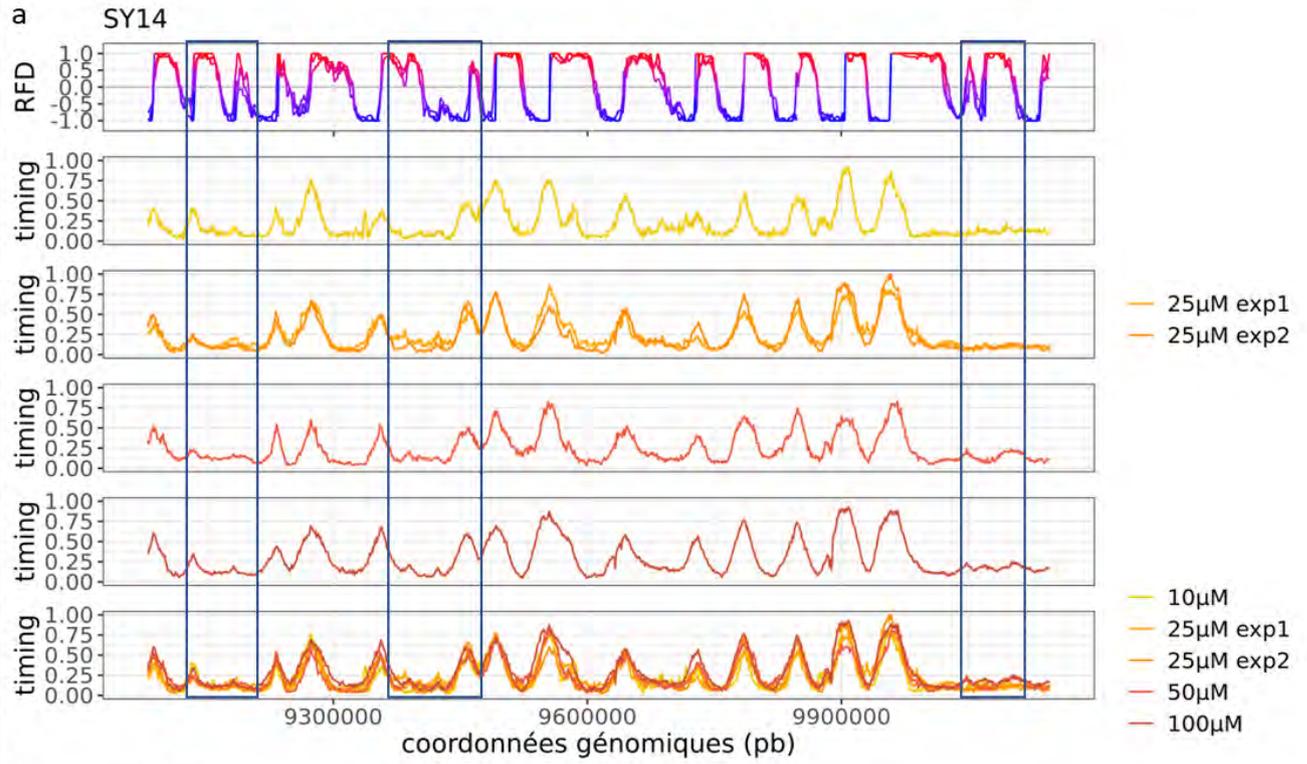
Les cadres bleus représentent des régions où les pics de *timing* sont mieux définis pour le profil de réplifications obtenues avec un marquage à 5 $\mu$ M.

#### V.2.2 Adaptation de la méthode dans la SY14 et la BY4742

La BY4742 et la SY14 montrent une incorporation de BrdU plus faible que la BT1. Il était donc possible que le marquage d'un cycle cellulaire avec une dose de 5 $\mu$ M de BrdU ne permette pas d'obtenir un profil temporel de réplification. C'est pourquoi nous avons dans un premier temps réalisé trois marquages au BrdU avec 10 $\mu$ M et 25 $\mu$ M, puis trois autres marquages à 25 $\mu$ M, 50 $\mu$ M et 100 $\mu$ M. Comme nous avons déjà réalisé les profils RFD pour nos souches, nous avons décidé de déterminer la meilleure dose en comparant les profils obtenus au profil RFD, afin de vérifier si toutes les zones d'initiation en RFD correspondaient à des pics de *timing* (Figure V-7). Les marquages à 10, 25, 50 et 100 $\mu$ M permettent d'obtenir des profils satisfaisants et similaires dans la SY14. Néanmoins, certaines zones d'initiation clairement visibles en RFD, mais localisées dans des régions tardives, ne sont pas toujours clairement visualisées sous forme de pics d'initiation tardifs (Figure V-7a cadre bleu). La qualité de détection de ces zones varie en fonction des profils. Le profil obtenu à 10 $\mu$ M ne permet pas de visualiser les pics attendus pour la SY14 (Figure V-7a). Pour la BY4742, au-delà de 25 $\mu$ M, les pics commencent à s'élargir et la définition du profil diminue, ce qui suggère que la trop forte concentration de BrdU sature le système (Figure V-7b). Pour la BY4742 la dose à 25 $\mu$ M lors de la première expérience permet d'obtenir la meilleure définition des régions tardives, mais lors de la seconde expérience, elle montre des pics légèrement élargis (Figure V-7b cadre bleu). A l'issue de ce test, nous avons décidé de réaliser systématiquement deux marquages à 25 et 50 $\mu$ M pour la SY14 et 10, et 25 $\mu$ M pour la BY4742, puis de sélectionner le meilleur profil. La répétition des marquages a montré que pour la BY4742, la meilleure dose semblait toujours être 10 $\mu$ M et pour la SY14, 25 $\mu$ M (Figure V-7).

### Figure V-7 : Profils Nanotiming à 10, 25, 50 et 100 $\mu$ M de BrdU et profil RFD pour le chromosome VII dans la SY14 et la BY4742.

a) Profils de la SY14. Les cadres bleus indiquent les régions où le profil à 10 $\mu$ M ne permet pas d'obtenir un pic correspondant à la zone d'initiation visible en RFD, tandis que les profils réalisés avec une dose supérieure ou égale à 25 $\mu$ M le permettent. b) Profils dans la BY4742. Les cadres bleus indiquent les régions tardives où le profil à 10 $\mu$ M ne permet pas d'obtenir un pic correspondant à la zone d'initiation visible en RFD, tandis que le profil à 25 $\mu$ M de la première expérience le permet.

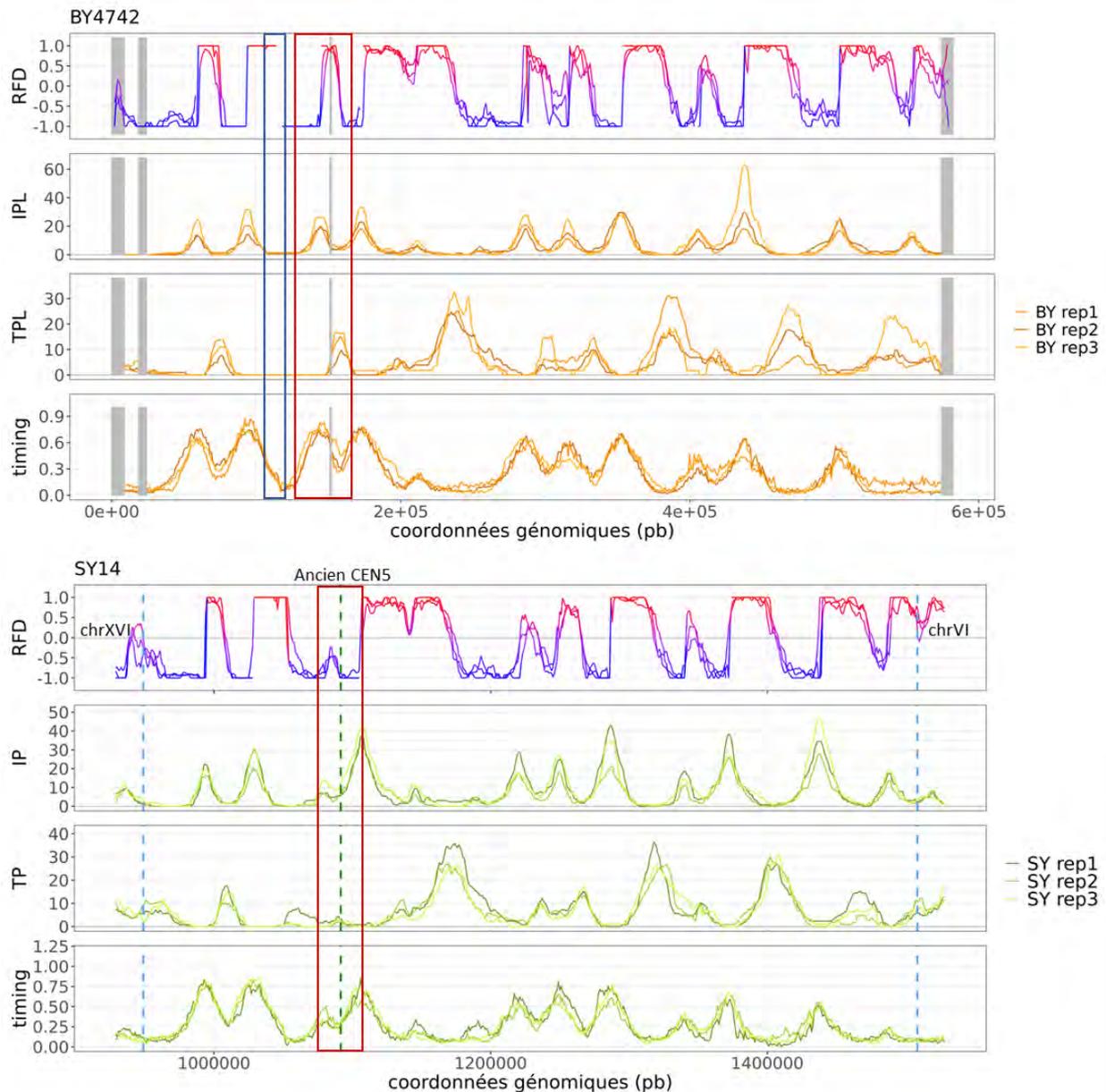


## V.3 Analyses des profils de réplication

### V.3.1 Comparaison des différentes analyses

La dynamique de réplication chez la BY4742 et la SY14 est analysée au travers de quatre profils différents : la RFD, l'IP, le TP et le *timing* de réplication. Les trois premiers sont obtenus par NFS et le dernier par une méthode indépendante, nanotiming. Globalement les résultats obtenus pour chacun des profils sont très similaires. La RFD est basée sur l'analyse collective de toutes les fourches détectées par NFS, donc elle détecte des changements de direction moyenne de réplication, alors que l'IP et le TP sont basés sur la couverture normalisée des segments compris entre deux fourches divergentes ou deux fourches convergentes, donc ils détectent la fréquence d'évènements individuels d'initiation ou de terminaison. Le nanotiming, quant à lui, détermine le signal BrdU moyen obtenu lors d'un marquage durant un cycle cellulaire, sachant que plus une région est répliquée de manière précoce, plus le signal BrdU moyen est élevé. La Figure V-8 présente les différents profils obtenus pour le chromosome V. Chaque pente ascendante de RFD est également détectée par un pic IP et nanotiming. Les informations du TP correspondent aux pentes descendantes observées en RFD et aux vallées observées en IP et nanotiming. Le centromère (Figure V-8, cadre rouge) montre dans la BY4742 un *timing* de réplication précoce. La zone d'initiation correspondante en RFD possède un  $\Delta RFD$  de -2, ce qui signifie que l'origine émet 100% de fourches vers la droite ( $RFD=1$ ) et 100% de fourche vers la gauche ( $RFD=-1$ ), autrement dit, qu'elle n'est pas répliquée de manière passive par une origine voisine. Le pic IP (à l'origine) est de 20, ce qui signifie que de nombreux segments bordés par deux fourches divergentes ont été détectés (Figure V-8 ,cadre rouge). Comme cette origine est très efficace, nous détectons de nombreux segments bordés par deux fourches divergentes. La valeur du TP (à droite de l'origine) est également élevée. À gauche de l'origine centromérique, la zone de terminaison du RFD n'est pas visible, tout comme le pic TP, en raison d'une absence de données (Figure V-8 , cadre bleu). Le manque de données provient d'une absence de détection des fourches par NFS. L'examen visuel des lectures dans cette région montre exclusivement des évènements de terminaison, non détectés par NFS en raison d'une trop grande proximité des positions X1.

Dans la SY14, le *timing* de réplication de cette région centromérique est plus tardif. Le  $\Delta RFD$  est réduit à 0,5, ce qui signifie que l'origine est répliquée passivement par les origines voisines. Le pic IP est diminué, tout comme le pic TP. Cet exemple illustre la concordance de ces quatre analyses, même si les informations utilisées sont différentes. Globalement, les profils de réplication dans la BY4742 et la SY14 sont similaires (Figure V-8), ce qui signifie que la réorganisation spatiale du génome n'a pas complètement bouleversé le programme de réplication. Néanmoins, de petites modifications sont visibles, c'est pourquoi nous avons décidé d'employer une analyse statistique pour les détecter automatiquement.

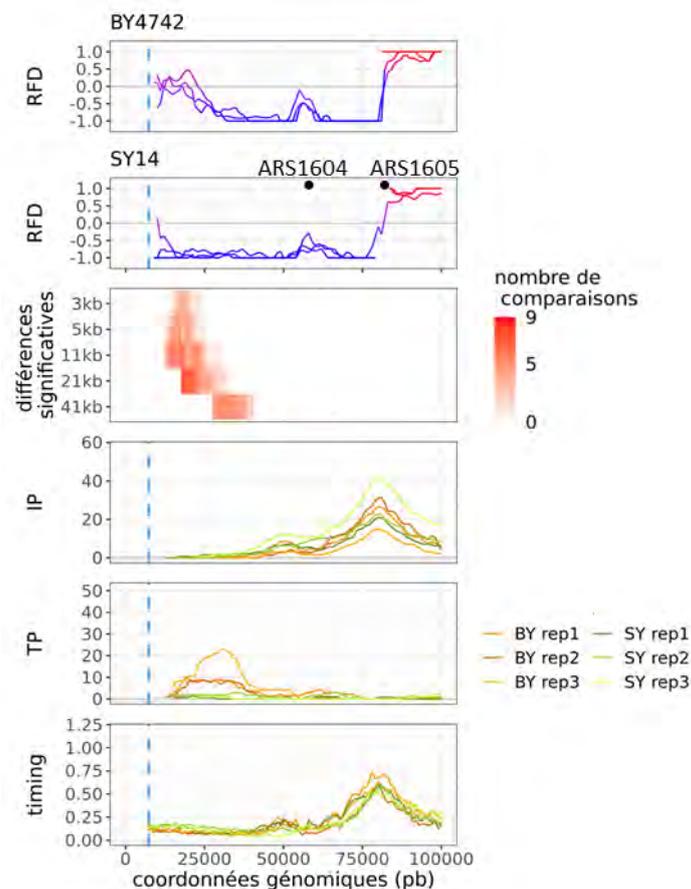


**Figure V-8 : Profils de répliation du chromosome V dans la BY4742 et la SY14.**

a-b) Le cadre rouge indique la région modifiée dans la SY14 à proximité du centromère. a) Profils de répliation dans la BY4742. Les régions supprimées dans la SY14 sont représentées en gris. b) Profils de répliation de la SY14, les zones de fusions sont représentées par des pointillés bleus et le centromère supprimé, par des pointillés verts.

Pour comparer les profils de la BY4742 et de la SY14, les données de la BY4742 ont été réalignées sur le génome de la SY14. Les analyses statistiques ont été réalisées sur l'intégralité des données dans le cas du RFD, mais restreintes aux origines actives montrant une pente ascendante en RFD ou un pic IP ou un pic de *timing* - pour l'IP et le *timing* de répliation. Aucune analyse statistique n'a été réalisée sur les profils TP, mais ces données ont été exploitées visuellement dans quelques cas. Les différentes

analyses statistiques ont permis de mettre en évidence 71 origines modifiées significativement (Figure supplémentaire VIII-2 à VIII-16). Parmi ces origines, une située sur le télomère du chromosome XVI a été exclue (Figure V-9, à droite des pointillés bleus figurant la position du TEL16L). En effet, le génome de la SY14, mais pas celui de la BY4742, présente en amont du télomère (à gauche des pointillés bleus) 7 kb de séquences provenant d'une autre partie du génome de la levure. Une incertitude subsiste sur le fait que cela puisse résulter d'un réarrangement intempestif lors de la construction de la souche ou d'un problème d'alignement. Les profil RFD suggèrent qu'une origine très proche du télomère (position 12000), active dans la BY4742, a été inactivée dans la SY14. L'incertitude sur l'assemblage de cette région du génome nous a décidé à ne pas la prendre en considération dans nos analyses



**Figure V-9 : Modification significative du TEL16L.**

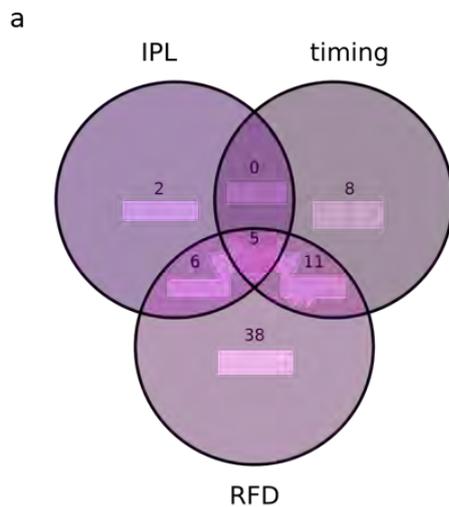
La SY14 possède en amont du TEL16L, représenté par les pointillés bleus, une région de 7kb d'origine incertaine. Les profils RFD et TP de la BY4742 montrent une zone de terminaison (coordonnées 12000-25000) formée par des fourches qui proviendraient d'une origine très proche du télomère (pointillés bleus). Cette zone de terminaison est absente des profils TP et RFD de la SY14. Le nom des ARS est indiqué uniquement par leur numéro.

Les analyses RFD, IP et *timing* sont en accord complet pour seulement 5 origines. Ce faible nombre peut s'expliquer par la différence de puissance statistique (Figure V-10a, Figure supplémentaire VIII-

2). Premièrement, l'IP prend en compte uniquement les lectures portant deux fourches divergentes, tandis que le RFD agrège les informations provenant de toutes les fourches de réplication. Deuxièmement, la hauteur d'un pic de *timing* de réplication dépend à la fois de l'efficacité d'une origine et de son moment de déclenchement, et permet d'estimer l'efficacité d'une origine moins précisément qu'un profil RFD. De plus, la modification d'une zone d'initiation en RFD, ou de l'efficacité d'une origine en IP, peut être la conséquence d'une modification des origines voisines. Par exemple, une origine possédant un délai d'activation peut entraîner une augmentation de l'efficacité des origines voisines due à leur moindre passivation. 7 origines apparaissent significativement modifiées sur le RFD et/ou l'IP en raison d'une modification du *timing* d'une origine voisine (Figure V-10b, Tableau supplémentaire VII-3, Figure supplémentaire VIII-3, VIII-4, VIII-5). 2 origines montrent une modification significative de leur *timing* de réplication qui impacte une zone de terminaison sur le RFD (Figure V-10b, Tableau supplémentaire VII-3, Figure supplémentaire VIII-6). De plus, 26 origines modifiées en RFD et/ou en IP sont dans des régions tardives (Figure V-10b b, Tableau supplémentaire VII-3, Figure supplémentaire VIII-7, VIII-8 et VIII-9). Pour prendre en compte les différences de puissances statistiques de chaque analyse, une rapide inspection visuelle des origines a été effectuée afin de déterminer si une modification visible mais non significative était présente sur les autres profils (Figure V-10b, Tableau supplémentaire VII-3, Figure supplémentaire VIII-8, VIII-10, VIII-11, VIII-12, VIII-13). 4 origines sont modifiées sur le RFD et le *timing* sans modification visible de l'IP (Figure V-10b, Tableau supplémentaire VII-3, Figure supplémentaire VIII-14). 6 origines sont détectées uniquement en RFD (Figure V-10b, Tableau supplémentaire VII-3, Figure supplémentaire VIII-15) et 3 origines sont détectées uniquement sur le *timing* de réplication (Figure V-10b, Tableau supplémentaire VII-3, Figure supplémentaire VIII-16).

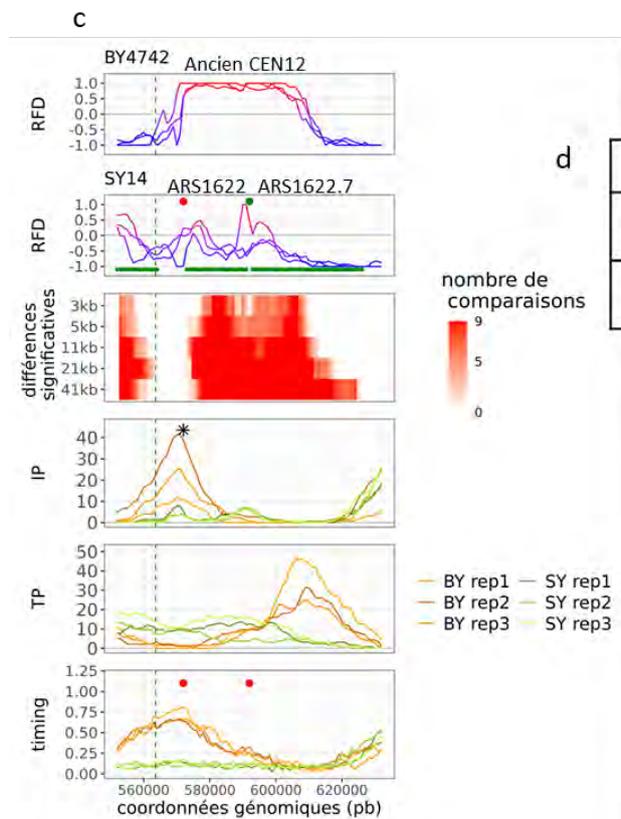
Pour la suite de l'analyse, nous avons considéré les 70 origines comme modifiées. Ces origines ont été classées en deux catégories, en fonction du type de modification. Une origine possédant un *timing* de réplication plus tardif et/ou un  $\Delta$ RFD plus faible et/ou un pic IP plus faible sera considérée comme « diminuée », tandis qu'une origine possédant un *timing* de réplication plus précoce et/ou un delta RFD plus fort et/ou un pic IP plus important sera considérée comme « augmentée ». Il est nécessaire de garder à l'esprit que la modification d'une origine peut être la conséquence de la modification des origines voisines, c'est pourquoi nous avons aussi déterminé si une origine diminuée était répliquée passivement par une origine voisine possédant un *timing* plus précoce et si une origine augmentée bénéficiait d'une moindre passivation par les origines voisines en raison de leur *timing* plus tardif (Tableau supplémentaire VII-3). Il existe une origine qui possède un *timing* de réplication plus tardif mais une activité significativement augmentée en RFD (Figure V-10c). Dans cette région, deux origines montrent un *timing* de réplication très retardé dans la SY14. L'origine de gauche voit son activité

diminuée en RFD, mais celle de droite voit son activité augmentée en RFD. Cette origine est considérée comme diminuée dans la suite de l'analyse, l'effet majeur étant un fort retard de déclenchement et l'augmentation de l'activité en RFD étant, sans doute, un effet secondaire de la diminution d'activité de l'origine de gauche. Au total, 47 origines sont considérées comme diminuées et 23 comme augmentées (Figure V-10d).



b

	Nombre d'origines
Voisin	7
TER	2
Tardive	26
IPL non significatif	19
Timing non significatif	6
RFD non significatif	5
IPL seul	0
Timing seul	3
RFD seul	6



d

Augmentée	Modifiées		Non modifiées
	Augmentée	Diminuée	
24	46	238	

### Figure V-10 : Résultats des différentes analyses statistiques.

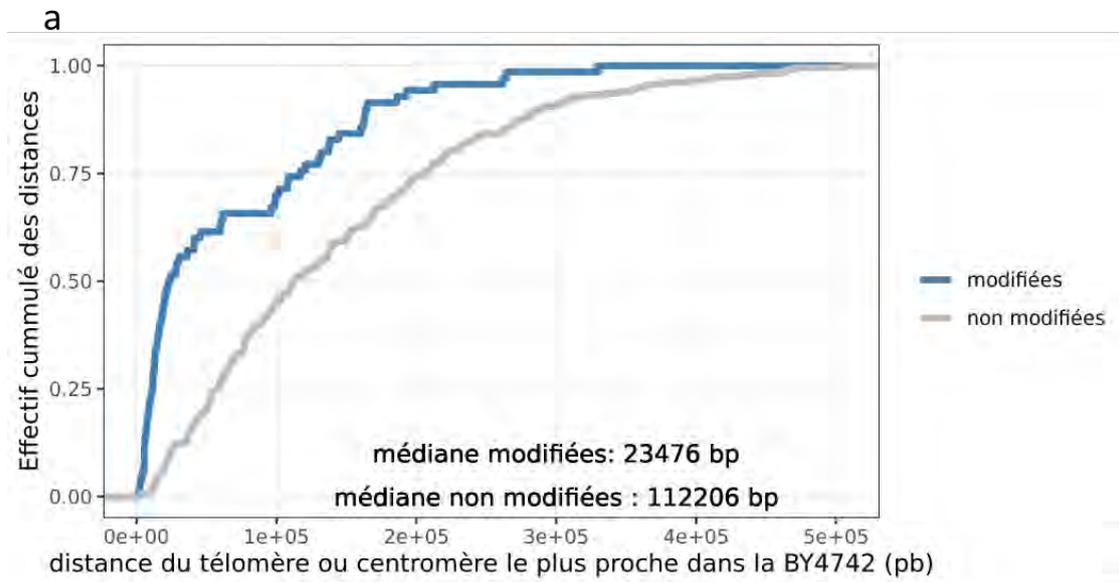
a) Diagramme de Venn représentant le nombre d'origines détectées par chaque analyse statistique ainsi que le nombres d'origines communément détectées par les analyses. b) Tableau représentant le nombre d'origines en fonction de différentes catégories. Voisin : nombre d'origines modifiées en RFD et/ou IP possédant une modification du *timing* de réplication d'une origine voisine. TER : nombre d'origines modifiées en *timing* engendrant uniquement sur le RFD la modification d'une zone de terminaison. Tardive : origines modifiées en IP et/ou RFD situées dans une région tardive. Modification IP/*timing*/RFD : modification perceptible visuellement. IP/*timing*/RFD seul : origines étant modifiées significativement sur un des profils seulement, aucune modification n'est visible sur les autres profils. c) Origine possédant un *timing* retardé et une efficacité augmentée en RFD. Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercle rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les zones de terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro. d) Tableau représentant le nombre d'origines détectées comme « augmentées », « diminuées » et « non modifiées ».

#### V.3.2 La majorité des changements sont centromériques ou télomériques

Lors de la construction de la SY14, les régions centromériques et télomériques ont été supprimées. On s'attend donc à observer des changements au voisinage de ces régions, mais pas forcément à plus grande distance. Pour tester cette hypothèse, nous avons déterminé la distance entre les origines actives, le centromère et les télomères ou fusion chromosomique de leur chromosome d'appartenance dans la BY4742, avant de calculer la fonction de répartition empirique des distances à l'élément le plus proche (Figure V-11a). Les origines non modifiées sont plus éloignées de ces éléments (distance médiane 112 kb) que les origines modifiées (distance médiane 23kb). La majorité des modifications observées semble donc liée à un effet de la suppression des centromères et des télomères qui se propagent sur une distance d'environ 20kb.

Afin de regarder la répartition des origines modifiées et non modifiées en fonction de leur distance à l'élément le plus proche, les origines ont été classées en trois groupes d'effectifs similaires définis par les quantiles de distances à 0,33 et 0,66. La première classe comprend les origines situées à une distance inférieure ou égal à 55kb d'un centromère ou d'un télomère. Cette classe regroupe 58 origines non modifiées, 24 origines diminuées et 19 origines augmentées. La seconde classe comprend les origines situées à une distance supérieure à 55kb et inférieure ou égal à 151kb. Elle regroupe 84 origines non modifiées, 13 origines diminuées et 3 origines augmentées. La dernière classe comprend toutes les origines situées à une distance supérieure à 151kb. Elle regroupe 93 origines non modifiées, 9 origines diminuées et 2 origine augmentée (Figure V-11). Ces résultats

montrent que certaines modifications dans la SY14 peuvent avoir lieu à une distance élevée des télomères et des centromères, même si la majorité des modifications est observée à proximité de ces éléments.



**b**

	Modifiées		Aucune modification	Total
	Augmentées	Diminuée		
<b>Proches</b>	<b>19</b>	<b>24</b>	<b>58</b>	<b>101</b>
<b>Moyennes</b>	<b>3</b>	<b>13</b>	<b>84</b>	<b>100</b>
<b>Eloignées</b>	<b>2</b>	<b>9</b>	<b>93</b>	<b>104</b>

**Figure V-11 : Distance entre les origines modifiées, les centromères et les télomères.**

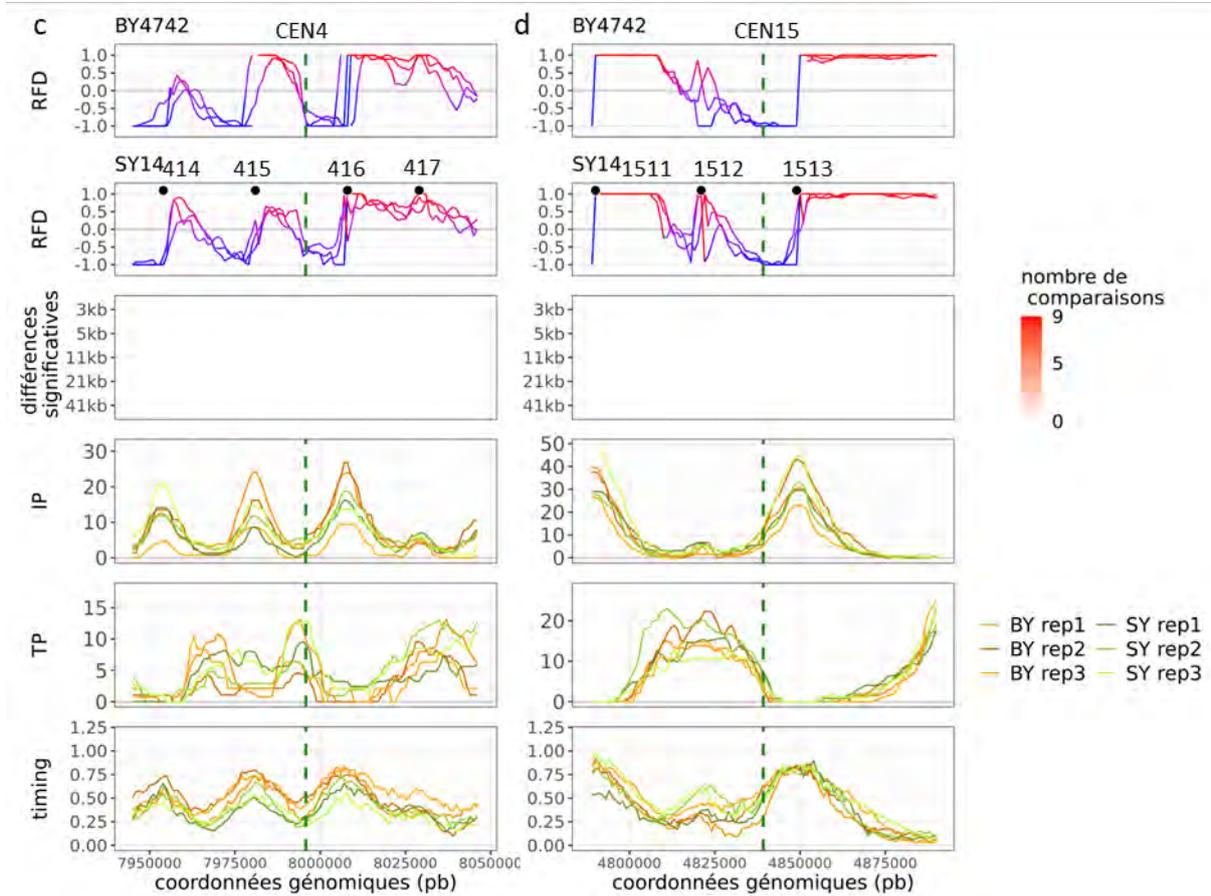
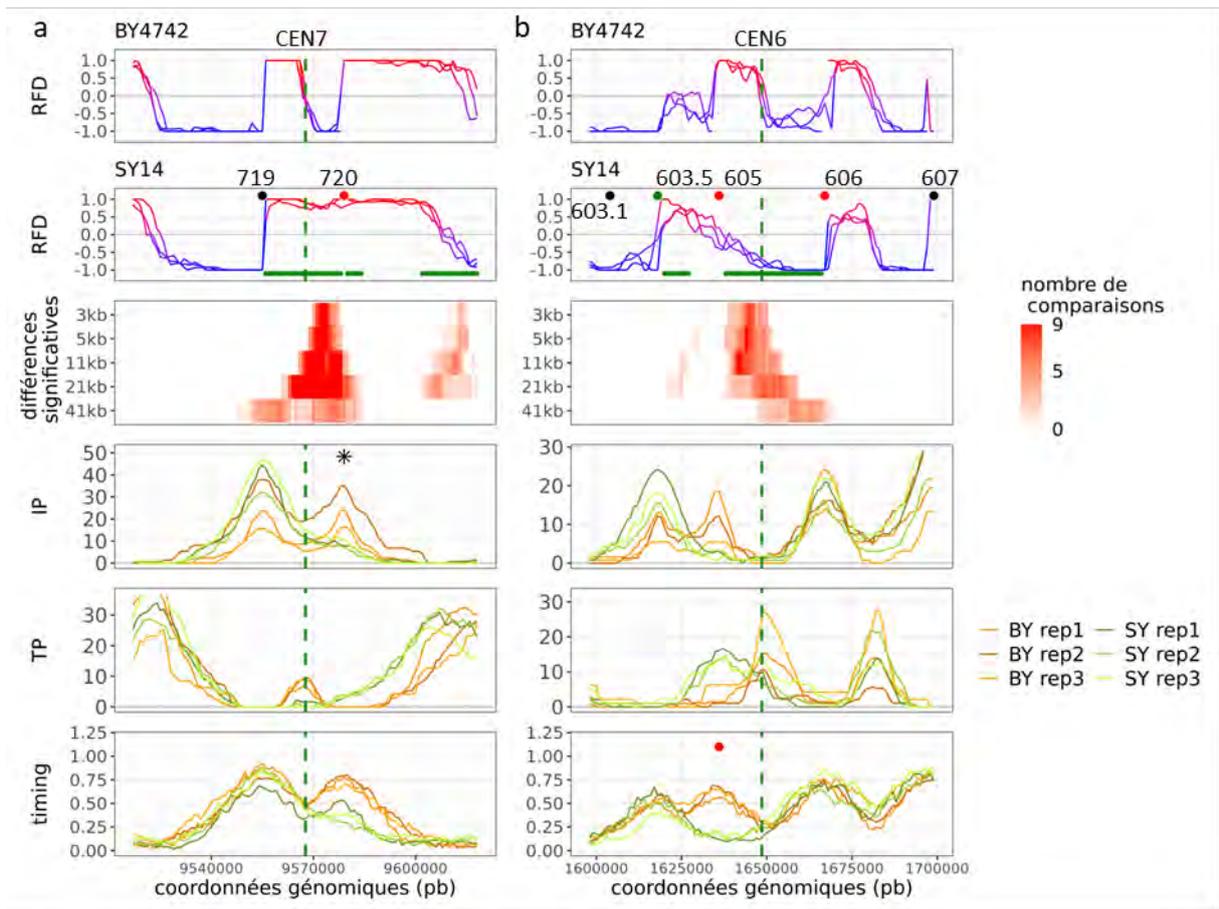
a) Fonction de distribution empirique de la distance entre les origines et le télomère ou le centromère le plus proche. Origines modifiées n=70, origines non modifiées n=235. La distance au centromère ou télomère le plus proche a été calculée sur le génome de la SY14 en prenant par chromosome la distance entre chaque origine et la coordonnée du début du centromère convertie en coordonnées SY14 ou la position des fusions chromosomiques, ou des télomères non modifiés. La distance la plus courte a ensuite été choisie. b) Répartition des modifications des origines chez la SY14 en fonction de la distance au centromère ou télomère le plus proche. Les origines actives ont été classées en fonction de leur distance au centromère ou télomère, en trois groupes comprenant chacun 1/3 des origines. Ainsi les origines proches sont à une distance inférieure à 55kb, les origines moyennes à une distance comprise entre 55 et 151 kb et les origines éloignées à une distance supérieure à 151kb.

### V.3.2.1 Modifications des origines à proximité des centromères.

La classe d'origines « proches » comprend 52 origines à proximité d'un centromère, dont 6 possèdent une activité augmentée et 19 une activité diminuée (Tableau V-1, Figure supplémentaire VIII-17). Il semblerait donc que la suppression des centromères ait diminué l'efficacité d'origine proche et retardé leur activation (Figure V-12b-c). Les 6 origines plus actives dans la SY14 sont toutes à proximité d'une origine inhibée. Il est très probable que l'augmentation de l'activité de ces origines soit due à leur moindre passivation par les origines retardées (6/19, Figure V-12c). Néanmoins, les deux centromères *CEN15* et *CEN4* ne montrent pas de modifications significatives des origines dans leur proximité (Figure V-12d). Le *CEN15* est toujours présent dans la SY14, donc l'absence de modification est attendue et implique qu'il n'y a pas d'effet en *trans* de la suppression des autres centromères sur le centromère restant. L'absence de modifications significatives du *CEN4* pourrait s'expliquer par une régulation différente de celles des autres centromères.

	Modifiées		Non modifiées	Total
	Augmentées	Diminuées		
Proches	6	19	27	52

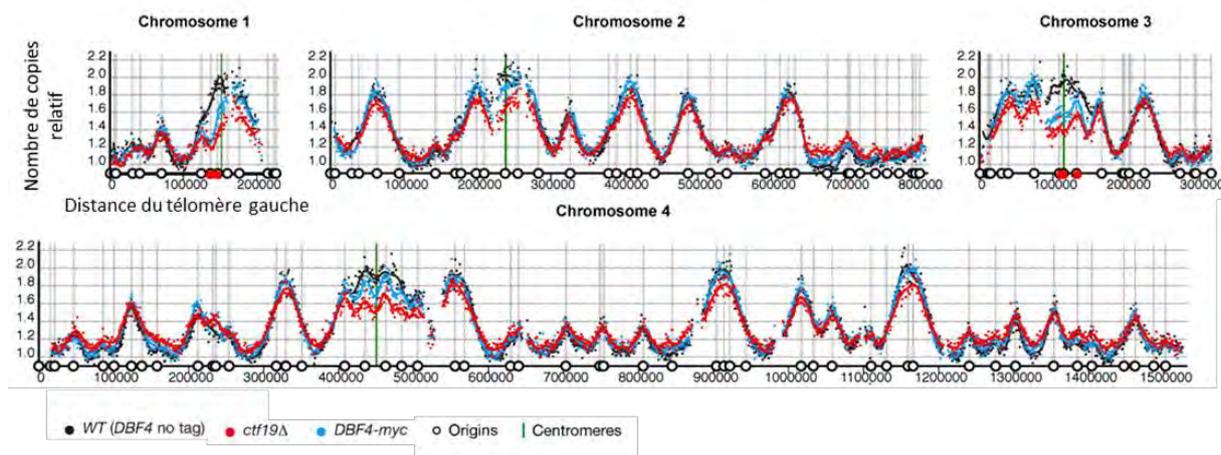
Tableau V-1 : Modifications des origines à proximité des centromères.



### Figure V-12 : Exemple de modifications au niveau des centromères.

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un timing de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un timing significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les zones de terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro. a) Exemple de délai dans l'activation d'une origine centromérique associé à une diminution de son efficacité. b) Exemple de trois origines voisines modifiées au voisinage d'un centromère. L'ARS603.5 voit son activité stimulée, l'ARS605 abolie et possède un retard significatif de son moment de réplication, et l'ARS606 possède une activité diminuée. c) CEN4 qui ne montre aucune modification significative, cependant le  $\Delta$ RFD de l'ARS415 est plus faible dans la SY14 et son timing, légèrement plus tardif. d) CEN15 encore présent dans la SY14. Il ne présente aucune modification significative, cependant le  $\Delta$ RFD de l'ARS1512 semble être plus élevé.

Afin de déterminer si la régulation du *timing* de réplication des origines péricentromériques est différente pour le CEN4, nous avons comparé les origines impactées dans la SY14 avec les origines impactées par la suppression de Ctf19. Il s'agit d'une protéine du kinétochore impliquée dans le maintien d'un *timing* de réplication précoce des origines centromériques (Natsume et al, 2013). Nous avons supposé que la suppression des centromères dans la SY14 entraînerait un phénotype répliatif similaire à celui d'une souche Ctf19 $\Delta$ . Afin de déterminer les origines modifiées en absence de Ctf19, j'ai utilisé les données de Natsume et al (2013). Dans un premier temps, à l'aide des figures représentant le *timing* en fonction de la distance du télomère gauche, j'ai considéré qu'une origine était modifiée par Ctf19 si en absence de Ctf19, et dans la souche Dbf4-myc (qui bloque la fixation de Dbf4 à Ctf19), le *timing* était visiblement retardé. 25 origines sont modifiées selon ce critère qualitatif (Figure supplémentaire VIII-18, origines en rouges). Parmi ces origines, 24 sont des origines actives dans la SY14.



**Figure V-13 : Exemple de modifications de *timing* observées dans la souche *Ctf19Δ* (Natsume et al., 2013).**

*Timing* de réplication des chromosomes I, II, III et IV adapté de Natsume et al, 2013. Les origines considérées comme modifiées sont en rouge. L'origine à droite du centromère du chromosome II n'est pas considérée comme modifiée, en raison de la forte variabilité dans les points correspondant au *timing* de la souche sauvage et du mutant *Dbf4*. Les deux origines bordant le centromère du chromosome IV ne sont pas considérées comme modifiées à cause de la faible différence entre la souche *DBF4-myc* et la souche sauvage.

Comme attendu, la majorité des origines diminuees dans la SY14 à proximité des centromères présente une inhibition en absence de Ctf19 (18/19, Tableau VI-2, Tableau supplémentaire VII-4). Les origines à proximité du CEN4 qui ne sont pas significativement modifiées dans la SY14 ne le sont pas non plus en absence de Ctf19 (Tableau supplémentaire VII-4, Figure V-13). Il est donc possible que ces origines possèdent un *timing* de réplication précoce indépendant d'un effet de Ctf19 et du centromère. 6 origines activées par Ctf19 ne sont pas modifiées dans la SY14. Parmi celles-ci, on retrouve comme attendu deux origines situées à proximité du CEN15 qui n'a pas été éliminé dans la SY14. Les autres sont situées au voisinage des CEN11, CEN8 et CEN14 (Tableau supplémentaire VII-4). Ces résultats suggèrent que certaines origines ne sont pas impactées de la même manière par la perte du centromère et la mutation de Ctf19. Néanmoins, l'analyse des origines modifiées dans la souche *Ctf19Δ* a été effectuée visuellement, contrairement à l'analyse des modifications dans la SY14. De plus, les souches de levures utilisées sont différentes, Natsume et al (2013) ayant utilisé la souche W303 pour leur étude.

Pour conclure, la majorité des origines péricentromériques montrent un délai d'activation et une diminution d'efficacité induits par la suppression des centromères et potentiellement l'absence de site de fixation pour Ctf19. Cette moindre efficacité entraîne une moindre passivation de certains origines voisines leur permettant de devenir plus efficaces.

	<b>Proche Centromères</b>		
	Augmentées	Diminuées	Non modifiées
Non régulées	6	1	21
Activées	0	18	6

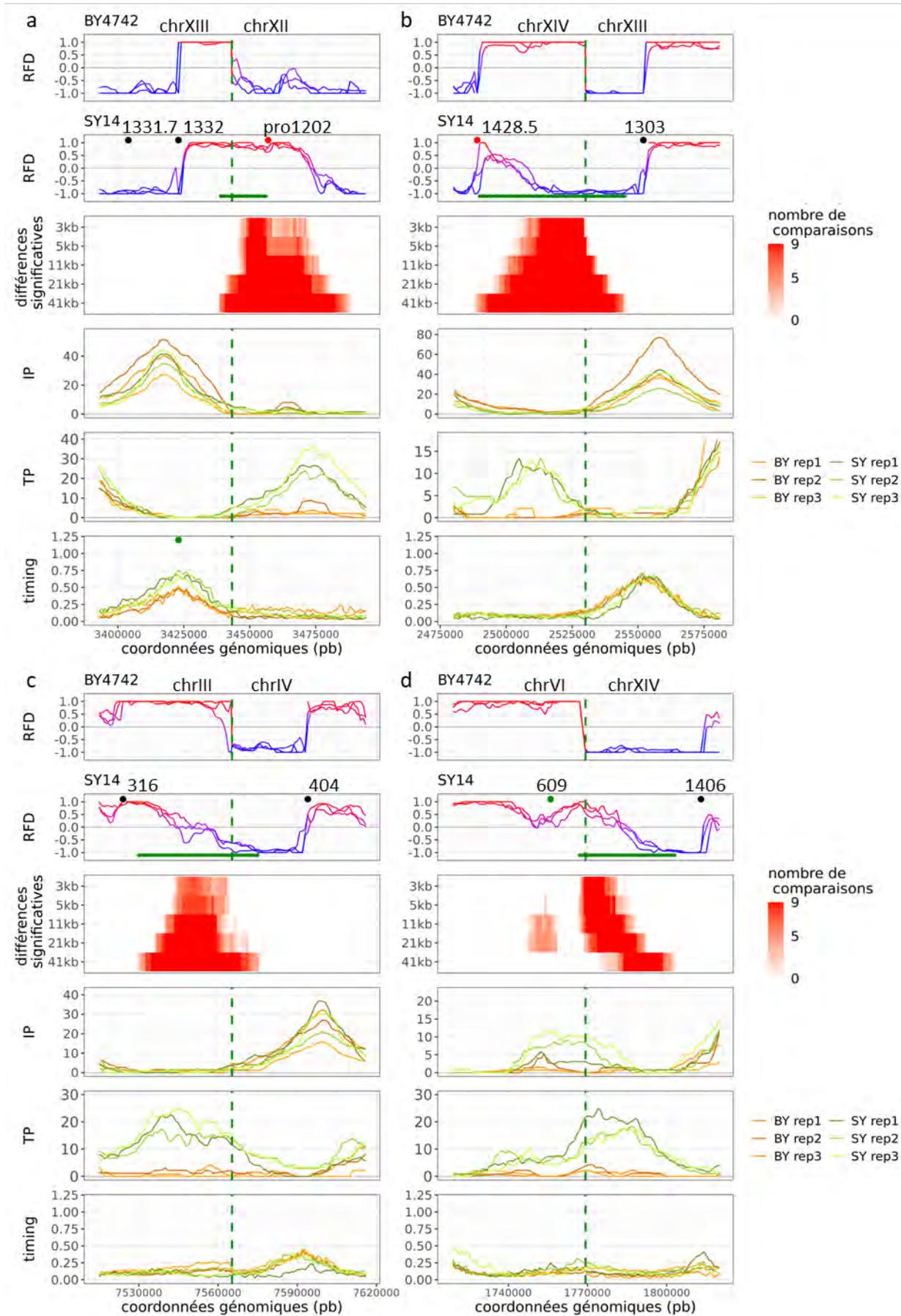
**Tableau V-2 : Tableau récapitulatif du nombre d'origines activées et non modifiées par Ctf19 dans la catégorie « proches » à proximité d'un centromère.**

### V.3.2.2 Modification des télomères

La classe d'origines « proches » comprend 49 origines à proximité d'un télomère (Tableau V-3), dont 5 possèdent une activité diminuée dans la SY14 et 13, une activité augmentée. La diminution de l'activité observée dans la SY14 provient majoritairement de la réplication passive des origines tardives proches des télomères par une origine précoce localisée sur le chromosome adjacent (4/5, Figure IV-14a,b , Figure supplémentaire VIII-19). En effet la fusion chromosomique permet à une fourche émanant de cette origine de traverser la jonction. 11 jonctions (11/15) montrent une traversée des fourches de réplication (Figure IV-14, Figure supplémentaire VIII-19).

	<b>Modifiées</b>		<b>Non Modifiées</b>	<b>Total</b>
	Augmentées	Diminuées		
Proches	13	5	31	49

**Tableau V-3 : Modification des origines situées à des télomères.**



### Figure V-14 : Modifications des origines télomériques.

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les zones de terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro. a) Exemple de fusion chromosomique entraînant l'avancée du *timing* de réplication d'une origine (ARS1332) et la réplication passive d'une autre origine (proARS1202) située de l'autre côté de la jonction. b) Exemple de fusion chromosomique entraînant la réplication passive d'une origine (ARS1428.5) par les fourches provenant de l'origine de l'autre côté de la fusion (ARS1303). c) Exemple de fusion chromosomique entraînant uniquement le passage des fourches d'un chromosome à l'autre, déplaçant la zone de terminaison. d) Exemple de fusion chromosomique entraînant une augmentation de l'activité d'une origine.

La majorité des modifications au voisinage des télomères consiste en une augmentation de l'activité des origines. Les télomères sont des régions de réplication tardive. Leur absence dans la SY14 semble donc associée à une activation plus efficace et/ou plus précoce des origines de leur voisinage (Figure V-14d). Parmi les protéines impliquées dans le maintien d'une réplication tardive des télomères, on trouve Rif1 et yKU70 (Cosgrove et al., 2002; Davé et al., 2014; Hafner et al., 2018; Lian et al., 2011; Peace et al., 2014). Rif1 se lie à Rap1 au niveau des télomères et inhibe la réplication de ces régions en recrutant la phosphatase PP1 qui contrecarre la phosphorylation des MCM par Cdc7, étape moléculaire importante dans l'activation des origines (Hiraga et al., 2014; Mattarocci et al., 2014). L'impact d'yKU70 sur la réplication est indirect. En absence de cette protéine, les télomères deviennent plus courts, ce qui induit leur réplication précoce. On peut donc supposer que l'absence de sites de fixation pour Rap1 due à la suppression des télomères ne permet plus à Rif1 d'inhiber les origines proches des télomères, ou que l'absence de télomères induit un mécanisme d'avancement de la réplication similaire à celui observé pour une souche possédant des petits télomères.

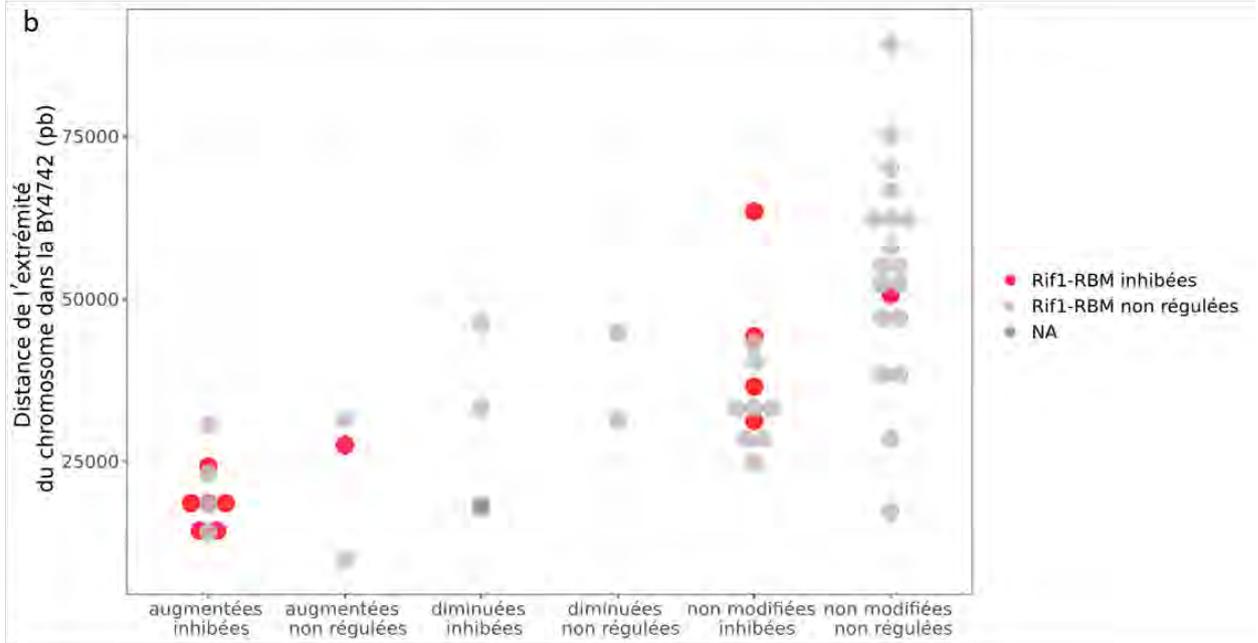
Les origines à proximité d'une fusion chromosomique ne sont pas toutes modifiées. Nous avons donc cherché ce qui différencie les 13 origines augmentées des autres. Dans un premier temps, nous avons comparé les origines impactées dans la SY14 avec les origines impactées par la suppression de Rif1 ou yKU70. Hafner et al (2018) ont analysé la réplication dans une souche *Rif1Δ* et identifié 54 origines inhibées par Rif1, dont 26 font partie des origines actives dans la SY14. Les données utilisées pour l'analyse d'yKU70 sont celles de Das et al (2015), qui mettent en évidence l'inhibition par cette protéine de 51 origines, dont 18 sont actives dans la SY14. Parmi les 13 origines augmentées dans la SY14, 10 origines sont inhibées par Rif1 et/ou yKU70 (Figure V-15a, Tableau supplémentaire VII-5). 12 origines qui ne sont pas modifiées sont également inhibées par Rif1 et/ou yKU70 (Figure V-15a,

Tableau supplémentaire VII-5). L'inhibition par Rif1 et/ou  $\gamma$ KU70 n'est donc pas un critère suffisant pour expliquer l'augmentation d'activité des origines dans la SY14.

Nous nous sommes alors intéressés à la distance entre les origines et l'extrémité des chromosomes avant suppression des télomères (Figure V-15b). La majorité des origines inhibées par Rif1 et/ou  $\gamma$ KU70 et augmentées dans la SY14 est située à moins de 25kb de l'extrémité d'un chromosome dans la BY4742 (9/10), tandis que les origines inhibées par Rif1 et/ou  $\gamma$ KU70 et non modifiées dans la SY14 et la majorité des origines augmentées et non régulées par Rif1 et/ou  $\gamma$ KU70 sont situées à une distance supérieure à 25kb. Il existe une origine augmentée située à moins de 25kb de l'extrémité du chromosome avant la suppression des télomères, qui n'est pas régulée par Rif1. Nous avons vérifié que cette origine ne possédait pas de site de fixation pour Rif1. Hafner et al (2018) ont identifié les sites de fixation de Rif1 par ChEC-seq. Nous avons vérifié la présence de site de fixation pour Rif1 pour l'ensemble des 49 origines proches d'un télomère. Pour cela nous avons utilisé les coordonnées des origines provenant de la base de données OriDB (Siow et al) et nous avons regardé les sites de fixations pour Rif1 à +/- 2.5kb de l'ARS (Figure supplémentaire VIII-20). Les 13 origines augmentées dans la SY14 possèdent un site de fixation pour Rif1 à proximité de l'ARS, mais de nombreuses origines non régulées et non modifiées par Rif1 possèdent aussi des sites de fixation pour cette protéine. Il semblerait donc qu'il existe un lien entre la position de l'origine et l'augmentation de son activité dans la SY14. Il est possible que les origines les plus proches de l'extrémité des chromosomes soient plus sensibles à la liaison entre Rif1 et Rap1 quant à leur inhibition. Hafner et al (2018), ont utilisé un mutant de Rif1 incapable de se lier à Rap1 (*Rif1-RBM*) afin de déterminer quelles origines étaient dépendantes de la liaison entre Rif1 et Rap1 pour leur inhibition. Si les origines augmentées dans la SY14 sont dépendantes de la liaison entre Rif1 et Rap1, elles ne devraient pas être inhibées par ce mutant. Or la proportion d'origines augmentées régulées et non régulées par *Rif1-RBM* est similaire (respectivement 7/13 et 6/13, Figure V-15c, Tableau supplémentaire VII-5), et la majorité des origines non modifiées ne sont pas affectées par *Rif1-RBM* (27/32, Figure V-15c). Il est donc peu probable que la suppression des sites de liaison pour Rap1 suite à la suppression des télomères soit responsable de la modification de l'activité des origines proches des télomères.

a

	Proches Télomères			
	Augmentées	Diminuées	Non modifiées	Total
Non régulées	3	2	19	24
Inhibées	10	3	12	25



c

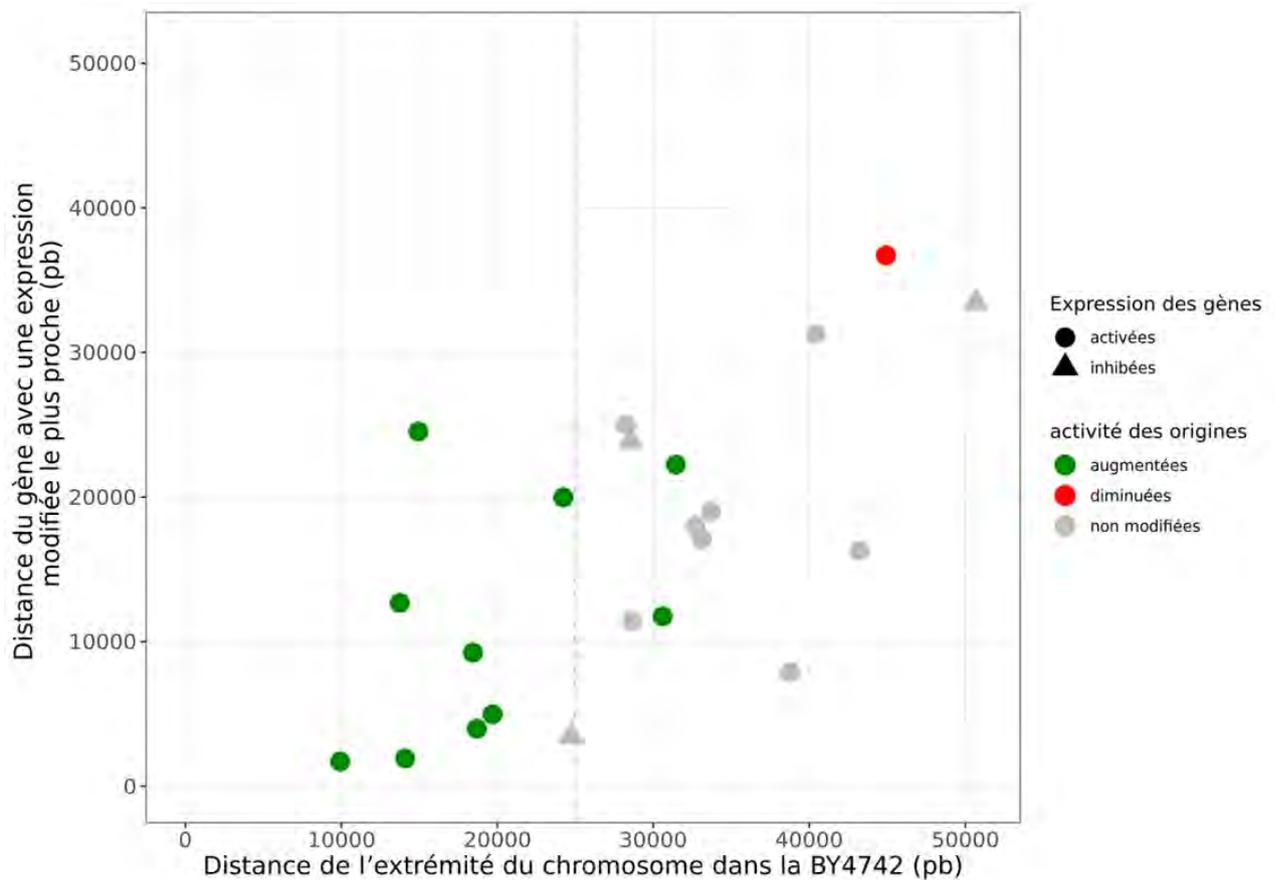
	Proches Télomères			
	Augmentées	Diminuées	Non modifiées	Total
Non régulées	6	4 + 1	26	37
Inhibées	7	0	5	12

**Figure V-15 : Modifications des origines télomériques par Rif1, yKU70 et Rif1-RBM.**

a) Tableau récapitulatif du nombre d'origines inhibées et non modifiées par Rif1 et/ou yKU70 dans la catégorie « proches » à proximité d'une fusion chromosomique ou d'un télomère. b) Distribution des origines par Rif1 et yKU70 en fonction de la distance entre l'origine et l'extrémité du chromosome avant la suppression des télomères. c) Tableau récapitulatif du nombre d'origines inhibées et non modifiées par Rif1-RBM dans la catégorie « proches » à proximité d'une fusion chromosomique ou d'un télomère, les chiffres en gris représente le nombre d'origines absentes des données de Hafner et al (2018).

Les protéines Sir, impliquées dans l'inhibition de l'expression des gènes situés dans les régions télomériques, sont elles aussi responsables d'une inhibition des origines à proximité des télomères (Stevenson et Gosthchling). Shao et al. (2018) ont observé une augmentation de l'expression de certains gènes situés à proximité des fusions chromosomiques. Il est donc possible que la modification

de l'expression des gènes soit corrélée à une activation plus précoce des origines. Parmi les origines augmentées dans la SY14 à proximité des fusions chromosomiques, 10 sont situées à moins de 30kb d'un gène possédant une expression augmentée à proximité d'un télomère (Tableau supplémentaire VII-5, Figure V-16). Et l'origine à moins de 25kb des télomères qui n'est pas inhibée par Rif1, mais qui possède un site de fixation pour la protéine, est à proximité d'un gène activé dans la SY14. De plus, l'origine qui n'est pas modifiée dans la SY14 mais est située à moins de 25kb de l'extrémité du chromosome est à proximité d'un gène inhibé dans la SY14. L'absence d'activation de cette origine pourrait donc être liée à un contexte répressif à proximité de cette fusion.



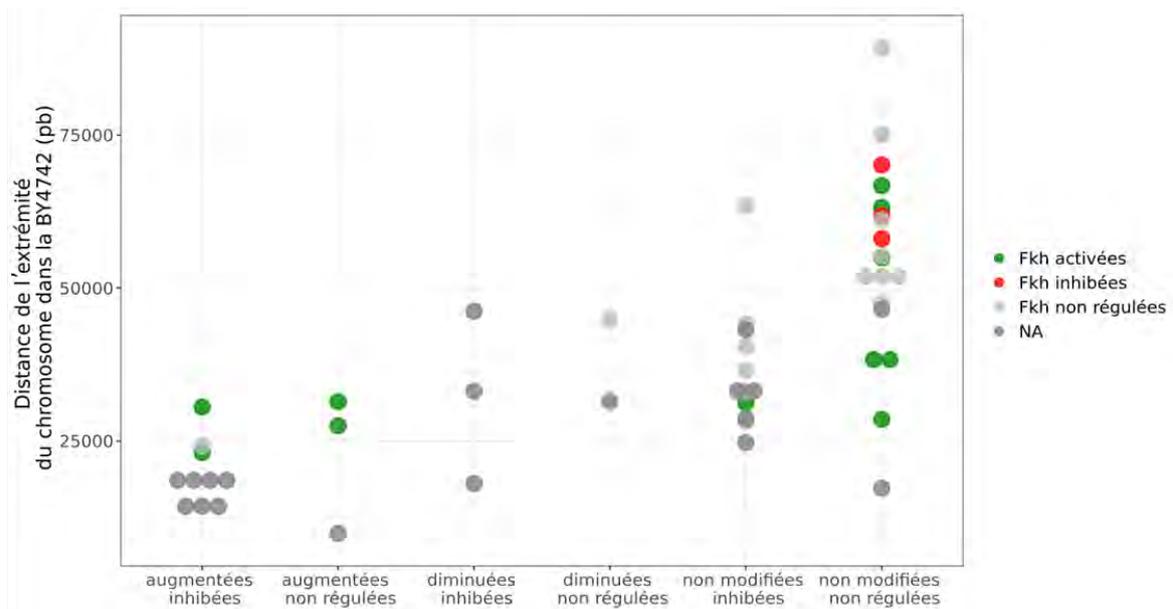
**Figure V-16 : Modification des origines à proximité des télomères et modification de l'expression des gènes.**

Seuls les gènes situés à moins de 25kb d'un télomère et modifiés dans la SY14 ont été utilisés. La distance entre les origines et les gènes modifiés a été calculée pour les deux extrémités du gène et la distance la plus courte a été utilisée.

Pour conclure, l'augmentation de l'activité des origines proches des fusions chromosomiques semble liée à leur proximité aux télomères, à une perte d'inhibition des origines par Rif1 et à une perte du contexte répressif imposé par les protéines Sir. Ces résultats montrent que la perte du contexte

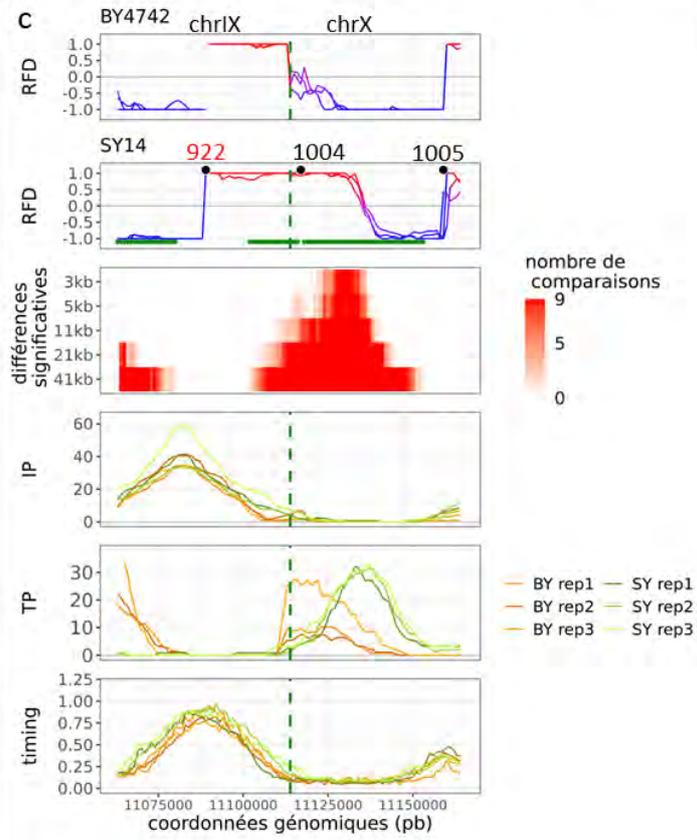
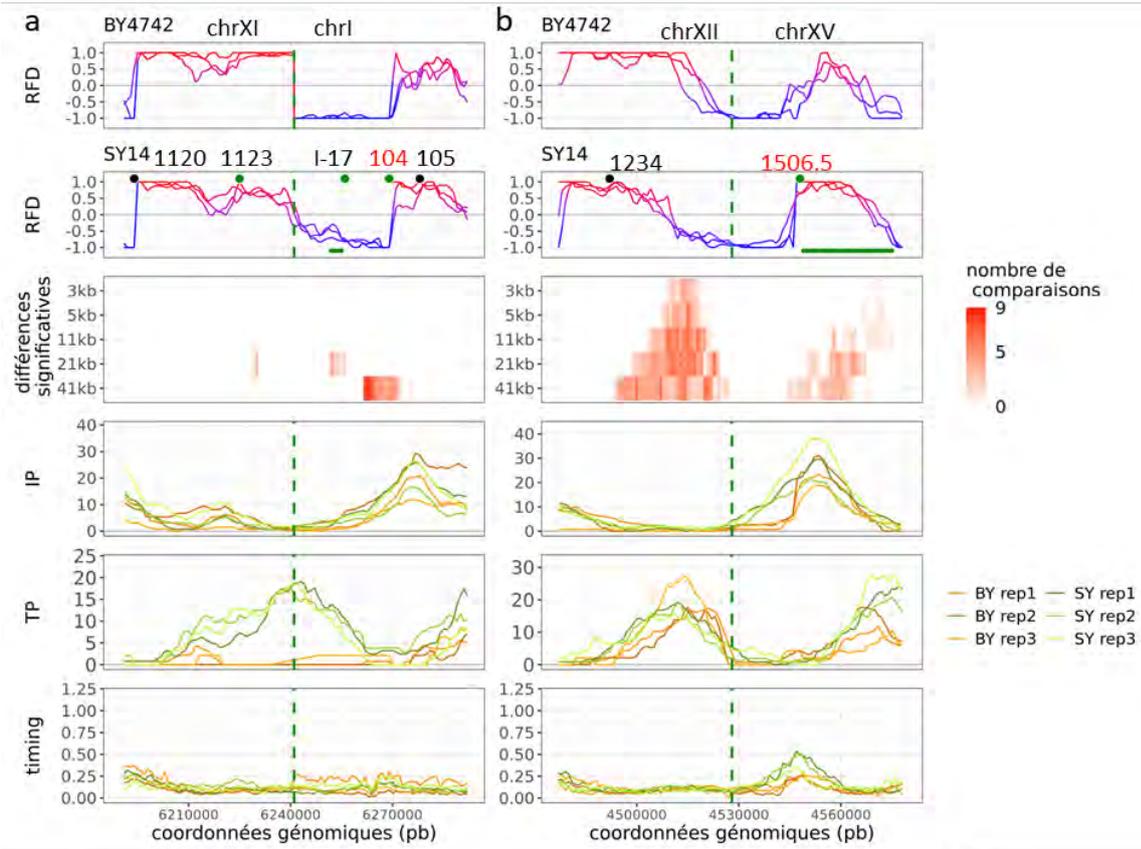
télomérique global a impacté la régulation des origines. Nous pouvons aussi envisager que la modification de ces origines est entraînée par leur relocalisation. L'absence du contexte périnucléaire habituel pourrait créer une zone où Rif1 et les protéines Sir ont plus de difficultés à agir.

On note que 3 origines sont augmentées à proximité des télomères et se situent à plus de 25kb de l'extrémité d'un chromosome dans la SY14. Ces origines sont activées par Fkh1-Fkh2 (Knott et al, 2012, Figure V-17, Tableau supplémentaire VII-5). Cependant, deux origines avec une proximité similaire, activées par Fkh1/2, ne sont pas modifiées dans la SY14. Ces origines possèdent une efficacité maximale en RFD ainsi qu'un *timing* de 1 (Figure VI-14b ARS1303, Figure V-18c, origine en rouge), tandis que les trois autres origines possèdent un *timing* inférieur à 0,5 (Figure V-14a ARS1332, Figure V-18a, b origine en rouge). Il est possible que l'effet de Fkh1-Fkh2 soit donc déjà maximal pour certaines origines et que la suppression des télomères n'ait pas entraîné un effet plus important contrairement aux autres origines. Il semblerait donc que la suppression des télomères et la fusion des chromosomes ont permis à Fkh1-Fkh2 d'avoir une activité activatrice plus importante sur les origines à proximité des télomères.



**Figure V-17 : Modifications des origines à proximité des télomères par Fkh1/2.**

Modifications des origines par Fkh1/2 en fonction de la distance entre l'origine et l'extrémité du chromosome avant la suppression des télomères, de leur modification par Rif1 et/ou yku70, inhibées ou non régulées, et de leur modification dans la SY14 augmentées, diminuées ou non modifiées.



**Figure V-18 : Représentation des origines comprise entre 25kb et 35kb de l'extrémité d'un chromosome dans la BY4742.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiqués par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les zones de terminaisons modifiées sont représentées par des lignes vertes. Les origines à moins de 35kb de l'extrémité d'un chromosome dans la BY4742, activées par Fkh1/2 sont en rouge. Le nom des ARS est indiqué uniquement par leur numéro. a) L'ARS104 est activée par Fkh1/2 et son activité est augmentée dans la SY14. b) L'ARS1506.5 est activée par Fkh1/2 et possède une activité augmentée dans la SY14. c) L'ARS922 est activée par Fkh1/2, mais n'est pas modifiée dans la SY14, cependant son *timing* est de 1 et son  $\Delta$ RFD de 2.

**V.3.3 Régulation des origines à une distance éloignée ou moyenne.**

27 origines modifiées dans la SY14 sont classées parmi les origines « moyennes » ou « éloignées » des centromères et des télomères. Parmi celles-ci, 5 sont augmentées et 22 sont diminuées (Tableau V-4). Aucune n'est répliquée passivement par une origine voisine plus active et aucune origine n'est plus efficace en raison de sa moindre passivation par une origine voisine.

	Modifiées		Aucune modification	Total
	Augmentées	Diminuée		
<b>Moyennes</b>	<b>3</b>	<b>13</b>	<b>84</b>	<b>100</b>
<b>Éloignées</b>	<b>2</b>	<b>9</b>	<b>93</b>	<b>104</b>

**Tableau V-4 : Récapitulatif du nombre d'origines modifiées et non modifiées dans les classes de distances « moyennes » et « éloignées » des centromères et des télomères.**

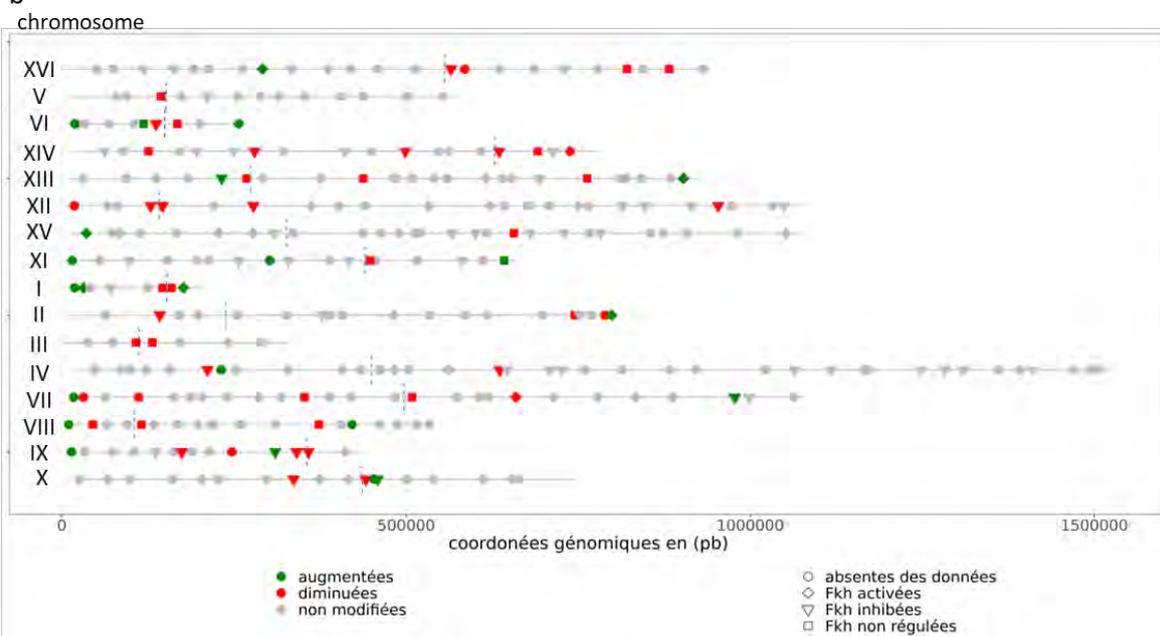
Trois origines modifiées à proximité des télomères sont activées par Fkh1/2. Nous avons dans un premier temps regardé la régulation des 27 origines modifiées dans la SY14 par Fkh1/2. Les données utilisées pour l'analyse de Fkh1 Fkh2 sont celles de Knott et collaborateurs obtenues en analysant la réplication dans des souches Fkh1 $\Delta$ , Fkh2 $\Delta$ , Fkh1 $\Delta$ Fkh2 $\Delta$  (Knott et al., 2012). Comme le double mutant montre une croissance très ralentie, les auteurs ont aussi analysé cette souche en surexprimant une protéine Fkh2 avec sa partie C-terminale tronquée. Les données des différentes souches ont été regroupées sous la forme d'une seule modification, peu importe le mutant impliqué. Knott et al., (2012) ont identifié 110 origines activées par Fkh1/2 dont 66 font partie des origines actives dans la

SY14. Ils ont aussi identifié 91 origines inhibées par Fkh1/2, dont 64 font partie des origines actives dans la SY14. Fkh1/2 active seulement une origine augmentée dans la SY14 parmi les classes de distances « moyennes » et « éloignées » (Figure V-19a, Tableau supplémentaire VII-6). En revanche, 9 origines diminuées dans la SY14 sont inhibées par Fkh1/2, et 37 origines non modifiées sont inhibées par Fkh1/2, parmi les classes de distances « moyennes » et « éloignées » (Figure V-19a, Tableau supplémentaire VII-6). La Figure V-19b représente les différentes origines dans la SY14 en fonction de leur modification par Fkh1/2. Les origines régulées par Fkh1/2 et modifiées dans la SY14 ne semblent pas se regrouper ou présenter une localisation précise. L'ensemble de ces données suggère que Fkh1/2 ne sont pas responsables de la modification des origines « moyennes » et « éloignées » dans la SY14.

a

	Moyennes			Eloignées			Total
	Augmentées	Diminuées	Non modifiées	Augmentées	Diminuées	Non modifiées	
Non régulées	2	6 + 2	26 + 12	1	4	26 + 12	62 + 29
inhibées	1	5	12	0	4	25	73
activées	0	0	34	1	1	30	65

b



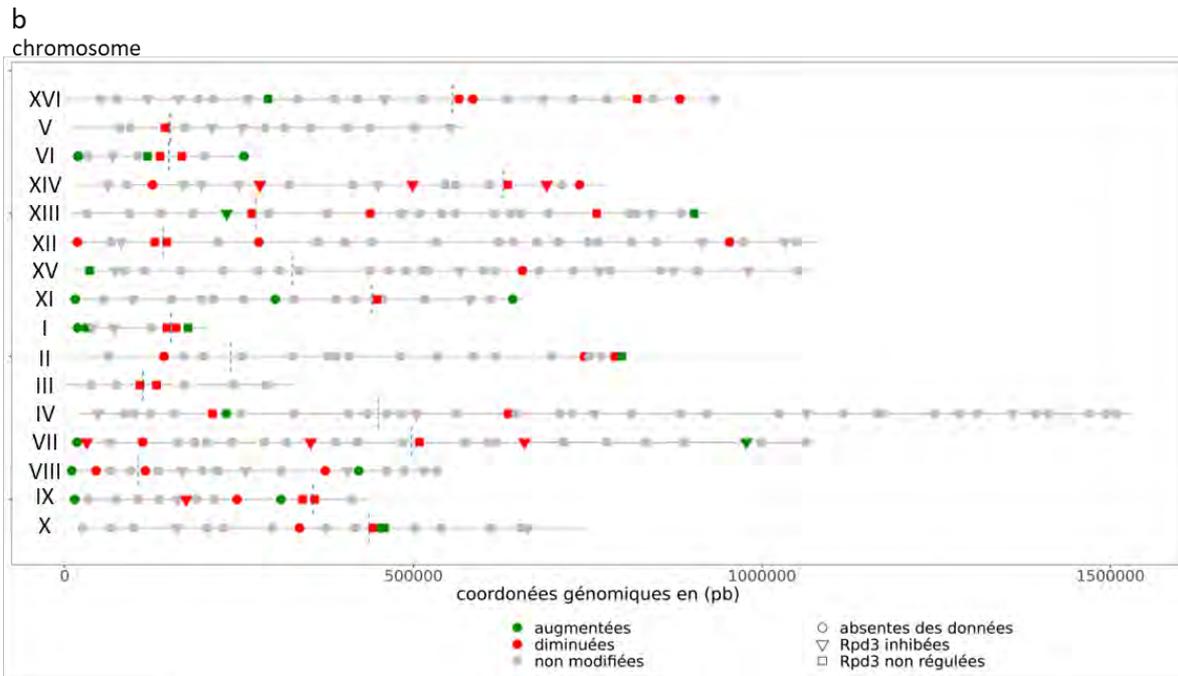
**Figure V-19 : Modification des origines à des distances « moyennes » et « éloignées » des centromères et des télomères par Fkh1/2.**

a) Tableau récapitulatif du nombre d'origines inhibées, activées et non modifiées par Fkh1/2 dans les classes de distances « moyennes » et « éloignées » des centromères et des télomères. Les chiffres en gris représentent le nombre d'origines absentes de données de Knott et al. (2012). b) Représentation des chromosomes de la BY4742 dans l'ordre du génome de la SY14. Haut : télomère gauche. Bas : télomère droit. Les centromères sont indiqués par les lignes pointillées. Les origines augmentées dans la SY14 sont en vert, les diminuées en rouge et les non modifiées en gris.

Rpd3 est une histone désacétylase impliquée dans l'inhibition de différentes origines. Afin de vérifier si les origines modifiées par cette protéine sont modifiées dans la SY14, nous avons comparé les origines inhibées par Rpd3 obtenues par Knott et al. (2009) aux origines modifiées dans la SY14. Knott et al. (2009) ont analysé 297 origines, dont 104 sont inhibées par Rpd3. Parmi les origines analysées, 191 sont actives dans la SY14. Seul une origine augmentée dans la SY14 est inhibée par Rpd3 (Figure V-20a). 6 origines diminuées dans la SY14 sont inhibées par Rpd3 et 34 origines non modifiées dans la SY14 sont inhibées par Rpd3 (Figure V-20a, Tableau supplémentaire VII-6). Comme pour Fkh1/2, les origines régulées et modifiées dans la SY14 ne semblent pas se regrouper ou présenter une localisation précise (Figure V-20b). L'ensemble de ces données suggère que Rpd3 n'est pas responsable de la modification des origines « moyennes » et « éloignées » dans la SY14.

a

	Moyennes			Éloignées			Total
	Augmentées	Diminuées	Non modifiées	Augmentées	Diminuées	Non modifiées	
Non régulées	2	1 + 9	43 + 24	1	2 + 4	37 + 39	83 + 79
inhibées	1	3	17	1	3	17	42



**Figure V-20 : Modification des origines à une distance « moyennes » et « éloignées » des centromères et des télomères par Rpd3.**

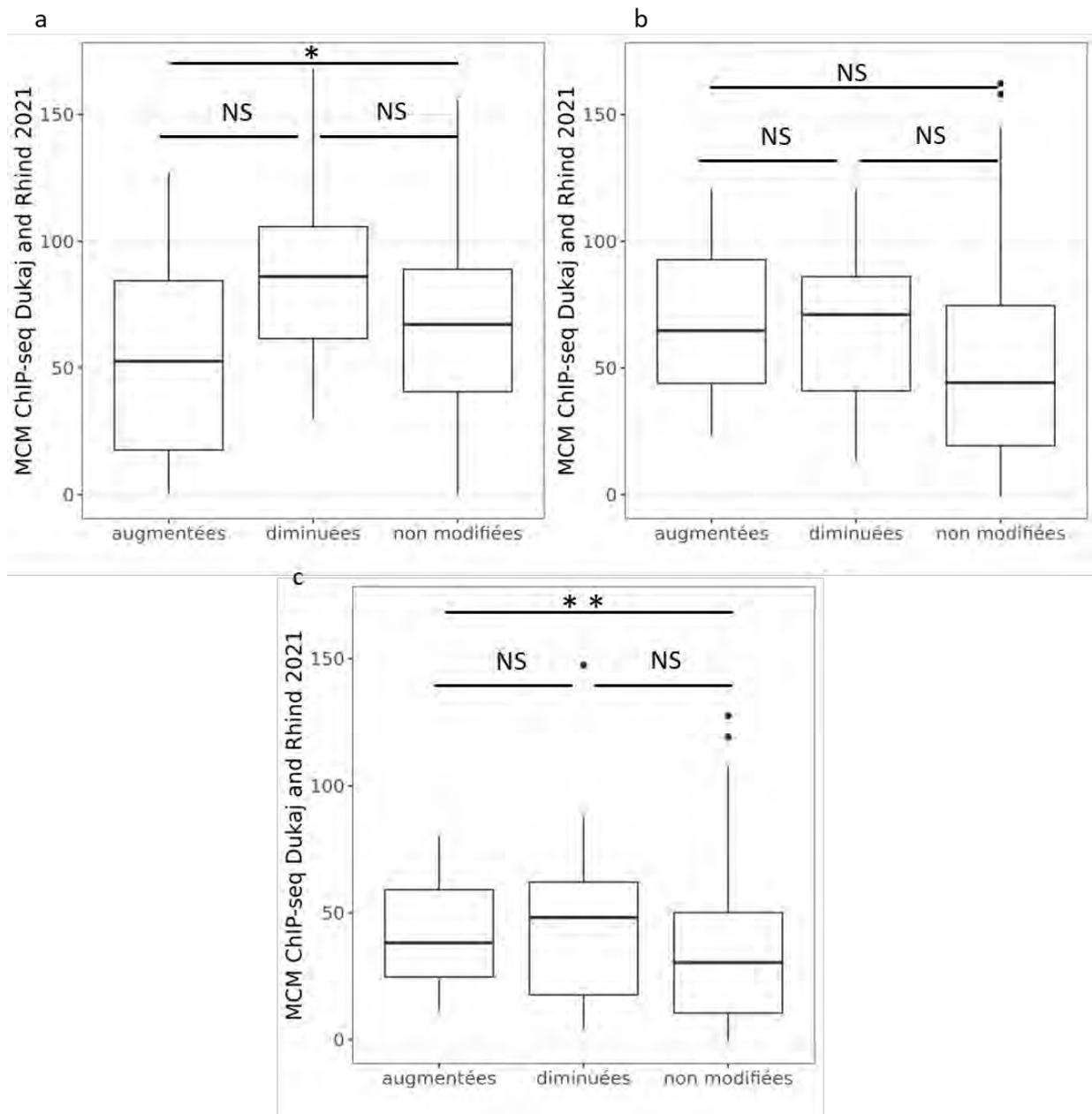
a) Tableau récapitulatif du nombre d'origines inhibées, et non modifiées par Rpd3 dans les classes de distances « moyennes » et « éloignées » des centromères et des télomères. Les chiffres en gris représentent le nombre d'origines absentes de données de Knott et al. (2009). b) Représentation des chromosomes de la BY4742 dans l'ordre du génome de la SY14. Haut : télomère gauche. Bas : télomère droit. Les centromères sont indiqués par les lignes pointillées. Les origines augmentées dans la SY14 sont en vert, les diminuées en rouge et les non modifiées en gris.

Dukaj et al. (2021) ont montré qu'il existait un lien entre *timing* de réplication et signal MCM-ChIP-seq. Ainsi, si une origine possède un signal plus fort en ChIP-seq, elle fixe plus de MCM et a tendance à être répliquée de manière plus précoce. Nous pouvons donc supposer que les origines diminuées dans la SY14 ont tendance à fixer moins de MCM. C'est pourquoi nous avons comparé les données de ChIP-seq obtenues par Dukaj et al. (2021) avec la modification des origines « éloignées » et « moyenne » dans la SY14. Les données sont disponibles pour 421 origines, dont 204 sont actives dans la SY14. Les valeurs de MCM-ChIP-seq pour les origines diminuées sont significativement plus

fortes que les valeurs de MCM-ChIP-seq pour les origines augmentées et non modifiées, lorsque la quantité de MCM intracellulaire est normale ou réduite de 93% (Figure V-21a, c). Cependant, lorsque la quantité de MCM intracellulaire est réduite de 86%, le signal MCM-ChIP-seq n'est pas significativement différent entre les trois catégories d'origines (Figure V-21b). Ces résultats suggèrent que malgré leur capacité à fixer un grand nombre de MCM dans une souche au caryotype normal, les origines diminuées dans la SY14 possèdent un défaut d'initiation de la réplication. Il serait intéressant de savoir si le nombre de MCM fixées diminue significativement ou pas dans la SY14 par rapport à la BY4742.

**Figure V-21 : Modification des origines à des distances « moyennes » et « éloignées » des centromères et des télomères en fonction du signal MCM-ChIP-seq (Dukaj and Rhind 2021).**

La modification de la distribution du signal MCM ChIP-seq a été vérifiée par un test Mann-Whitney-Wilcoxon, \* =  $p\text{-value} \leq 0.05$ , \*\* =  $p\text{-value} \leq 0.01$ , N.S= non significatif  $p\text{-value} > 0.05$ , augmentées n=5, diminuées n=22 et non modifiées n=117. a) Signal obtenu en conditions de culture standard. b) Signal obtenu en induisant la dégradation de 86% de la quantité de MCM intracellulaire par l'induction de la dégradation de MCM4 avec 30 $\mu$ M d'auxine. c) Signal obtenu en induisant la dégradation de 97% de la quantité du complexe MCM intracellulaire par l'induction de la dégradation de MCM4 avec 100 $\mu$ M d'auxine.



#### V.4 Analyse de la vitesse de progression des fourches de réplication.

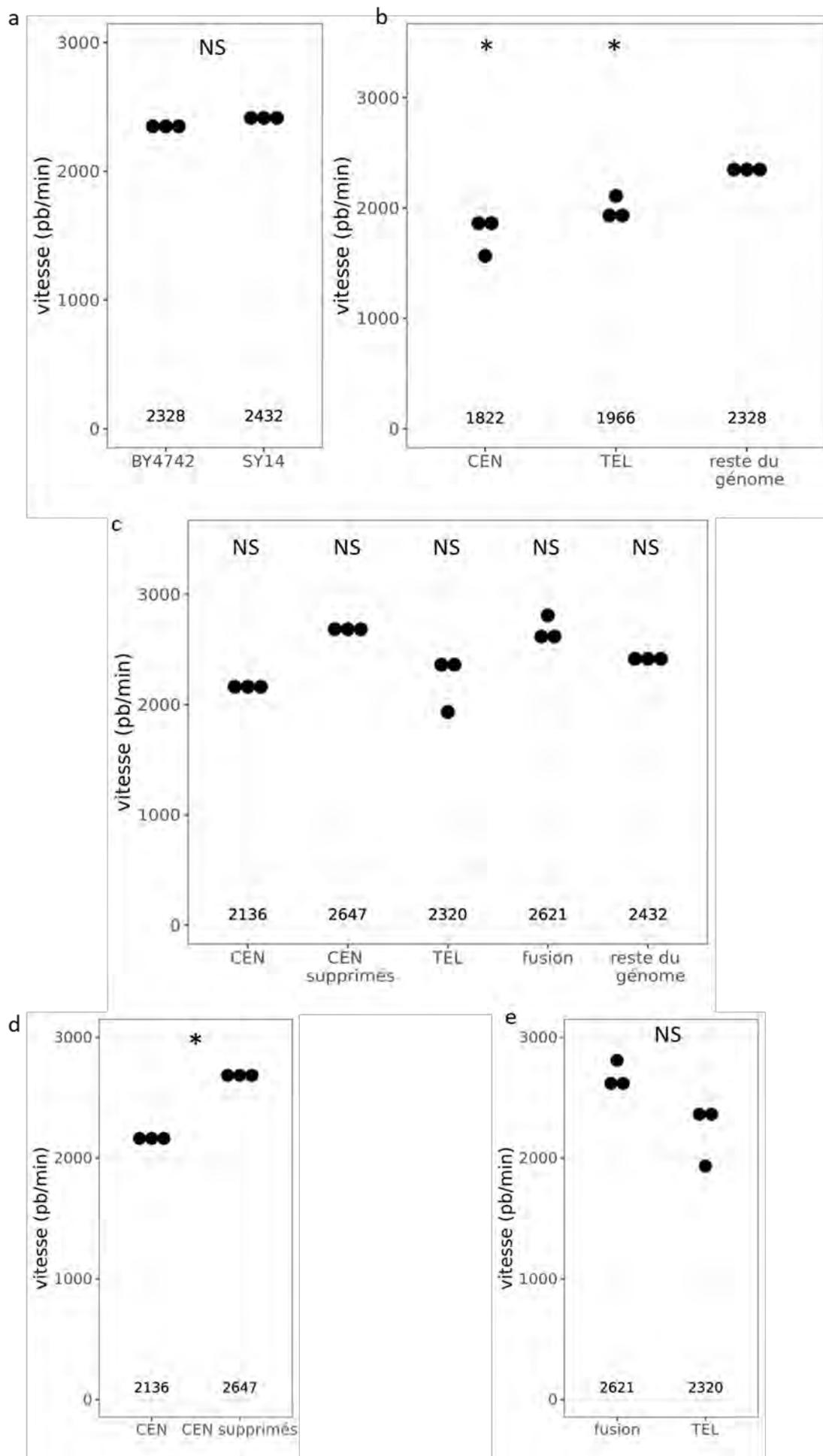
Nous avons comparé la vitesse de progression des fourches de réplication dans la BY4742 et la SY14 en retirant les fourches qui traversent les zones de ralentissement de la progression des fourches, c'est-à-dire l'ADNr, les ARN-t, les centromères, les télomères et les fusions chromosomiques (Theulot et al 2022, Figure V-22). La vitesse de progression des fourches est similaire entre la BY4742 et la SY14, avec respectivement une vitesse médiane de 2328pb/min et 2432pb/min (Figure V-22a).

Dans la BY4742, les fourches sont ralenties au niveau des centromères et télomères ( $p$ -value  $\leq 0,05$ , Figure V-22b). Dans la SY14, aucun ralentissement de la vitesse de progression des fourches

n'est observable pour les centromères et les télomères qui ne sont pas supprimés et pour les centromères supprimés et les zones de fusion (Figure V-22c). En revanche, la vitesse de progression des fourches qui traversent le centromère restant est diminuée par rapport aux fourches qui traversent les centromères supprimés. La vitesse des fourches qui traversent les deux télomères restants n'est pas significativement modifiée par rapport aux fourches qui traversent les fusions chromosomiques (Figure V-22d, e). Pour conclure, la suppression des centromères et des télomères dans la SY14 a aboli la diminution de la vitesse de progression des fourches observée en leur présence.

**Figure V-22 : Vitesse de progression des fourches dans le génome de la BY4742 et le génome de la SY14.**

a-b-c) La modification de la vitesse de progression des fourches a été analysée comme décrit par Theulot et al. (2022), NS : non significatif ( $p\text{-value} > 0,05$ ), \* :  $p\text{-value} \leq 0,05$ . a) Comparaison de la vitesse de progression des fourches de réplication dans la BY4742 et la SY14. b-c) Vitesse de progression des fourches en fonction de différents éléments génomiques connus pour ralentir les fourches. Les tests statistiques ont été réalisés comme décrits par Theulot et al. (2022) entre la vitesse moyenne sur l'ensemble du génome et les autres vitesses. b) Vitesse dans la BY4742. c) Vitesse dans la SY14. d) Comparaison de la vitesse de progression des fourches entre le centromère restant dans la SY14 et les centromères supprimés. e) Comparaison de la vitesse de progression des fourches entre les télomères et les fusions chromosomiques dans la SY14.



## VI. Discussion

### VI.1 Amélioration de l'analyse des profils de réplication

#### VI.1.1 Amélioration de la détection des fourches de réplication

La reproductibilité de nos analyses, ainsi que leur robustesse, dépendent en partie de la quantité de données utilisées. Par exemple, un profil RFD construit avec un nombre de fourches plus faible pourrait augmenter la divergence entre les réplicats biologiques et diminuer le nombre de régions détectées. C'est pourquoi nous avons décidé de réaliser les RFD à partir de séquençages PromethION.

Nous savons que NFS n'est pas optimal pour tirer le maximum d'informations de nos lectures nanopore. En effet, NFS ne peut pas détecter les initiations ou les terminaisons durant le *pulse*, ni les signaux de fourches incomplets. Fork-ML semble être un bon candidat pour améliorer la détection des fourches. Fork-ML étant basé sur l'entraînement d'un réseau de neurones, la détection dépend de l'annotation des lectures. Différentes annotations peuvent donc être envisagées selon que l'on souhaite (i) détecter les fourches pour construire les profils RFD, IP ou TP ; (ii) détecter les pauses dans la progression des fourches ; on pourrait par exemple annoter des *reads* provenant du ADNr où le site de pause est connu ; (iii) détecter uniquement le début et la fin du *pulse* pour déterminer la vitesse de progression des fourches.

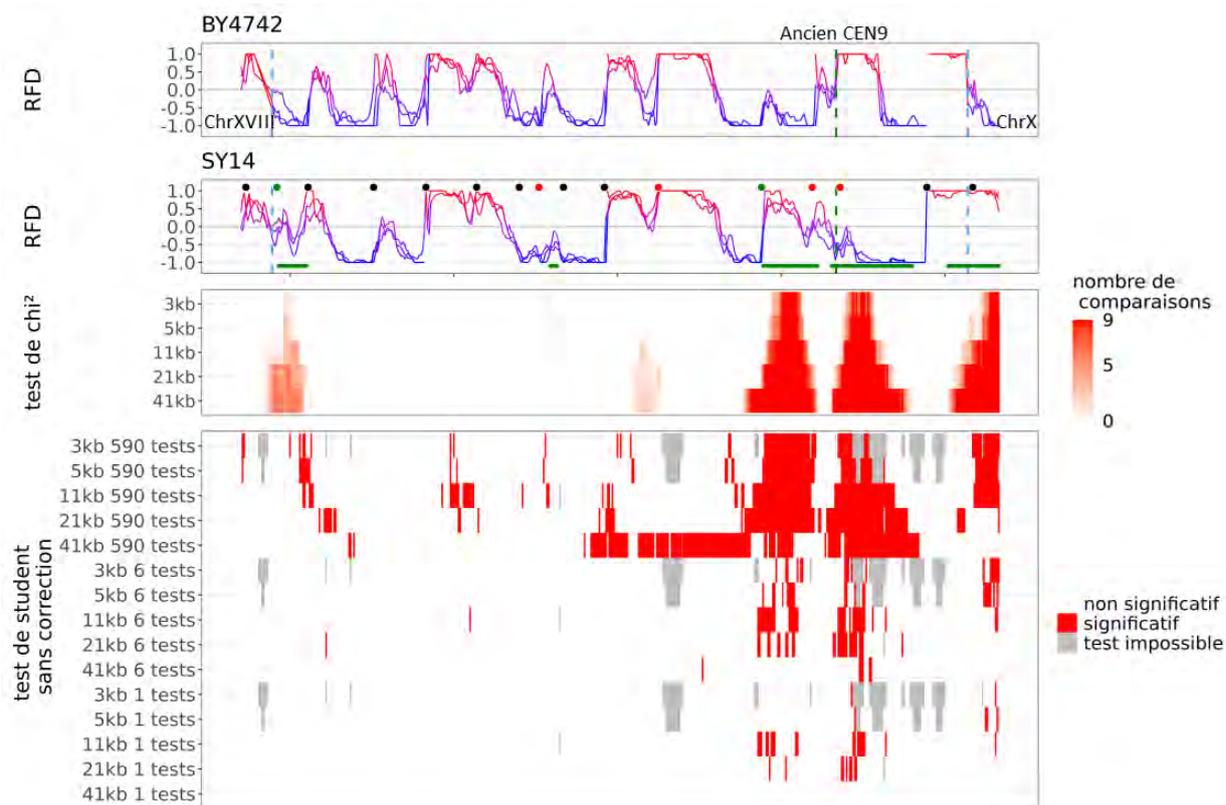
Avant de développer de tels modèles, il convient de régler le problème de surdétection des fourches. Généralement, ce type de problème provient d'un surentraînement du réseau. Brièvement, les paramètres d'un premier réseau de neurones sont déterminés aléatoirement. La qualité de la détection de ce premier réseau est mesurée, puis les paramètres sont ajustés et la qualité du second réseau créé est de nouveau mesurée. Le principal problème rencontré lors de l'entraînement de Fork-ML est la faible augmentation de la qualité du réseau avec les cycles d'entraînement successifs. Cela signifie que lorsque les paramètres sont choisis au hasard, le réseau est aussi performant que lorsque les paramètres sont ajustés après entraînement. La méthode de validation du réseau est peut-être à remettre en cause. Actuellement, le réseau est validé selon sa capacité à détecter les fourches droites, les fourches gauches et les régions sans fourches. La majorité des lectures annotées ne possède pas de fourches, donc on se repose fortement sur le réseau pour détecter correctement les régions sans fourche et ce dernier s'avère peu entraîné à détecter correctement les fourches. Pour y remédier, nous pourrions enrichir le jeu de données d'entraînement en lectures avec fourches. Une autre option serait de ne valider le réseau que sur les données fourche droite et fourche gauche.

## VI.1.2 Amélioration des analyses statistiques

### VI.1.2.1 Test de Student pour comparer les RFD

Dans un premier temps, nous avons envisagé d'utiliser un test de Student pour comparer la proportion de fourches droites entre les trois réplicats BY4742 et les trois réplicats SY14 sur des fenêtres de différentes tailles. Lorsque nous avons utilisé la correction de Benjamini et Yekutieli pour tenir compte de la multiplicité des tests, le nombre de régions significatives a trop fortement diminué, ce qui nous a conduit à écarter cette méthode. Comme nos deux échantillons sont petits, les *p-values* du test de Student sont faibles et la correction pour tests multiples élimine trop de régions. Néanmoins, il est difficile de décider comment modifier cette correction pour éviter trop de faux négatifs tout en réduisant le nombre de faux positifs. Le nombre de tests pourrait être réduit en n'analysant qu'une seule échelle, mais la multiplicité des échelles nous permet de mieux visualiser certains changements qui ne sont pas détectés sur toutes les échelles.

Pour utiliser le test de Student tout en prenant en compte le nombre de faux positifs, nous aurions pu déterminer un seuil  $\alpha$  plus faible. On s'attend à observer sur nos 59 000 tests environ 590 tests faussement positifs avec  $\alpha=0,01$ , 59 avec  $\alpha=0,001$ , et 6 tests avec  $\alpha=0,0001$ . Dans ce contexte, le choix du seuil  $\alpha$  par l'expérimentateur doit impérativement être défini avant de regarder les résultats, afin de ne pas biaiser le choix en fonction des résultats attendus. La Figure VI-1 présente les résultats pour le chromosome 9 sans correction pour tests multiples et en fonction du nombre de faux positifs attendus pour divers seuils  $\alpha$ . Les seuils de 6 tests faussement positifs ou un 1 test faussement positif mettent uniquement en évidence la modification importante de la région péricentromérique (Figure VI-1). Le seuil de 6 tests faussement positifs possède uniquement trois *bins* significatives supplémentaires (Figure VI-1). Le choix du seuil  $\alpha$ , et par extension le nombre de faux positifs que l'on souhaite accepter, est arbitraire. De plus, cette stratégie ne prend pas en compte la non-indépendance des tests pratiqués pour deux fenêtres adjacentes ou à différentes échelles. L'utilisation du test du Chi<sup>2</sup>, avec la correction de Benjamini-Yekutieli, nous a permis de nous affranchir des biais potentiels induits par la non-indépendance de certains tests, et du choix arbitraire d'un nombre acceptable de faux positifs.



**Figure VI-1 : Comparaison des résultats du test de Student pour différents seuil  $\alpha$ .**

Les profils RFD sont annotés en fonction des modifications détectées par le test du Chi<sup>2</sup> et l'analyse visuelle. Les origines actives sont indiquées par des cercles noir, les origines significativement diminuées sont indiquées par des cercles rouges, les terminaisons modifiées sont indiquées par des lignes vertes. Les résultats des tests de Student sont obtenus en considérant différents nombres de faux positifs attendus sur le génome pour différentes *p-values* : 590 tests pour une *p-value*  $\leq 0,01$  ; 6 tests pour une *p-value*  $\leq 0,0001$  ; 1 test pour une *p-value*  $\leq 0,00002$ . Les régions où le test est impossible sont celles où la variance entre les trois réplicats BY4742 ou les trois réplicats SY14 est nulle.

### VI.1.2.2 Test de Chi<sup>2</sup> pour comparer les RFD

#### VI.1.2.2.1 Synthèse des comparaisons

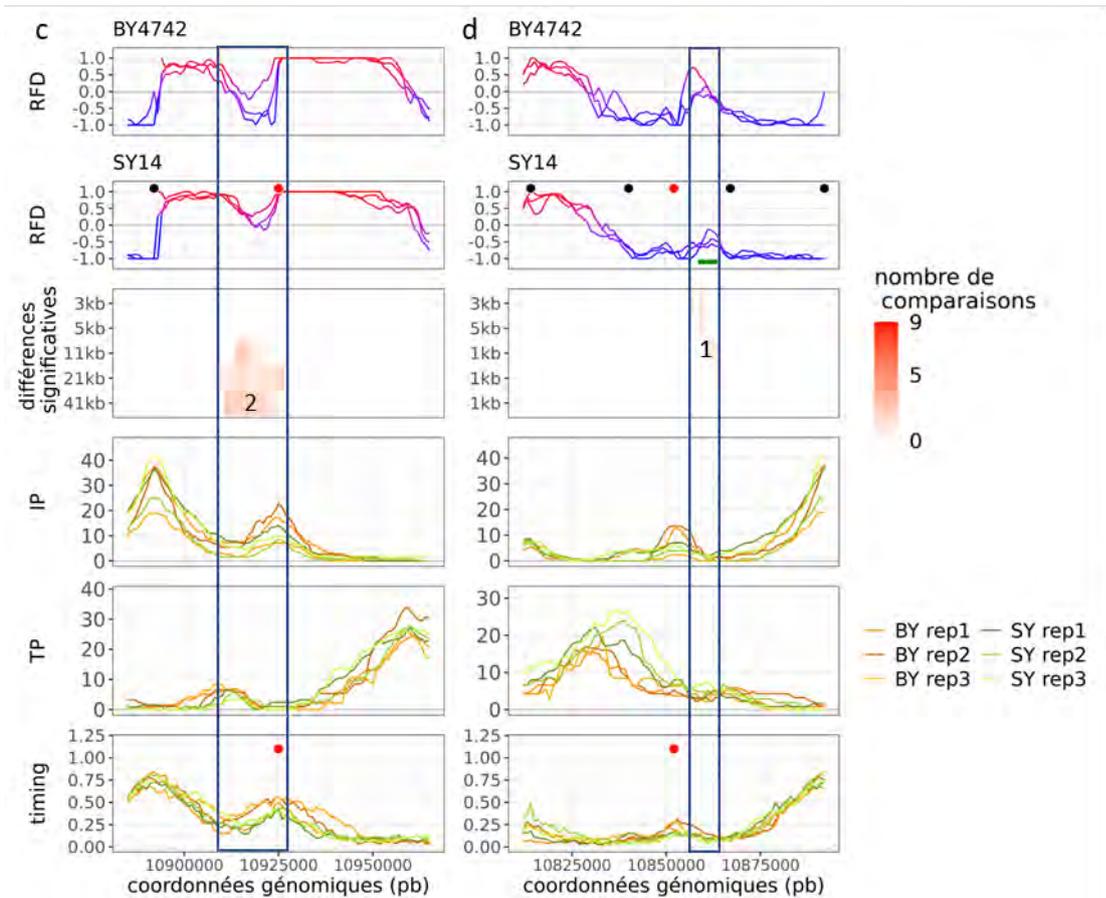
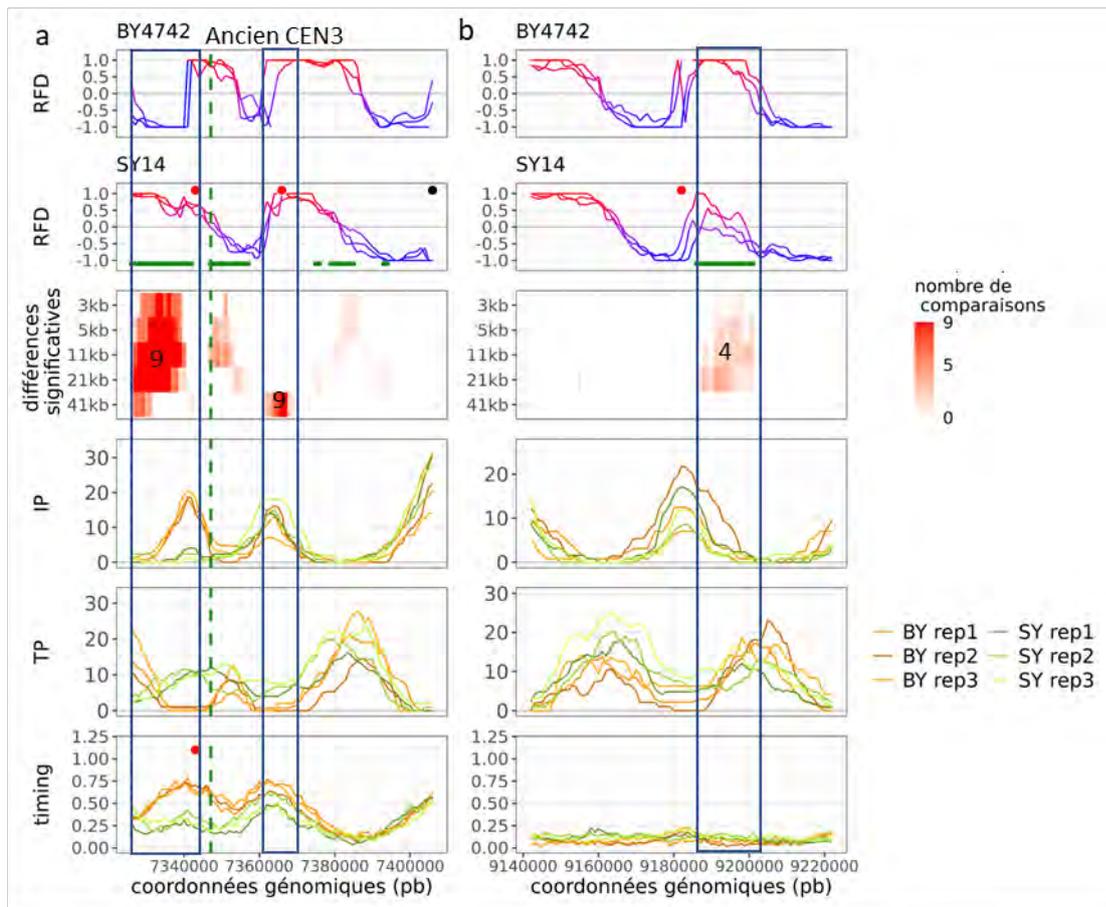
Afin de détecter les modifications des RFD, nous avons décidé de comparer des fenêtres chevauchantes de différentes tailles coulissant sur 1kb par un test de Chi<sup>2</sup> comparant le nombre de fourches gauches et de fourches droites entre deux profils. Le principal inconvénient de cette stratégie est la réalisation de comparaison deux à deux. Le test n'intègre pas la variabilité possible entre les réplicats, et il est nécessaire de déterminer un seuil de comparaison afin de décider si une région est significativement modifiée ou non.

J'ai synthétisé les résultats des différentes comparaisons en comptant le nombre de comparaisons significatives pour une bin, puis décidé de prendre en compte un test dès lors qu'une seule comparaison était positive sur au moins deux bins ou deux échelles de fenêtre. Le choix de ce seuil dépend de notre volonté à potentiellement exclure des faux négatifs ou garder des faux positifs. La méthode qui permet d'exclure le plus de faux positifs est celle qui consiste à garder uniquement les régions significatives dans les 9 comparaisons, et celle qui induit le plus de faux positifs consiste à garder les régions dès lors qu'une comparaison est significative. Ici, j'ai choisi de garder les régions dès qu'une comparaison était significative, car nous voulions nous intéresser à des régions de faible modification. De plus, nous disposions de deux autres jeux de données pour détecter des altérations de la réplication, les profils d'IP et de *timing* de réplication. Afin de vérifier si les modifications avec un nombre de comparaisons faible sont détectées aussi en *timing* et/ou en IP, j'ai associé à chaque origine modifiée en RFD la valeur du nombre maximal de comparaisons dans la zone significative ayant permis de détecter l'origine (Figure VI-2a, cadre bleu). J'ai ensuite vérifié si ces origines étaient aussi détectées sur le *timing* et/ou l'IP. Toutes les origines qui sont détectées par 9 comparaisons significatives ne sont pas détectées par le RFD et/ou l'IP, néanmoins plus le nombre de comparaisons significatives diminue, plus le nombre d'origines détectées aussi en IP et/ou en *timing* diminue (Tableau VI-1). Lorsque le nombre de comparaisons significatives est de 1, une modification en *timing* est aussi visible, ce qui confirme que les régions avec une seule comparaison significative ne représentent pas exclusivement des faux positifs (Figure VI-1d).

Nombre de comparaison significatives	Nombre d'origines	Détection Timing	Détection IPL	Régions tardive
0	245	8	2	1
1	10	1	0	6
2	10	1 + 1	0	7
3	1	0	0	1
4	2	0	0	2
5	3	0	0	2
6	3	1 + 1	1	1
7	0	0	0	0
8	0	0	0	0
9	31	13 + 5	10	6

**Tableau VI-1 : Evolution du nombre de comparaisons significatives pour les origines détectées en RFD en fonction de leur détection en IP et en *timing*.**

Le nombre de comparaisons significatives pour une origine est défini comme le nombre de comparaisons maximal pour la zone significative qui correspond à l'origine. Les chiffres en gris pour la détection par le *timing* correspondent aux modifications détectées sur les origines voisines, c'est-à-dire que la modification du RFD provient de la modification du *timing* de répllication d'une origine voisine.

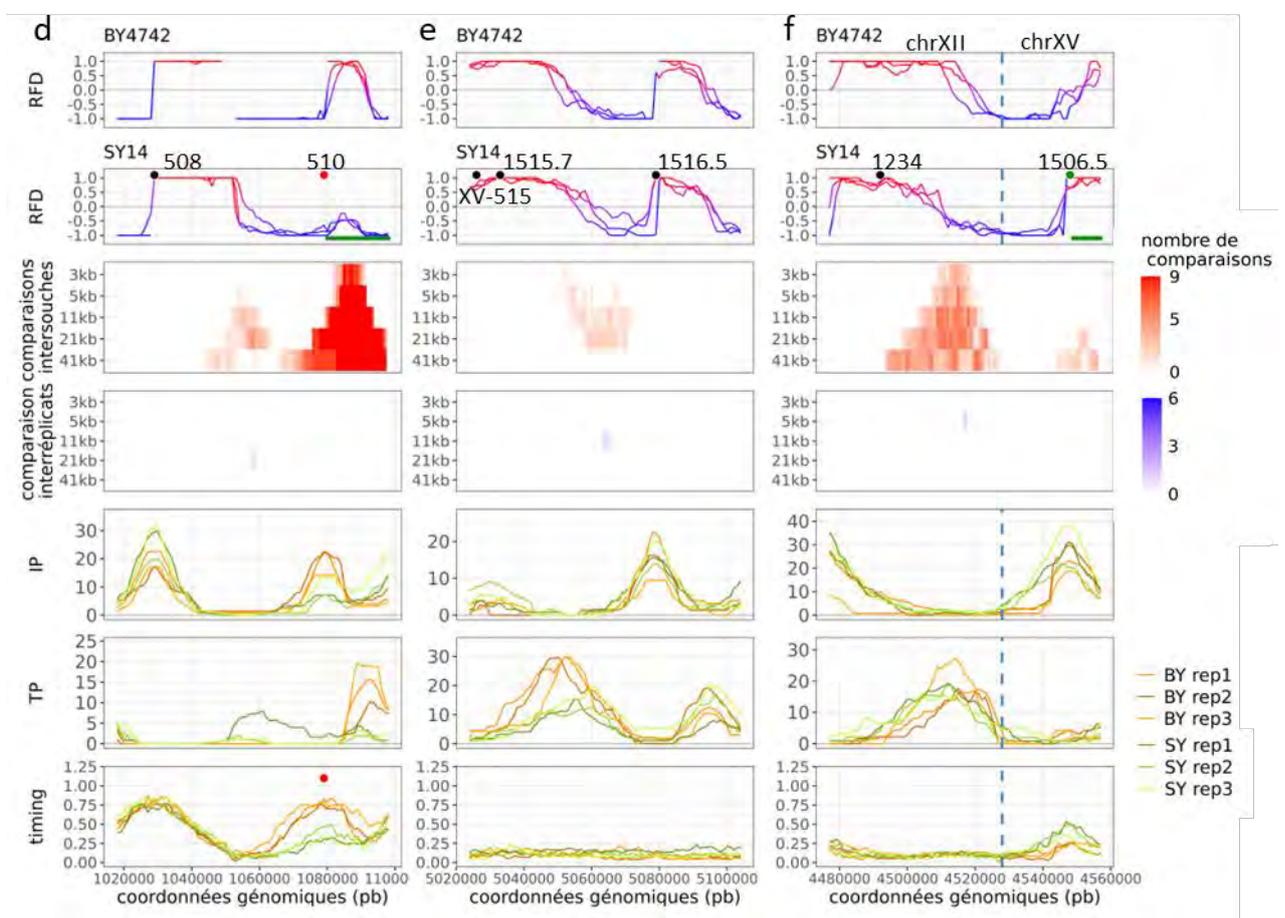
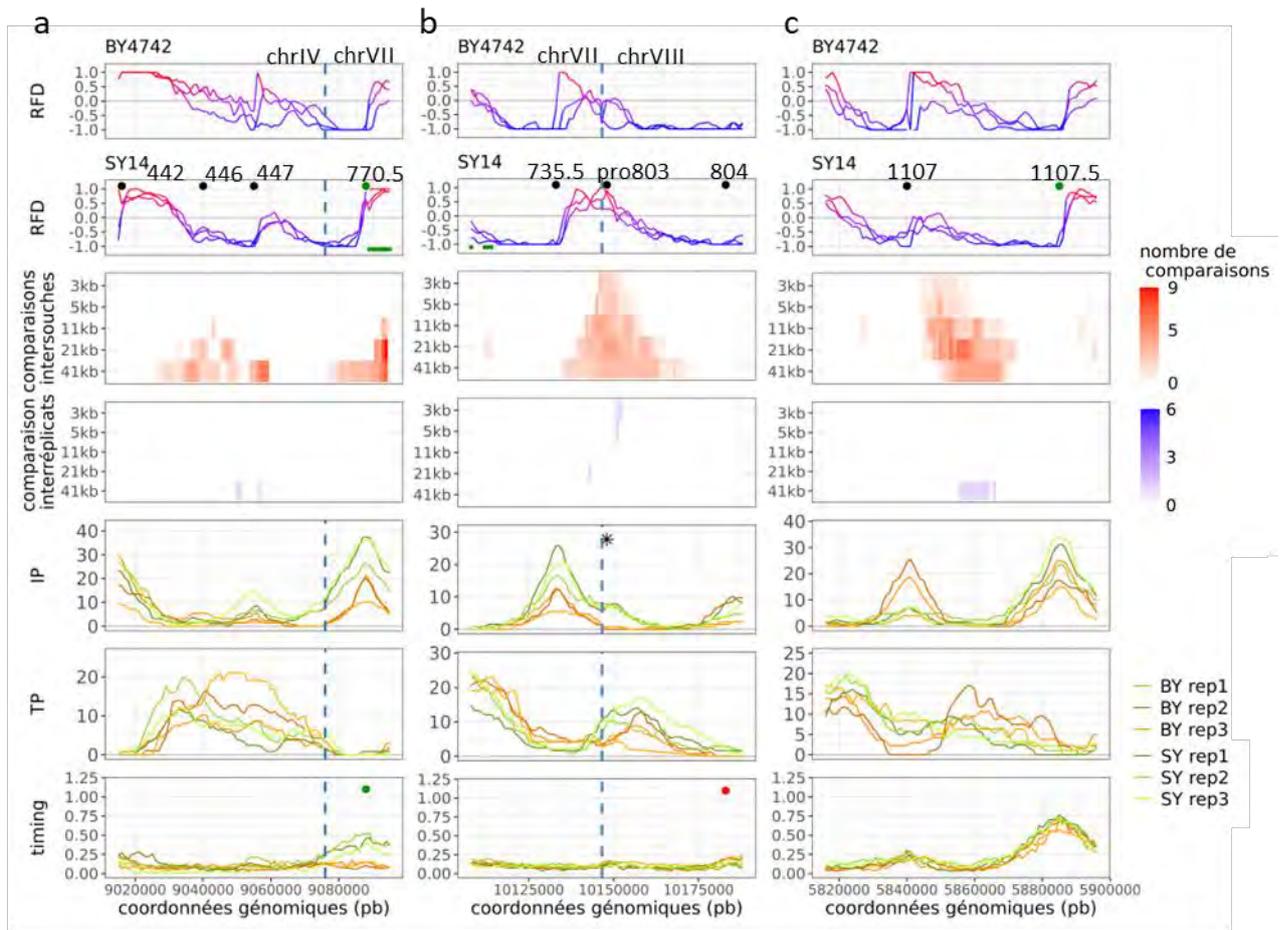


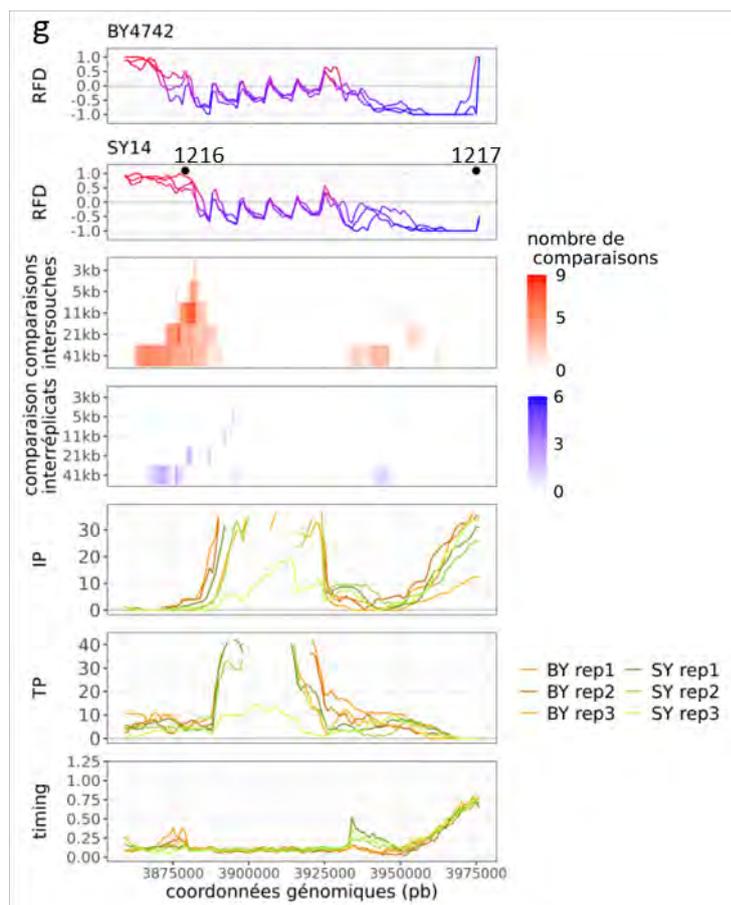
### Figure VI-2 : Exemple d'origines détectées par différents nombres de comparaison significatives.

Les cadres bleus représentent les régions où le nombre maximal de comparaisons significatives a été déterminé pour chaque origine, et ce nombre est indiqué. Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondant. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les zones de terminaisons sont indiquées par des lignes vertes.

#### VI.1.2.2.2 Exclusion des régions montrant des différences significatives entre les réplicats

Il existe 39 bins significatives entre réplicats dans des régions montrant des différences significatives entre les souches. Elles se regroupent en 7 régions que j'ai choisi d'exclure systématiquement de l'analyse, peu importe le nombre de bin concernées (Figure VI-3). Néanmoins, le faible nombre de bins significatives entre réplicats montre que le test du Chi<sup>2</sup> avec correction pour tests multiples détecte majoritairement des régions présentant exclusivement des différences entre les souches (2101 bins). Nous aurions donc pu considérer que toutes les régions significatives présentes sur le test du Chi<sup>2</sup> étaient des régions de modifications entre la BY4742 et la SY14, quels que soient les résultats du test du Chi<sup>2</sup> entre les réplicats. Néanmoins, cette stratégie nous aurait poussé à considérer comme significatives les bordures de l'ADNr alors qu'il est possible que les différences observées ne proviennent que d'un problème d'alignement dans cette région répétitive. De plus, seule une région exclue montre une modification significative de l'IP (Figure VI-3d). Les deux autres régions où l'initiation est modifiée en RFD ne montrent ni de modification de l'IP ni du *timing* de réplication (Figure VI-3e, f). Parmi les 3 régions qui ne montrent qu'une modification de la terminaison en RFD (Figure VI-3a, b, c), une seule montre une modification convaincante du TP (Figure VI-3c). Nous aurions pu décider de ne retenir parmi ces 7 régions que celles qui nous semblaient convaincantes, mais le choix de régions à conserver dépend de l'expérimentateur et peut être sujet à controverse. C'est pourquoi j'ai préféré adopter une stratégie plus stricte et indépendante de l'observateur, et donc d'exclure l'ensemble de ces 7 régions.





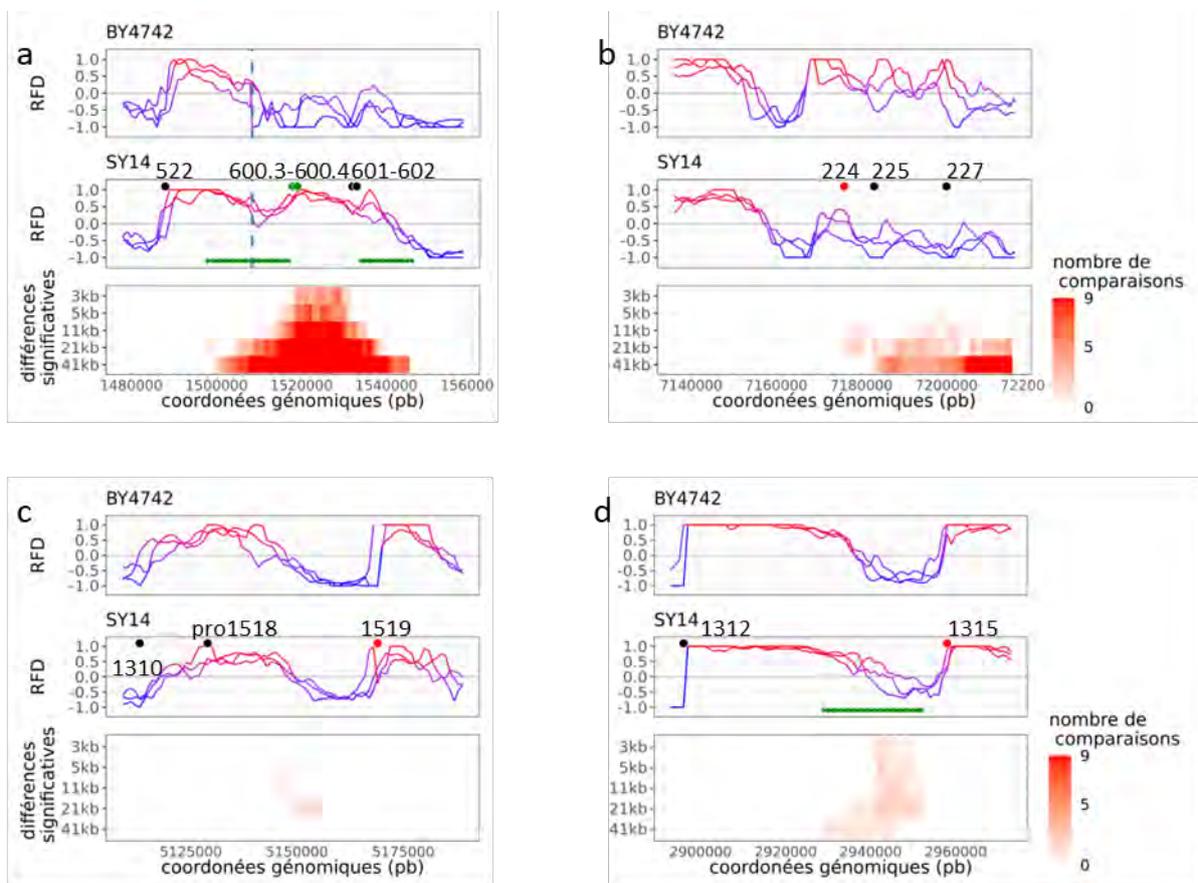
**Figure VI-3 : Modifications significatives entre les réplicats.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercle rouges sur les profils correspondant. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les zones de terminaisons sont indiquées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro

#### VI.1.2.2.3 Analyses visuelles

J'ai choisi d'utiliser les tests statistiques pour sélectionner les zones où les profils étaient significativement différents entre souches, puis de réaliser une analyse visuelle pour comprendre la nature des changements décelés et les interpréter de façon simple et cohérente en termes de changement d'efficacité des origines locales et de changements de position des évènements de terminaison qui en découlent. J'ai fait ce choix car mes compétences en programmation étaient trop faibles pour créer un algorithme permettant d'analyser les pentes de RFD afin d'obtenir les informations concernant les initiations et terminaisons qu'elles contiennent. L'analyse visuelle introduit inévitablement des biais liés à l'expérimentateur, d'autant plus importants que les

modifications sont faibles et complexes à interpréter. J'ai choisi de ne considérer que le RFD pour l'interprétation, afin de n'utiliser que les informations qui ont servi à la réalisation des tests. J'ai rencontré deux principales difficultés lors de l'analyse visuelle : (i) les régions où le RFD varie entre 0 et 1 dans une souche et -1 et 0 dans l'autre, car la différence de RFD entre le début et la fin de la pente correspondant à une initiation est l'indicateur de l'efficacité de l'origine. Il était plus difficile de comparer ces informations dans ces régions (Figure VI-4a et b). (ii) Pour les régions où l'origine modifiée n'est pas dans la zone significative (Figure VI-4c et d), la difficulté résidait dans l'annotation de la terminaison, de l'initiation ou des deux. Pour limiter les biais induits par l'analyse visuelle, j'ai réalisé plusieurs fois l'analyse du génome entier, afin de vérifier si mon interprétation était toujours la même.



**Figure VI-4 : Exemple de régions où l'interprétation du RFD est difficile.**

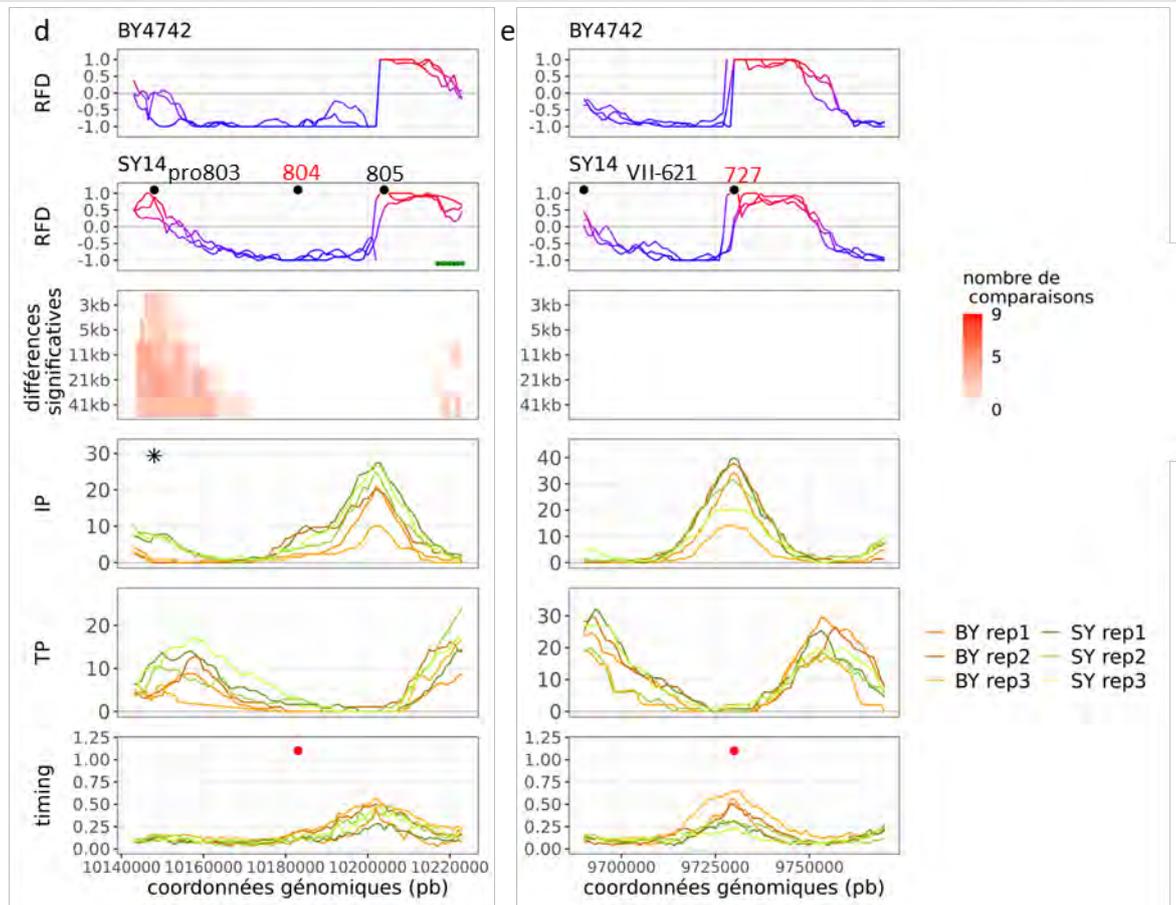
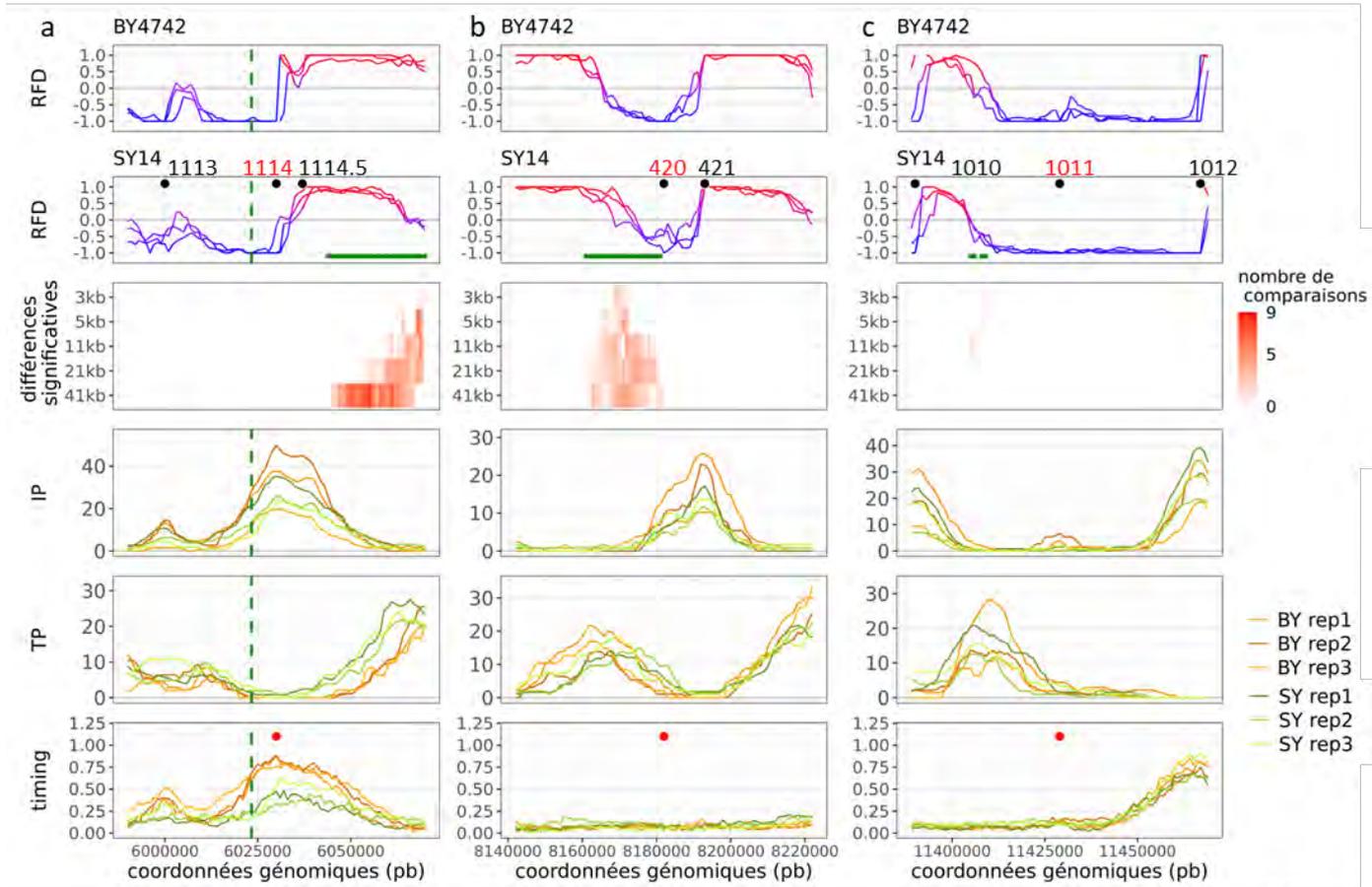
Les origines actives non modifiées sont indiquées par des cercles noirs sur le profil RFD SY14. Les origines significativement diminuées en RFD sont indiquées par des cercles rouges. Les origines significativement augmentées en RFD, par des cercles verts. Les terminaisons modifiées sont indiquées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro. a-b) Régions où le RFD varie entre 1 et 0 dans une souche et -1 et 0 dans l'autre. c-d) Régions où la modification est sur une zone de terminaison. Le choix de l'annotation de la terminaison ou de la zone d'initiation à proximité est laissé à l'observateur.

### VI.1.2.3 *Timing* de réplication

J'ai effectué un test de Student uniquement sur les régions encadrant +/- 1,5 kb des ARS détectées comme actives, afin de pouvoir directement détecter en nanotiming les modifications des origines sans passer par une étape d'analyse visuelle. Le nombre de tests est très grand, car ils sont réalisés sur des fenêtres chevauchantes de 1kb coulissant sur 10pb. Nous avons donc choisi de ne pas réaliser de correction pour tests multiples, mais de prendre en compte le nombre de tests significatifs par origine, et avons considéré une ARS comme significativement modifiée dès lors que 25% des tests de cette région étaient significatifs. Avec ce seuil, 24 origines sont considérées comme significativement modifiées, dont 16 détectées aussi en RFD et/ou en IP. Deux origines supplémentaires sont associées à des changements significatifs de zones de terminaisons en RFD. Le Tableau VI-2 montre que plus le seuil est strict, plus le nombre d'origines détectées en *timing* baisse, mais plus la fraction de ces origines qui est confirmée en RFD et/ou IP augmente. Lorsque le seuil dépasse 75%, le nombre d'origines détectées en *timing* et RFD et/ou IP diminue, ce qui montre que ces seuils sont trop stricts. Seulement 5 origines supplémentaires sont détectées lorsque le seuil passe de 50% à 25% (Tableau VI-2), trois sont liées à une modification significative d'une zone de terminaison (Figure VI-5a, b et c), ce qui montre que le choix d'un seuil à 25% plutôt qu'à 50% a ajouté des régions qui corroborent les modifications significatives du RFD.

Pourcentage de tests positifs	Nombre d'origines modifiées en Timing	Nombre d'origines détectées en timing et en RFD et/ou IPL
100%	10	10
≥ 75%	13	12
≥ 50%	19	16
≥ 25%	24	16
≥ 10%	25	16
≥ 5%	32	16
1 test	38	18

Tableau VI-2 : Nombre d'origines détectées comme modifiées en *timing* de réplication et en RFD et/ou IP en fonction du % de tests significatifs sur le *timing* retenu comme seuil.



**Figure VI-5 : Origines significativement modifiées en *timing* avec un % de modification significative compris entre 25 % et 50%.**

Les origines actives non modifiées sont indiquées par des cercles noirs sur le profil RFD SY14. Les origines significativement diminuées en RFD ou retardées en *timing* de réplication sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou avancées en *timing* sont indiquées par des cercles rouges. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont indiquées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro.

#### VI.1.2.4 Conclusion sur les analyses statistiques

Nous souhaitons déterminer si le profil de réplication de la SY14 diffère du profil de réplication de la BY4742. Les zones de modifications les plus importantes sont évidentes à l'œil et l'approche statistique n'est pas nécessaire à leur identification. Cependant, il existe de nombreuses zones de modification faible où il était nécessaire d'employer une approche statistique, afin de distinguer les différences biologiques des différences liées à la fluctuation expérimentale. Nous avons mené différentes approches statistiques pour analyser les initiations des profils RFD, IP et *timing*, en essayant, à chaque étape, de réduire le nombre de faux positifs tout en nous permettant de détecter des modifications faibles. Ces choix nous ont permis de révéler un grand nombre de modifications faibles, mais il reste plausible qu'une petite fraction de ces régions corresponde à des faux positifs. Cependant, il est peu probable que cela affecte nos conclusions, qui sont restées générales.

#### VI.1.3 Analyses complémentaires

##### VI.1.3.1 Études de la modification des zones de terminaison dans la SY14 et la BY4742

Dans un premier temps, j'ai choisi de concentrer mon analyse sur les origines de réplication, car les modifications des zones de terminaison découlent de celles des origines. Les profils RFD et TP permettent une analyse relativement facile des zones de terminaison. Les profils de *timing* sont moins informatifs à cet égard, car des modifications des vallées du profil, où se dispersent les événements de terminaison, sont moins faciles à détecter que des modifications des pics, où se concentrent les événements d'initiation. Les tests statistiques réalisés sur les RFD ont parfois identifié des zones de terminaison modifiées sans pour autant que les origines qui les bordent ne soient elles-mêmes significativement modifiées. Plusieurs éléments peuvent expliquer ce phénomène : (i) si les origines bordant la zone de terminaison sont dans tous les cas efficaces à 100%, une modification de leur *timing* relatif ne se traduira sur le profil RFD que par un déplacement de la zone de terminaison vers l'origine devenue relativement plus tardive ; (ii) s'il y a une modification de la vitesse de progression des fourches entre les deux origines, par exemple si la fourche provenant de l'origine A s'arrête alors la

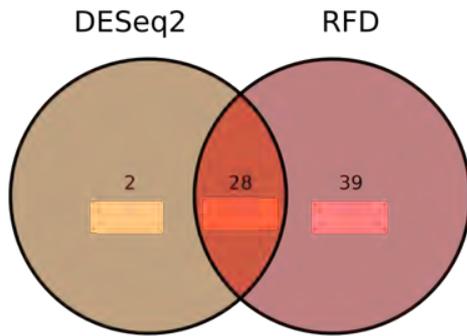
zone de terminaison sera plus proche de l'origine A, car la fourche provenant de l'origine B va continuer sa route jusqu'à rencontrer la fourche A qui est arrêtée. Notons toutefois que nos mesures des vitesses des fourches de réplication n'ont pas décelé de différence significative de leur valeur moyenne entre les deux souches.

Pour l'analyse du profil TP, il n'était pas possible d'utiliser la même stratégie que pour l'IP. Lors de l'analyse de l'IP, nous avons considéré les ARS comme des gènes pour lesquels on analyse la couverture en ARNm, or les terminaisons n'ont pas lieu en des localisations définies que nous aurions pu utiliser pour regarder la couverture normalisée des segments. Une stratégie plus générale consiste à comparer la couverture en segments situés entre deux fourches convergentes ou divergentes sur l'ensemble des bins du génome. Cette analyse, réalisée par Gaël Millot avec le logiciel DESeq2 pour le TP, permet de mettre en évidence des zones de modifications importantes au niveau des centromères et des télomères (Figure VI-6c). 28 zones de terminaisons sont communes à la détection par DESeq2 et le RFD, 2 sont détectées uniquement par l'analyse du TP (Figure VI-6b, d) et 39 uniquement par l'analyse du RFD (Figure VI-6a). Ces résultats confirment la plus grande puissance de détection des modifications en RFD. Néanmoins, une des modifications visibles uniquement en TP est intéressante. Le profil TP présente deux pics dans la BY4742, tandis qu'il n'en possède qu'un dans la SY14. Sur les profils RFD et IP, on peut voir la disparition d'une zone d'initiation qui induit la fusion des deux pics TP en un seul dans la SY14 (Figure VI-6d).

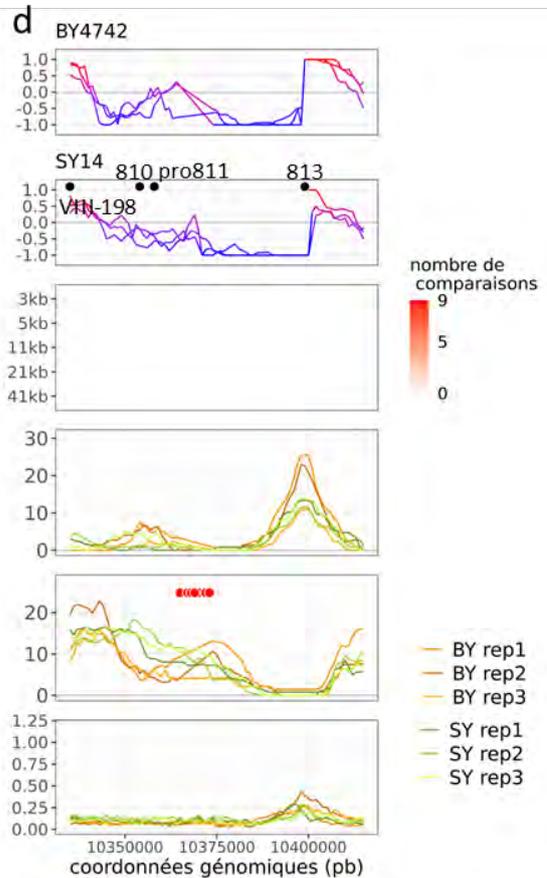
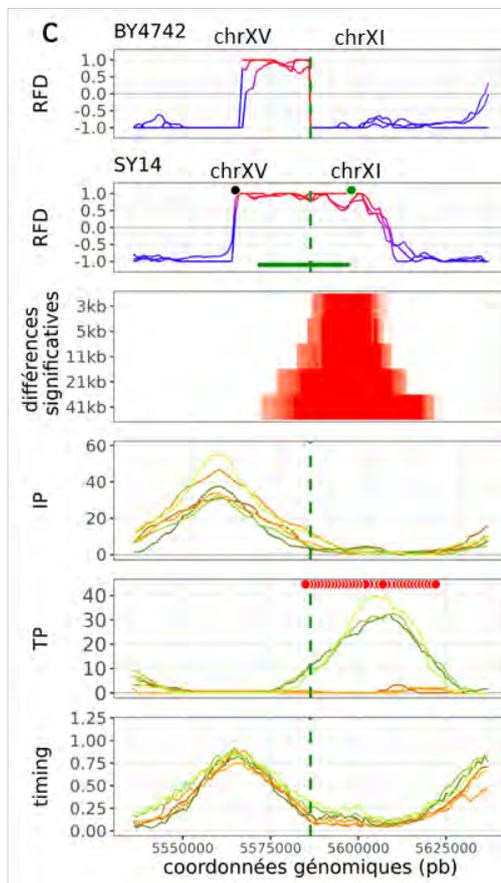
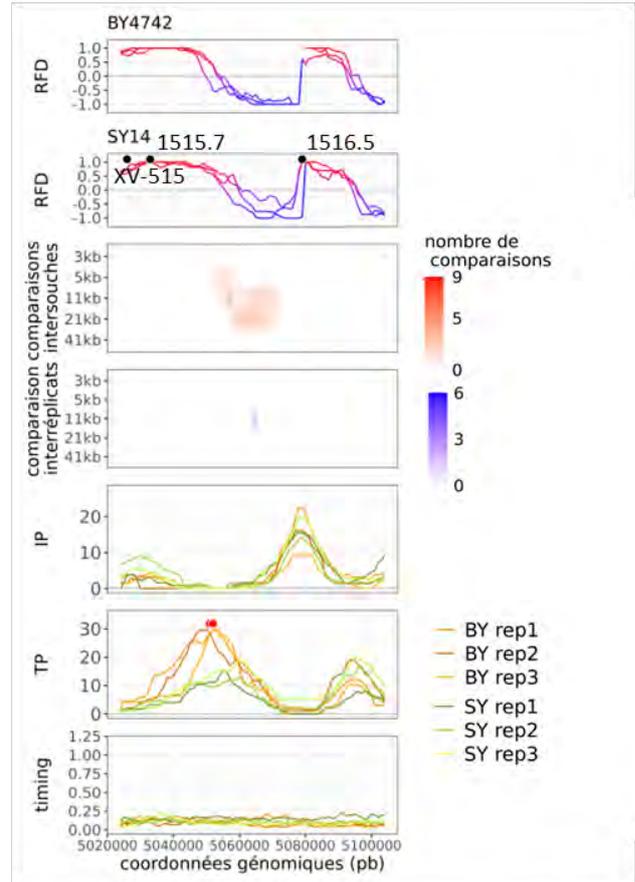
#### **Figure VI-6 : Analyses statistiques des TP par DESeq2.**

a) Diagramme de Venn représentant le nombre de terminaisons détectées par l'analyse du RFD et du TP. b-c-d) Les origines actives non modifiées sont indiquées par des cercles noirs sur le profil RFD SY14. Les origines significativement diminuées en RFD ou retardées en *timing* de réplication sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou avancées en *timing* sont indiquées par des cercles rouges. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les zones de terminaison considérées comme augmentées et diminuées par DESeq2 sont respectivement représentées en rouge et en vert sur le profil TP. Le nom des ARS est indiqué uniquement par leur numéro. b) Exemple de modification détectée uniquement par l'analyse du TP avec DESeq2. Cette modification était détectée par le RFD, mais elle a été exclue de l'analyse car elle comportait une région de différence significative entre réplicats. c) Exemple de fusion chromosomique où la création d'une zone de terminaison est détectée par l'analyse du TP. d) Exemple de modification de distribution des événements de terminaison, détectée uniquement par l'analyse du TP. Cette modification résulte de l'inactivation d'une origine dans la SY14, non détectée par les autres méthodes bien que visible à l'œil sur le profil RFD.

a



b



### VI.1.3.2 Analyse de la vitesse de progression des fourches dans la SY14 et la BY4742

#### VI.1.3.3 Analyser des modifications de vitesse localement

Une partie des modifications observées dans la SY14 pourrait s'expliquer par une modification de la vitesse des fourches. Par exemple, une augmentation de la vitesse entre deux origines pourrait augmenter la fréquence de leur réplication passive et donc diminuer l'efficacité apparente de la plus lente à s'activer. On pourrait envisager de comparer la vitesse des fourches entre les origines qui encadrent une origine modifiée ou une zone de terminaison modifiée, pour voir si les modifications observées peuvent s'expliquer par un changement de vitesse, puis de nous intéresser aux causes de ce changement.

#### VI.1.3.4 Réaliser une carte des vitesses dans la SY14 et la BY4742

Dans un premier temps, il est possible d'utiliser la même stratégie que Theulot et al. (2022) pour obtenir une carte des vitesses pour la BY4742 et la SY14, et donc d'identifier les zones de ralentissement ou d'accélération des fourches dans la BY4742 et dans la SY14. Par exemple, la carte devrait mettre en évidence les centromères comme site de pause dans la BY4742, mais pas dans la SY14. Cependant, nous risquons de rater des zones de ralentissement faible dans la BY4742 non significative, qui correspondent à des zones d'accélération faible dans la SY14 non significative ou inversement. Cette stratégie permet d'identifier des zones de modifications au sein du génome, mais pas entre les souches. La meilleure stratégie pour obtenir cette information serait de comparer la vitesse de progression des fourches entre des régions similaires pour la BY4742 et la SY14. L'analyse de la vitesse de progression des fourches réalisée par Theulot et al. (2022), compare la carte des vitesses entre la fenêtre de test et la distribution des vitesses du génome par un test de Mann-Whitney-Wilcoxon. Dans notre cas, il faudrait comparer la distribution des vitesses pour la fenêtre de test entre la BY4742 et la SY14. Cette stratégie devrait permettre de déterminer les éventuelles zones de modification de la vitesse des fourches entre la BY4742 et la SY14.

## VI.2 La réplication de l'ADN chez la levure est faiblement affectée par l'organisation tridimensionnelle

Nous avons montré que la réorganisation du génome dans la SY14 n'avait pas engendré de réorganisation complète du programme de réplication, mais un mélange de modifications importantes localisées au niveau des télomères et des centromères, et de modifications plus faibles dispersées tout le long du génome. Néanmoins, nous ne pouvons pas exclure d'autres mécanismes que la réorganisation spatiale du génome proprement dite pour expliquer ces changements.

### VI.2.1 Les origines à proximité des centromères

Le mécanisme majeur imposant un *timing* de réplication précoce aux origines péricentromériques est le recrutement de Dbf4 par la protéine du kinétochore Ctf19. Les effets observés dans la SY14 sont relativement similaires à ceux observés dans une souche *Ctf19Δ*. Néanmoins, Lazar-Stefanita et al. (2022), qui ont étudié la réplication dans la souche de levures à deux chromosomes (Luo et al 2018), suggèrent que les effets observés sur les origines centromériques sont liés à l'absence de regroupement des centromères. L'analyse d'un mutant *Ctf19Δ* dans la BY4742 devrait permettre de conclure. Si le *timing* des centromères dans la BY4742 *Ctf19Δ* se révèle similaire à celui de la SY14, il n'est pas nécessaire d'invoquer un défaut de regroupement des centromères pour expliquer les effets observés dans la SY14. On pourrait aussi vérifier par ChIP-seq si la fixation de Ctf19 est bien abolie dans les régions où le centromère est supprimé. Si ce n'est pas le cas, on pourrait imaginer que l'absence de regroupement des centromères est responsable de l'inactivité de Ctf19. Un autre moyen de vérifier si le regroupement des centromères impacte le *timing* de réplication, et non l'action unique de Ctf19, serait de vérifier par Hi-C si les centromères sont toujours regroupés dans une souche *Ctf19Δ*. Cela permettrait de dissocier le mécanisme d'action de Ctf19 du regroupement des centromères. Enfin, la réalisation d'un double mutant SY14 *Ctf19Δ* permettrait de déterminer si Ctf19 exerce une action de stimulation des origines péricentromériques ne nécessitant pas les séquences centromériques supprimées lors de la construction de la SY14.

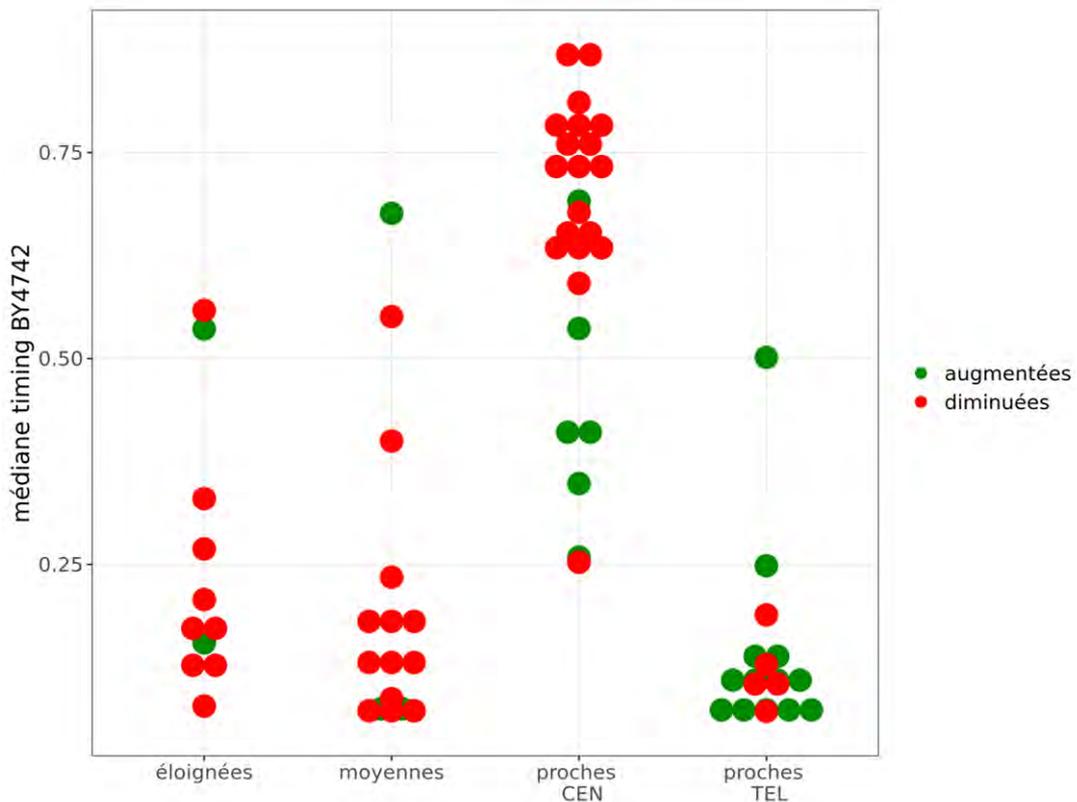
### VI.2.2 Les origines à proximité des télomères

Nous avons montré que les origines situées à moins de 25kb de l'extrémité d'un chromosome dans la BY4742, et dont l'activité augmente dans la SY14, sont normalement inhibées par Rif1, indépendamment de son interaction avec Rap1, et/ou sont situées à proximité d'un gène télomérique dont l'expression est augmentée dans la SY14 (Tableau supplémentaire VII-5). Plusieurs mécanismes peuvent être proposés pour expliquer cette observation. (i) La taille des télomères régule le *timing* de réplication (Bianchi and Shore, 2007). Il est donc possible que l'absence de télomères dans la SY14 avance le moment de déclenchement et/ou l'efficacité de certaines origines. Néanmoins, deux origines positionnées sur les chromosomes XVI et X, à une distance inférieure à 25kb de l'extrémité du chromosome dans la BY4742, ne montrent pas d'augmentation de leur activité dans la SY14. Il est possible que le raccourcissement des télomères de ces deux chromosomes soit insuffisant pour affecter ces deux origines. Nous pourrions envisager de raccourcir la taille des télomères dans la BY4742, d'observer quelles origines sont impactées et s'il s'agit des mêmes origines que dans la SY14. (ii) Ces origines sont majoritairement inhibées par Rif1 (Hafner et al., 2018) et/ou sont à proximité d'un gène dont l'expression est augmentée dans la SY14 (Shao et al., 2018), on peut donc supposer que la

perte des télomères dans la SY14 entraîne un défaut de fixation de Rif1 et des protéines Sir. La comparaison des sites de fixation de Rif1 et des protéines Sir qui sont responsables de l'inhibition de l'expression des gènes télomériques permettrait de déterminer si les changements observés peuvent ou non s'expliquer par une modification de la fixation de ces protéines. (iii) Les contraintes spatiales imposées par la nouvelle organisation du génome pourraient forcer ces origines à résider à l'intérieur du noyau, ce qui contrebalancerait les mécanismes d'inhibition de ces origines par une plus grande disponibilité en facteurs limitants. De plus, les chromosomes XVI et X possèdent deux origines positionnées à moins de 25kb de l'extrémité d'un chromosome dans la BY4742, d'activité inchangée dans la SY14. Ces deux chromosomes possèdent encore un télomère à leur autre extrémité, qui réside à la périphérie nucléaire. Il est possible que l'intégralité de ces chromosomes réside à la périphérie nucléaire et par conséquent, que les deux origines non modifiées y résident aussi. Pour le vérifier, il serait possible de placer un marqueur fluorescent à côté de ces origines afin de visualiser leur localisation nucléaire. (iv) Des études suggèrent un mécanisme de rétrocontrôle de l'action de Rif1 médié par sa phosphorylation par DDK inhibant sa liaison à PP1 (Hiraga et al., 2014 ; Mattarocci et al., 2014). Le mécanisme suggéré par les auteurs est une inhibition de Rif1 en fin de phase S par DDK pour rendre les origines actives. Il est donc possible que Dbf4, qui ne se lie plus aux origines centromériques en absence de Ctf19, contrecarre davantage l'action de Rif1. Nous pourrions vérifier ce mécanisme en déterminant si la fixation de Rif1 est inchangée, mais la fixation de Dbf4 augmentée, aux origines augmentées dans la SY14.

### VI.2.3 Diminution de l'activité des origines éloignées des centromères et des télomères

Les modifications de la SY14 consistent en 47 origines diminuées et 23 origines augmentées. Ce résultat est contre-intuitif, car l'inhibition des origines centromériques permet de libérer des facteurs limitants ce qui pourrait prédire une activité globalement plus importante des origines dans la SY14, or ce n'est pas ce qui est observé. De plus, les origines diminuées dans la SY14 possèdent en majorité un *timing* de réplication tardif (Figure VI-7). Les origines inhibées dans la SY14 qui ne sont pas à proximité des centromères et des télomères semblent être capable de fixer plus de complexe MCM dans une souche au caryotype normal d'après les données de Dukaj and Rhind (2021). Il faudrait pour interpréter cette observation savoir si l'inhibition de ces origines s'accompagne ou non d'une diminution de leur chargement en MCM. Si les origines diminuées possèdent moins de MCM dans la SY14 que dans la BY4742, alors l'inhibition de ces origines sera sans doute liée à un défaut de chargement des MCM. Cependant si les origines diminuées possèdent autant de MCM dans la BY4742 et la SY14 cela suggérera que ces origines possèdent un défaut d'activation des MCM.



**Figure VI-7 : Timing de réplication dans la BY4742 pour les origines modifiées dans la SY14 en fonction de la proximité des origines avec les télomères et les centromères.**

Les origines proches sont situées à une distance inférieure ou égale à 55kb d'un centromère ou d'un télomère. Ces origines sont ensuite classées dans la catégorie CEN ou TEL en fonction de l'élément pour lequel elles sont le plus proches. Les origines moyennes sont à une distance comprise entre 55kb et 151kb des centromères et télomères et les origines éloignées sont situées à une distance supérieure à 151kb des centromères et des télomères.

### VI.2.3.1 Candidats potentiels pour l'inhibition des origines dans la SY14

#### VI.2.3.1.1 Rif1

Rif1 n'étant plus séquestré au niveau des télomères, on pourrait imaginer que les origines diminuées dans la SY14 sont inhibées par Rif1. Cependant cette catégorie d'origine n'est pas enrichie en origines inhibées par Rif1 ou par un mutant Rif1 dépourvu du site de liaison pour Rap1 qui le séquestre aux télomères (Rif1-RBM, Tableau supplémentaire VII-7). Il est toutefois possible qu'un grand nombre de sites de fixation pour Rif1 soient perdus dans la SY14 et que Rif1 soit davantage disponible pour inhiber des origines qu'il ne contrôle pas habituellement. Dans ce cas, l'analyse de la fixation de Rif1 dans la SY14 pourrait permettre de mettre en évidence une redistribution de ses sites de fixation, et l'analyse de la réplication dans une souche SY14-*Rif1Δ* devrait déceler une augmentation

de l'activité de ces origines. Il est aussi possible d'envisager un mécanisme intermédiaire entre organisation nucléaire et régulation par Rif1, où les origines diminuées dans la SY14 résident à proximité de l'enveloppe nucléaire où Rif1 est séquestré par sa palmitoylation (Park et al., 2011). Ainsi, des origines habituellement inaccessibles à Rif1, en raison d'une localisation intranucléaire, pourraient devenir plus accessibles à Rif1 et être inhibées si elles peuvent accéder à ce compartiment dans la SY14. Déterminer la colocalisation des origines modifiées avec Rif1 dans la SY14 et la BY4742 pourrait permettre de tester ce mécanisme.

#### VI.2.3.1.2 Les histone désacétylases Sir2 et Sir3.

Sir2 et Sir3 sont deux histone désacétylases liées aux régions télomériques. La suppression des télomères dans la SY14 augmente potentiellement la quantité de ces protéines disponible pour leurs autres sites de fixation dans le génome. Doughty et al., (2021) ont montré que la SY14 possédait un métabolisme fermentatif perturbé, liée à une diminution de l'expression de certains gènes impliqués dans le passage de la respiration à la fermentation. En présence de nicotinamide, un inhibiteur des désacétylases de la famille Sir, la capacité de la SY14 à croître en présence d'éthanol est rétablie. Ceci suggère que dans cette souche, les Sir inhibent tout ou partie des gènes impliqués dans le passage de la respiration à la fermentation. Les auteurs ont suggéré une redistribution des protéines Sir en absence de leur site de fixation télomérique pour expliquer l'inhibition de ces gènes. Comme Sir2 et Sir3 sont responsable de la désacétylation de nucléosomes bordant les origines de réplication, ce qui réduit le recrutement du complexe MCM (Hoggard et al., 2018), il est possible qu'une redistribution de protéines Sir soit responsable de l'inhibition des origines observées dans la SY14. Nous pourrions tester cette hypothèse en analysant la fixation de Sir2 et Sir3 par CHIP-seq dans la BY4742 et la SY14.

#### VI.2.3.2 Modification des origines et organisation nucléaire

##### VI.2.3.2.1 Fkh1/2

Le mécanisme d'action de Fkh1/2 et l'organisation du génome sont intimement liés, car Fkh1/2 imposent aux origines qu'ils régulent un positionnement au centre du noyau. Cependant, ce positionnement ne semble pas essentiel à leur action, car lorsqu'une origine activée par Fkh1/2 est positionnée à la périphérie nucléaire, elle reste activée de manière précoce (Ebrahimi et al, 2014). Nos résultats montrent que les origines activées par Fkh1/2 ne sont pas modifiées dans la SY14. On aurait pu s'attendre à une stimulation de ces origines par la suppression de la compétition pour le recrutement des facteurs limitants avec les centromères.

Le résultat le plus curieux est que nombre d'origines inhibées par Fkh1/2 voient leur activité encore diminuée dans la SY14. En présence de Fkh1/2, ces origines sont défavorisées dans la

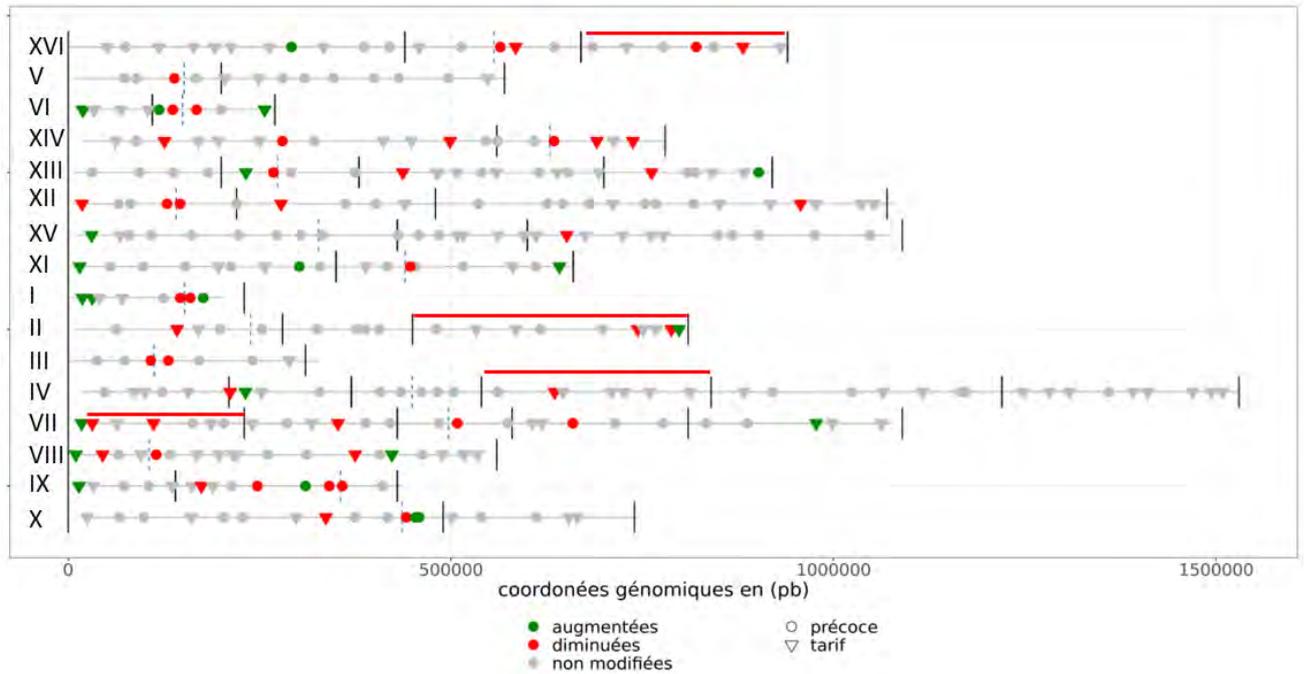
compétition pour l'accès aux facteurs limitants, peut-être parce que le regroupement des origines stimulées par Fkh1/2 relègue ces origines dans une région nucléaire moins favorable. Il est possible que la réorganisation du génome dans la SY14 leur impose une localisation encore plus défavorable que dans la BY4742. Pour évaluer cette hypothèse, il faudrait examiner la localisation des origines inhibées par Fkh1/2 en présence et en absence de ces protéines dans la SY14 et la BY4742.

#### VI.2.3.2.2 Les TADs

Eser et al. (2017) ont identifié des TADs dans la levure et ont montré qu'ils comportaient des origines avec un *timing* de réplication similaire. Les interactions intrachromosomiques étant renforcées dans la SY14, on pourrait supposer que les TADs et leurs différences de *timing* sont renforcés. La Figure VI-8 présente les TADs observés par Eser et al. (2017) en fonction du *timing* de réplication dans la BY4742 (précoce  $> 0,5$  et tardif  $\leq 0,5$ ), ainsi que leur modification dans la SY14. Certains TADs comprenant essentiellement des origines tardives contiennent des origines diminuées dans la SY14. Par exemple, le chromosome IV comporte un TAD tardif, contenant une seule origine précoce, et dont certaines origines sont diminuées dans la SY14 (Figure VI-8, lignes rouges). Néanmoins, il se pourrait que les TADs de la SY14 diffèrent de ceux de la BY4742. Une analyse des données de conformation chromosomique de la BY4742 et de la SY14 serait nécessaire pour déterminer s'ils sont différents entre la BY4742 et la SY14 et si ces différences sont corrélées aux changements d'activité des origines.

#### **Figure VI-8 : TADs chez la levures et origines modifiées dans la SY14.**

Les bordures des TADs sont représentées par les lignes noires. Les centromères sont représentés par les lignes pointillées bleues. Les lignes rouges représentent des exemples de TADs avec une majorité d'origines tardives et une ou plusieurs origines diminuées dans la SY14.



### VI.2.3.2.3 Organisation des origines au sein du noyau

Les origines précoces ont tendance à se regrouper avec les origines péricentromériques et les origines activées par Fkh1/2 (Duan et al., 2010; Knott et al., 2012). Ce regroupement semble contraint par l'attachement des télomères à la périphérie nucléaire et l'attachement des centromères au corpuscule polaire (Tjong et al., 2012). Ces événements étant perturbés dans la SY14, il est possible que le regroupement d'origines précoces ne soit pas maintenu et que cela entraîne un retard de leur activation. Néanmoins, la majorité des modifications observées dans la SY14 concernent des origines tardives. L'analyse des données de conformation chromosomique obtenues dans la SY14 pourrait nous permettre de déterminer si ce regroupement a toujours lieu. Si le regroupement des origines précoces n'a plus lieu dans la SY14, cela signifierait que ce regroupement n'est pas nécessaire pour induire la réplication précoce. Si le regroupement a toujours lieu dans la SY14, cela signifierait que le regroupement est indépendant des contraintes physiques imposées par la localisation des centromères et des télomères.

La diminution de l'activité des origines dans la SY14 pourrait être induite par un positionnement moins favorable au sein du noyau, par une pauvreté locale en facteurs limitants, ou en raison du recrutement de PP1 par Rif1. Le marquage des origines diminuées dans la SY14 avec une protéine fluorescente pourrait permettre de comparer leur localisation avec celle constatée dans la BY4742. Néanmoins, toutes les analyses de localisation des origines prennent en compte un modèle où la quantité de facteurs limitants est plus importante au centre du noyau et à proximité des origines

péricentromériques, et moins importante à la périphérie nucléaire, mais nous n'avons aucun élément pour attester d'une répartition similaire des facteurs limitants dans la BY4742 et la SY14. Cette éventuelle hétérogénéité est difficile à tester expérimentalement.

#### VI.2.4 Conclusion et perspectives

L'analyse de la réplication dans une souche de levure à un chromosome a permis de montrer que les origines péricentromériques étaient impactées d'une manière similaire à celle observée dans une souche Ctf19 $\Delta$ , ce qui suggère que seule la présence de Ctf19 régule l'activité des origines péricentromériques. Les régions péritélomériques sont quant à elles impactées par la réorganisation du génome, principalement via la traversée des fusions chromosomiques par les fourches de réplication. Néanmoins la suppression des télomères a aussi entraîné une avancée du *timing* de réplication ou une augmentation de l'efficacité de certaines origines situées à moins de 25kb des fusions chromosomiques. Pour finir, nous avons observé dans la SY14 une modification d'activité, le plus souvent inhibitrice, de plusieurs origines éloignées des centromères et des télomères. Ces résultats suggèrent un impact de la réorganisation du génome sur la réplication au-delà des effets locaux au voisinage des régions supprimées ou réarrangées. De plus, il semblerait qu'il existe une interconnexion entre les protéines régulant la réplication et l'organisation du génome. De plus amples analyses sont nécessaires pour déterminer la nature des connexions entre ces deux mécanismes.

## VII. Table supplémentaire

**Tableau supplémentaire VII-1 : Liste des échantillons biologiques utilisés.**

Ce tableau référence les échantillons biologiques utilisés dans ce manuscrit, la souche utilisée, l'appareil de séquençage utilisé, la concentration en BrdU lors du marquage, et le type d'expérience réalisée. Ce tableau indique par une croix les échantillons qui ont permis de générer les différents profils de réplication utilisés pour l'analyse de la réplication dans la souche SY14.

<b>Souche</b>	<b>Appareil</b>	<b>Concentration BrdU</b>	<b>Expérience</b>	<b>Utilisation des échantillons pour les profils de réplication</b>
SY-pBL-AUR1	MinION	1mM	NFS	
BY-pBL-AUR1	MinION	1mM	NFS	
SY-pBL-URA3	MinION	1mM	NFS	
BY-pBL-URA3	MinION	1mM	NFS	
SY-pBL-AUR1	MinION	100µM	NFS	
BY-pBL-AUR1	MinION	100µM	NFS	
SY-pBL-URA3	MinION	100µM	NFS	
BY-pBL-URA3	MinION	100µM	NFS	
SY-pBL-URA3	promethION	1mM	NFS	<b>X</b>
BY-pBL-URA3	promethION	1mM	NFS	<b>X</b>
SY-pBL-URA3	promethION	1mM	NFS	<b>X</b>
BY-pBL-URA3	promethION	1mM	NFS	<b>X</b>
SY-pBL-URA3	promethION	1mM	NFS	<b>X</b>
BY-pBL-URA3	promethION	1mM	NFS	<b>X</b>
BY-pBL-URA3	MinION	10µM	Nanotiming	
BY-pBL-URA3	MinION	25µM	Nanotiming	<b>X</b>
SY-pBL-URA3	MinION	10µM	Nanotiming	
SY-pBL-URA3	MinION	25µM	Nanotiming	<b>X</b>
BY-pBL-URA3	MinION	25µM	Nanotiming	
BY-pBL-URA3	MinION	50µM	Nanotiming	
BY-pBL-URA3	MinION	100µM	Nanotiming	
SY-pBL-URA3	MinION	25µM	Nanotiming	
SY-pBL-URA3	MinION	50µM	Nanotiming	
SY-pBL-URA3	MinION	100µM	Nanotiming	

BY-pBL-URA3	MinION	10µM	Nanotiming	<b>X</b>
BY-pBL-URA3	MinION	25µM	Nanotiming	
BY-pBL-URA3	MinION	10µM	Nanotiming	<b>X</b>
BY-pBL-URA3	MinION	25µM	Nanotiming	
SY-pBL-URA3	MinION	25µM	Nanotiming	<b>X</b>
SY-pBL-URA3	MinION	50µM	Nanotiming	
SY-pBL-URA3	MinION	25µM	Nanotiming	<b>X</b>
SY-pBL-URA3	MinION	50µM	Nanotiming	

### Tableau supplémentaire VII-2 : Table de conversion des données.

Cette table présente les coordonnées dans la SY14 et la BY4742 des régions similaires définies par alignement des génomes. Le chromosome XVI est le premier chromosome du mégachromosome de la SY14 et possède son télomère gauche. Le chromosome X est le dernier chromosome du mégachromosome de la SY14 et possède son télomère droit.

Chromosome BY4742	Début	Fin	Chromosome SY14	Début	Fin
chrI	3471	152032	chrI	6241034	6389597
chrI	152308	203742	chrI	6389597	6441032
chrII	9217	237896	chrI	6441032	6669721
chrII	238906	805482	chrI	6669720	7236312
chrIII	1176	111875	chrI	7236312	7347009
chrIII	112556	330823	chrI	7347016	7565287
chrIV	18914	449099	chrI	7565287	7995495
chrIV	449601	1530098	chrI	7995495	9075991
chrV	8532	18519	chrI	948857	958844
chrV	23899	151178	chrI	958844	1086126
chrV	151938	574254	chrI	1086126	1508448
chrVI	9138	149215	chrI	1508447	1648526
chrVI	149530	270371	chrI	1648526	1769230
chrVII	9767	496927	chrI	9080295	9567467
chrVII	497240	1076306	chrI	9567466	10146542
chrVIII	8464	105688	chrI	10146542	10243770
chrVIII	106255	227297	chrI	10243770	10364817
chrVIII	235490	538708	chrI	10374877	10678101
chrIX	11093	355649	chrI	10689002	11033559
chrIX	356048	436400	chrI	11033558	11113912
chrX	21647	745649	chrI	11113912	11837734
chrXI	3321	439685	chrI	5586383	6022753
chrXI	440399	658678	chrI	6022753	6241034
chrXII	4793	140583	chrI	3443721	3579061
chrXII	140848	507310	chrI	3579503	3945965
chrXII	507311	1082395	chrI	3949651	4527981

chrXIII	9300	273700	chrI	2530060	2794477
chrXIII	274491	923278	chrI	2794477	3443271
chrXIV	18095	629034	chrI	1769229	2380172
chrXIV	629520	779405	chrI	2380172	2530061
chrXV	10572	13248	chrI	4527979	4534724
chrXV	13249	21902	chrI	4534725	4543388
chrXV	30958	1073938	chrI	4543406	5586383
chrXVI	3	556254	chrI	7329	563610
chrXVI	557039	942256	chrI	563631	948853

### Tableau supplémentaire VII-3 : Liste des origines significativement modifiées dans la SY14.

X : origine significative dans l'analyse

- : origine modifiée dans l'analyse sans que la modification soit significative

Voisin : l'origine voisine est modifiée dans l'analyse

TER : une des zones de terminaisons est significativement modifiée dans l'analyse du RFD, sans que la zone d'initiation le soit.

Tardive : le *timing* de répllication de l'origine est tardif.

La régulation par les origines voisines a été définis visuellement, une origine est répliquée passivement lorsque l'origine voisine possèdent un *timing* de répllication avancé ou lorsque l'origine est impliquée dans une fusion chromosomique avec une origine plus précoce. Les origines opportunistes sont celles qui bénéficie d'une moindre passivation par une origine voisine en raison d'un délais d'activation de cette dernière.

ARS	IPL	Nanotiming	RFD	Régulation par les origines voisines
ARSI_17		tardive	X	
ARS104		tardive	X	
ARS108	-	X	X	
ARS109			X	
ARS110	X		-	
ARS206	-	tardive	X	
ARS224		tardive	X	
ARS228	X	tardive	X	répliquée passivement
ARS229		-	X	
ARS307	-	X	X	
ARS309		-	X	
ARS409		tardive	X	
ARS409.3	-	tardive	X	
ARS420		X	TER	
ARS510	-	X	X	
ARS600.3	-	tardive	X	
ARS600.4	-	tardive	X	
ARS603.5	-	voisin	X	opportuniste
ARS605	-	X	X	
ARS606			X	
ARS609	-	tardive	X	
ARS700.5	-	X	X	
ARSVII_33		voisin	X	repliquée passivement
ARS704		tardive	X	
ARS716		tardive	X	
ARS720	X	-	X	
ARS727		X		
ARS733	-	tardive	X	
proARS803	X	tardive	-	
ARS804		X	-	

proARS806	-	X	X	
ARS816		X	-	
ARS818.5		tardive	X	
proARS904	X	tardive	X	
ARS912	-	X	X	
ARS915		X	X	
ARS916	X	voisin	X	opportuniste
ARS919		X	X	
ARS920	-	X	X	
ARS1011	-	X	TER	
ARS1015	X	X	X	
ARS1016		voisin	X	opportuniste
ARS1017		voisin	X	opportuniste
ARSXI_16		-	X	
ARS1107.5			X	
ARS1114		X	-	
ARS1123		tardive	X	
proARS1202		voisin	X	répliquée passivement
proARS1207	X	-	X	
ARS1209	X	X	X	
ARS1212	-	tardive	X	
ARS1229		tardive	X	
ARSXIII_227	X	voisin	X	opportuniste
ARS1309	X	X	X	
ARS1315	-	tardive	X	
ARS1327		tardive	X	
ARS1332		X		
ARS1410		tardive	X	
ARS1414			X	
ARS1420			X	
ARS1426	X	X	X	
ARS1427		tardive	X	
ARS1428.5		tardive	X	répliquée passivement
ARS1506.5		-	X	
ARS1519		tardive	X	
ARS1614		X		
ARS1622	X	X	X	
ARS1622.7	-	X	X	
ARS1627			X	
ARS1630		tardive	X	

**Tableau supplémentaire VII-4 : Régulation des origines modifiées proches des centromères par Ctf19.**

+ : Origines activées par Ctf19

<b>ARS</b>	<b>activité</b>	<b>Ctf19</b>
ARS1622	diminuée	+
ARS1622.7	diminuée	+
ARS510	diminuée	+
ARS603.5	augmentée	
ARS605	diminuée	+
ARS1426	diminuée	+
ARSXIII_227	augmentée	
ARS1309	diminuée	+
proARS1207	diminuée	+
ARS1209	diminuée	+
ARS108	diminuée	+
ARS109	diminuée	+
ARS110	augmentée	
ARS307	diminuée	+
ARS309	diminuée	+
ARS720	diminuée	+
proARS806	diminuée	+
ARS916	augmentée	
ARS919	diminuée	+
ARS920	diminuée	+
ARS1015	diminuée	+
ARS1016	augmentée	
ARS1017	augmentée	
ARS606	diminuée	
ARS1114	diminuée	+

**Tableau supplémentaire VII-5 : Régulation des origines augmentées à proximité des télomères.**

- : inhibition par la protéine

+ : activation par la protéine

NA : origines absentes du jeu de données, Rif1 et Rif1-RBM (Hafner et al. 2018), KU (Das et al. 2015), Fkh1/2 (Knott et al. 2012).

Les origines à moins de 30kb d'un gène télomérique possédant une expression augmentée dans la SY14 sont représentés par une croix.

ARS	Rif1	KU	Rif1-RBM	Distance de l'extrémité du chromosome dans la BY4742 (pb)	Gène télomérique	Fkh1/2
ARS600.3	-	<b>NA</b>	-	18691	<b>X</b>	<b>NA</b>
ARS600.4	-	-	-	19691	<b>X</b>	<b>NA</b>
ARS609	-	-	-	13746	<b>X</b>	<b>NA</b>
ARS1332			-	27516		<b>+</b>
ARS1506.5		-		30593	<b>X</b>	<b>+</b>
ARSXI_16	-	<b>NA</b>	-	14938	<b>X</b>	<b>NA</b>
ARS1123	-	-	-	24232	<b>X</b>	
ARSI_17	-	<b>NA</b>		18437	<b>X</b>	<b>NA</b>
ARS104		<b>NA</b>		31437	<b>X</b>	<b>+</b>
ARS229	-	-		23127		<b>+</b>
ARS700.5	-	-	-	17472		<b>NA</b>
proARS803		<b>NA</b>		9922	<b>X</b>	<b>NA</b>
proARS904	-	<b>NA</b>		14091	<b>X</b>	<b>NA</b>

**Tableau supplémentaire VII-6 : Modifications des origines à une distance moyenne et éloignée des centromères et des télomères par Rpd3 et Fkh1/2.**

- : inhibition par la protéine

+ : activation par la protéine

NA : origines absentes du jeu de données, Rpd3 (Knott et al., 2006), Fkh1/2 (Knott et al., 2012).

Moyenne				Eloignées			
ARS	Fkh	Rpd3	activité	ARS	Fkh	Rpd3	activité
ARS1627			diminuée	ARS1614	+		augmentée
ARS1630		NA	diminuée	ARS1414	-	-	diminuée
ARS1410		NA	diminuée	ARS1315		NA	diminuée
ARS1420	-	-	diminuée	ARS1327			diminuée
ARS1427		-	diminuée	ARS1519		NA	diminuée
ARS1212	-	NA	diminuée	ARS409	-		diminuée
ARS1229	-	NA	diminuée	ARS409.3	NA	NA	diminuée
ARS1107.5	NA	NA	augmentée	ARS420	-	NA	diminuée
ARS206	-	NA	diminuée	ARS727	+	-	diminuée
ARS224		NA	diminuée	ARS816		NA	diminuée
ARS704	NA	NA	diminuée	ARS912	-	-	diminuée
ARS716		-	diminuée				
ARS733	-	-	augmentée				
ARS818.5	NA	NA	augmentée				
ARS915	NA	NA	diminuée				
ARS1011	-	NA	diminuée				

**Tableau supplémentaire VII-7 : Modifications des origines à une distance moyenne et éloignée des centromères et des télomères par Rif1 et Rif1-RBM.**

- : inhibition par la protéine

+ : activation par la protéine

NA : origines absentes du jeu de données (Hafner et al 2018)

<b>ARS</b>	<b>activité</b>	<b>Rif1</b>	<b>Rif1-RBM</b>
ARS1614	augmentée		
ARS1627	diminuée		
ARS1630	diminuée		
ARS1410	diminuée		
ARS1414	diminuée		
ARS1420	diminuée		
ARS1427	diminuée		<b>+</b>
ARS1315	diminuée		
ARS1327	diminuée		
ARS1212	diminuée	-	
ARS1229	diminuée		
ARS1519	diminuée		
ARS1107.5	augmentée	<b>+</b>	
ARS206	diminuée		<b>+</b>
ARS224	diminuée	-	
ARS409	diminuée		
ARS409.3	diminuée		
ARS420	diminuée		
ARS704	diminuée	-	<b>+</b>
ARS716	diminuée		-

ARS727	diminuée		
ARS733	augmentée		
ARS816	diminuée		
ARS818.5	augmentée		
ARS912	diminuée		
ARS915	diminuée	<b>NA</b>	<b>NA</b>
ARS1011	diminuée		

**Tableau supplémentaire VII-8 : Liste des souches utilisées dans ce manuscrit.**

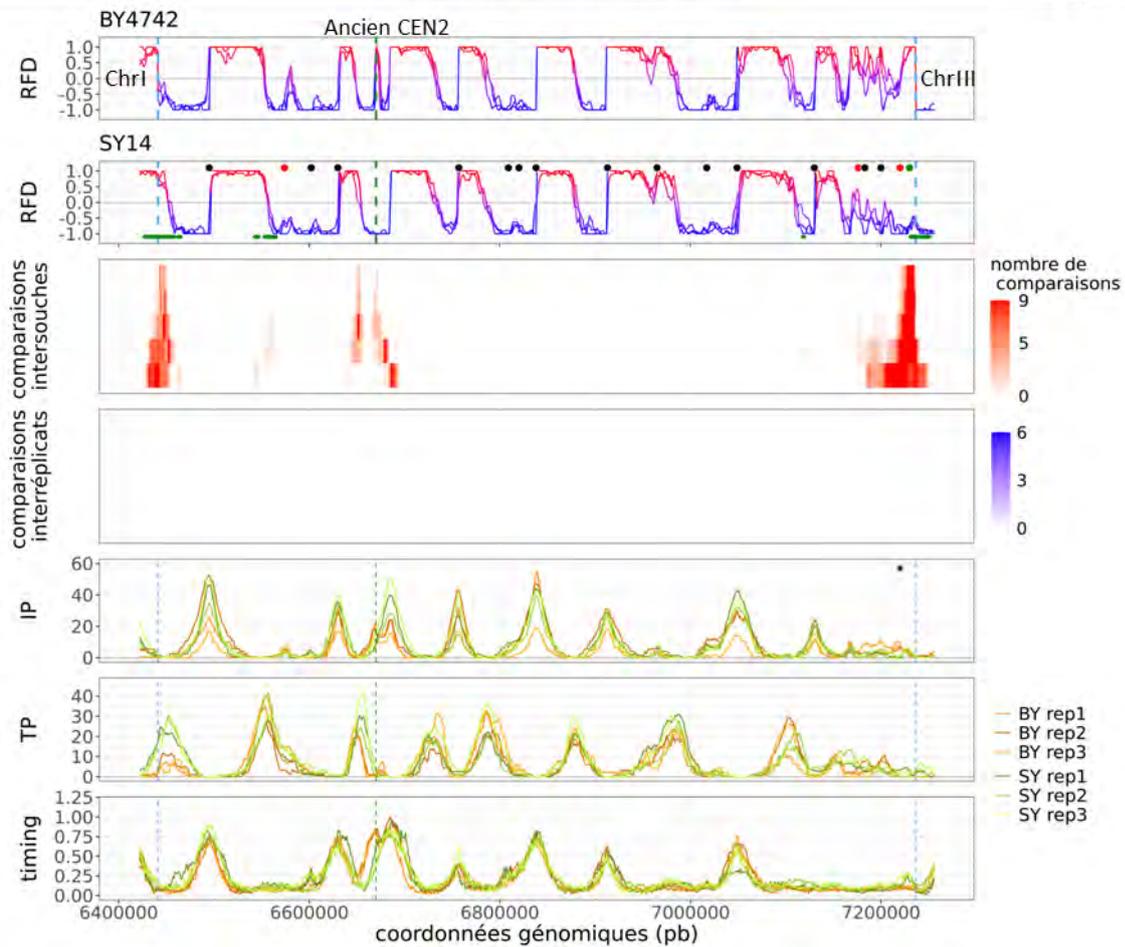
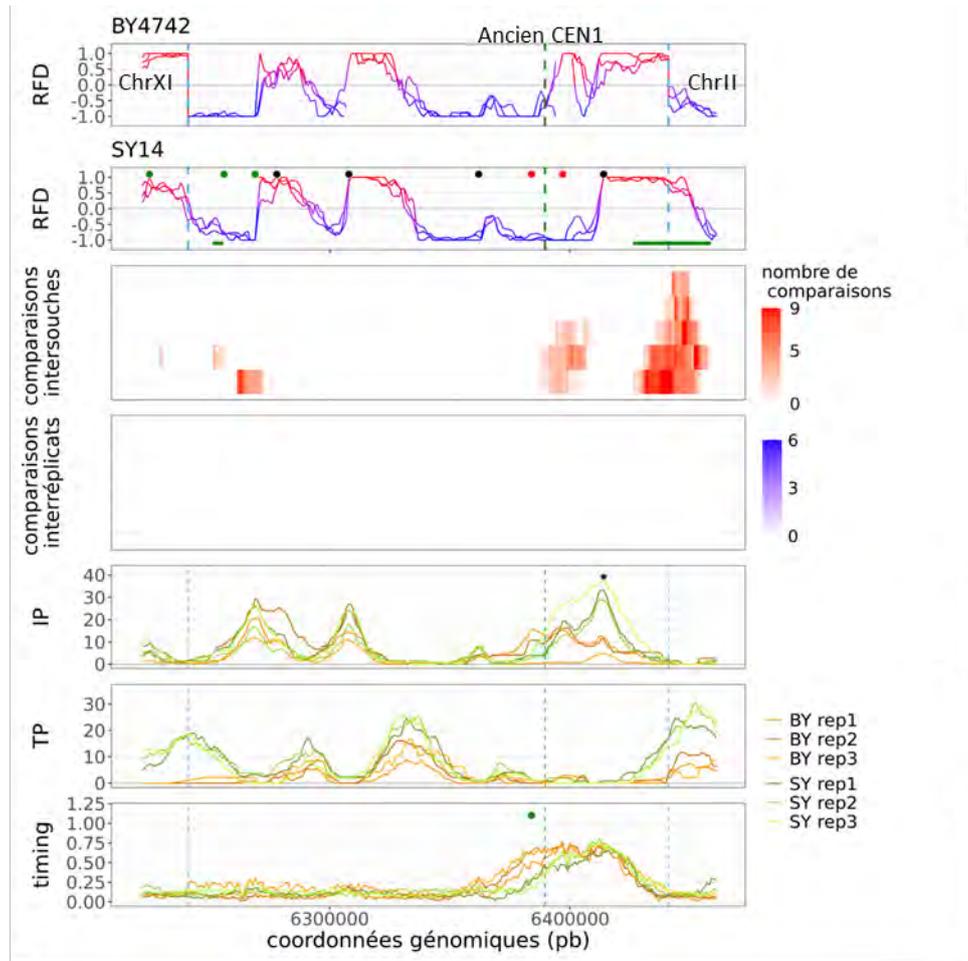
Souches	Génotypes
BT1	MATa trp1-1 leu2-3,112 his3-11,15 bar1::LEU2 ura3-1::URA3-GPD-hsvTK <sub>co</sub> -ADH-hENT1 <sub>co</sub>
W303	MATa trp1-1 leu2-3,112 his3-11,15 bar1::LEU2 ura3-1
BY4742	MATα ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0
SY14	MATα ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0
BY-pBL-AUR1	MATα ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 aur1:: AUR1-GPD-hsvTK <sub>co</sub> -ADH-hENT1 <sub>co</sub>
SY14-pBL-AUR1	MATα ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 aur1:: AUR1-GPD-hsvTK <sub>co</sub> -ADH-hENT1 <sub>co</sub>
W303-ura3	MATa trp1-1 leu2-3,112 his3-11,15 bar1::LEU2 URA3
BY4742-URA3	MATα URA3 leu2Δ0 his3Δ1 lys2Δ0
SY14-URA3	MATα URA3 leu2Δ0 his3Δ1 lys2Δ0
BY-pBL-URA3	MATα ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 URA3:: URA3-GPD-hsvTK <sub>co</sub> -ADH-hENT1 <sub>co</sub> -NOUR
SY14-pBL-URA3	MATα ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 URA3 :: URA3-GPD-hsvTK <sub>co</sub> -ADH-hENT1 <sub>co</sub> -NOUR

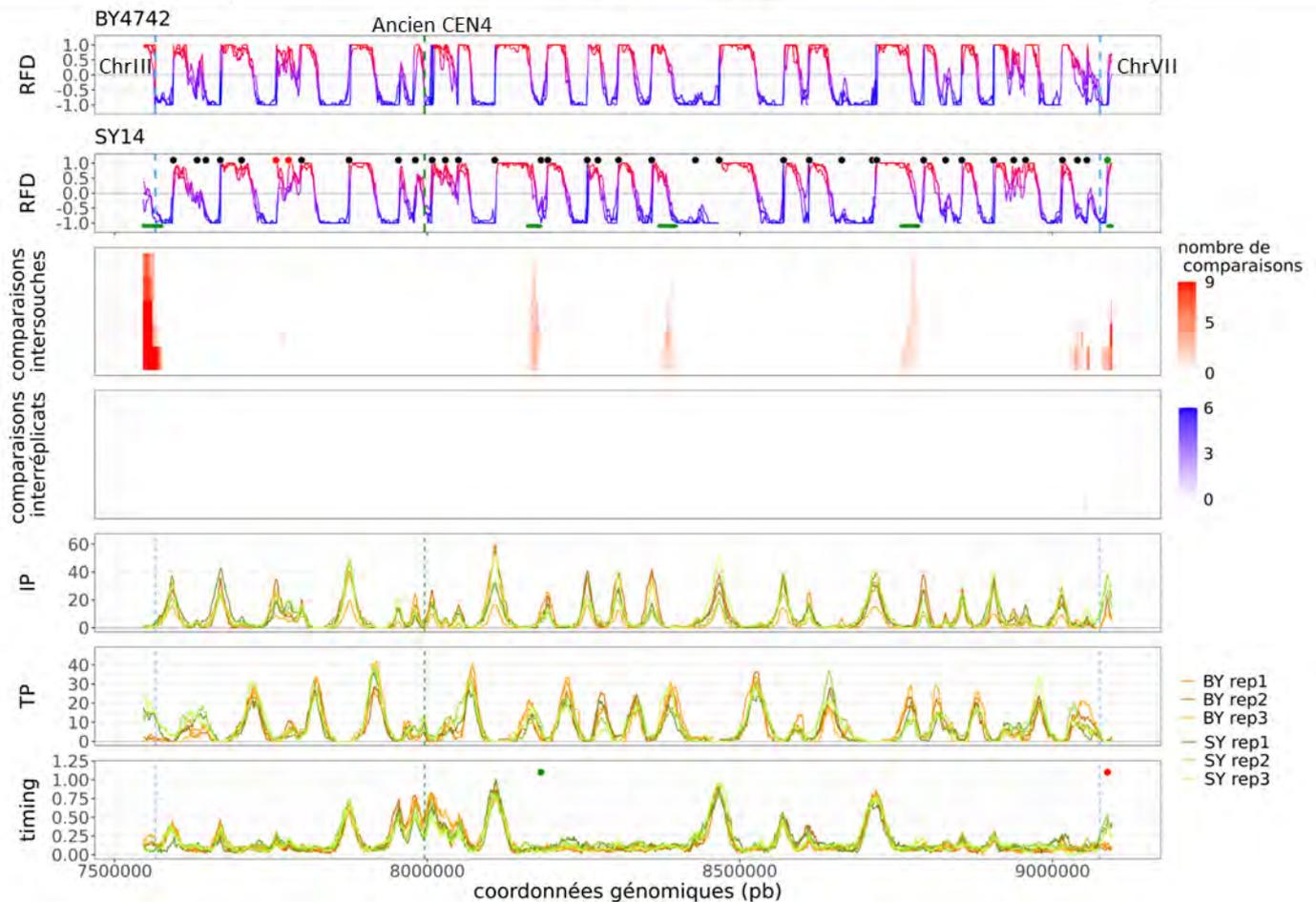
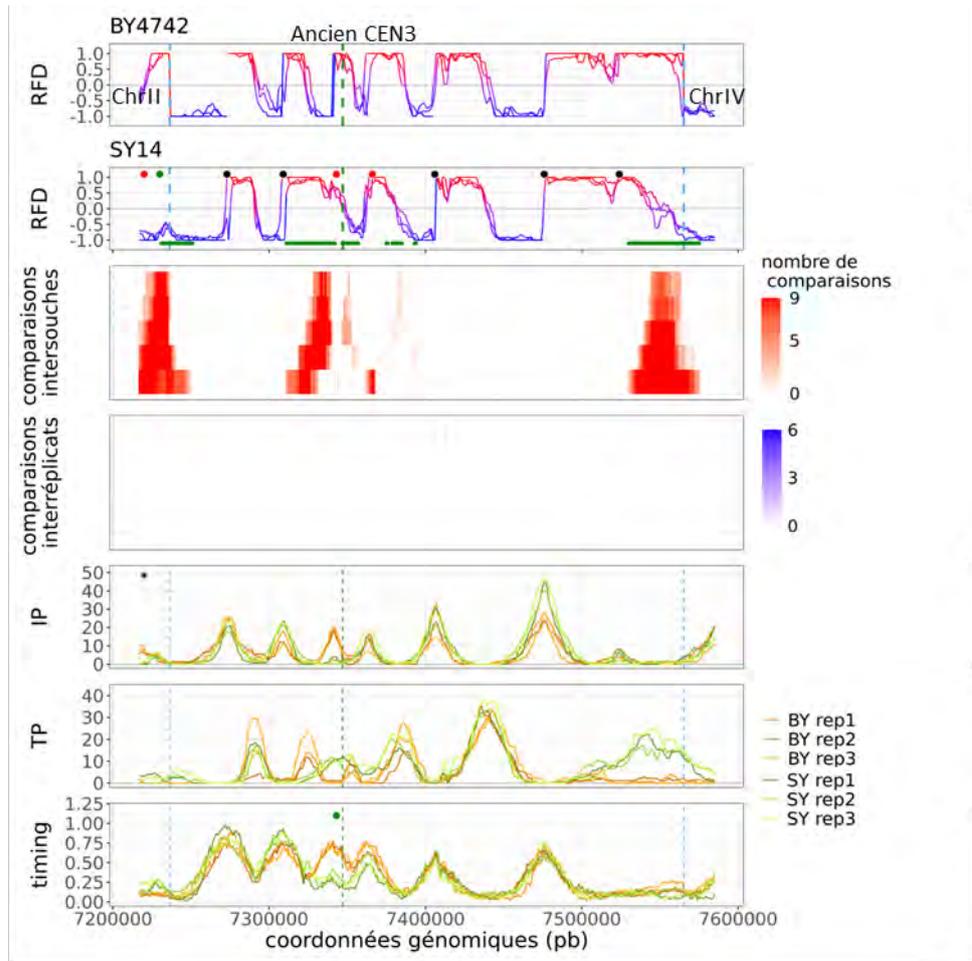
## VIII. Figures supplémentaires

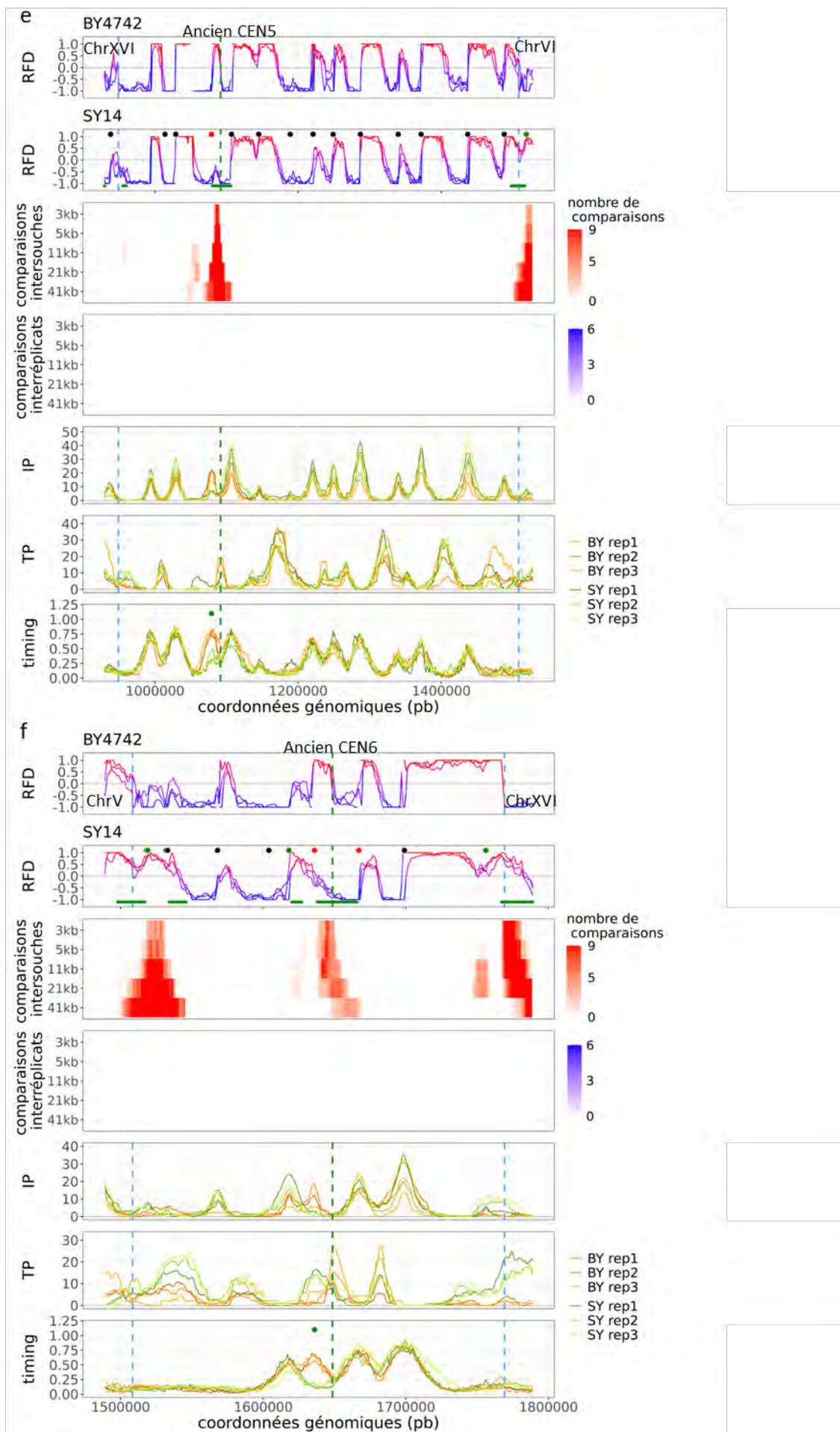
### **Figure supplémentaire VIII-1 : Comparaison des profils de réplication entre la BY4742 et la SY14, chromosomes entiers.**

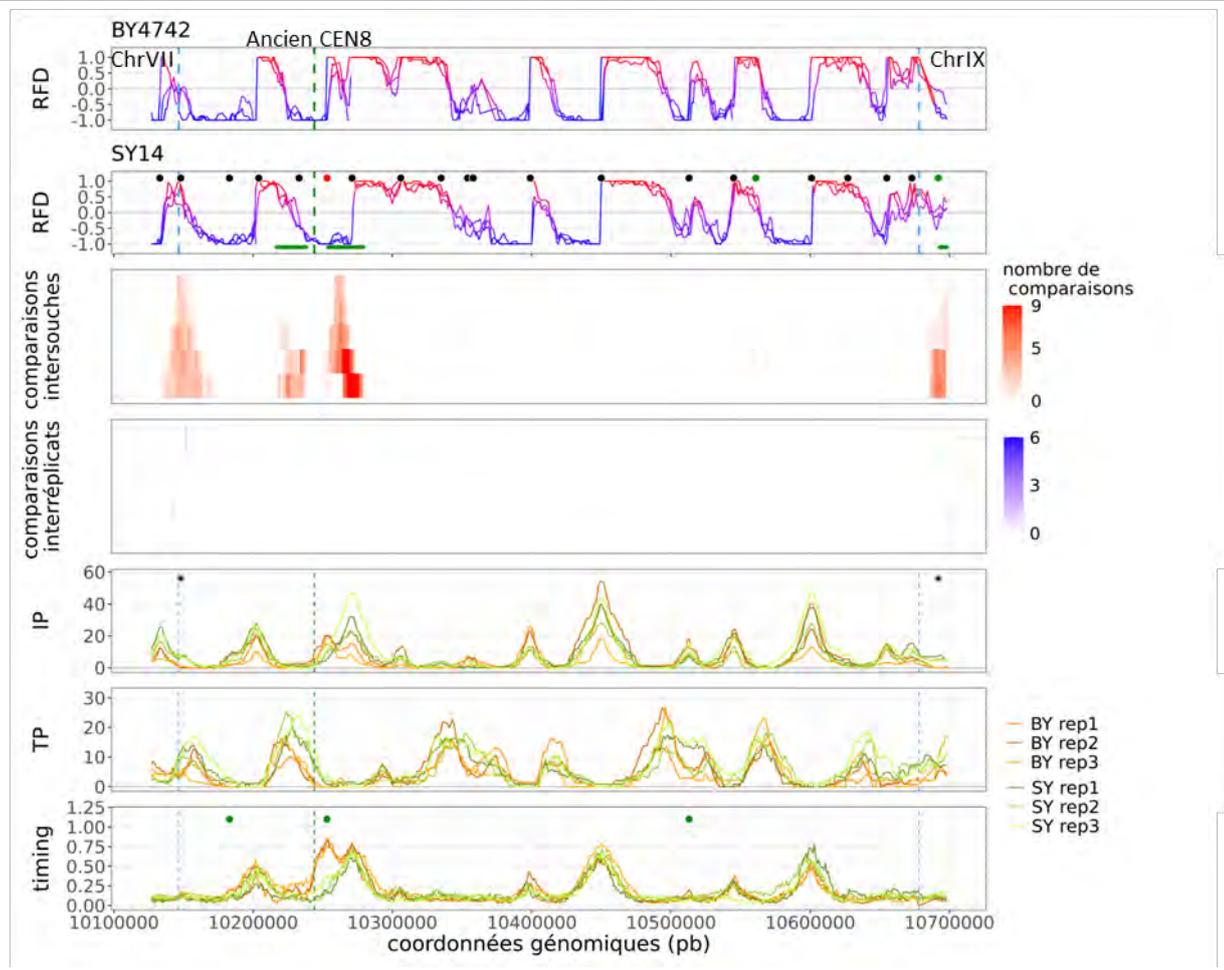
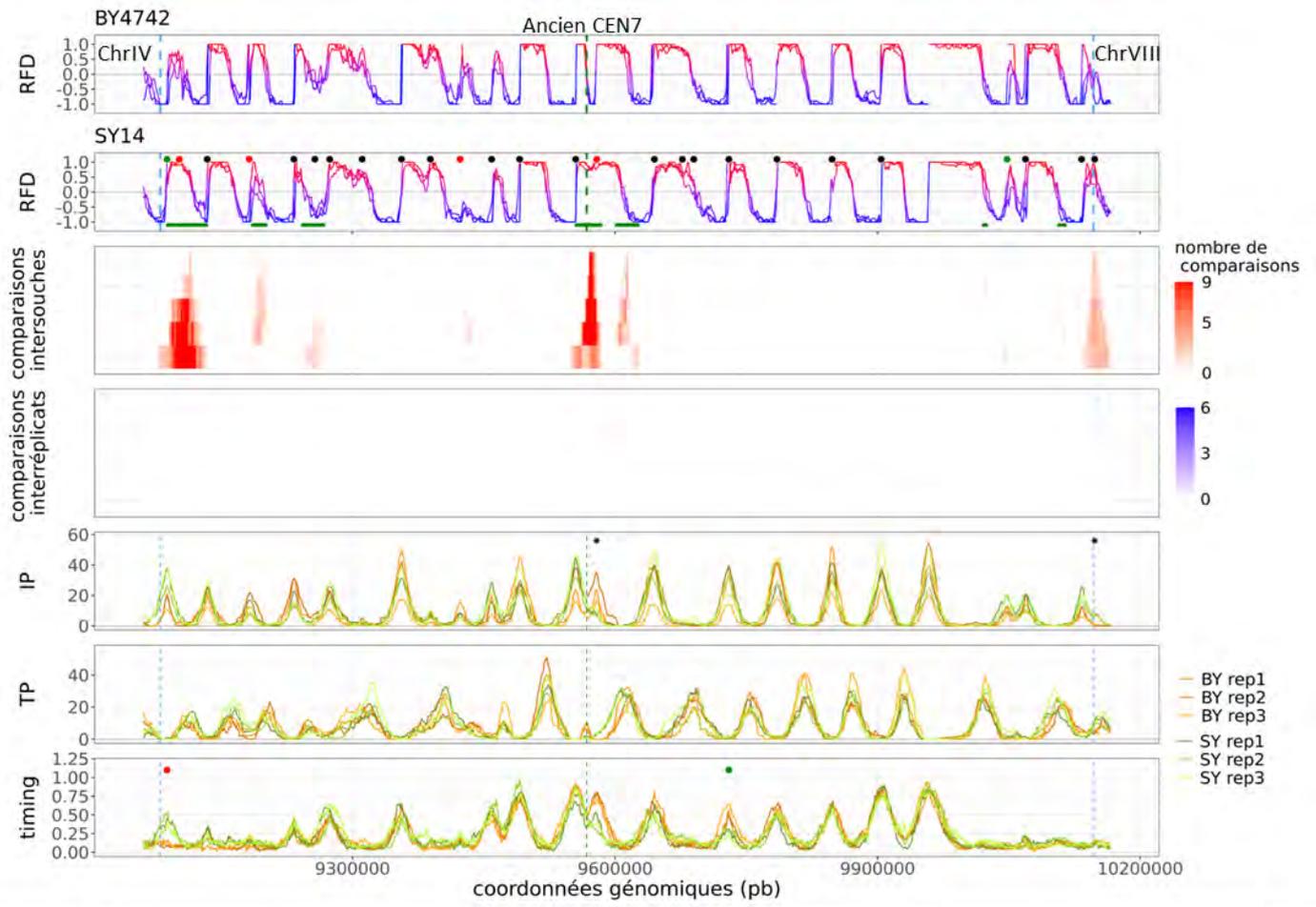
Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes.

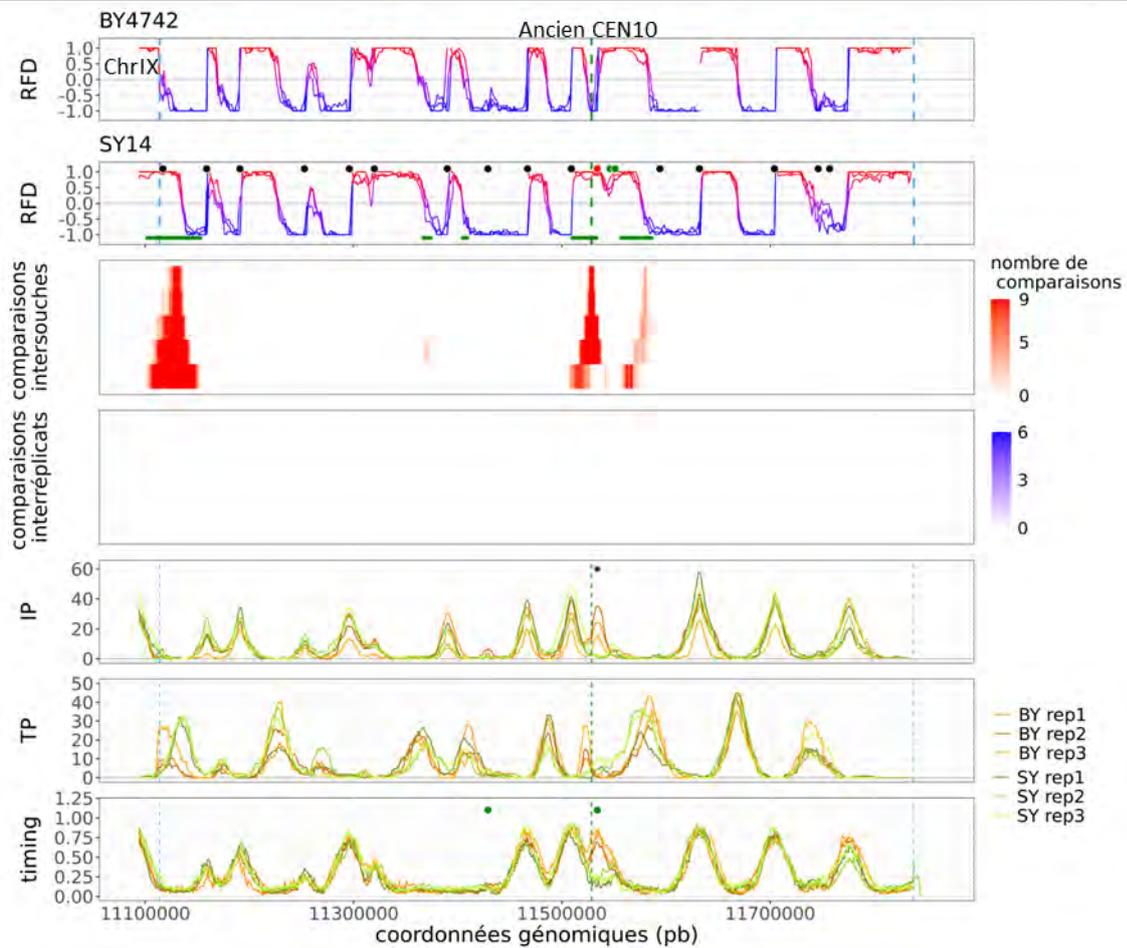
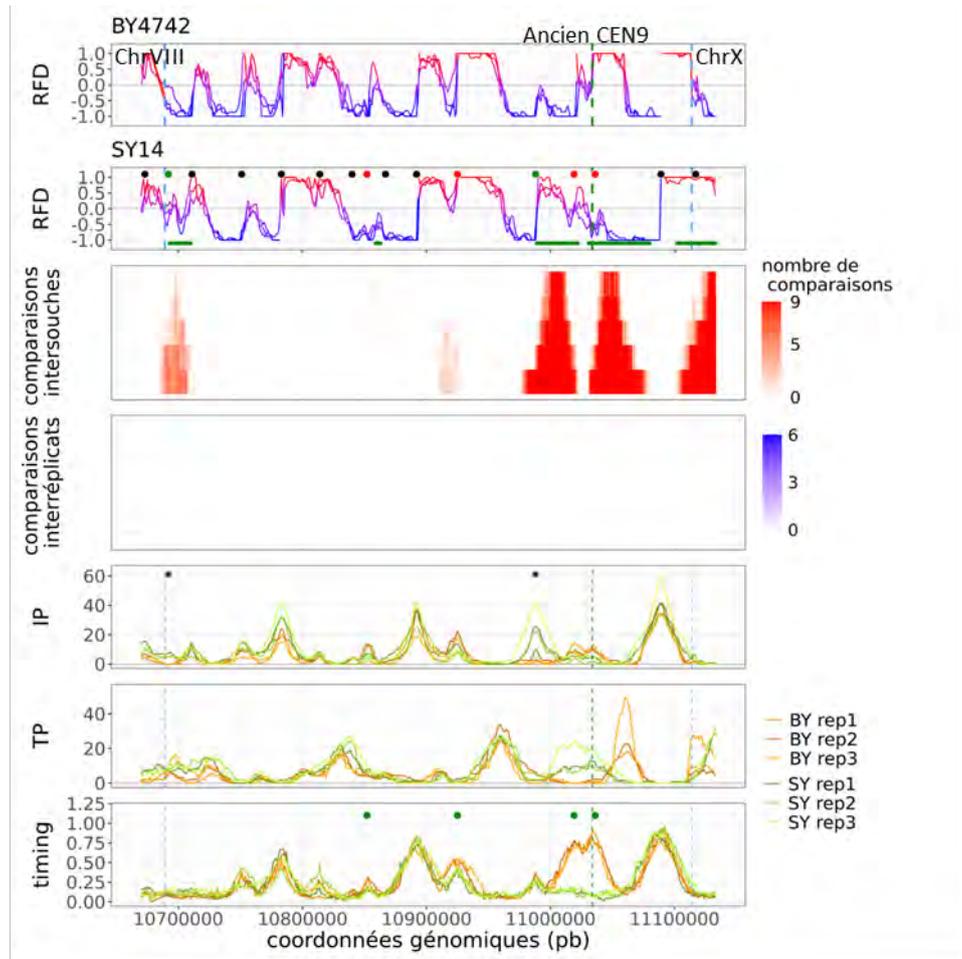
- a) Chromosome I
- b) Chromosome II
- c) Chromosome III
- d) Chromosome IV
- e) Chromosome V
- f) Chromosome VI
- g) Chromosome VII
- h) Chromosome VIII
- i) Chromosome IX
- j) Chromosome X
- k) Chromosome XI
- l) Chromosome XII
- m) Chromosome XIII
- n) Chromosome XIV
- o) Chromosome XV
- p) Chromosome XVI

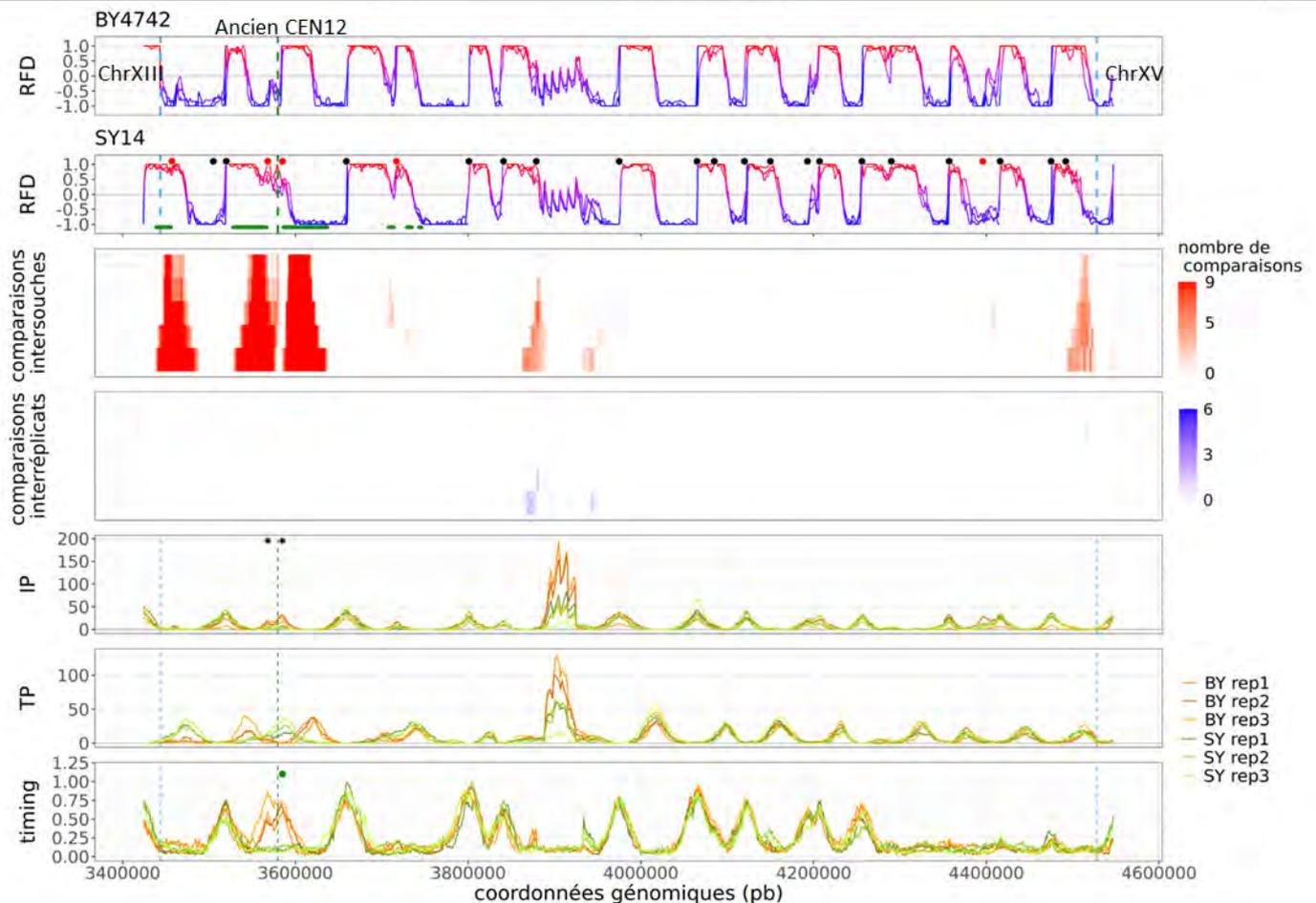
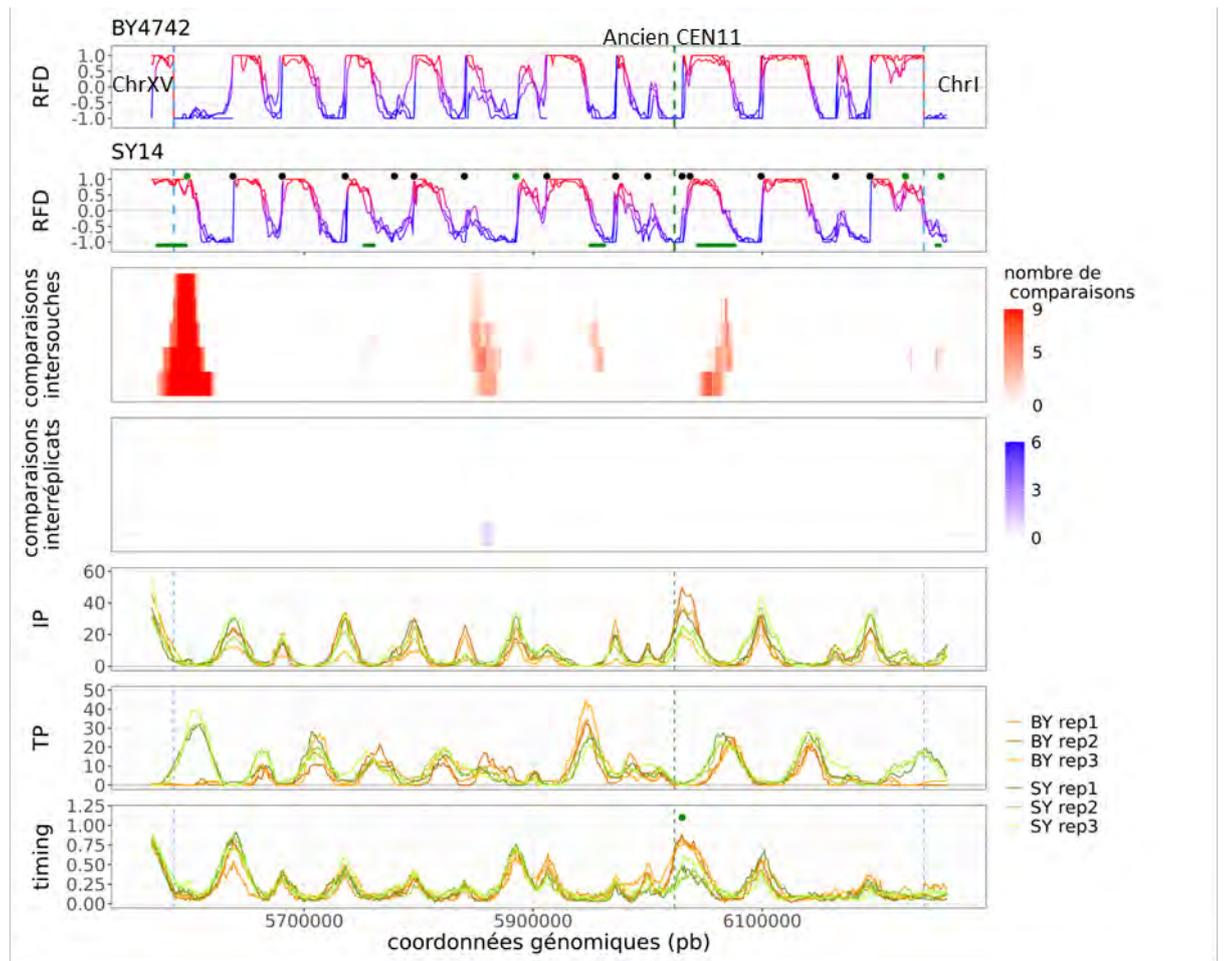


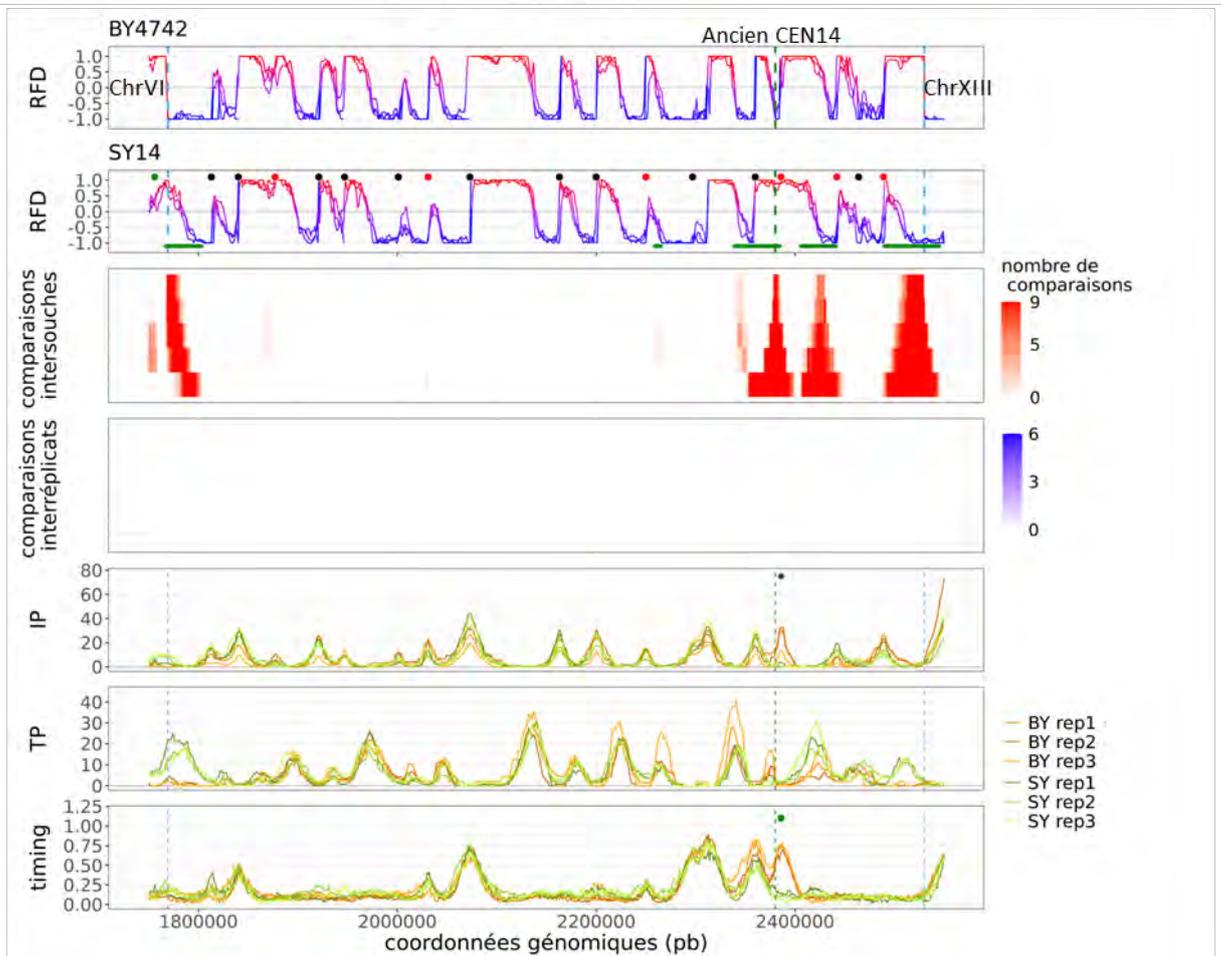
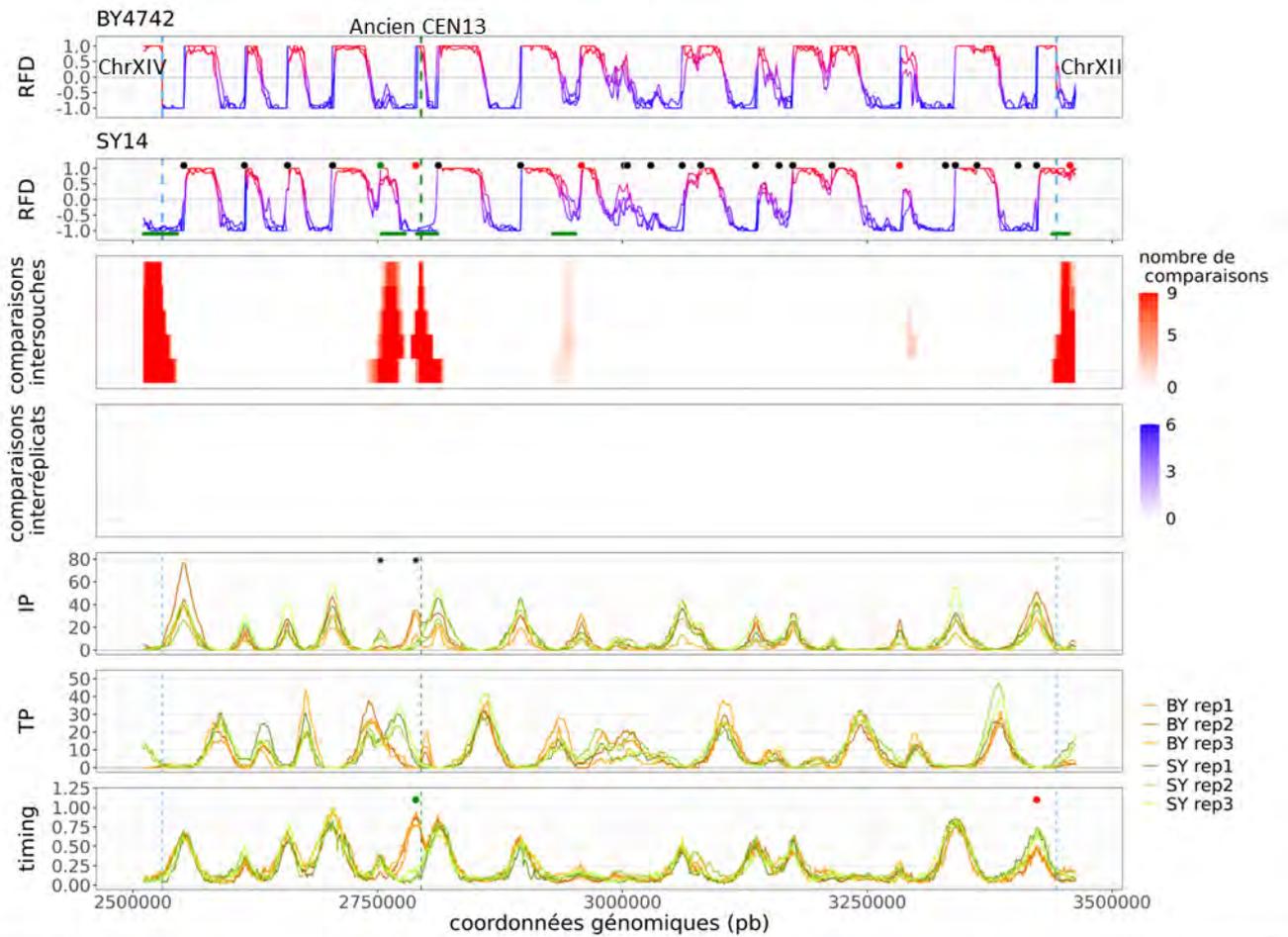


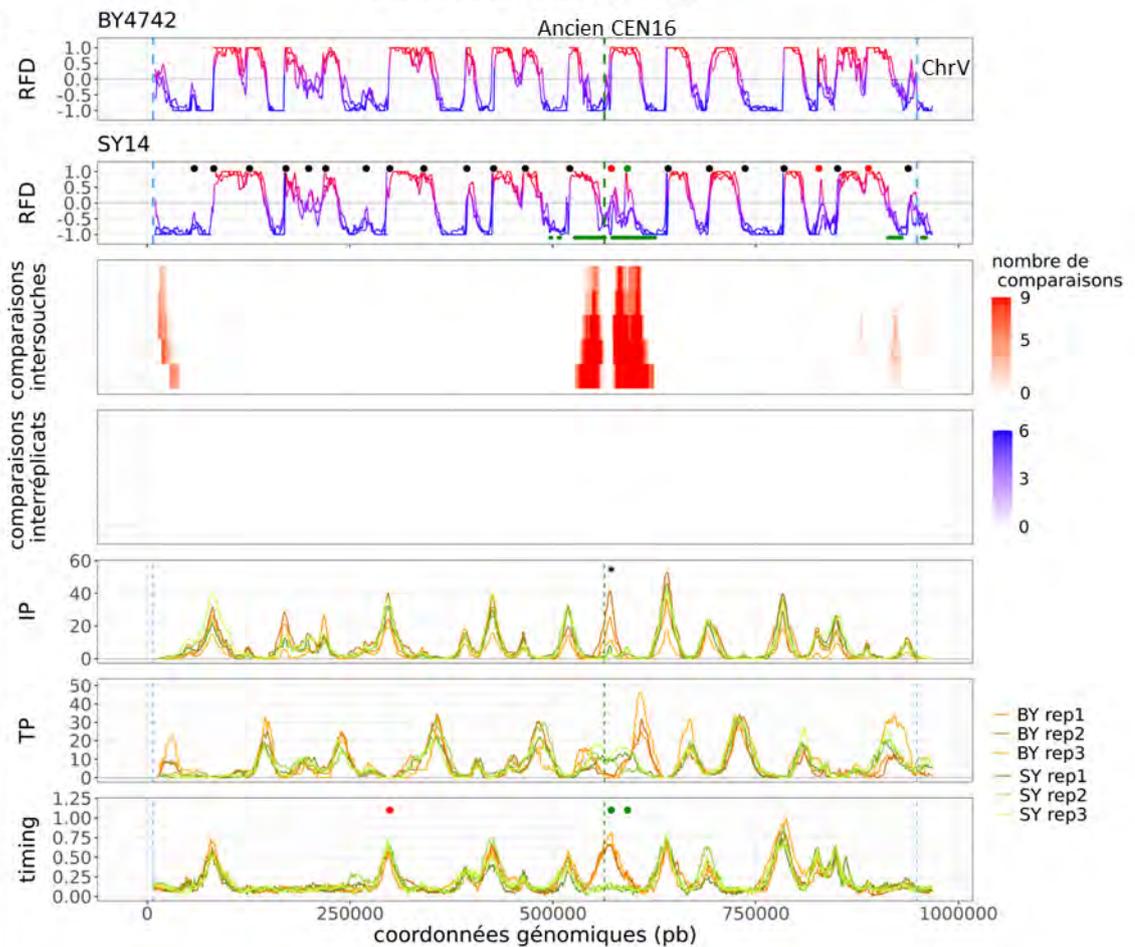
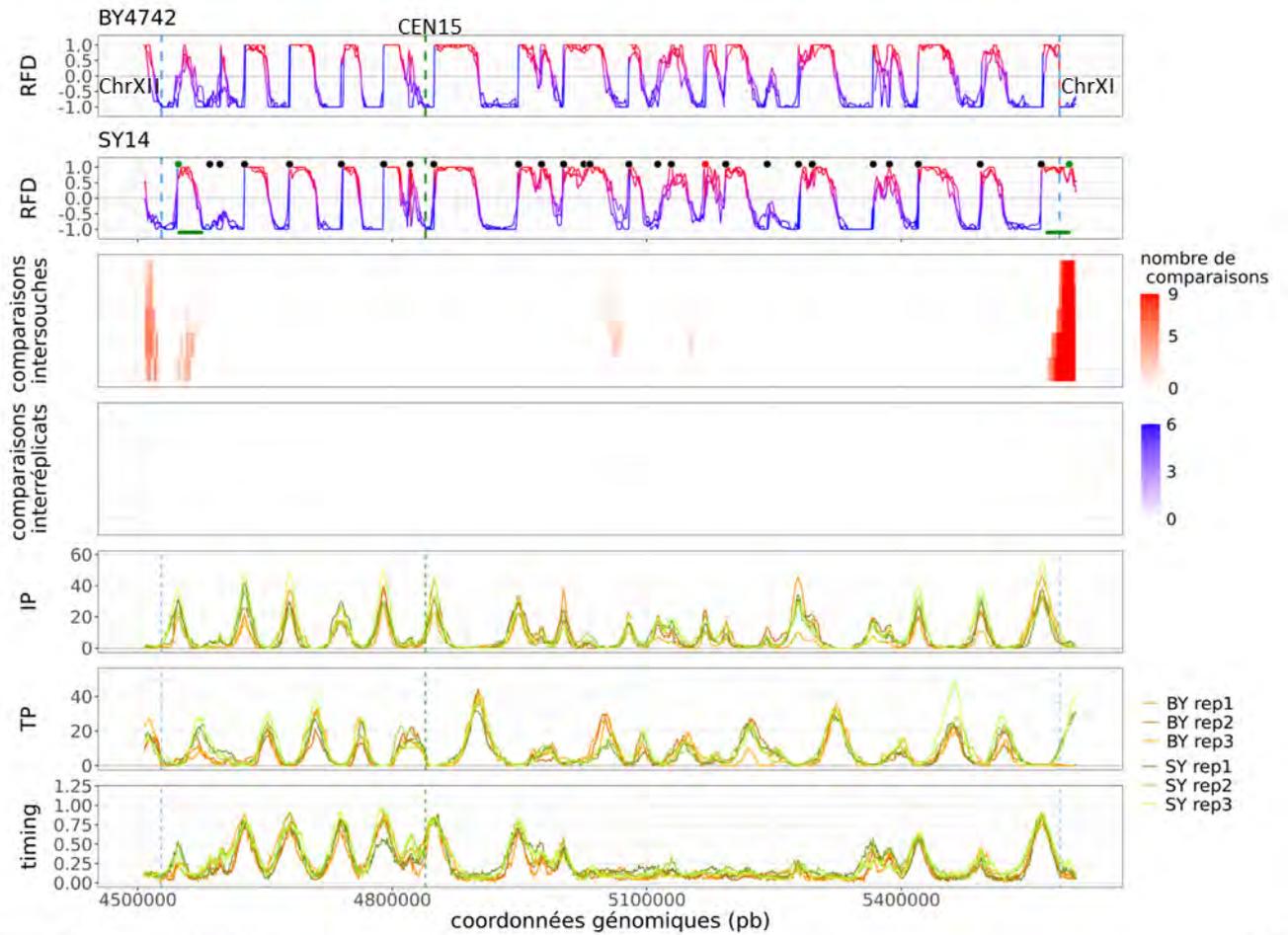








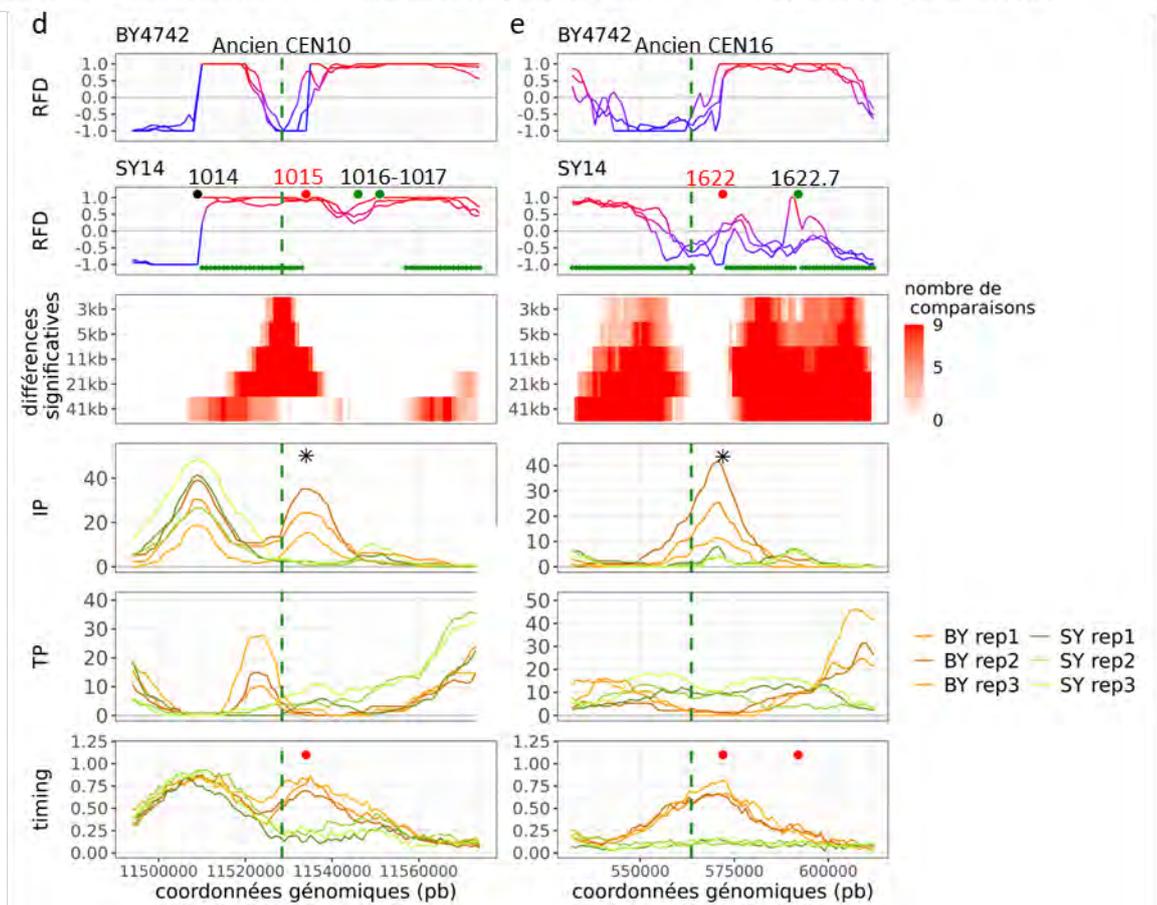
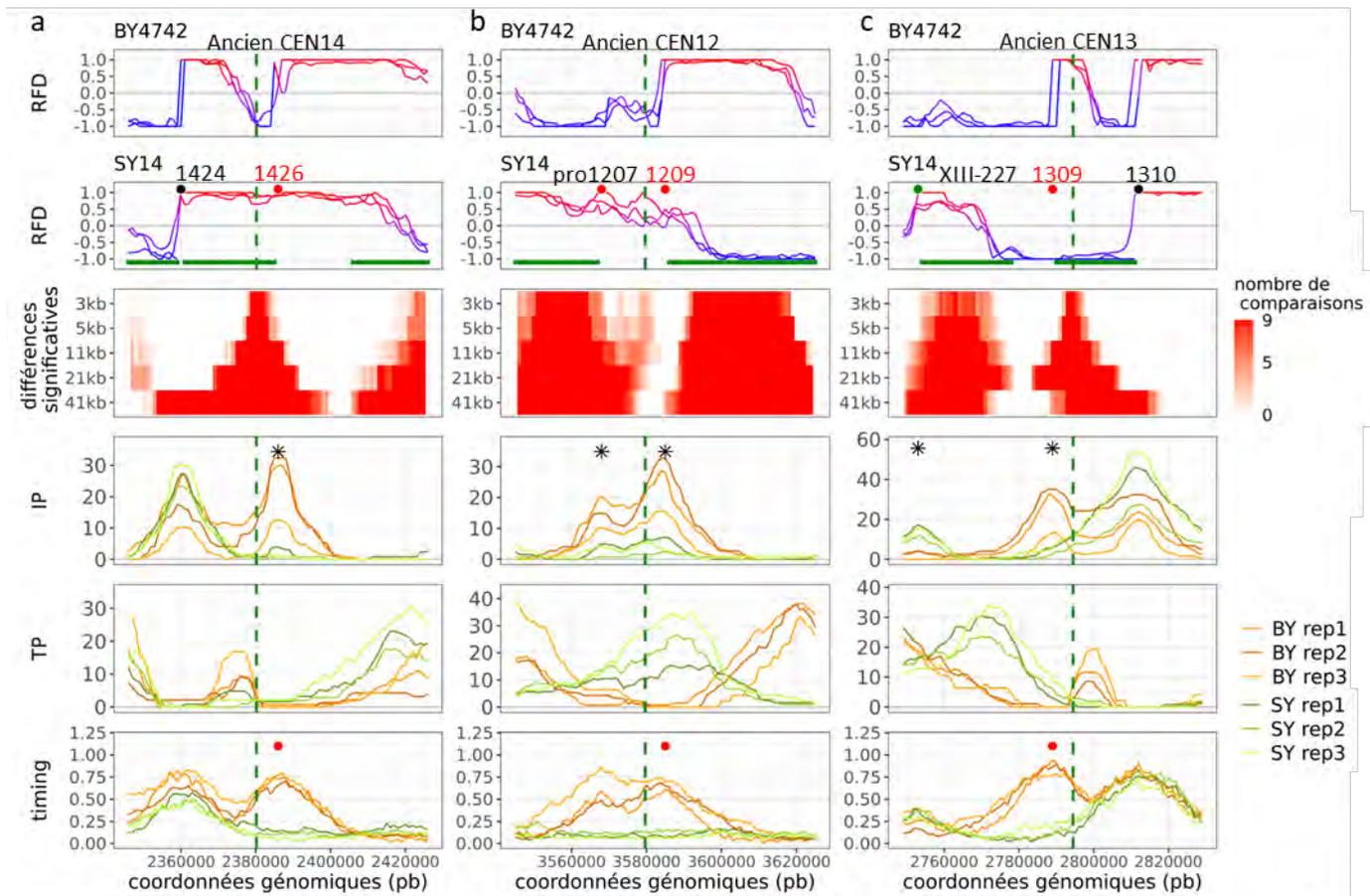




**Figure supplémentaire VIII-2 : Origines détectées comme significativement modifiées par le RFD, l'IP et le *timing*.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

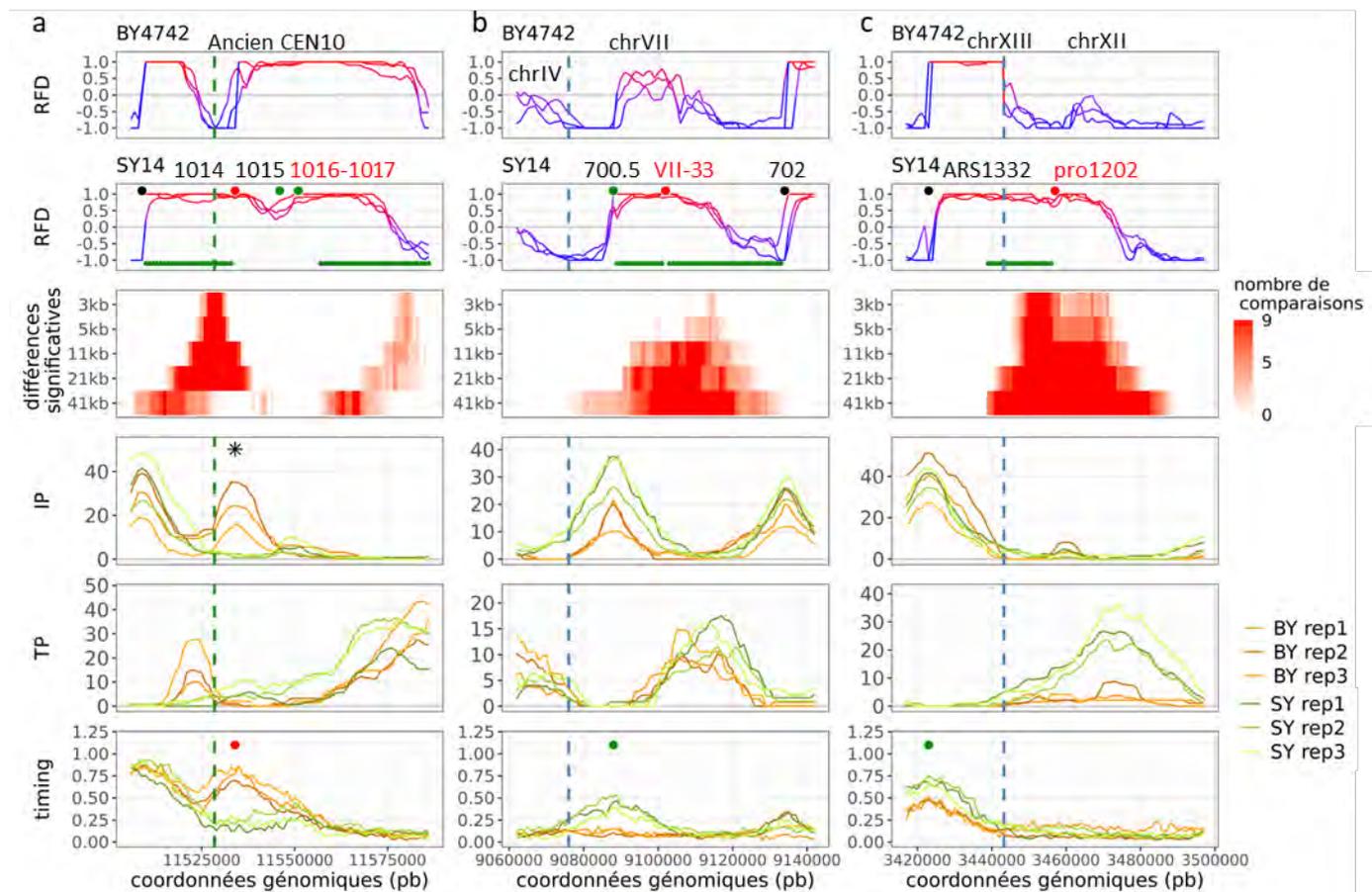
- a) ARS1426
- b) ARS1209
- c) ARS1309
- d) ARS1015
- e) ARS1622



**Figure supplémentaire VIII-3 : Origines détectées comme significativement modifiées uniquement sur le RFD dont l'origine voisine possède un *timing* de réplication modifié significativement.**

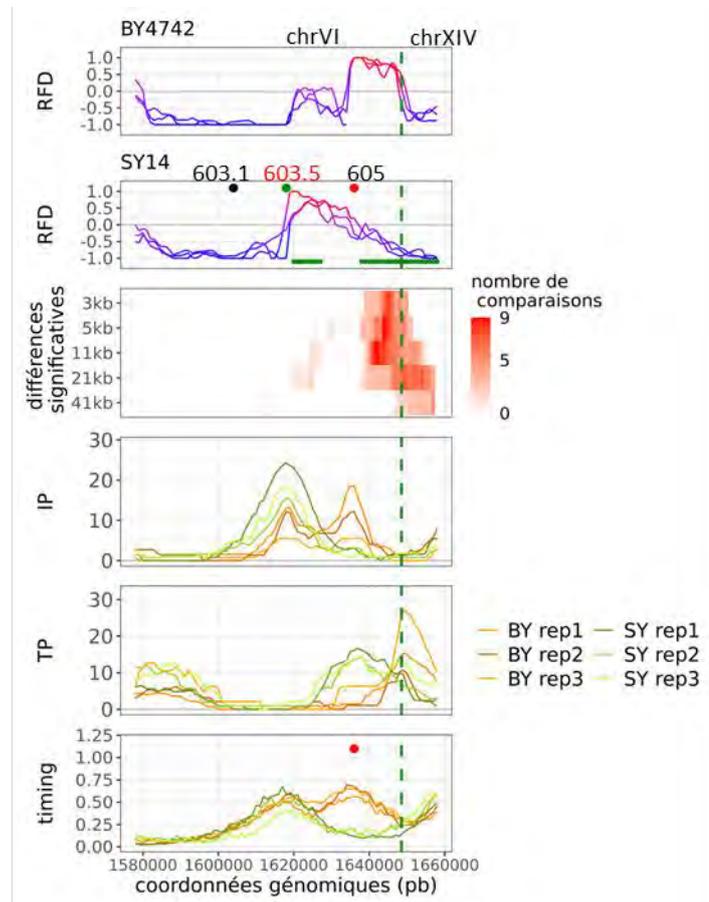
Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

- a) ARS1016 et ARS1017
- b) ARSVII-33
- c) proARS1202



**Figure supplémentaire VIII-4 : Origine détectée comme significativement modifiée uniquement sur le RFD, avec une modification visible de l'IP dont l'origine voisine possède un *timing* de réplication modifié significativement.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

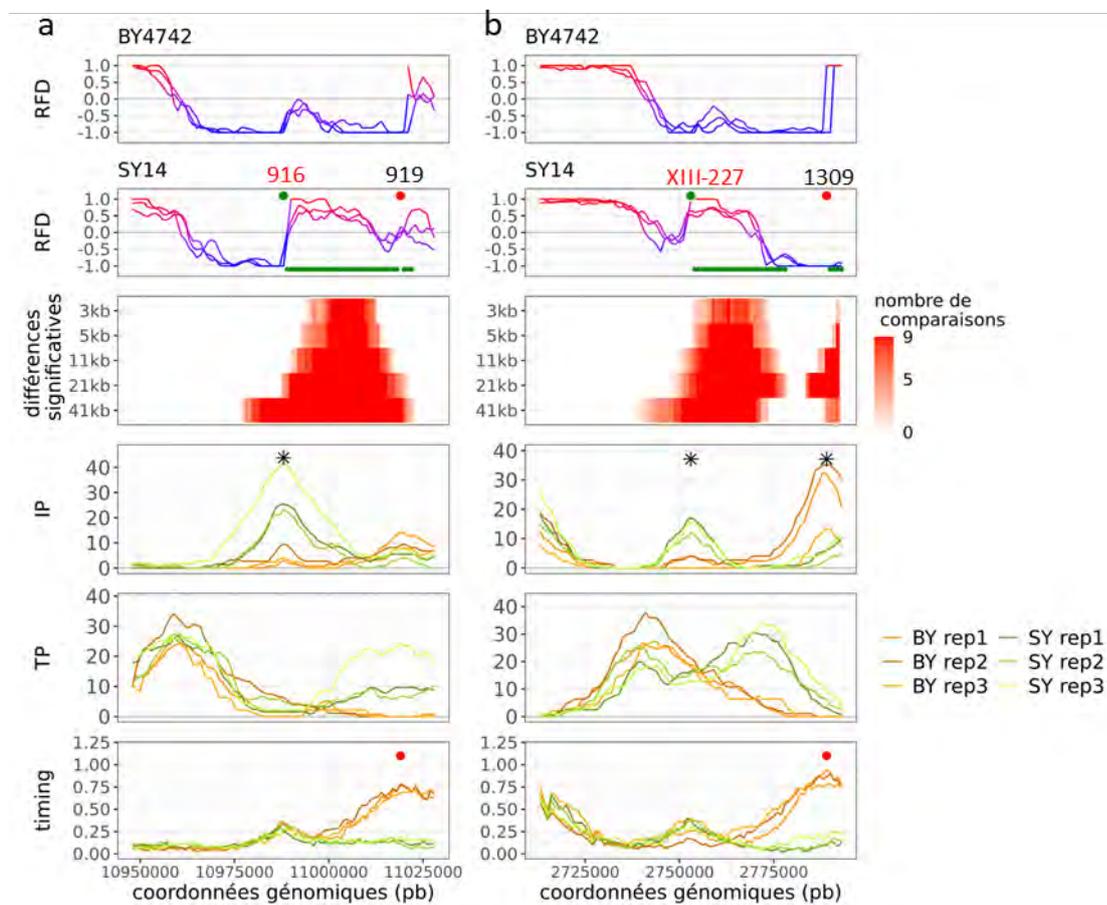


**Figure supplémentaire VIII-5 : Origines détectées comme significativement modifiées par l'IP et le RFD possédant une origine voisine modifiée significativement en *timing* de réplication.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge.

a) ARS916

b) ARSXIII-227

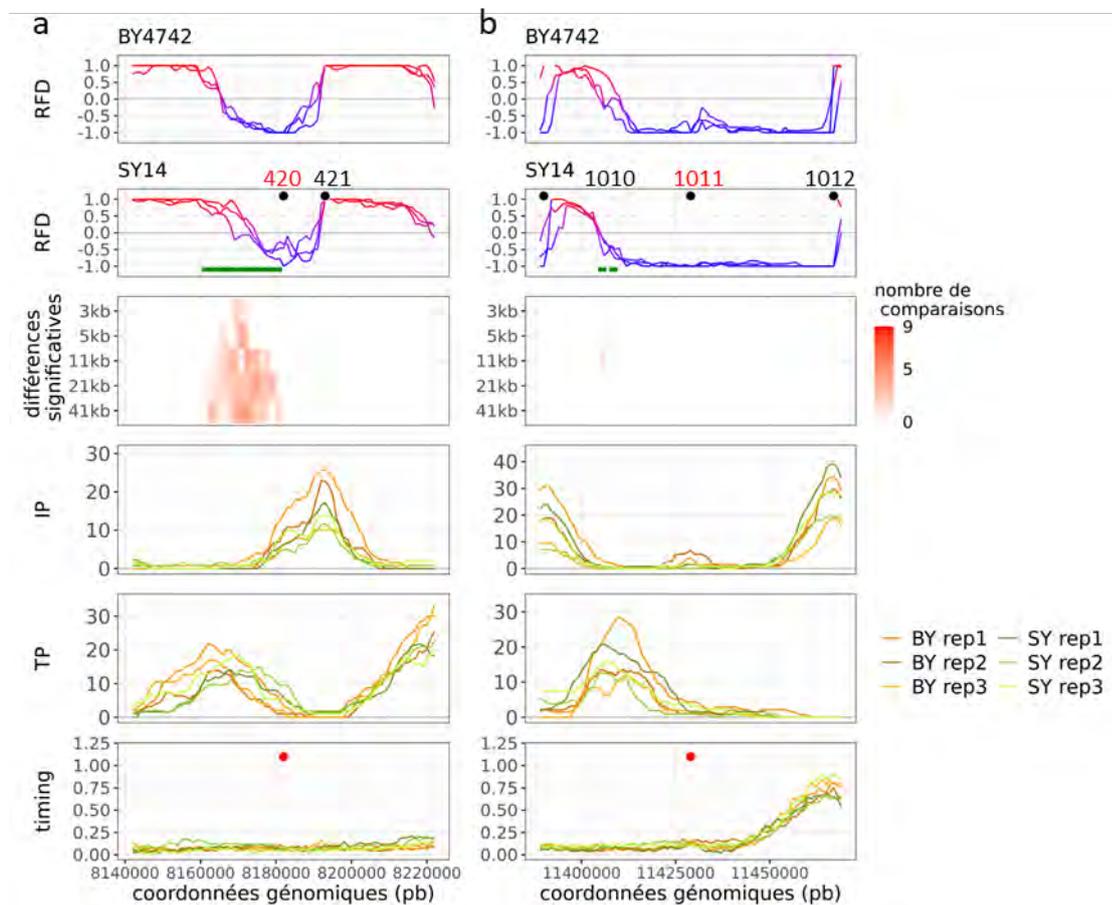


**Figure supplémentaire VIII-6 : Origines détectées comme significativement modifiées par le *timing* de réplication impactant une zone de terminaison sur le RFD.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge.

a) ARS420

b) ARS1011

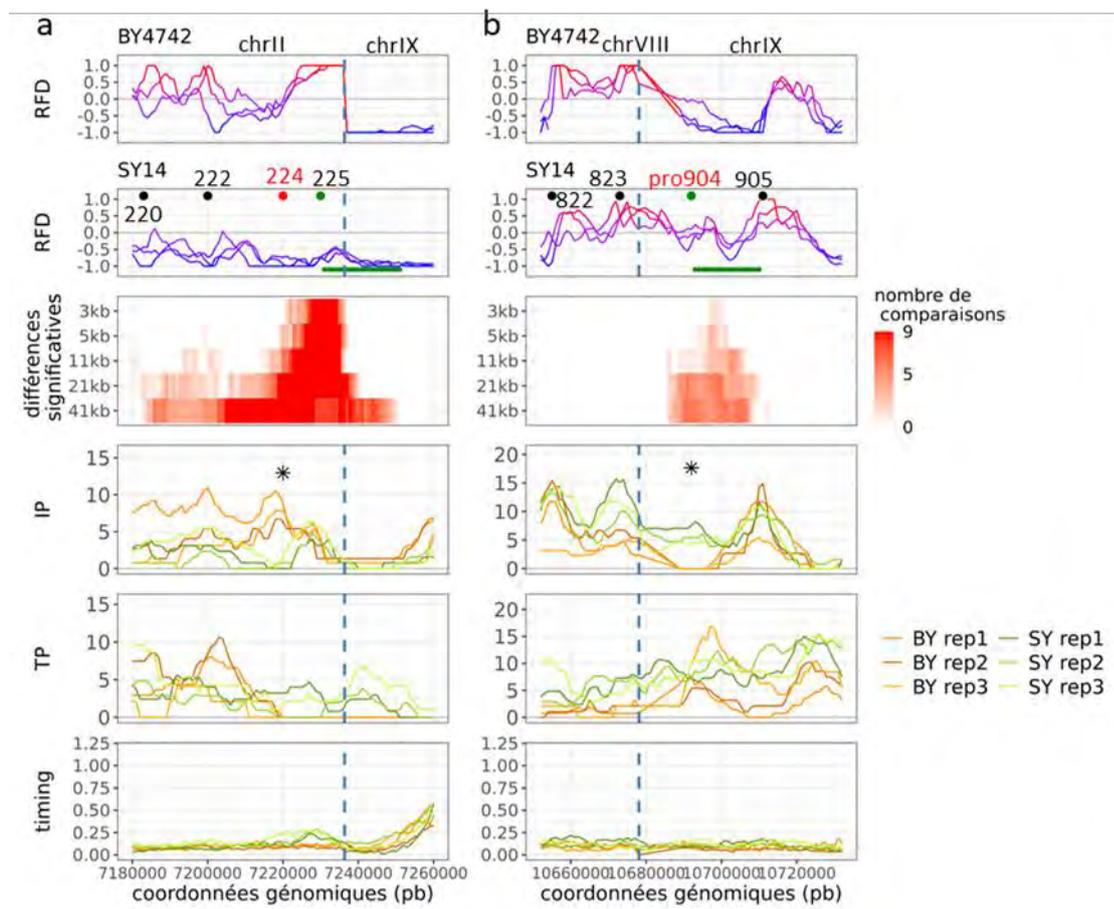


**Figure supplémentaire VIII-7 : Origines détectées comme significativement modifiées par l'IP et le RFD dans une région tardive.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des fusions chromosomiques est indiquée par des pointillés bleus.

a) ARS224

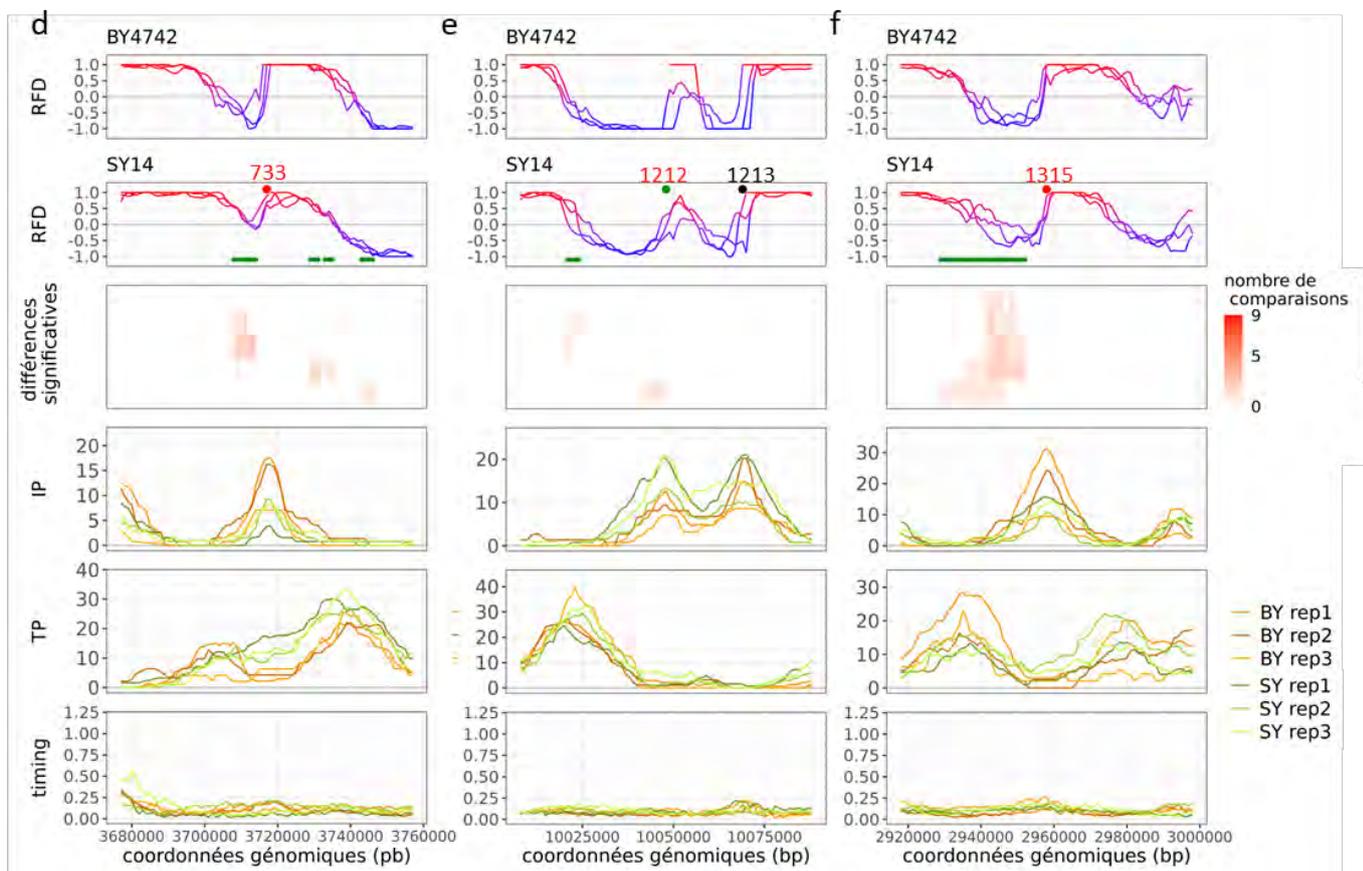
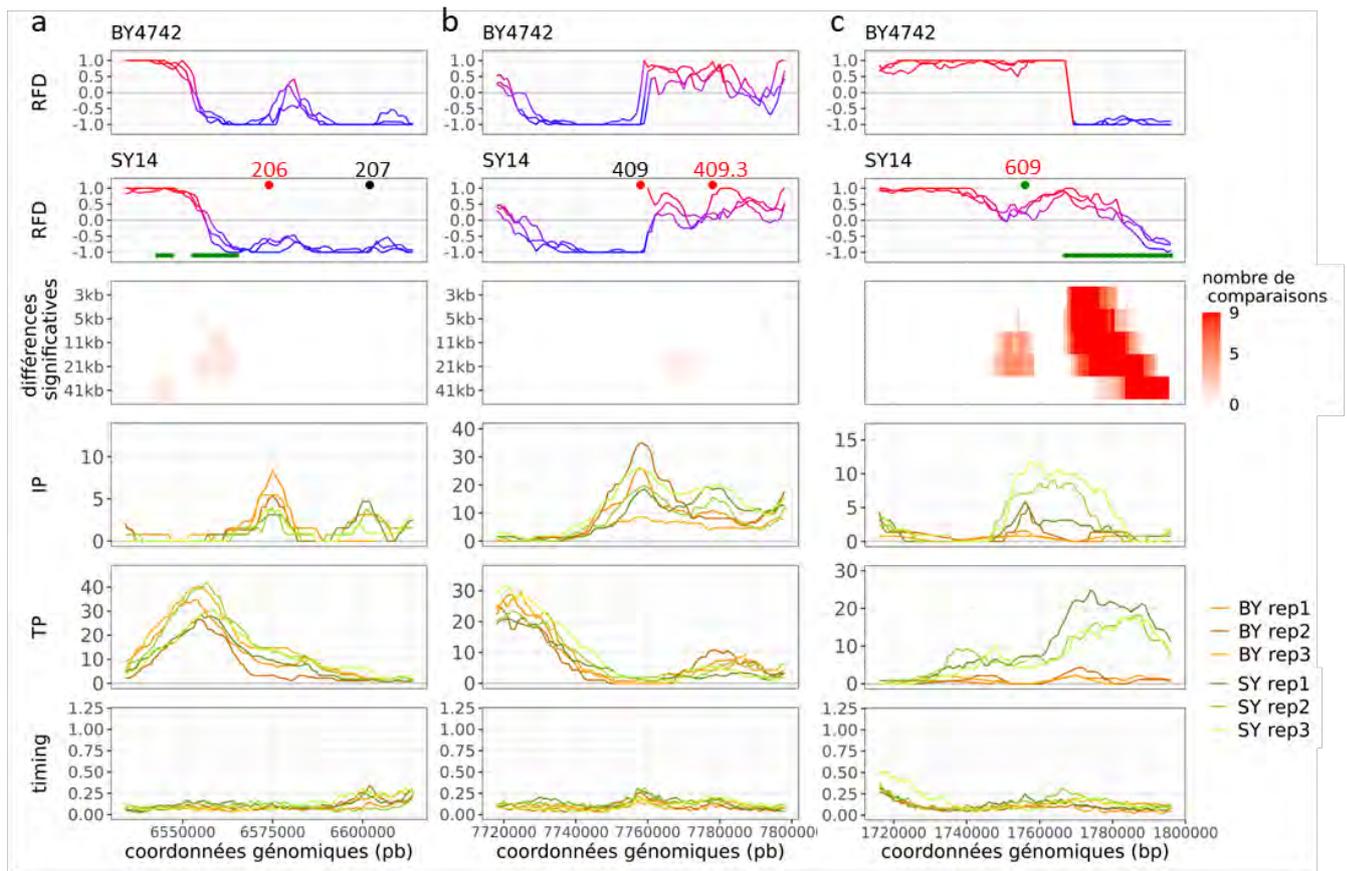
b) proARS904

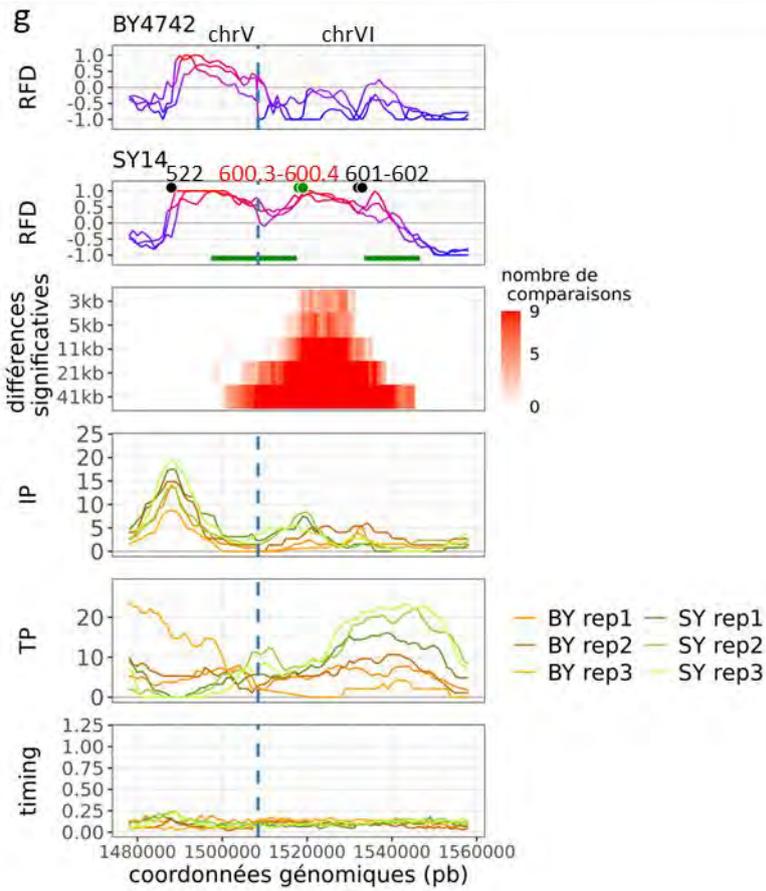


**Figure supplémentaire VIII-8 : Origines détectées comme significativement modifiées uniquement sur le RFD étant dans une région tardive et possédant une modification visible mais non significative de l'IP.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des fusions chromosomiques est indiquée par des pointillées bleus.

- a) ARS206
- b) ARS409.3
- c) ARS609
- d) ARS733
- e) ARS1212
- f) ARS1315
- g) 600.3-600.5

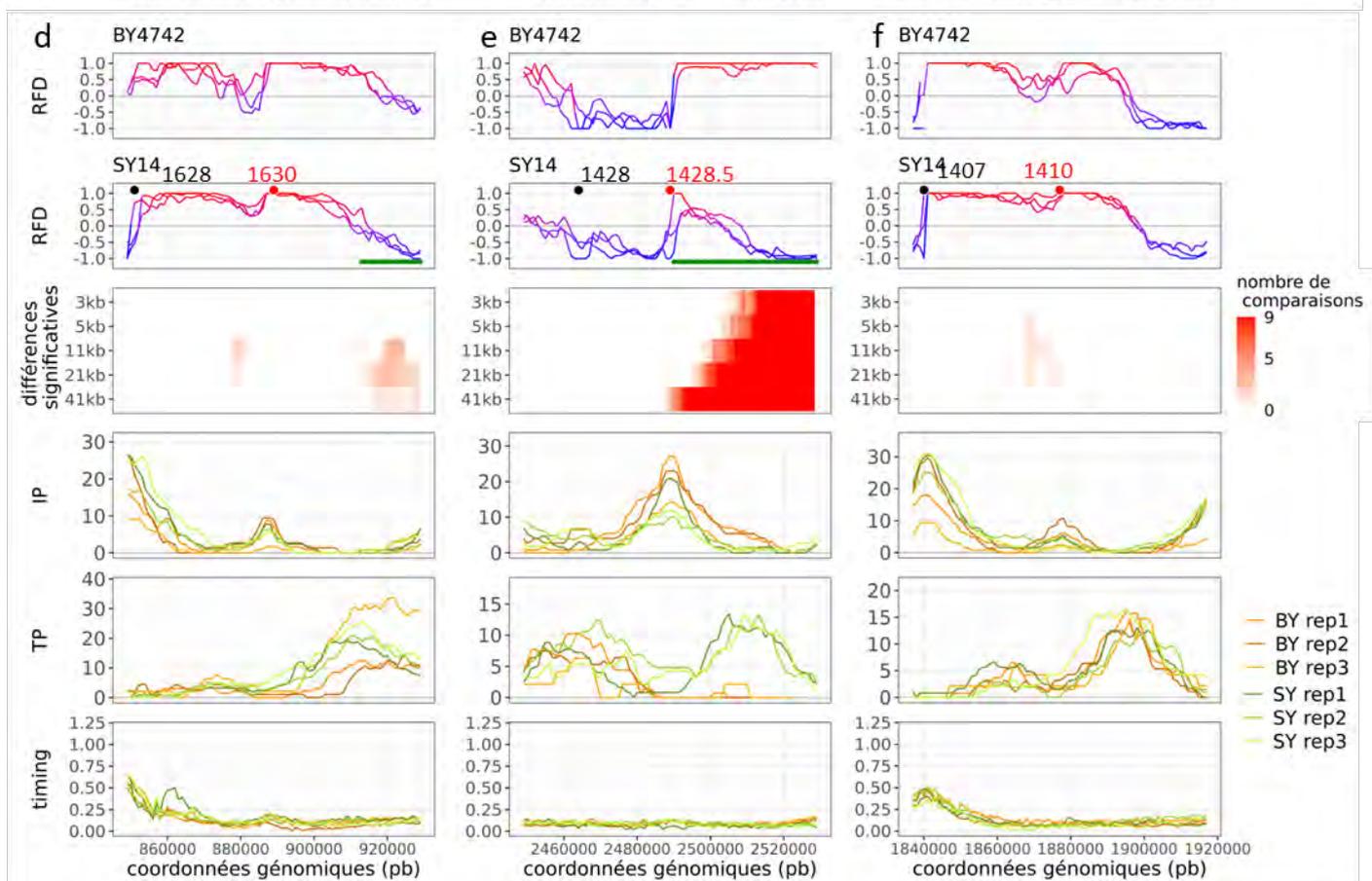
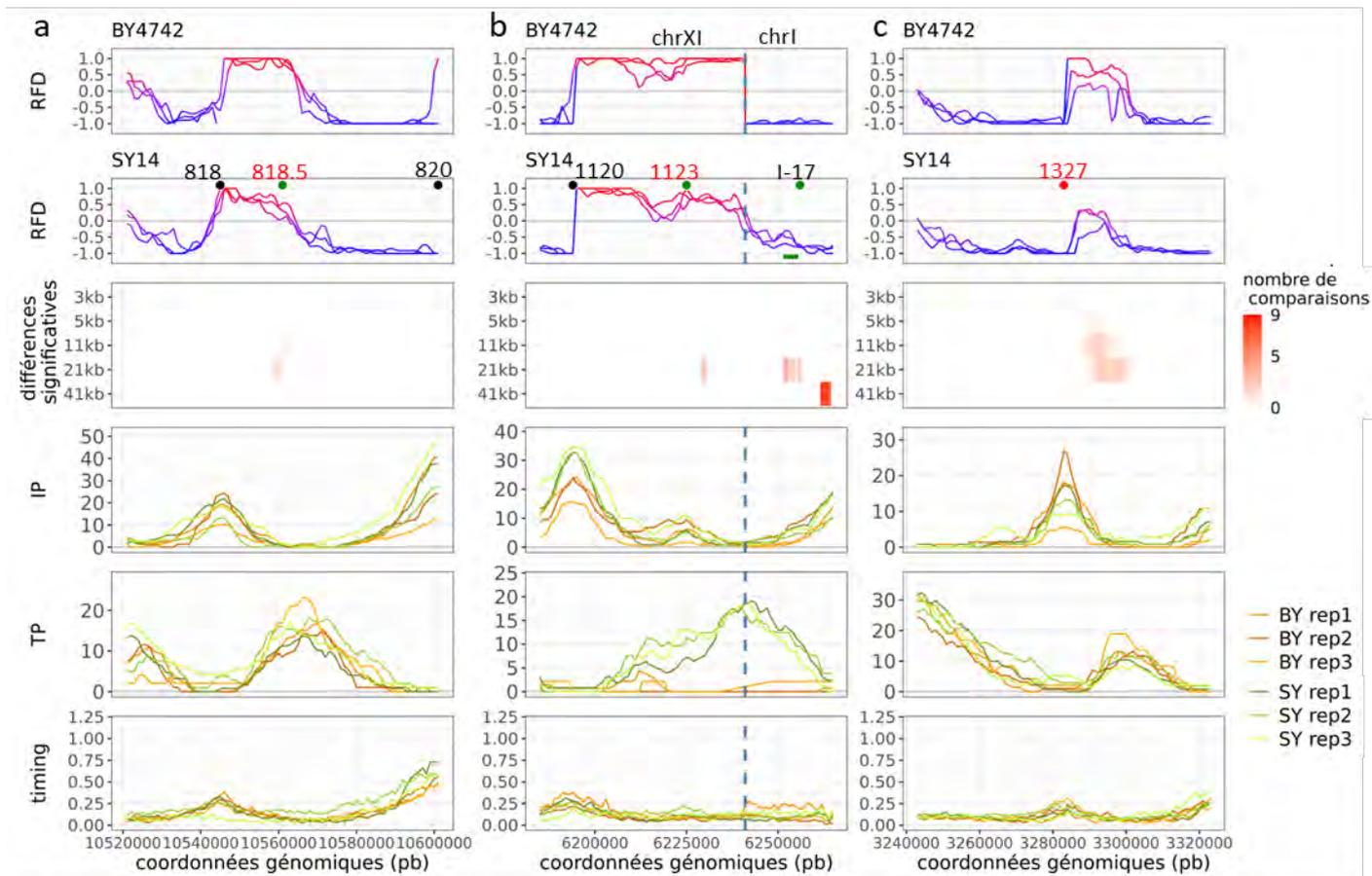


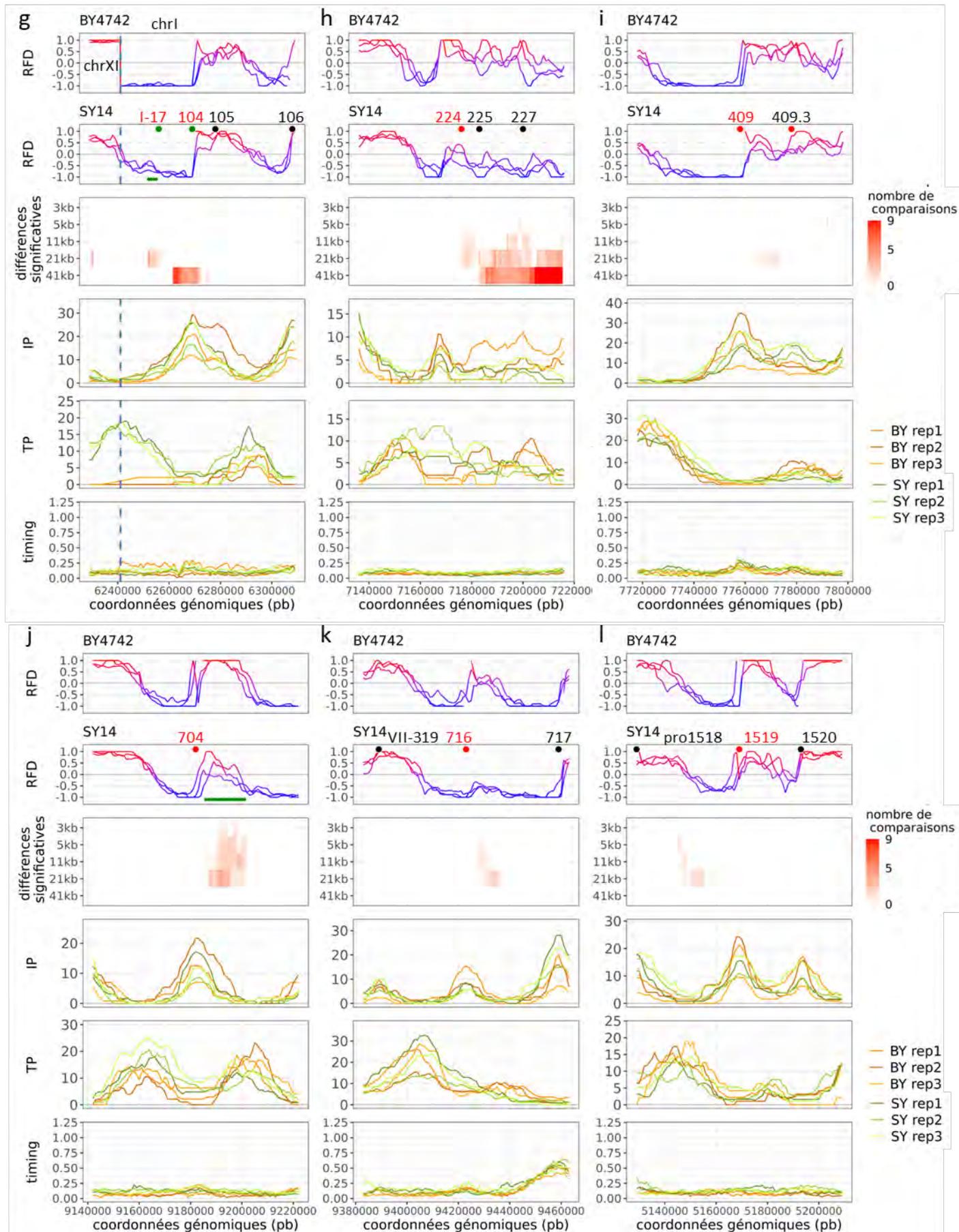


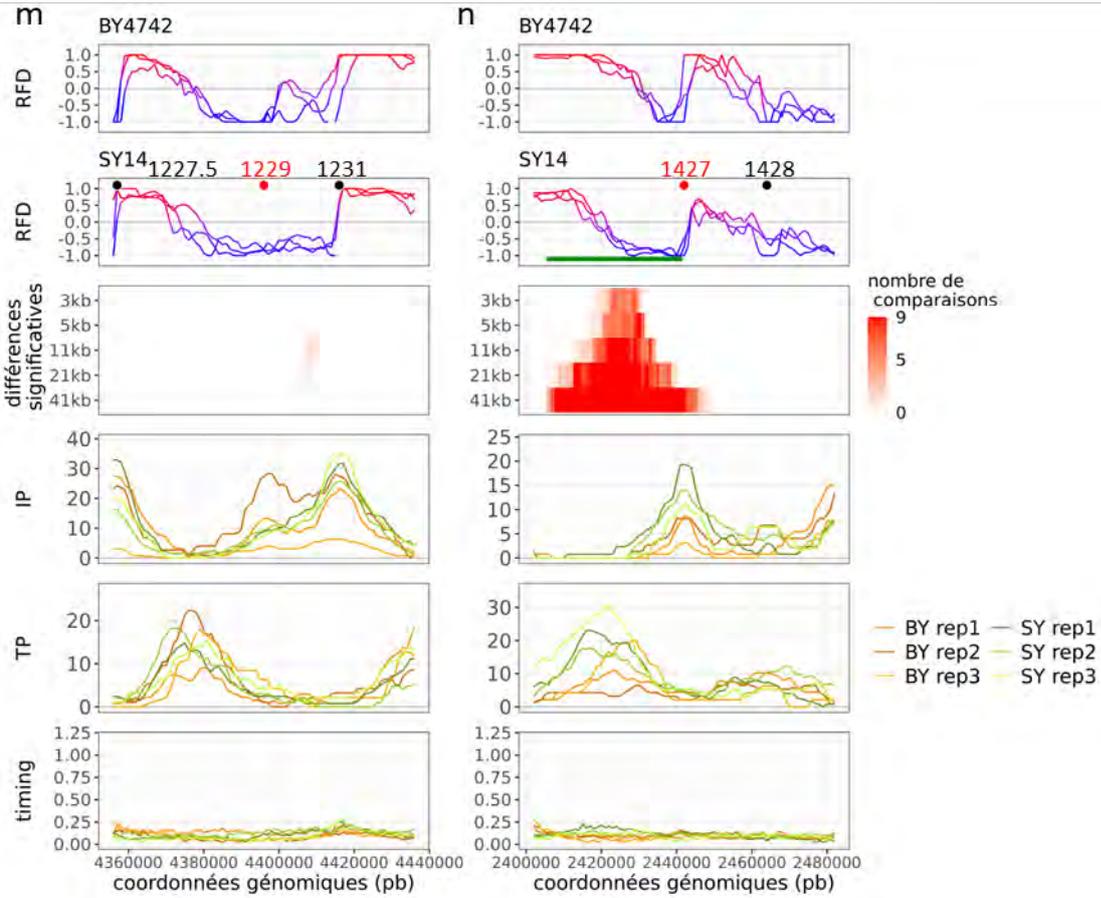
**Figure supplémentaire VIII-9 : Origines détectées comme significativement modifiées uniquement sur le RFD étant dans une région tardive et ne possédant pas de modification visible de l'IP.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des fusions chromosomiques est indiquée par des pointillées bleus.

- a) ARS818.5
- b) ARS1123
- c) ARS1327
- d) ARS1630
- e) ARS1428.5
- f) ARS1410
- g) ARS1-17 et ARS104
- h) ARS224
- i) ARS409
- j) ARS704
- k) ARS716
- l) ARS1519
- m) ARS1229
- n) ARS1427



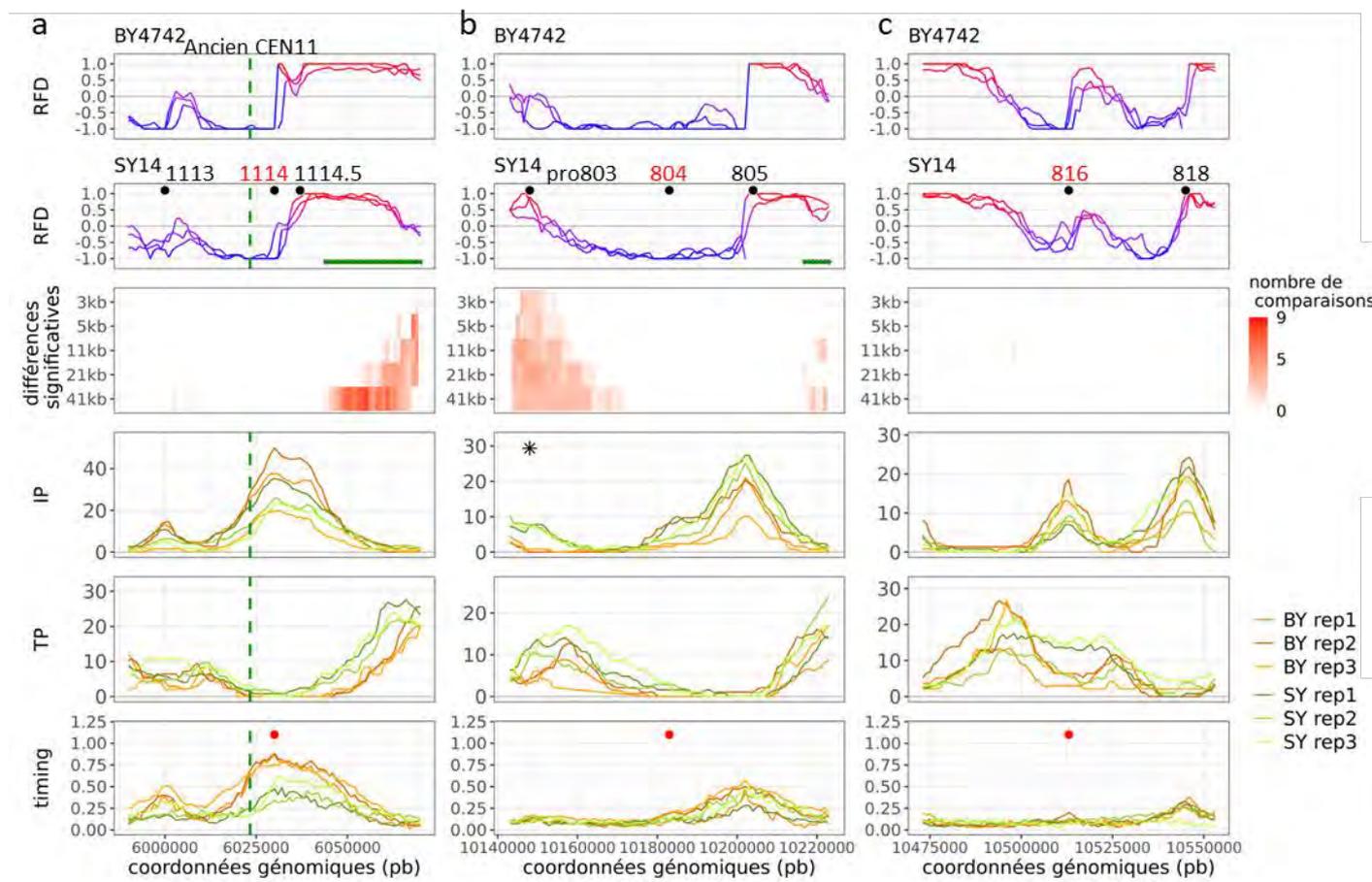




**Figure supplémentaire VIII-10 : Origines détectées comme significativement modifiées uniquement sur le *timing* de réplication mais possédant une modification visible sur l'IP et/ou RFD.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

- a) ARS114
- b) ARS804
- c) ARS816

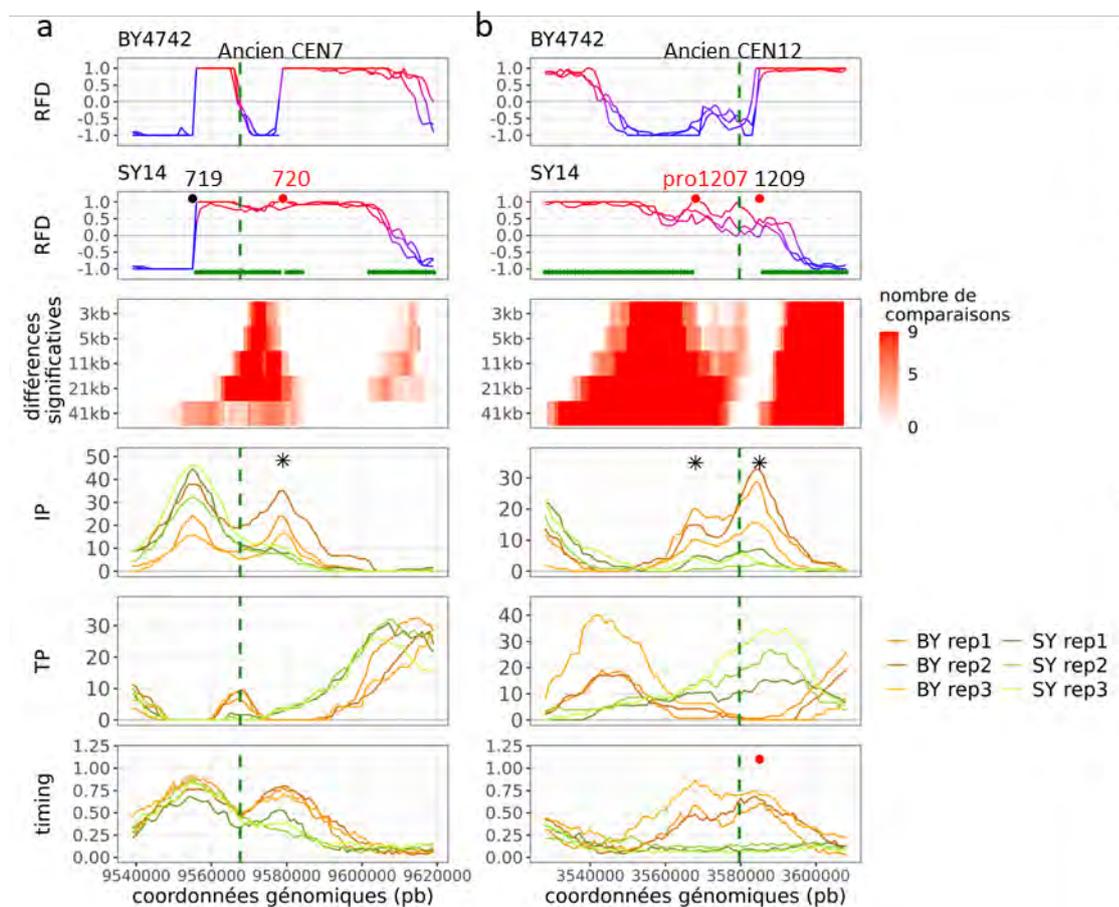


**Figure supplémentaire VIII-11 : Origines détectées comme significativement modifiées par l'IP et le RFD et possédant une modification visible en *timing* mais non significative.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

a) ARS720

b) proARS1207



**Figure supplémentaire VIII-12 : Origines détectées comme significativement modifiées par le *timing* et le RFD possédant une modification visible de l'IP mais non significative.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des fusions chromosomiques est indiquée par des pointillées bleus et celle des centromères par des pointillées verts.

a)ARS108

b)ARS307

c)ARS510

d)ARS700.5

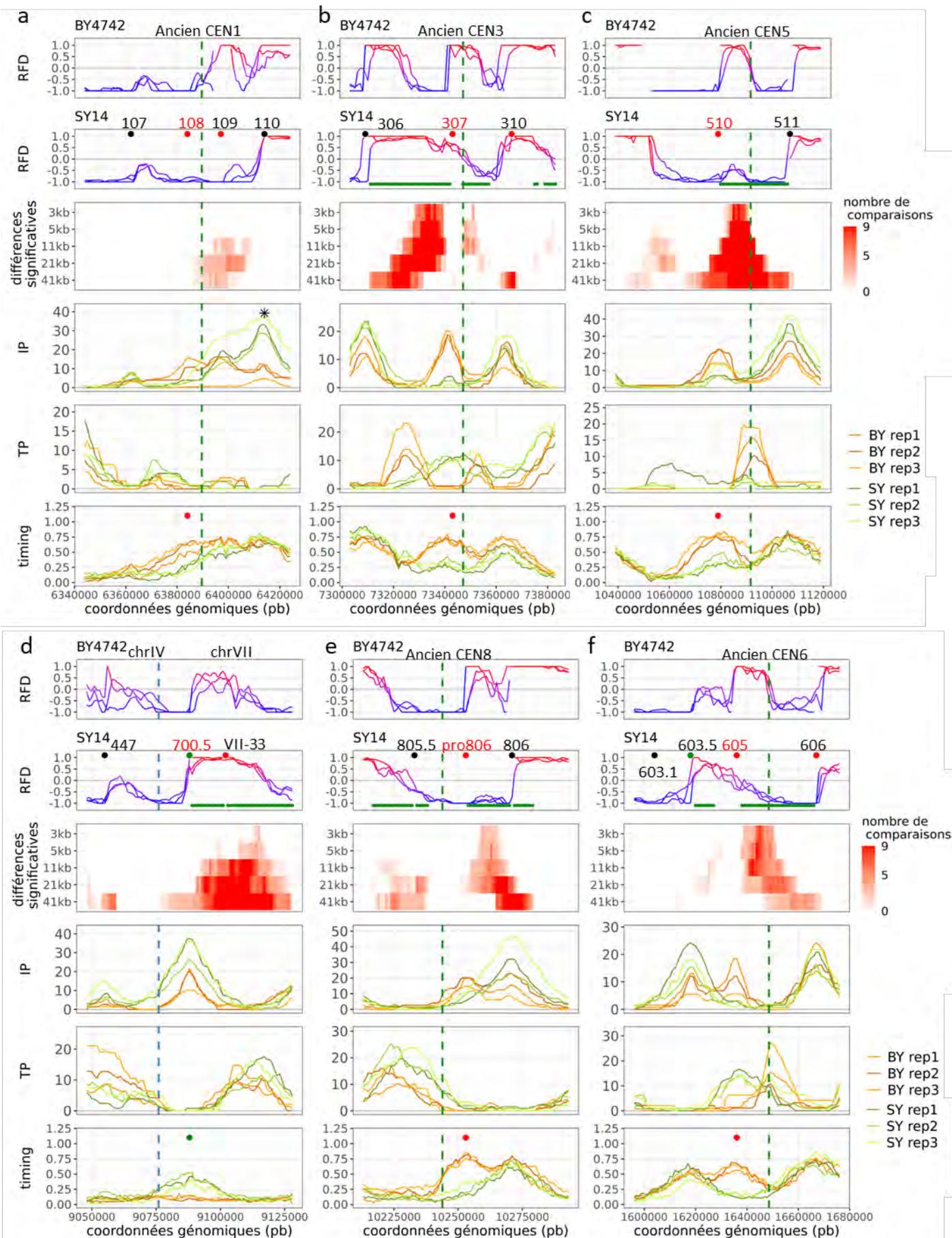
e)proARS806

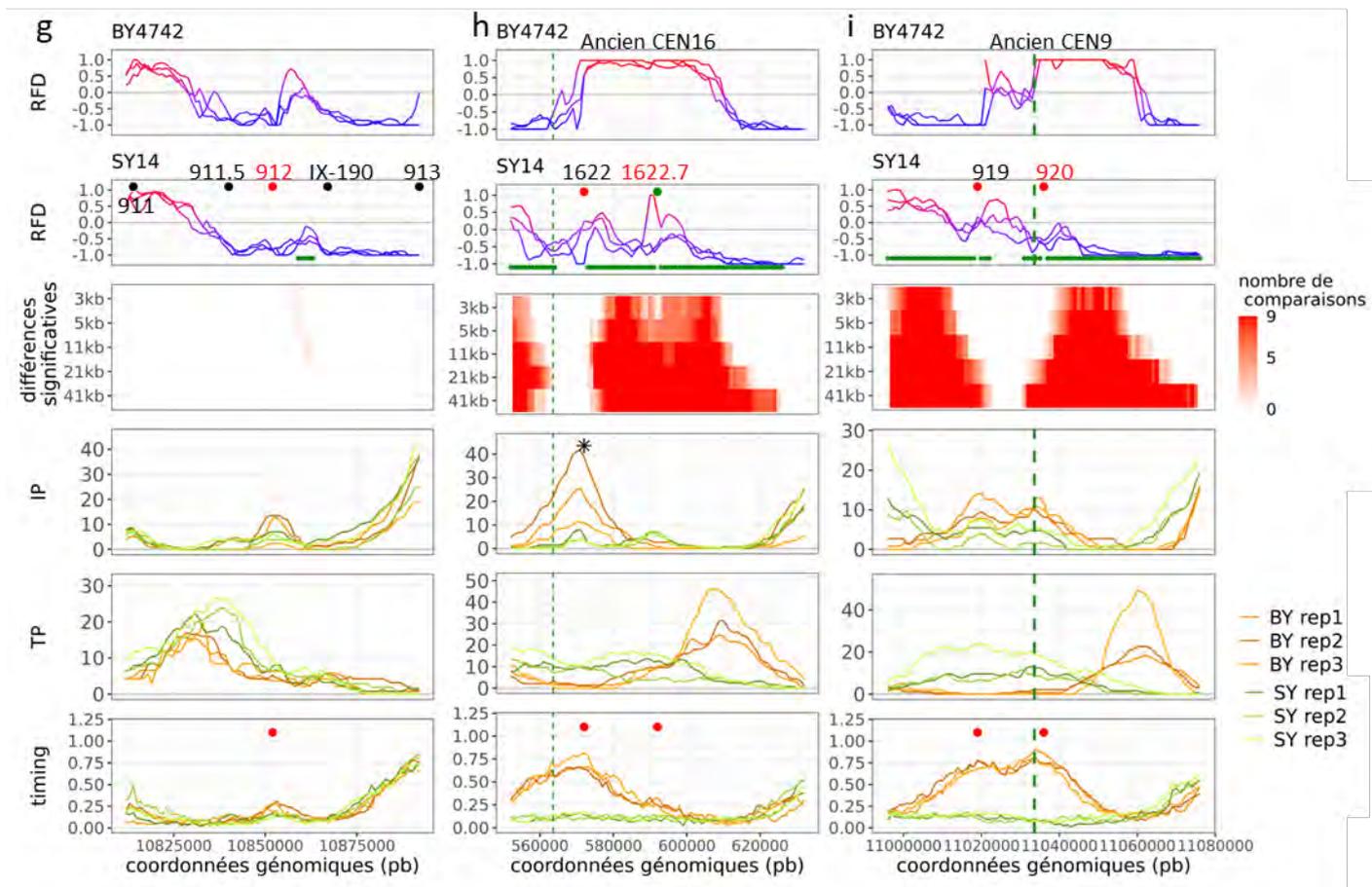
f)ARS605

g)ARS912

h)ARS1622.7

i)ARS920



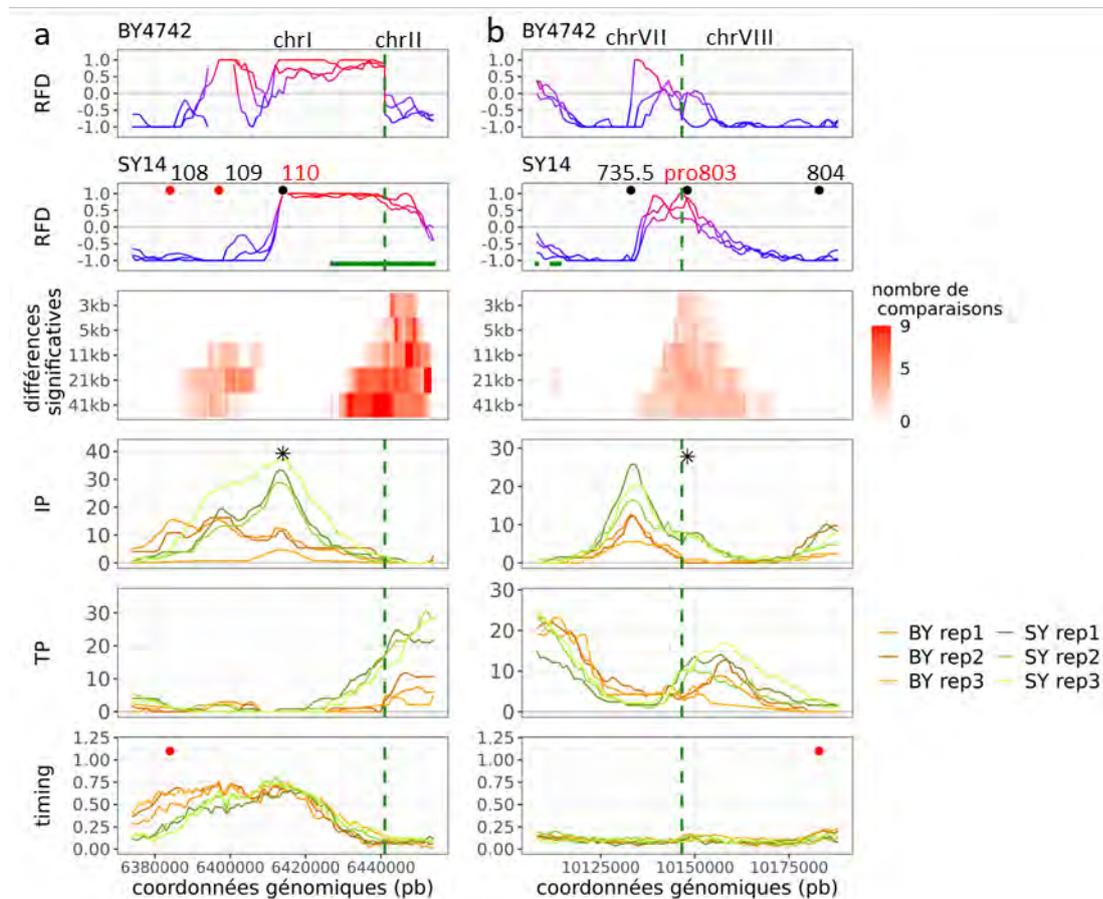


**Figure supplémentaire VIII-13 : Origines détectées comme significativement modifiées uniquement sur l'IP mais possédant une modification visible sur le RFD.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

a) ARS110

b) proARS803, la modification du RFD est significative mais cette région a été exclue car elle possédait une zone significative lors de la comparaison entre les réplicats

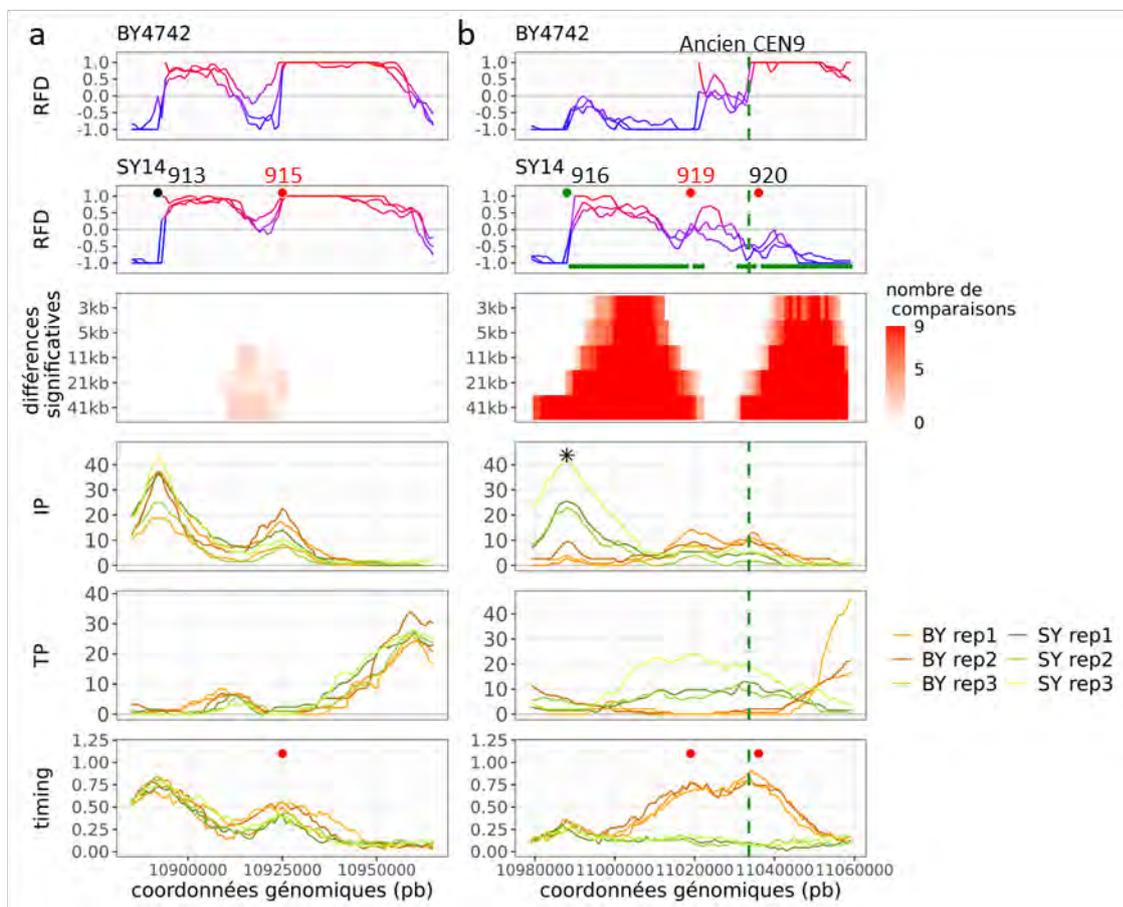


**Figure supplémentaire VIII-14 : Origines détectées comme significativement modifiées par le *timing* et le RFD ne possédant pas de modification visible de l'IP.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

a) ARS915

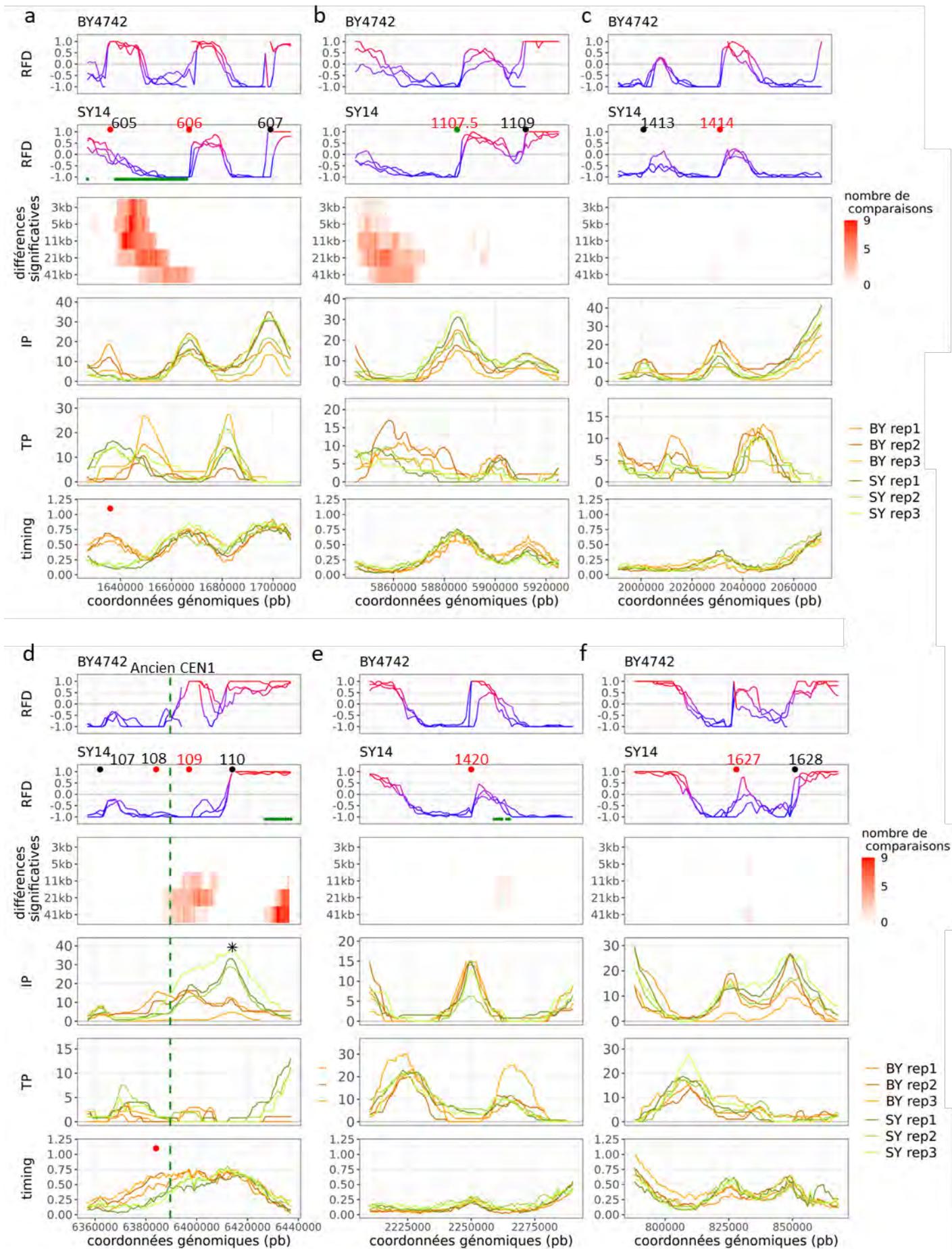
b) ARS919



**Figure supplémentaire VIII-15 : Origines détectées comme significativement modifiées uniquement sur le RFD.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

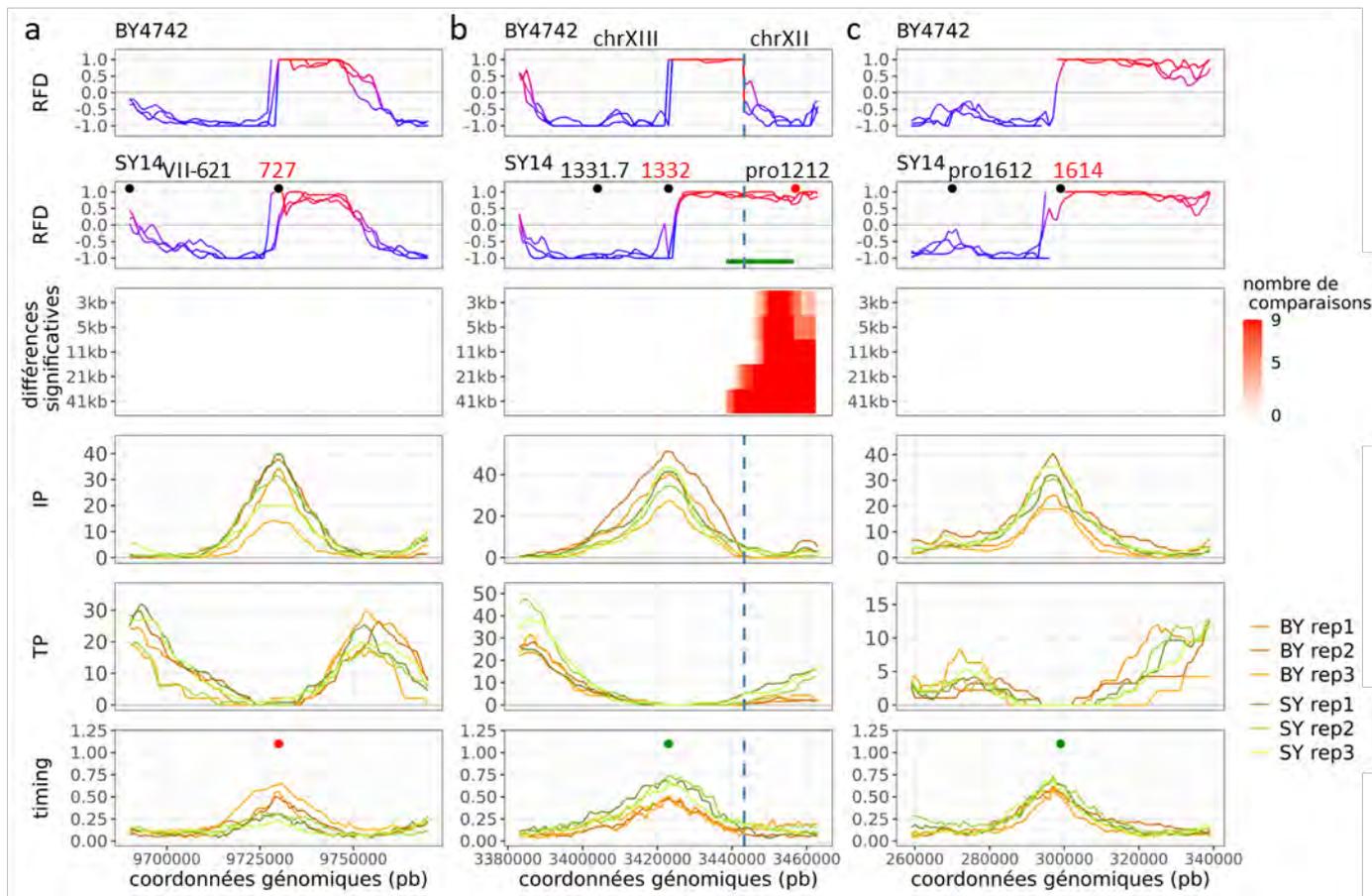
- a) ARS606
- b) ARS1107.5
- c) ARS1414
- d) ARS109
- e) ARS1420
- f) ARS1627



**Figure supplémentaire VIII-16 : Origines détectées comme significativement modifiées uniquement sur le *timing* de réplication.**

Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des fusions chromosomiques est indiquée par des pointillées bleus.

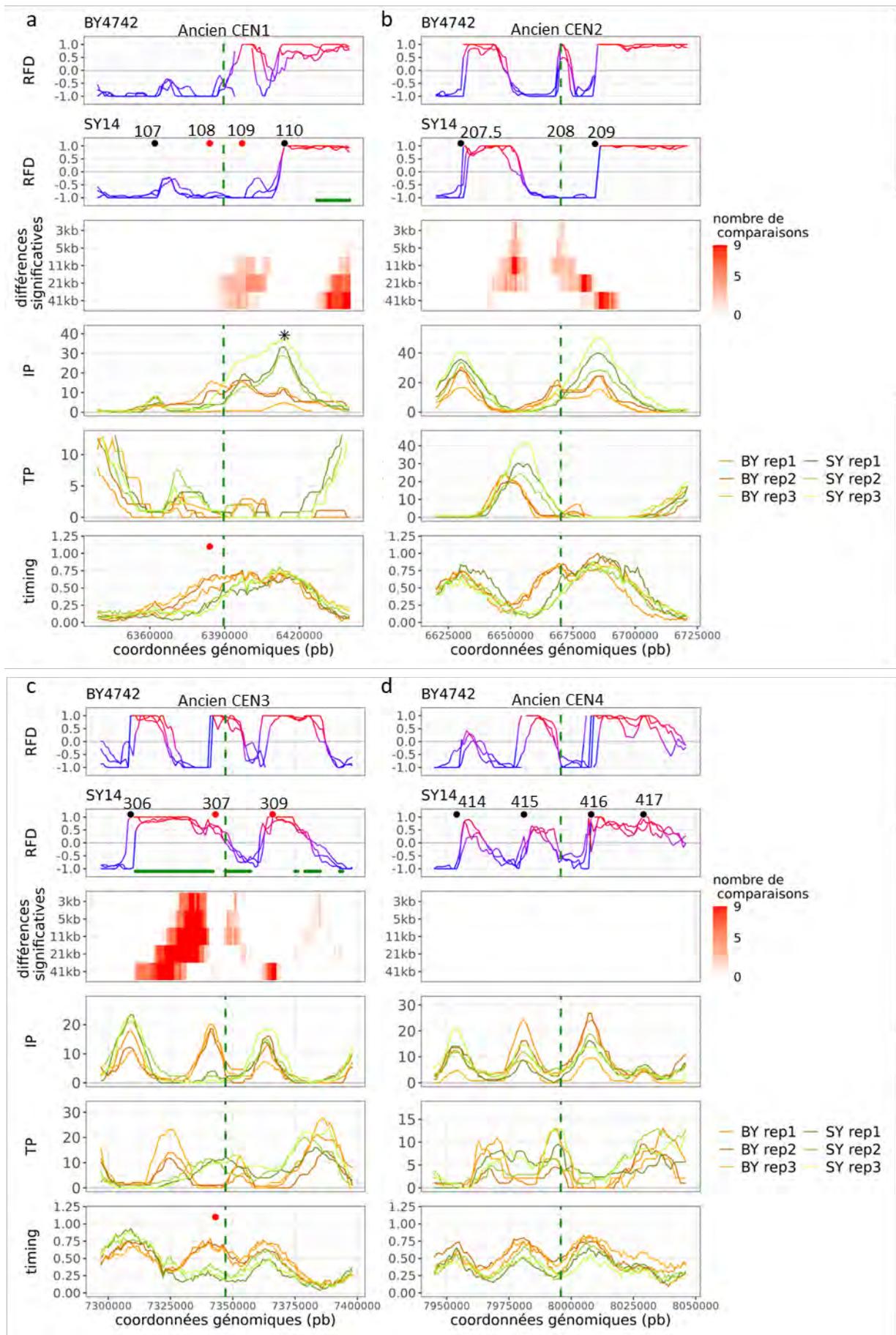
- a) ARS727
- b) ARS1323
- c) ARS1614

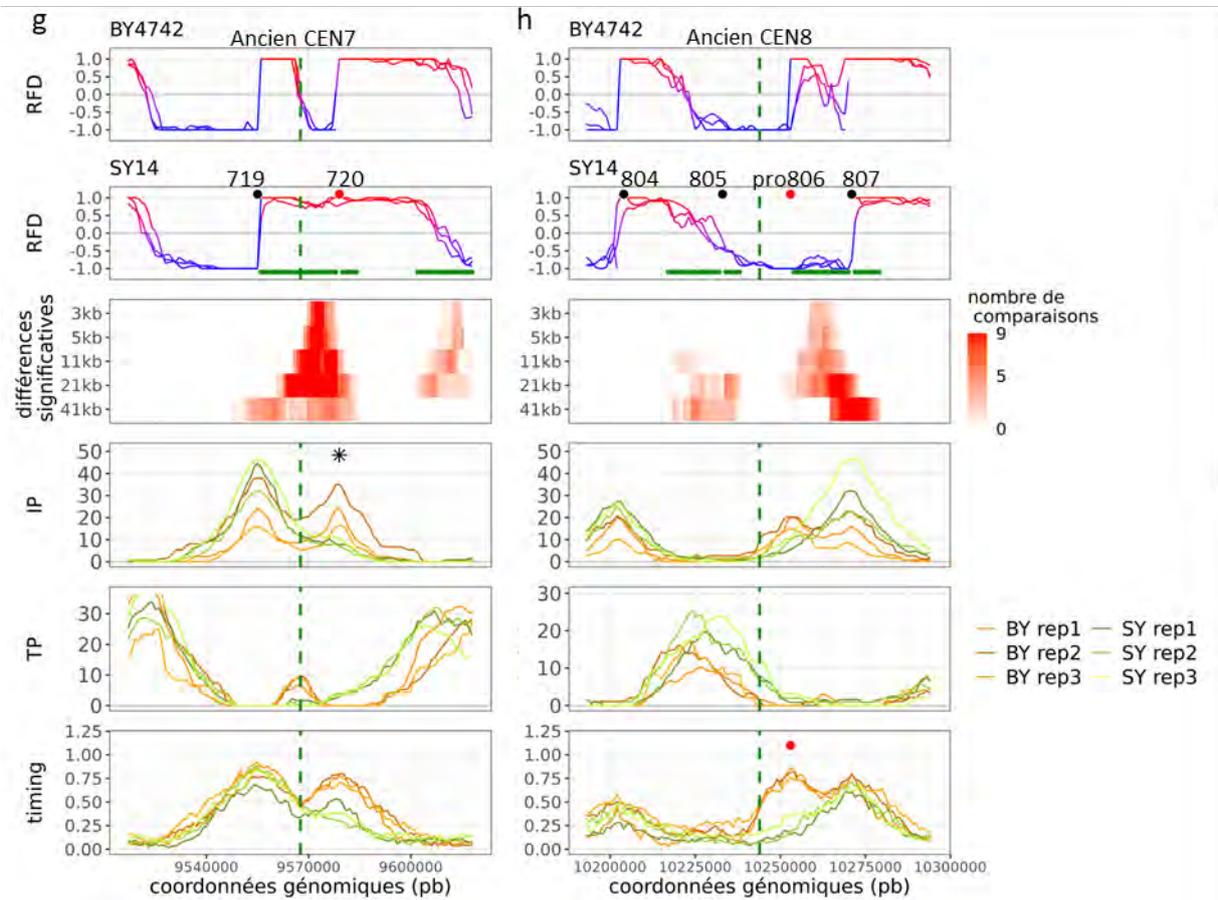
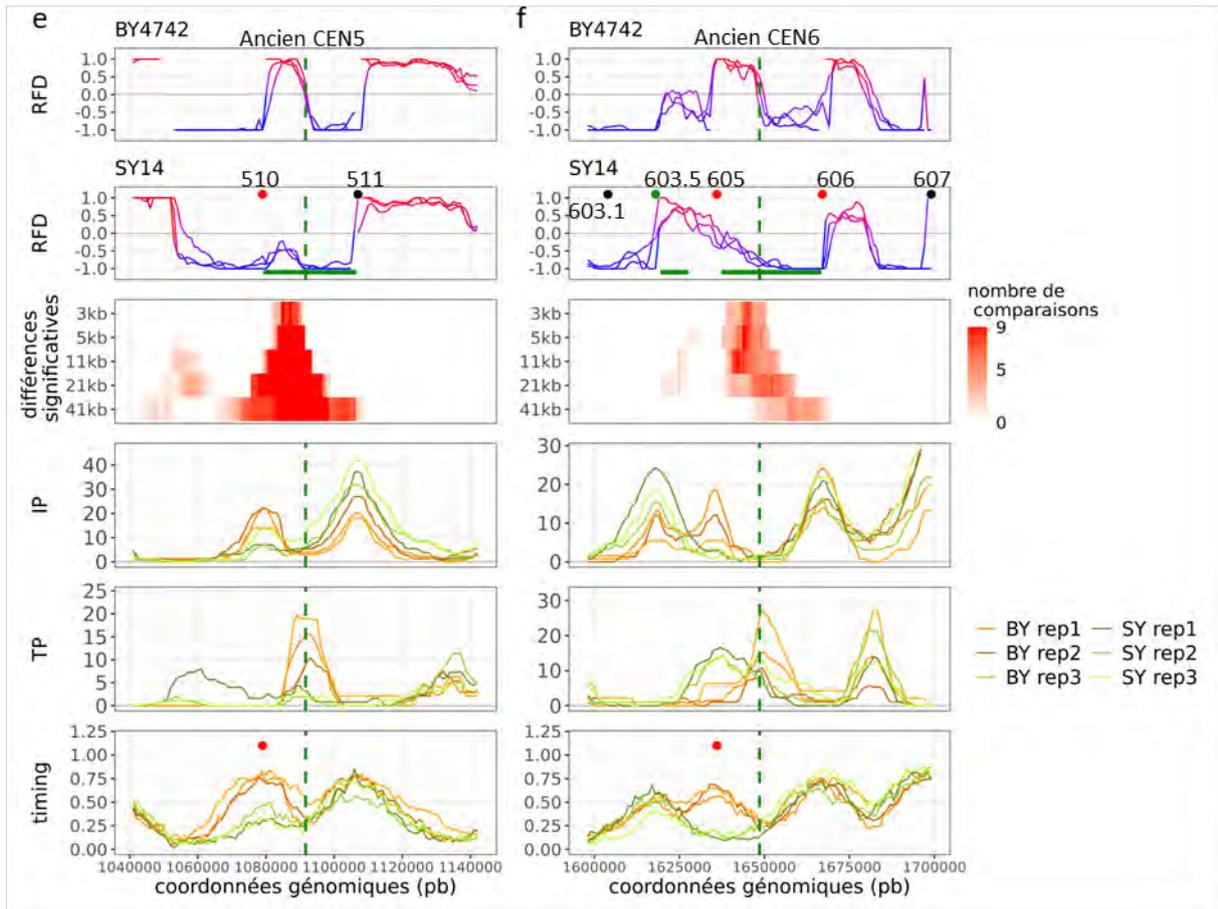


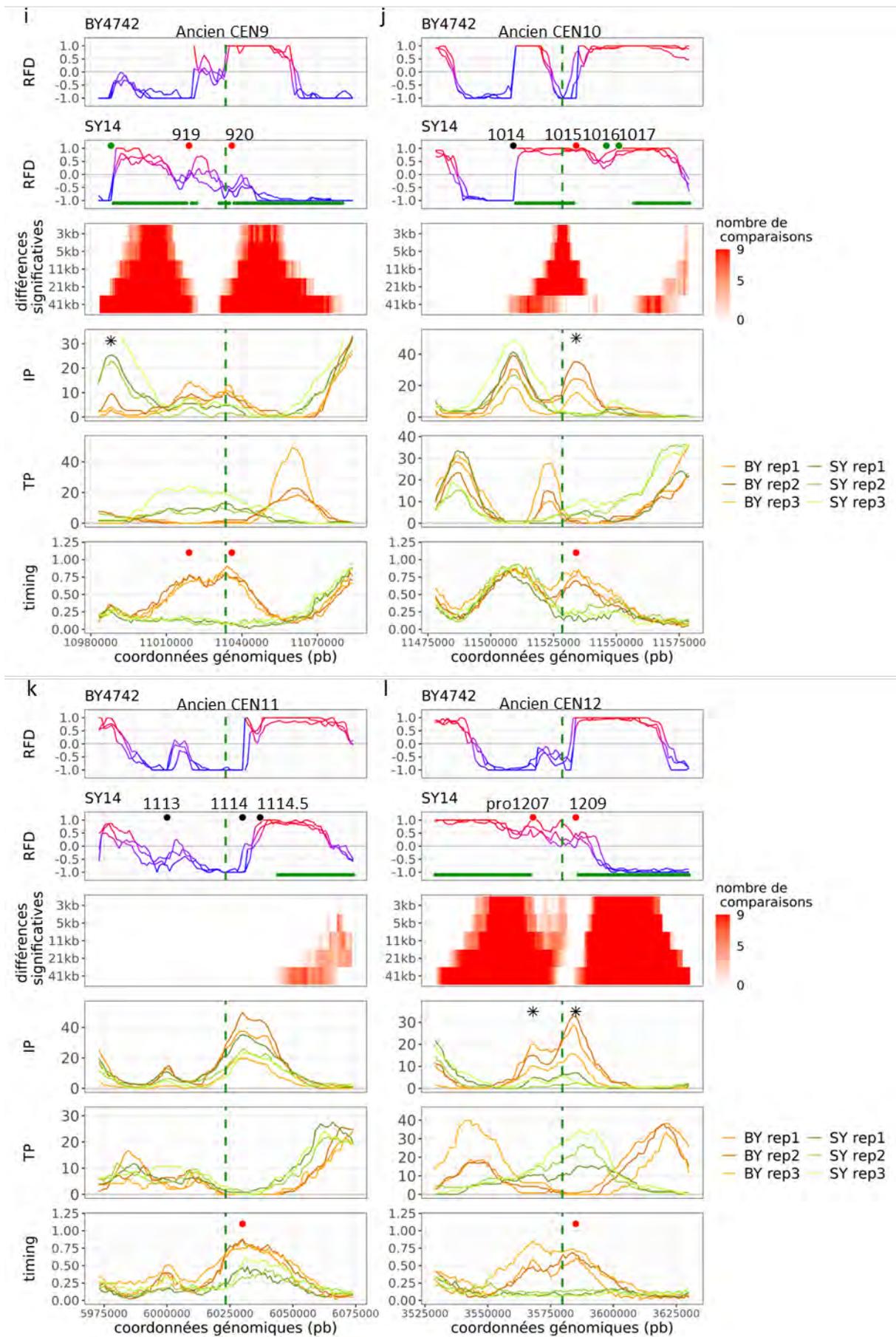
**Figure supplémentaire VIII-17 : Comparaison des profils de réplication entre la BY4742 et la SY14 pour les régions péri-centromériques.**

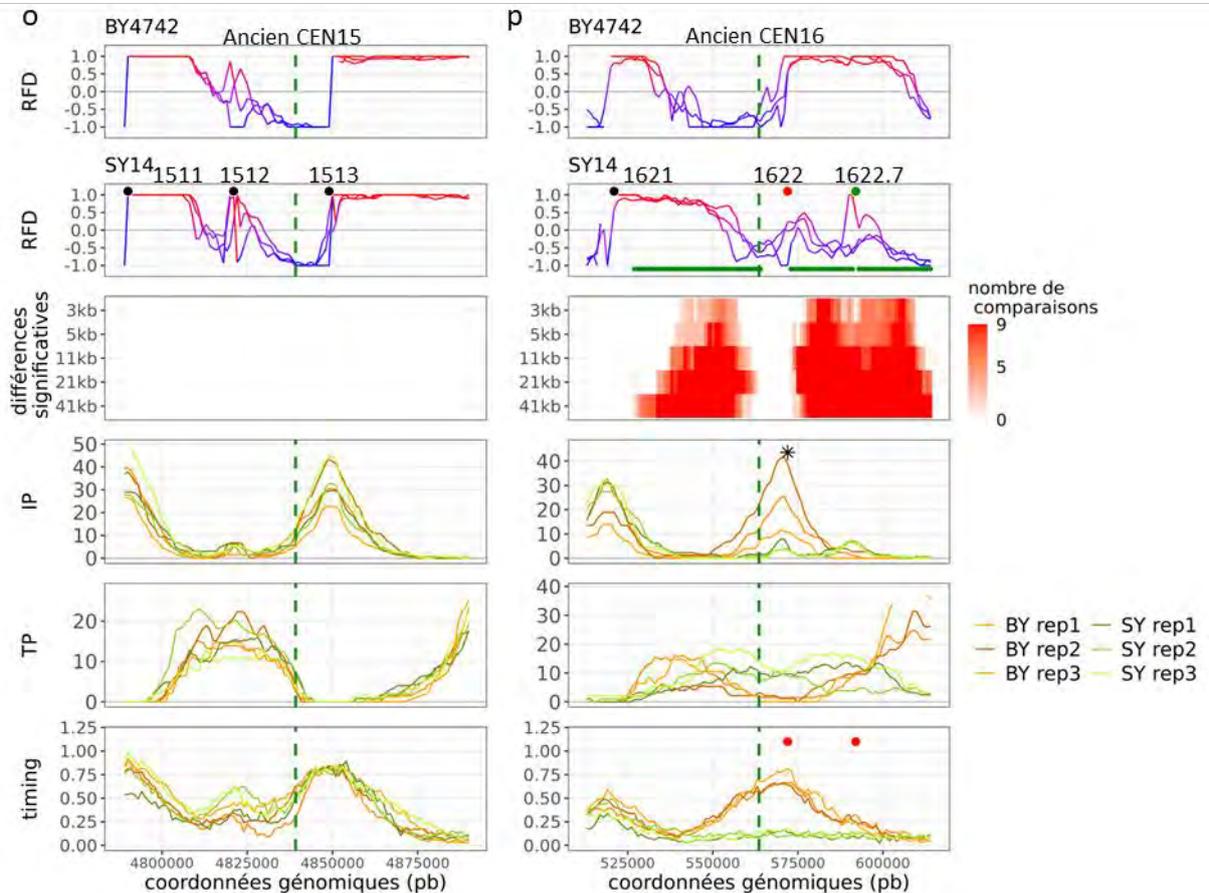
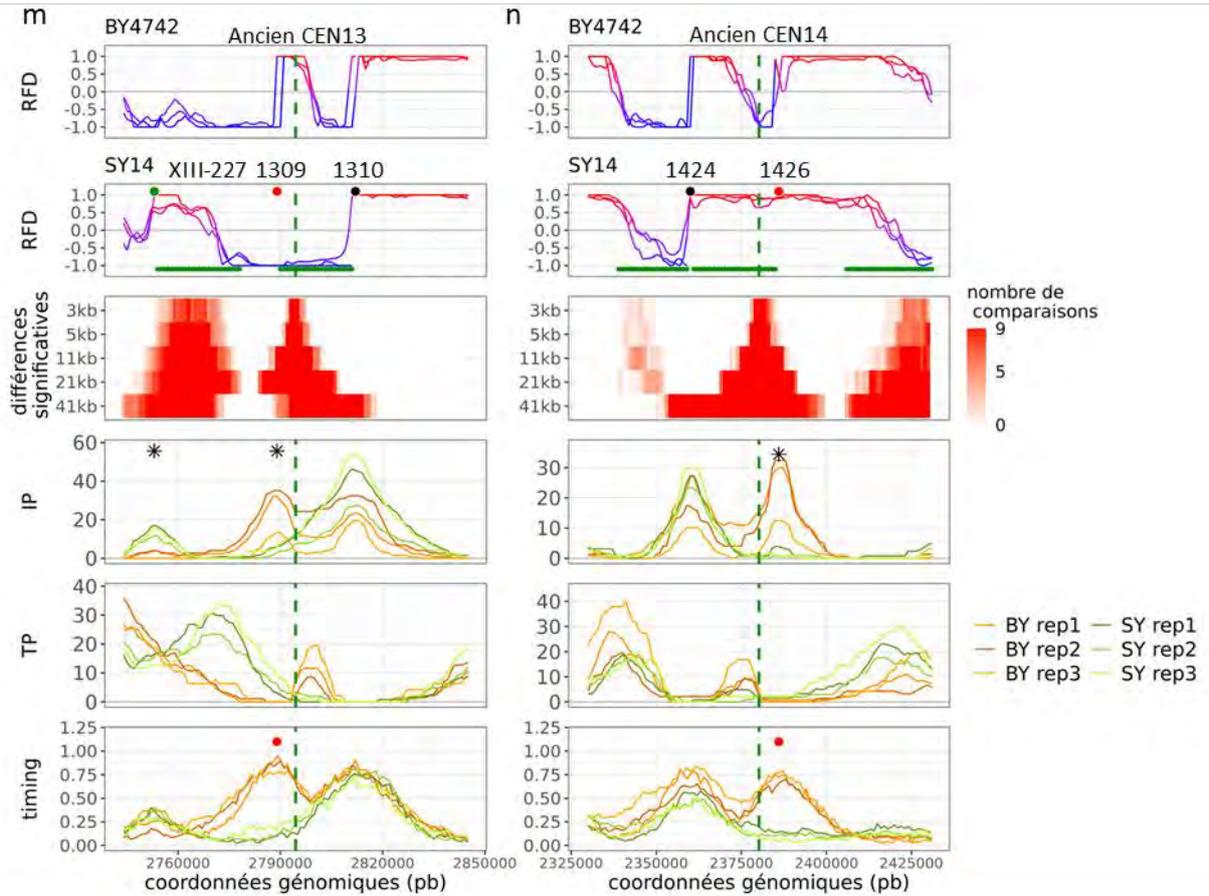
Ces profils de réplication ont été établis à +/- 51kb de la position du début du centromère dans la BY4742 convertis en coordonnées SY14. Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro, et l'origine d'intérêt est indiquée en rouge. La position des centromères est indiquée par des lignes pointillées vertes.

- a) CEN1
- b) CEN2
- c) CEN3
- d) CEN4
- e) CEN5
- f) CEN6
- g) CEN7
- h) CEN8
- i) CEN9
- j) CEN10
- k) CEN11
- l) CEN12
- m) CEN13
- n) CEN14
- o) CEN15
- p) CEN16



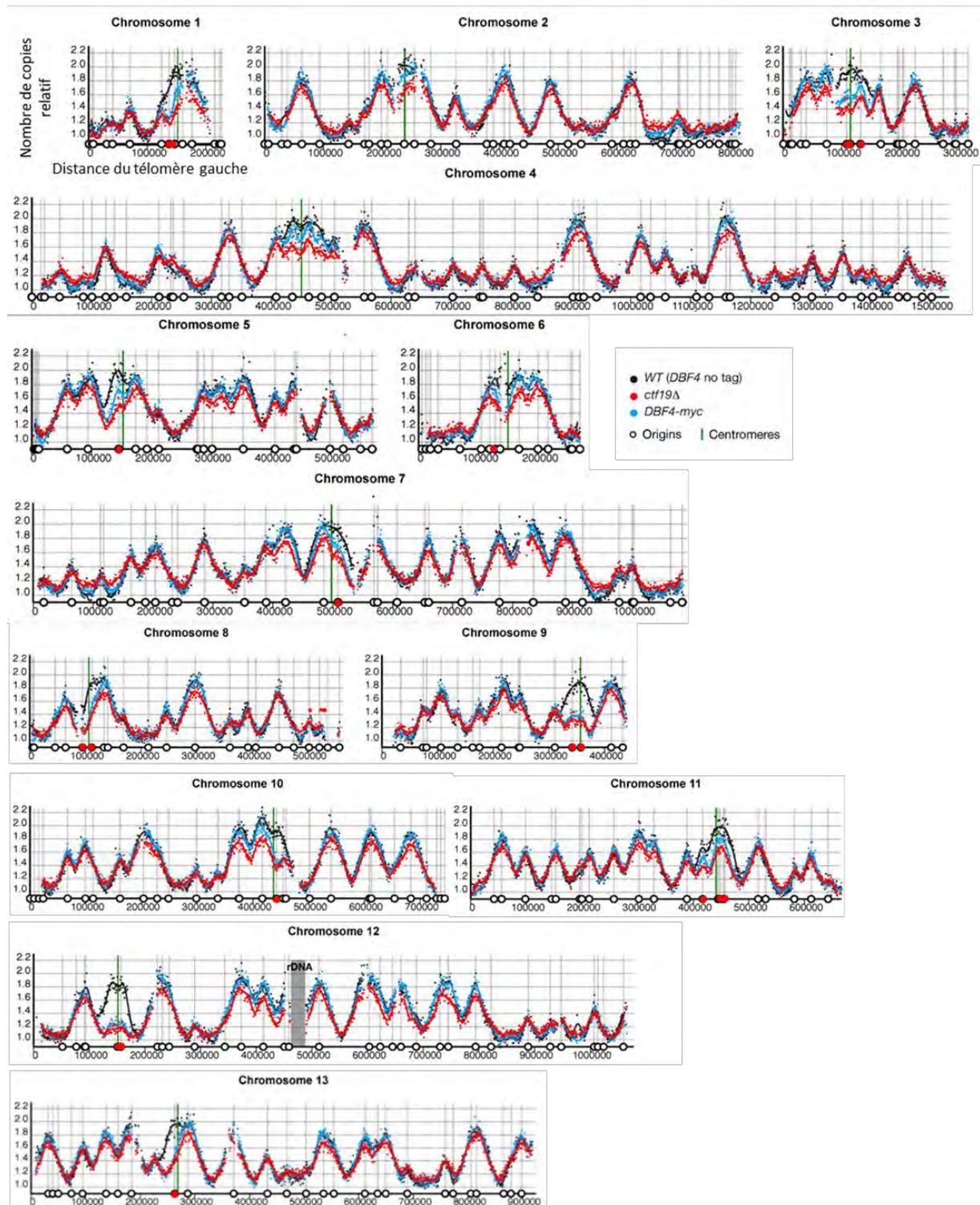


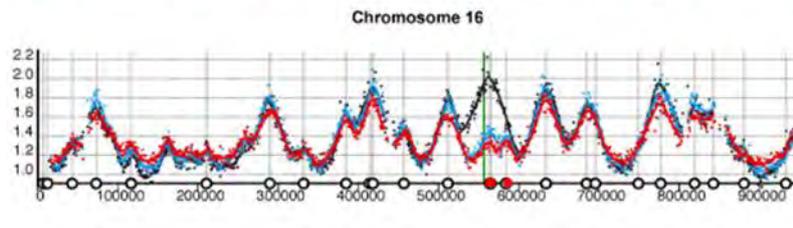
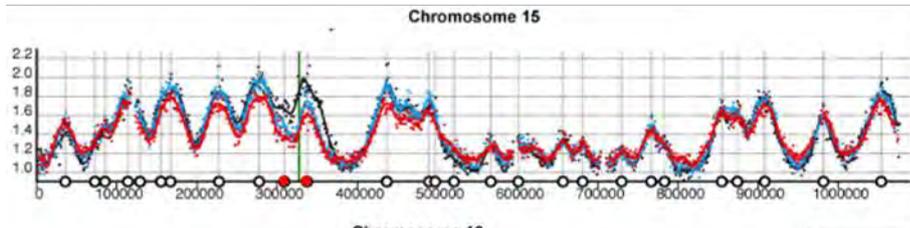
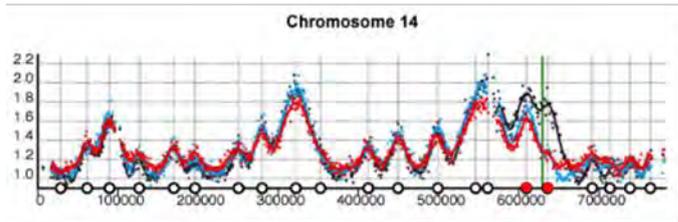




### Figure supplémentaire VIII-18 : Modifications du *timing* de réplication en absence de Ctf19.

Cette figure est adaptée de Natsume et collaborateurs (2013). Les origines considérées comme modifiées en absence de Ctf19 sont indiquées par des points rouges.

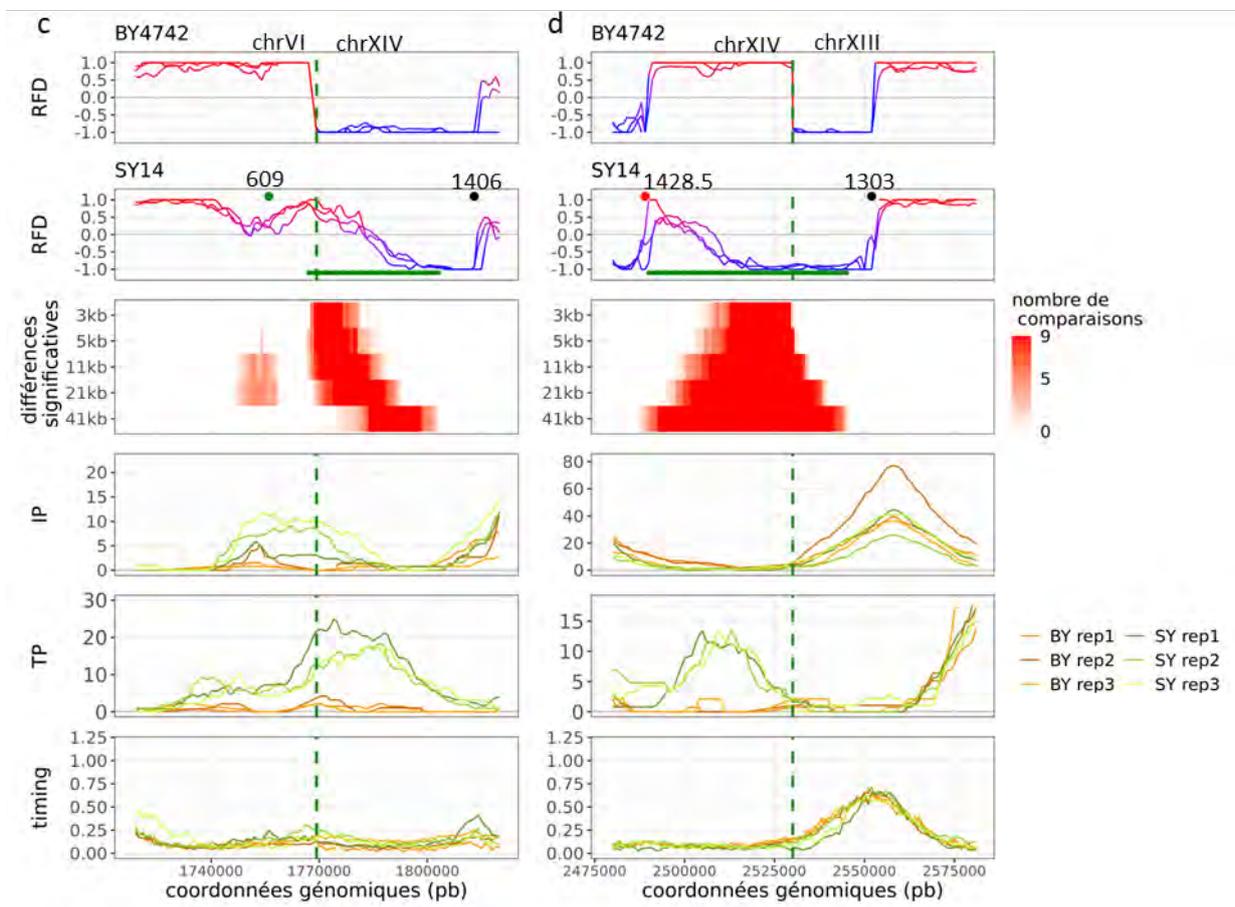
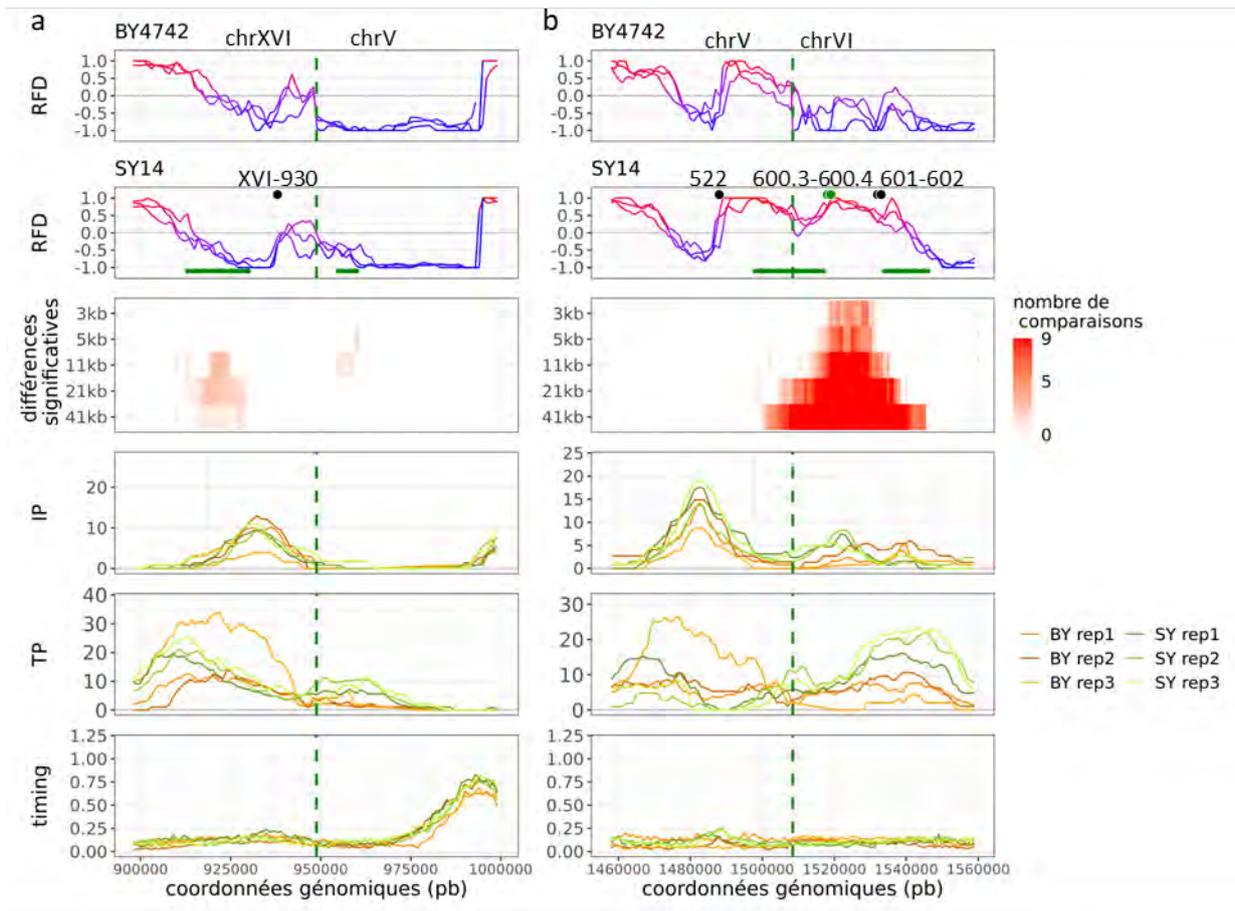


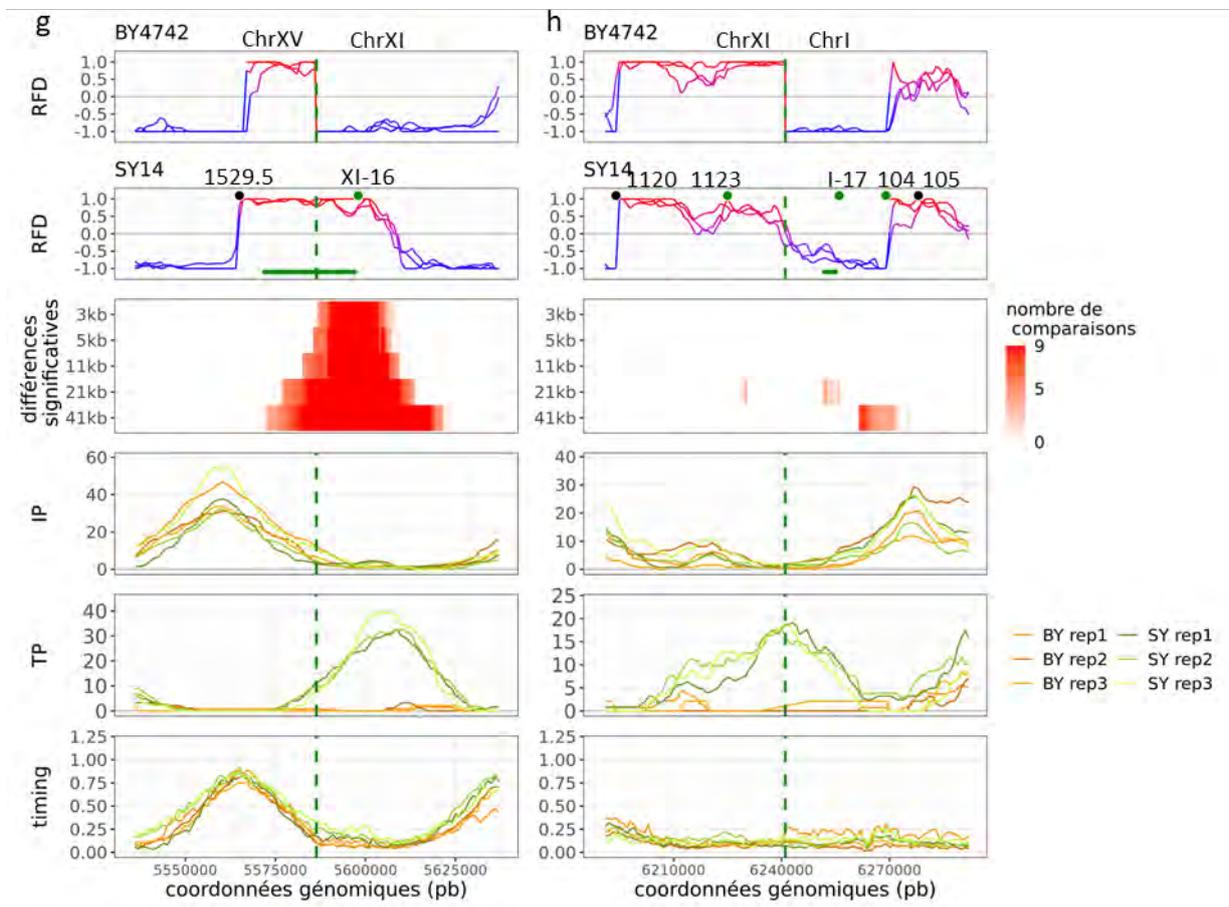
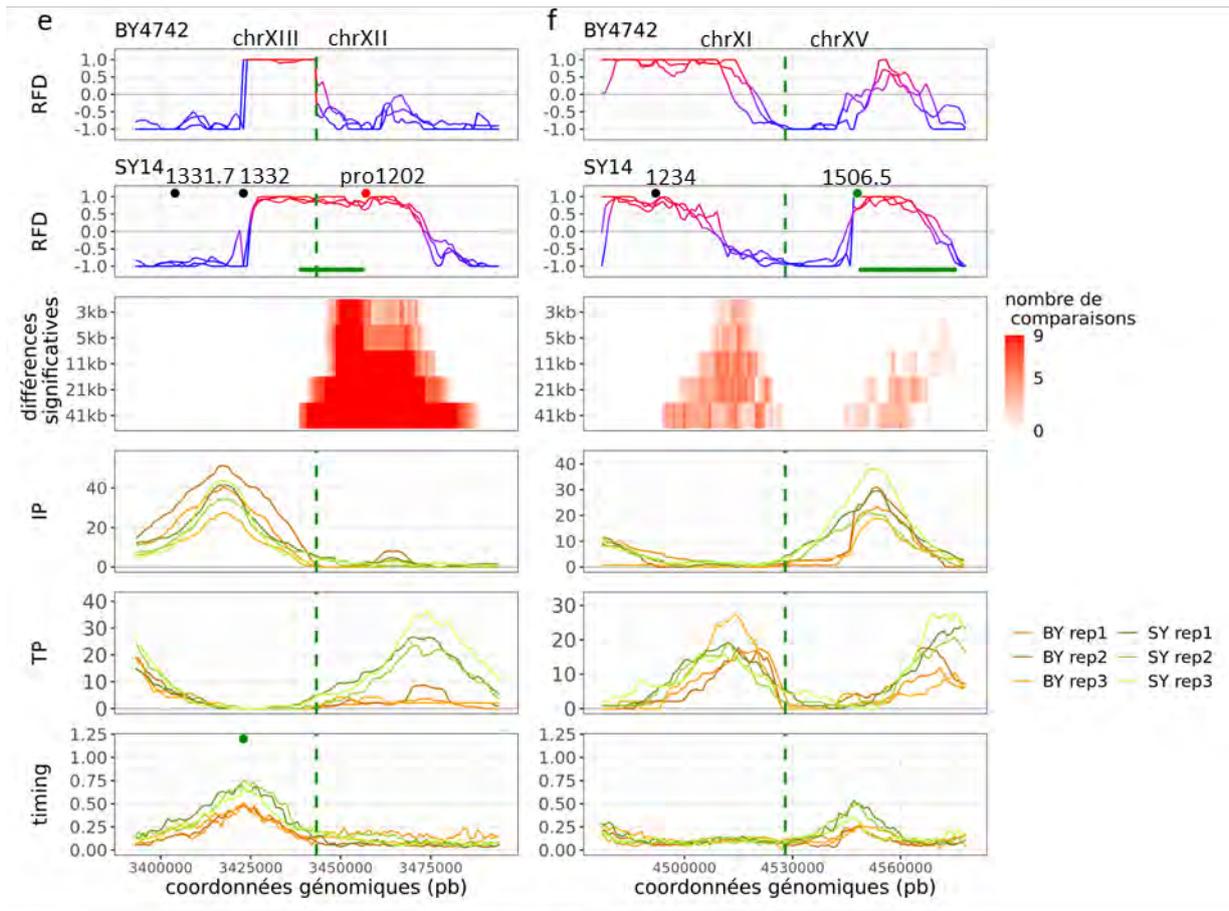


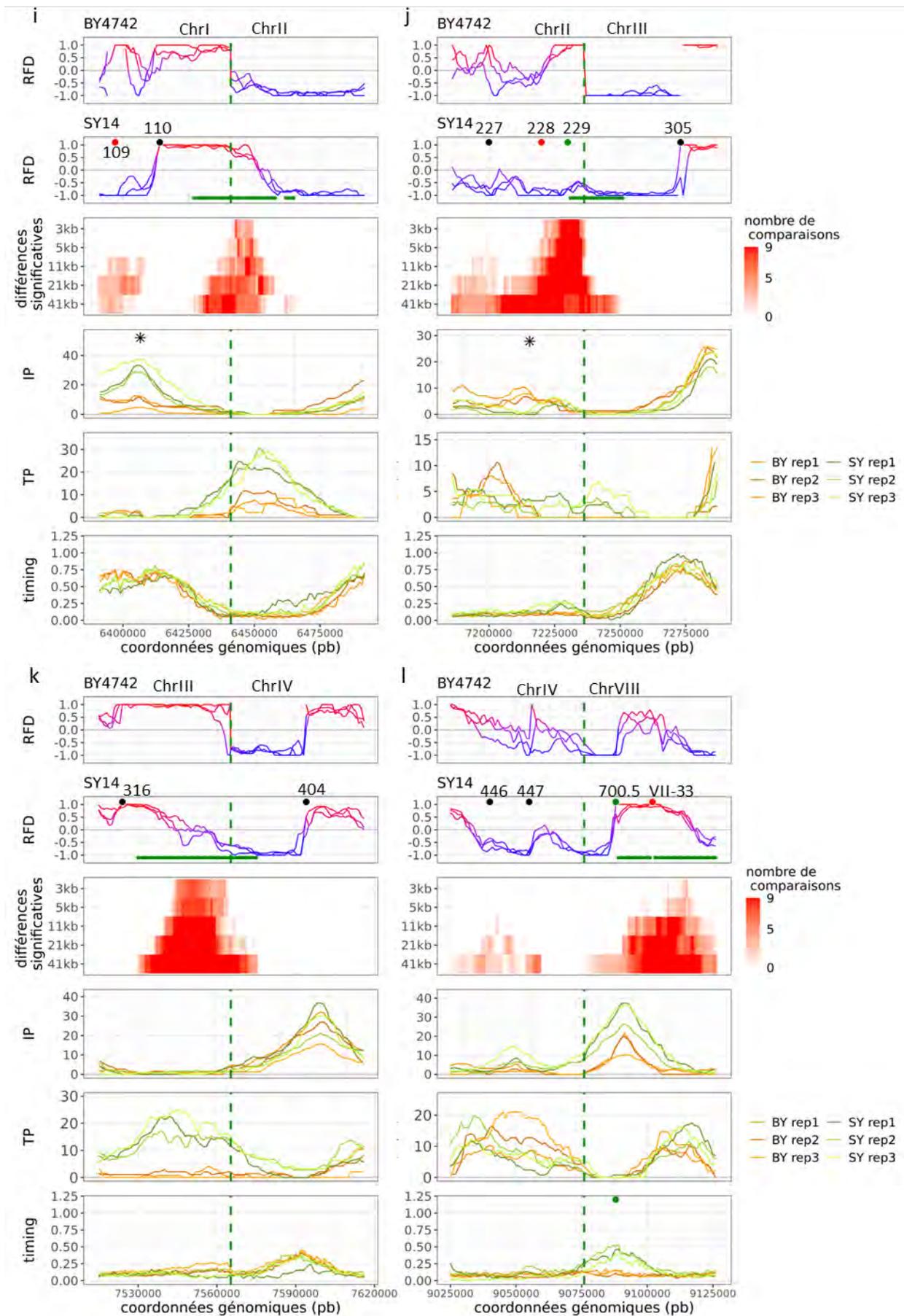
**Figure supplémentaire VIII-19 : Comparaison des profils de réplication entre la BY4742 et la SY14 pour les fusions chromosomiques.**

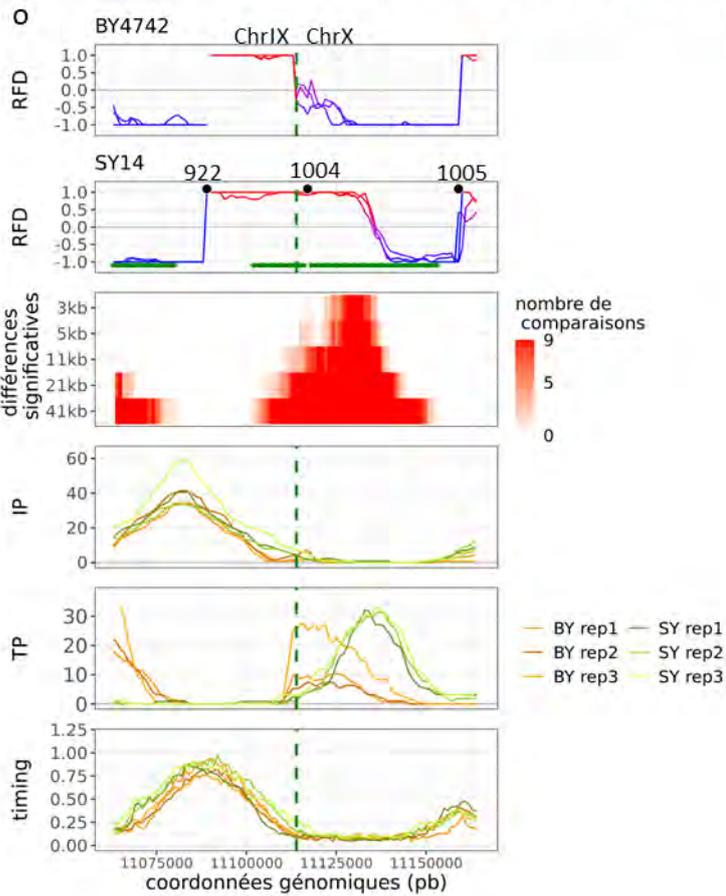
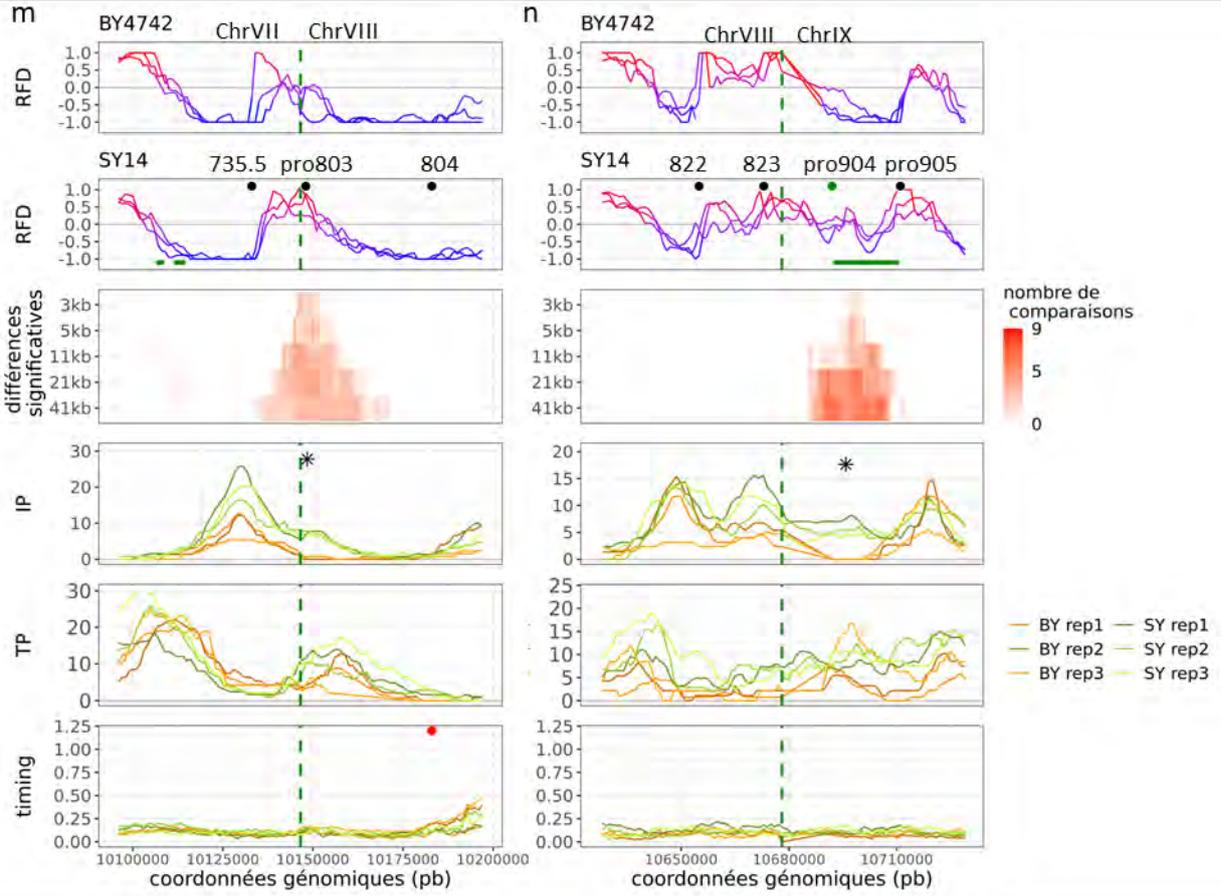
Ces profils de réplication ont été établis à +/- 51kb de la position de la fusion chromosomique, définis par les coordonnées de l'extrémité des chromosomes de la BY4742 s'alignant sur le génome de la SY14. Les origines actives sont indiquées par des cercles noirs sur le profil RFD. Les origines significativement diminuées en RFD ou possédant un *timing* de réplication retardé sont indiquées par des cercles rouges sur les profils correspondants. Les origines significativement augmentées en RFD ou possédant un *timing* significativement avancé sont indiquées par des cercles verts. Les origines significativement modifiées en IP sont indiquées par des étoiles. Les terminaisons modifiées sont représentées par des lignes vertes. Le nom des ARS est indiqué uniquement par leur numéro. La position des fusions chromosomiques est indiquée par des pointillés bleus.

- a) Fusion chromosome XVI chromosome V
- b) Fusion chromosome V chromosome VI
- c) Fusion chromosome VI chromosome XIV
- d) Fusion chromosome XIV chromosome XIII
- e) Fusion chromosome XIII chromosome XII
- f) Fusion chromosome XII chromosome XV
- g) Fusion chromosome XV chromosome XI
- h) Fusion chromosome XI chromosome I
- i) Fusion chromosome II chromosome III
- j) Fusion chromosome III chromosome IV
- k) Fusion chromosome IV chromosome VII
- l) Fusion chromosome VII chromosome VIII
- m) Fusion chromosome VIII chromosome IX
- n) Fusion chromosome IX chromosome X





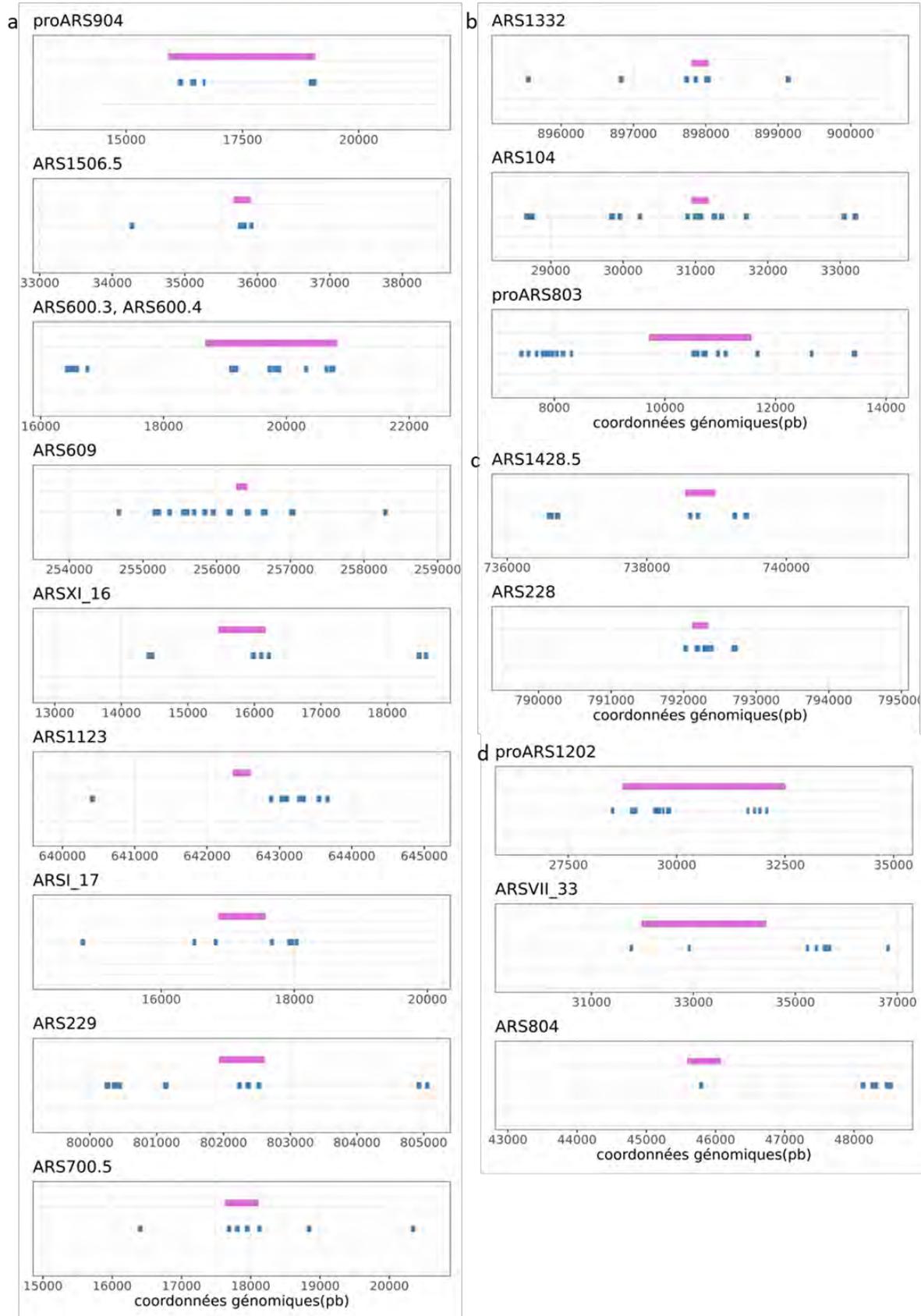




**Figure supplémentaire VIII-20 : Fixation de Rif1 à proximité des origines « proches » des télomères.**

Les sites de fixation de Rif1 obtenus par Hafner et al (2018) sont représentés en bleu, les ARS sont représentés en rose.

- a) Origines inhibées par Rif1 avec une activité « augmentée » dans la SY14
- b) Origines non régulées par Rif1 avec une activité « augmentée » dans la SY14
- c) Origines inhibées par Rif1 avec une activité « diminuée » dans la SY14
- d) Origines non régulées par Rif1 avec une activité « diminuée » dans la SY14
- e) Origines régulées par Rif1, non modifiée dans la SY14
- f) Origines non régulées par Rif1, non modifiée dans la SY14

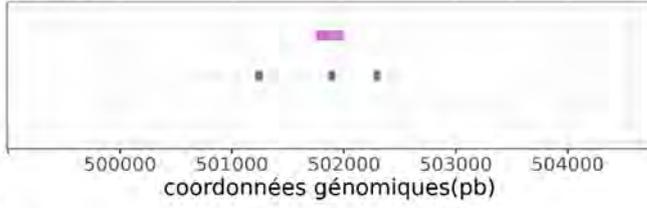




f



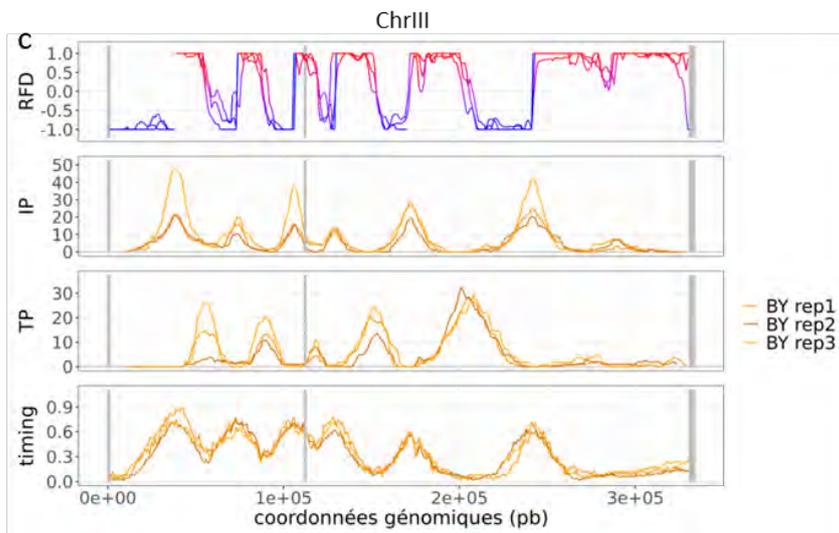
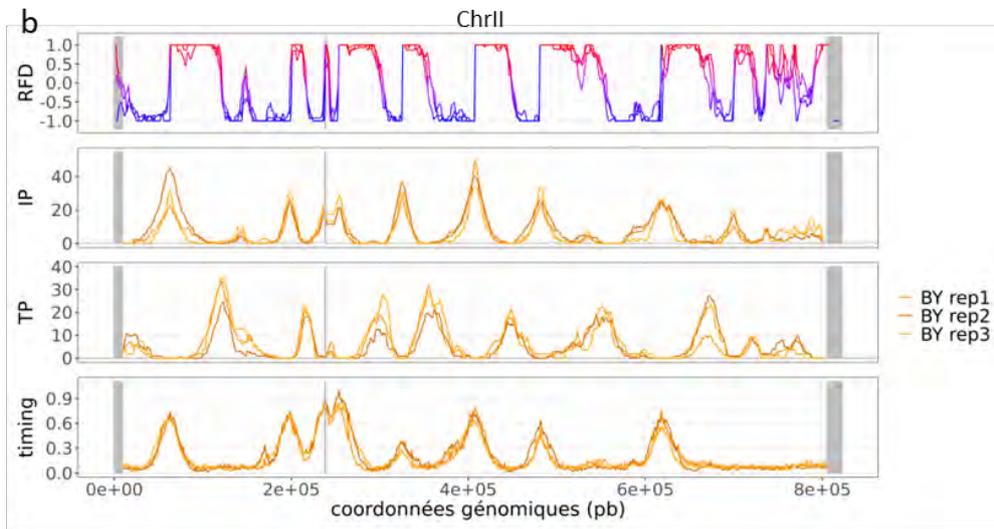
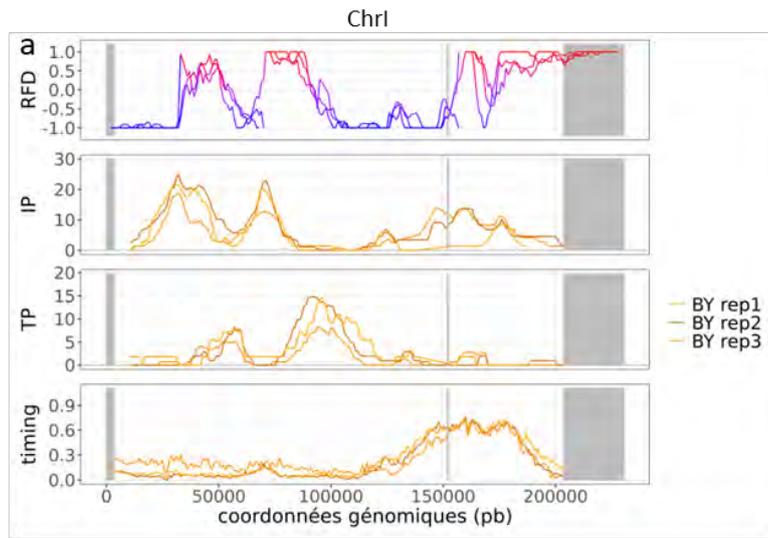
ARS822

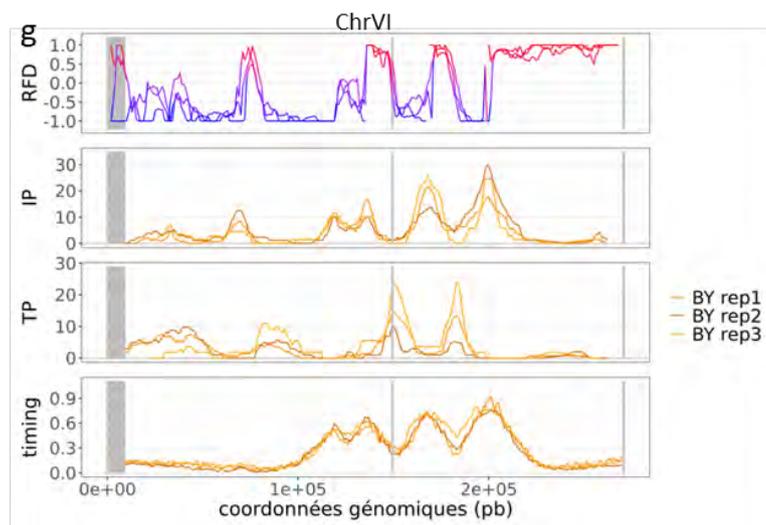
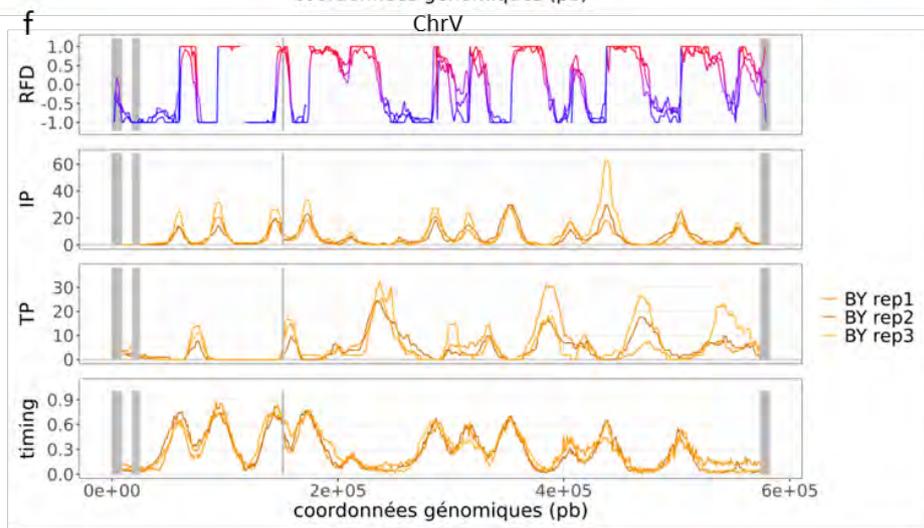
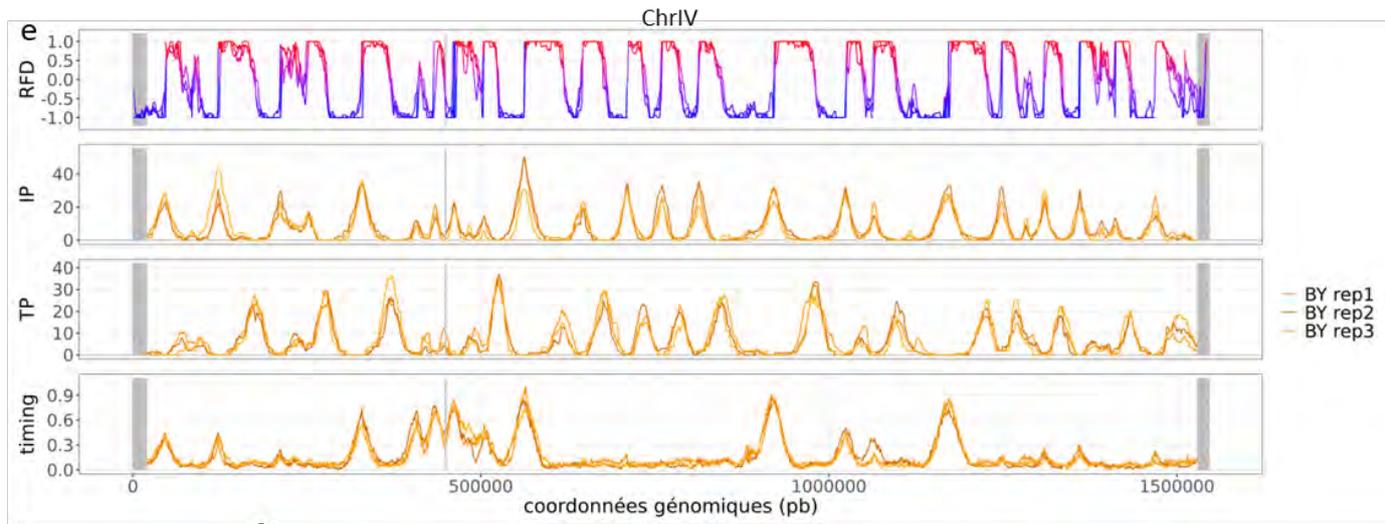


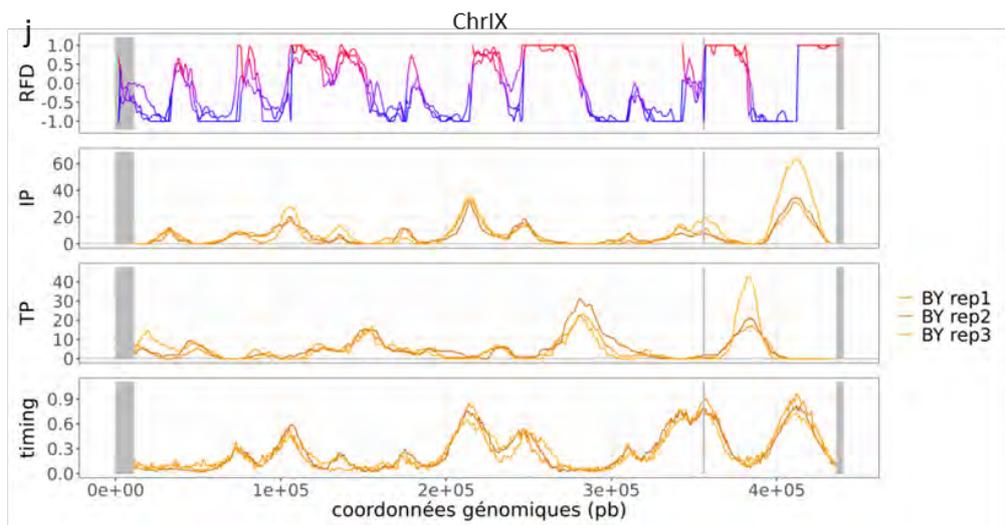
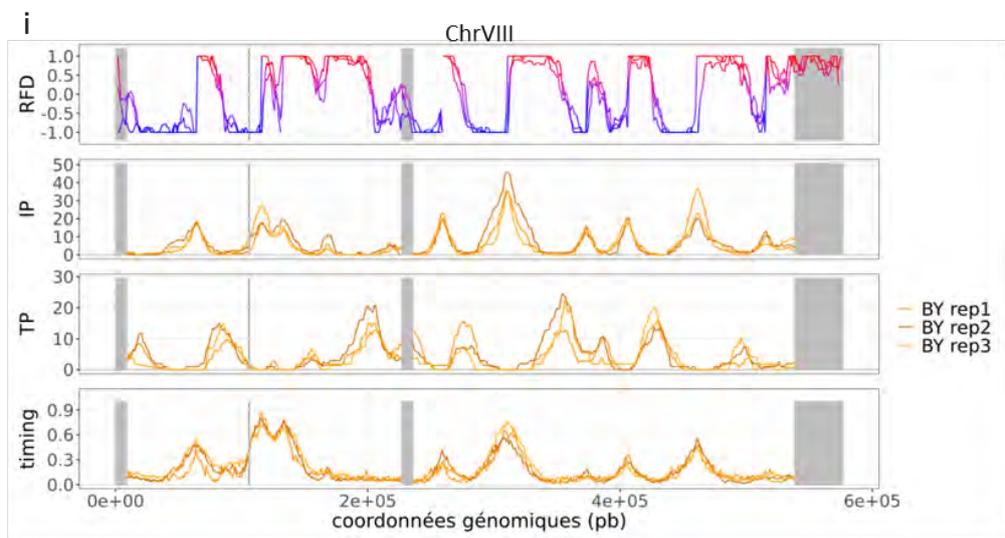
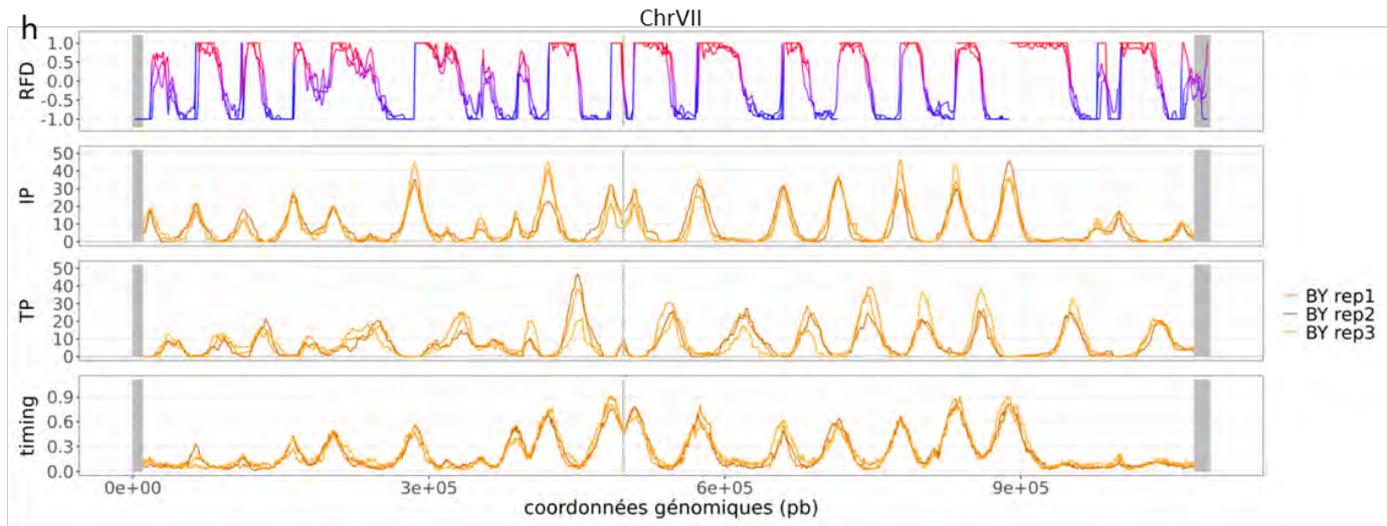
### Figure supplémentaire VIII-21 : Profil de réplication de la BY4742.

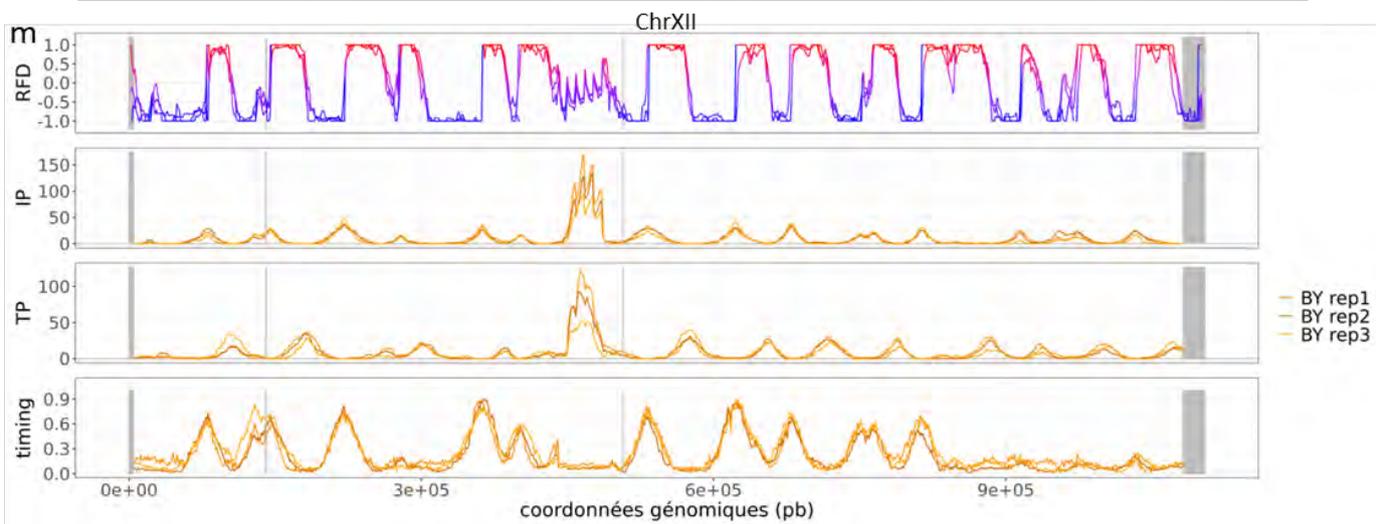
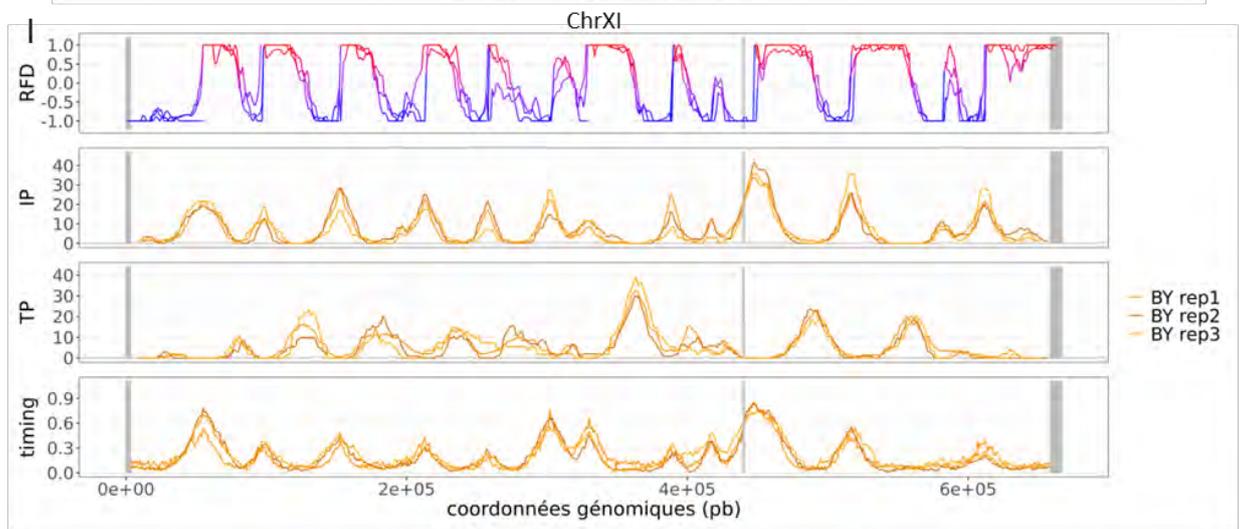
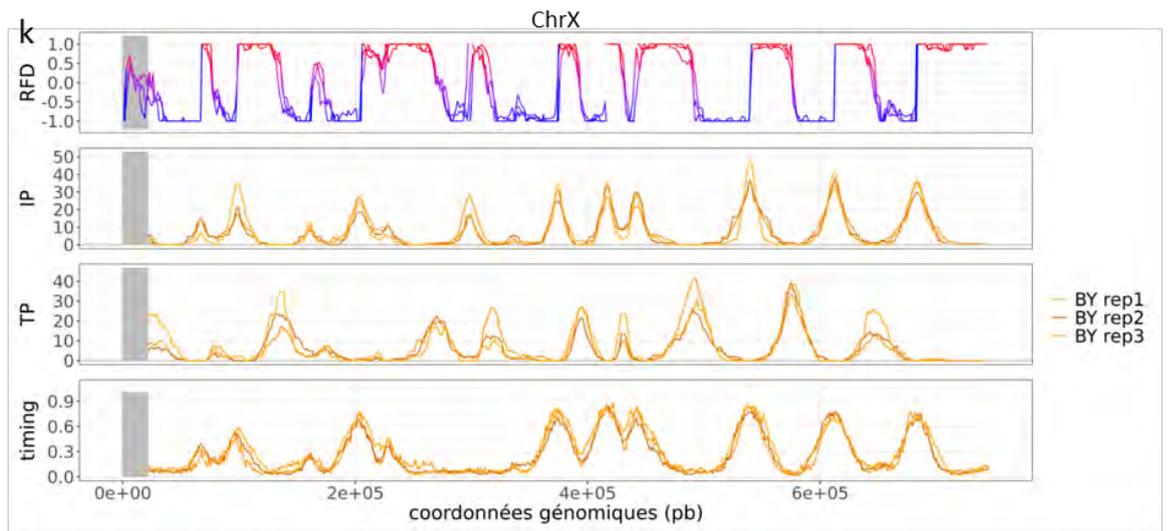
Les zones supprimées dans la SY14 dont les coordonnées ont été obtenues à la suite de l'alignement des génomes sont représentées en gris.

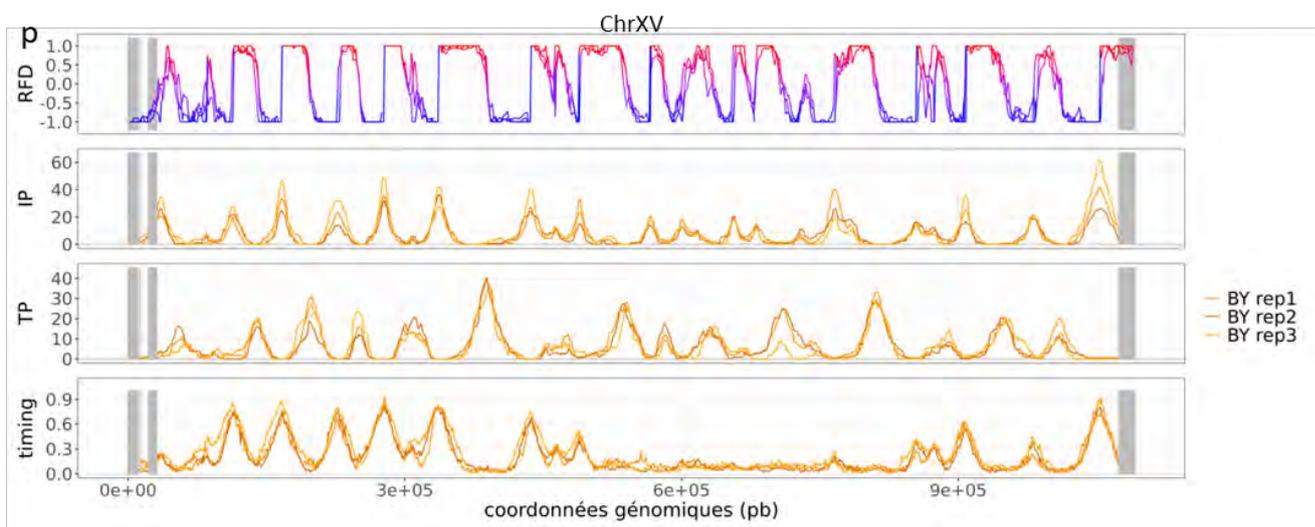
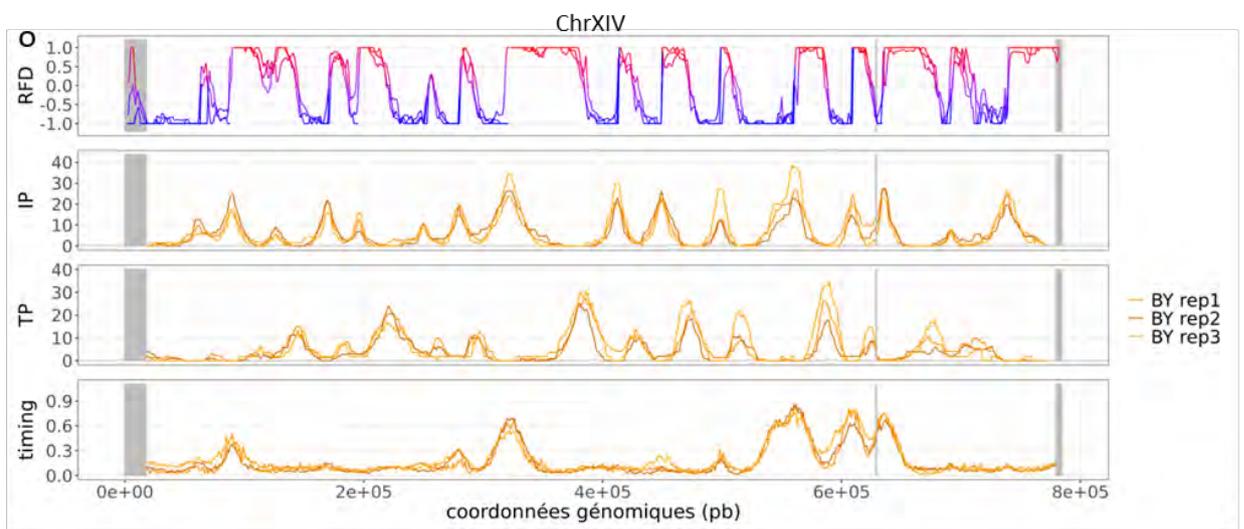
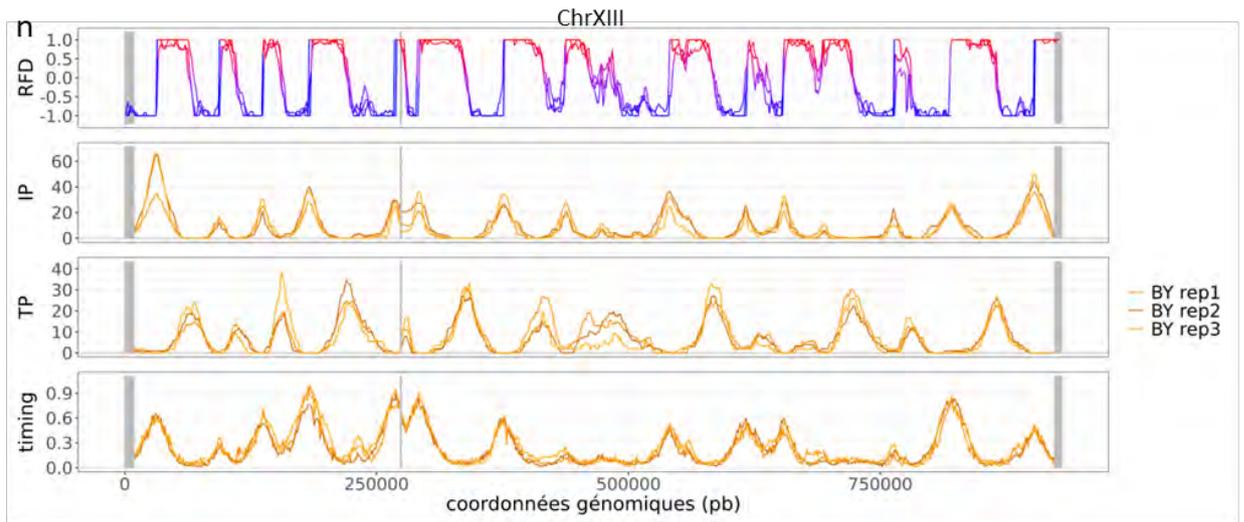
- a) Chromosome I
- b) Chromosome II
- c) Chromosome III
- d) Chromosome IV
- e) Chromosome V
- f) Chromosome VI
- g) Chromosome VII
- h) Chromosome VIII
- i) Chromosome IX
- j) Chromosome X, la suppression du centromère n'est pas visible car seulement 196pb ont été retirées.
- k) Chromosome XI
- l) Chromosome XII
- m) Chromosome XIII
- n) Chromosome XIV
- o) Chromosome XV
- p) Chromosome XVI

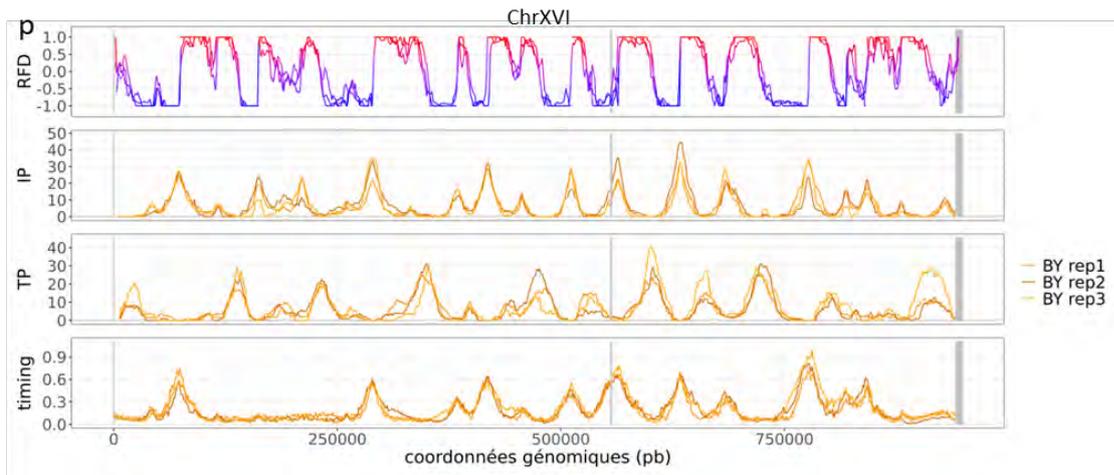












## IX. Annexe : Genome-wide mapping of individual replication fork velocities using nanopore sequencing

J'ai contribué à cet article en réalisant des analyses préliminaires à la construction de la BT1, ainsi qu'en réalisant le mutant *rtt109* dans cette souche.

# Genome-wide mapping of individual replication fork velocities using nanopore sequencing

Bertrand Theulot <sup>1,2,9</sup>, Laurent Lacroix <sup>1,9</sup>✉, Jean-Michel Arbona <sup>3</sup>, Gael A. Millot <sup>4</sup>, Etienne Jean<sup>1</sup>, Corinne Cruaud<sup>5</sup>, Jade Pellet<sup>1</sup>, Florence Proux<sup>1</sup>, Magali Hennion <sup>6</sup>, Stefan Engelen <sup>7</sup>, Arnaud Lemainque<sup>5</sup>, Benjamin Audit <sup>8</sup>, Olivier Hyrien <sup>1</sup>✉ & Benoît Le Tallec <sup>1</sup>✉

Little is known about replication fork velocity variations along eukaryotic genomes, since reference techniques to determine fork speed either provide no sequence information or suffer from low throughput. Here we present NanoForkSpeed, a nanopore sequencing-based method to map and extract the velocity of individual forks detected as tracks of the thymidine analogue bromodeoxyuridine incorporated during a brief pulse-labelling of asynchronously growing cells. NanoForkSpeed retrieves previous *Saccharomyces cerevisiae* mean fork speed estimates ( $\approx 2$  kb/min) in the BT1 strain exhibiting highly efficient bromodeoxyuridine incorporation and wild-type growth, and precisely quantifies speed changes in cells with altered replisome progression or exposed to hydroxyurea. The positioning of >125,000 fork velocities provides a genome-wide map of fork progression based on individual fork rates, showing a uniform fork speed across yeast chromosomes except for a marked slowdown at known pausing sites.

<sup>1</sup>Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, Université PSL, 46 rue d'Ulm, F-75005 Paris, France. <sup>2</sup>Sorbonne Université, Collège Doctoral, F-75005 Paris, France. <sup>3</sup>Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR5239, INSERM, U1293, Université Claude Bernard Lyon 1, 46 allée d'Italie, F-69364 Lyon, France. <sup>4</sup>Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France. <sup>5</sup>Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France. <sup>6</sup>Université Paris Cité, Epigenetics and Cell Fate, UMR7216, CNRS, Paris 75013, France. <sup>7</sup>Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ. Evry, Université Paris-Saclay, 91057 Evry, France. <sup>8</sup>ENSL, CNRS, Laboratoire de physique, F-69342 Lyon, France. <sup>9</sup>These authors contributed equally: Bertrand Theulot, Laurent Lacroix. ✉email: [laurent.lacroix@inserm.fr](mailto:laurent.lacroix@inserm.fr); [hyrien@bio.ens.psl.eu](mailto:hyrien@bio.ens.psl.eu); [letallec@bio.ens.psl.eu](mailto:letallec@bio.ens.psl.eu)

Efficient genome duplication in eukaryotes depends on the proper progression of multiple replication forks. Perturbations of fork movement lead to fork stalling and collapse, generating genomic instability that likely drives cancer development<sup>1</sup>. Despite this pivotal role, the determinants of fork progression are still elusive.

Single-molecule (SM) analyses of DNA replication by DNA fibre autoradiography and its fluorographic evolutions have revealed relatively constant mean fork speeds (1–2 kb/min) in eukaryotic cells together with a broad dispersion of individual fork velocities (0.5–4.0 kb/min) within a given cell line<sup>2–4</sup> and even within a single cell<sup>5</sup>, suggesting large fluctuations of replication fork speed along eukaryotic genomes. When combined with fluorescence in situ hybridization (FISH) of DNA probes, DNA fibre studies can unveil fork progression within specific regions<sup>6</sup>. In some cases no marked differences between local and genome-wide speeds were observed (e.g., refs. 6–9), but slower forks along the pericentromeric and centromeric portions of human chromosomes<sup>10</sup> and faster forks in long transcribed genes in chicken cells<sup>9</sup> have been reported. However, only a handful of loci could be analysed given the excessively low throughput of such approaches. To map replication genome-wide, Raghuraman and colleagues determined the mean replication timing (RT) profile of the entire *S. cerevisiae* genome and interpreted the slopes connecting peaks and valleys (i.e., regions of initiation and termination of DNA replication, respectively) as a proxy for local, population-averaged fork velocities, calculating a broad range of speeds depending on genomic location<sup>11</sup>. In sharp contrast, a time-course monitoring of replisome progression found that population-averaged fork velocity is homogeneous throughout yeast chromosomes<sup>12</sup>. It therefore remains unclear whether forks travel at variable or constant speed along eukaryotic genomes.

We have recently developed FORK-seq, a high-throughput, high-resolution, SM-based replication mapping technique relying on the detection by nanopore sequencing of 5-bromo-2'-deoxyuridine (BrdU), a thymidine analogue incorporated in replicating DNA<sup>13</sup>. Here we introduce NanoForkSpeed (NFS), a method capable of positioning, orienting and extracting the velocity of replication forks from BrdU tracks synthesized during a brief pulse-labelling of asynchronously growing cells. NFS allows the determination of the speed of single forks with unprecedented spatial and temporal resolutions, with an unparalleled throughput, and with remarkable simplicity in terms of sample preparation and analysis. Thanks to NFS, we generated in *S. cerevisiae* a genome-wide map of fork progression based on the measurement of individual fork velocities.

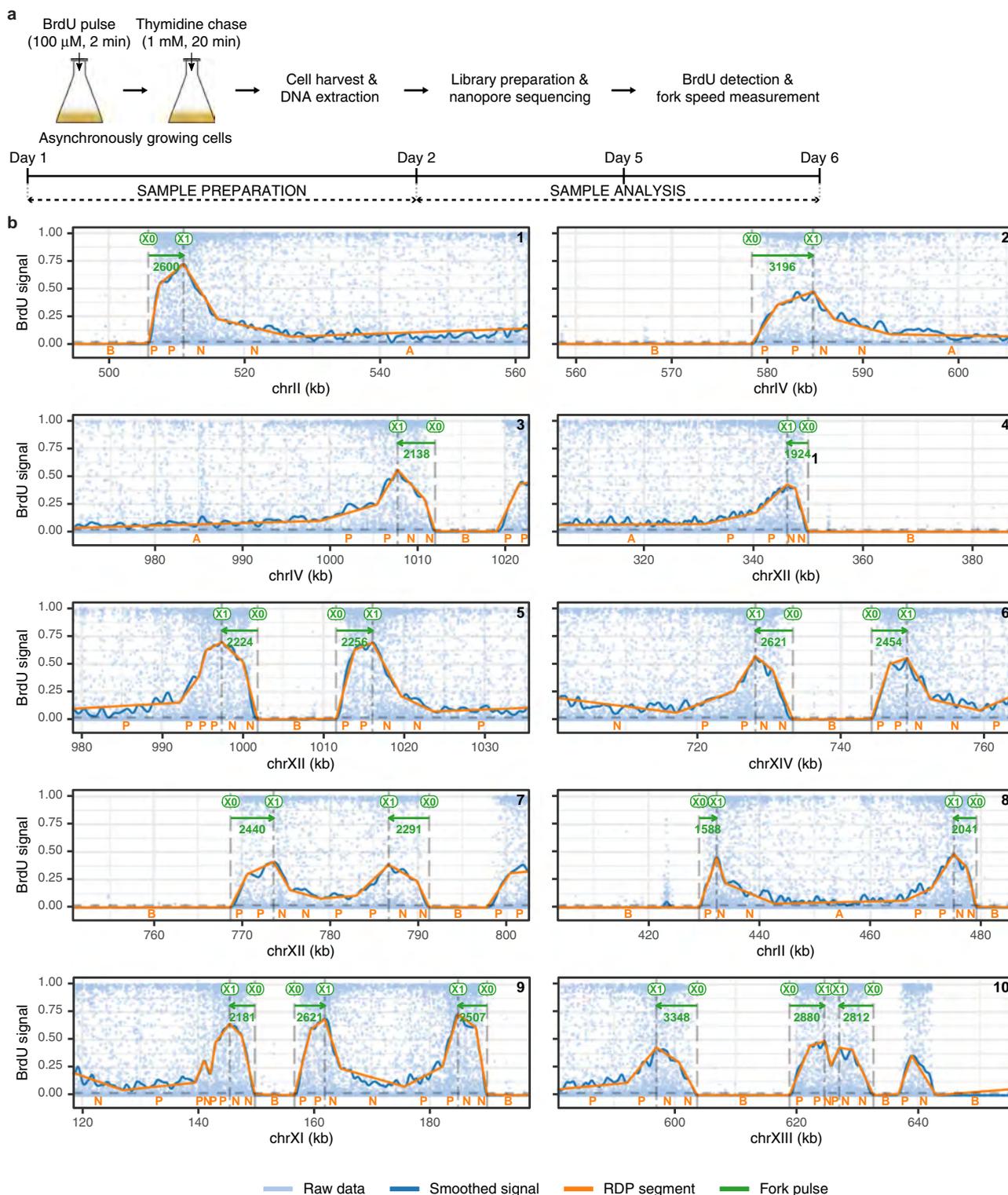
## Results

**Detection of BrdU incorporated into replicating DNA.** We previously developed the RepNano software to detect BrdU on sequencing reads produced by Oxford Nanopore Technologies (ONT) devices<sup>13</sup>. Here we anchored our BrdU detection process in ONT's Megalodon program (<https://github.com/nanoporetech/megalodon>) and estimated BrdU incorporation probability at each thymidine site rather than over 96 bp windows as in RepNano<sup>13</sup> (see the 'Methods' section). Megalodon combines Guppy (ONT) GPU-accelerated basecalling and read mapping to the reference genome into one single step, making it a fast, straightforward and easy-to-use pipeline. Its neural network was trained using nanopore reads of genomic DNA from thymidine-auxotroph MCM869 yeast cells<sup>14</sup> grown with 11 different proportions of BrdU in the culture medium, from 0 to 100% with increments of 10% ('Methods'). We first tested the ability of Megalodon to recover from nanopore reads the total BrdU content of these genomic DNA samples determined by mass spectrometry

(Supplementary Fig. 1a). RepNano<sup>13</sup> and other published BrdU basecallers, DNAscent<sup>15</sup> and DNAscent v2<sup>16</sup>, were also assessed for comparison. Megalodon exhibited the lowest background signal without BrdU and most closely paralleled the perfect correlation line, despite a slight tendency to underestimate BrdU content (Supplementary Fig. 1a). Megalodon also gave the most balanced estimates of BrdU proportion per 1 kb window over the entire range of tested BrdU contents, including the narrowest peaks of null BrdU content corresponding to parental DNA, showing again the lowest background of all basecallers (Supplementary Fig. 1b).

**BT1, a yeast strain with optimised BrdU incorporation and wild-type growth.** The budding yeast *S. cerevisiae* lacks a thymidine salvage pathway and is therefore unable to incorporate exogenous thymidine into DNA. Several strategies have been adopted to reconstitute this pathway in vivo in order to render yeasts amenable to DNA labelling with thymidine analogues. They converged towards the combined expression of human equilibrative nucleoside transporter 1 (hENT1), which improves exogenous thymidine uptake, and of Herpes simplex virus thymidine kinase (hsvTK) allowing the conversion of thymidine into thymidine monophosphate (dTMP) (e.g., refs. 17–19). Additional inactivation of the thymidylate synthase-encoding *CDC21* gene that is essential for de novo dTMP biosynthesis resulted in strains entirely dependent on external thymidine, or its analogues, for growth<sup>14,20</sup>. We previously analysed the replication of the yeast genome by FORK-seq using one such strain, namely MCM869<sup>13,14</sup>. BrdU incorporation is very high in MCM869 cells due to the absence of competing intracellular thymidine, enabling the detection of replication tracks synthesized during short pulse-chase experiments<sup>13</sup>. However, MCM869 yeasts grew at a reduced rate compared to wild-type (WT) cells even at saturating thymidine concentration (Supplementary Fig. 2a, b). To examine fork progression in WT conditions, we engineered the BT1 strain that retained a functional de novo pathway, and therefore had WT growth properties (Supplementary Fig. 2a–c), while containing codon-optimised *hsvTK* and *hENT1* genes for maximal protein expression and BrdU incorporation in yeast cells. BrdU content profiles of nanopore-sequenced genomic DNA from asynchronously growing BT1 cells pulsed with 100  $\mu$ M BrdU for 2 min followed by a 20 min chase with 1 mM thymidine showed that BrdU incorporation was of comparable efficiency as in MCM869 cells (Supplementary Fig. 2d). In contrast, replication tracks were hardly detectable on pulse-labelled DNA from thymidine-prototroph, BrdU-incorporating strains constructed using currently available *hsvTK* and *hENT1* integrative vectors<sup>17,19</sup> (BT2 and BT3 strains, Supplementary Fig. 2d). BT1 thus constitutes a potent tool to examine fork progression in physiologically relevant conditions.

**Fork orientation and fork speed measurement by NFS.** Typical BrdU signals in sequencing reads of genomic DNA from pulse-labelled BT1 cells (Fig. 1a) displayed an asymmetrical shape, consisting of a steep ascending slope starting from a segment of null BrdU content (BrdU signal increasing from 0 to  $\approx 0.5$ ), followed by a shallower decreasing slope (from  $\approx 0.5$  to  $\approx 0.1$ ) (Fig. 1b). These signals resembled those from MCM869 cells (Supplementary Fig. 2d), previously demonstrated to correspond to elongating replication forks, with the steep and shallow slopes reflecting BrdU incorporation during the pulse and the chase, respectively, and the signal asymmetry revealing fork direction<sup>13</sup>. Since fork speed can be measured as the length of the labelled track divided by the corresponding labelling time, we developed a pipeline named NFS to capture the track length synthesized during the BrdU pulse, that is the section between the starting



**Fig. 1 Replication fork speed measurement procedure by NFS.** **a** Scheme of the protocol for BrdU pulse-labelling of DNA replication in BT1 cells. The usual timeline is indicated. **b** BrdU content profiles of nanopore sequencing reads of genomic DNA from pulse-labelled BT1 cells processed by NFS. Panels show typical replication signals, namely rightward, leftward, diverging and converging forks. Light blue dots, raw data from Megalodon (dots represent the probability of BrdU at each thymidine position); blue curve, smoothed signal; orange lines, segments resulting from the piecewise linear simplification method using the Ramer-Douglas-Peucker algorithm (RDP) to detect and orient BrdU tracks (B, flat segments with background BrdU level; A, flat segments with a BrdU level above background; P, segments with a positive slope; N, segments with a negative slope); X0, estimated position of the start of BrdU incorporation; X1, estimated position of the start of the thymidine chase; green arrow, fork direction, with fork velocity (bp/min, in green) indicated below.

and ending points of the steep slope. A piecewise linear simplification method first converted reads into a sequence of segments classified into 4 categories based on their slope and mean BrdU signal: (i) flat segments with background BrdU level (B); (ii) flat segments with a BrdU level above background (A); (iii) segments with a positive slope (P); and (iv) segments with a negative slope (N) (Fig. 1b). We then identified two patterns for elongating forks depending on their direction. Rightward forks were preceded by a B segment which replicated before the BrdU pulse, then consisted in one or successive P segments, occasionally interrupted by A segments owing to noise, and followed by at least one N segment corresponding to DNA replicated during the chase (Fig. 1b, reads 1, 2). Consequently, we used a regular expression procedure to search for the “BP(P|A)\*N+” pattern. Leftward forks were recognized with the symmetrical pattern “P+(N|A)\*NB” (Fig. 1b, reads 3, 4). This procedure excluded both incomplete replication tracks (Fig. 1b, read 3, position 1020 kb, and read 7, position 800 kb) and symmetrical signals due to pairs of forks initiated after the start of the pulse or terminated before the end of the pulse, for which the actual labelling time could not be precisely estimated (Fig. 1b, read 10, position 640 kb). The precision of track detection and orientation by NFS was confirmed by (i) a virtually null false-positive rate ( $\approx 50$  mapped forks per 10 Gb of DNA with no BrdU labelling versus  $\approx 15,000$  mapped forks per 10 Gb of pulse-labelled DNA; Supplementary Data 1); (ii) the remarkable similarity of BT1 replication fork directionality (RFD) profile aggregating oriented replication tracks detected by NFS with MCM869 RFD profiles computed by FORK-seq or Okazaki fragment sequencing<sup>13</sup> (Spearman’s pairwise correlation coefficients of 0.83 and 0.88, respectively, Supplementary Fig. 3a, b) and (iii) the spatial coincidence between known yeast origins<sup>21</sup>, upward slopes (i.e., initiation regions) on BT1 RFD profile and individual initiation sites defined as the midpoints between diverging forks (Supplementary Fig. 3a, c).

For every read, NFS yielded the coordinates of each of the B/P/A/N segments decomposing replication tracks. The length of DNA replicated during the BrdU pulse was then readily computed as the distance between the B/P (X0, start of BrdU incorporation) and (P|A)/N transitions (X1, start of the thymidine chase) for rightward forks and between N/B (X0) and P/(A|N) (X1) transitions for leftward forks (Fig. 1b). Fork velocity was subsequently calculated as the ratio between track length and BrdU pulse duration. The latter was set to 2 min, which was sufficient for detection of replication signals while brief enough to maximize the probability of obtaining complete tracks along nanopore reads averaging  $\approx 15$  kb in length. We first conducted a pilot experiment in which BT1 cells were pulsed with BrdU doses ranging from 10  $\mu$ M to 1 mM and chased with a ten-fold excess of thymidine to determine the optimal labelling conditions. As anticipated, the maximal BrdU signal amplitude rose with increasing BrdU concentrations, whereas similar velocities were computed regardless of the BrdU dose (Supplementary Fig. 4). This indicated that our pipeline was functional for a broad range of signal amplitudes and that fork speed measurement by NFS was independent of BrdU concentration in the tested range. However, in line with reports that high doses of BrdU are toxic for TK-expressing *S. cerevisiae* strains<sup>20,22</sup>, we found that BrdU concentrations over 100  $\mu$ M both slowed down BT1 S phase and triggered checkpoint activation (Supplementary Fig. 5). Although these doses did not visibly impact fork speed (Supplementary Fig. 4), probably because of the very limited exposure to BrdU during the 2 min pulse, we opted for a pulse concentration of 100  $\mu$ M BrdU.

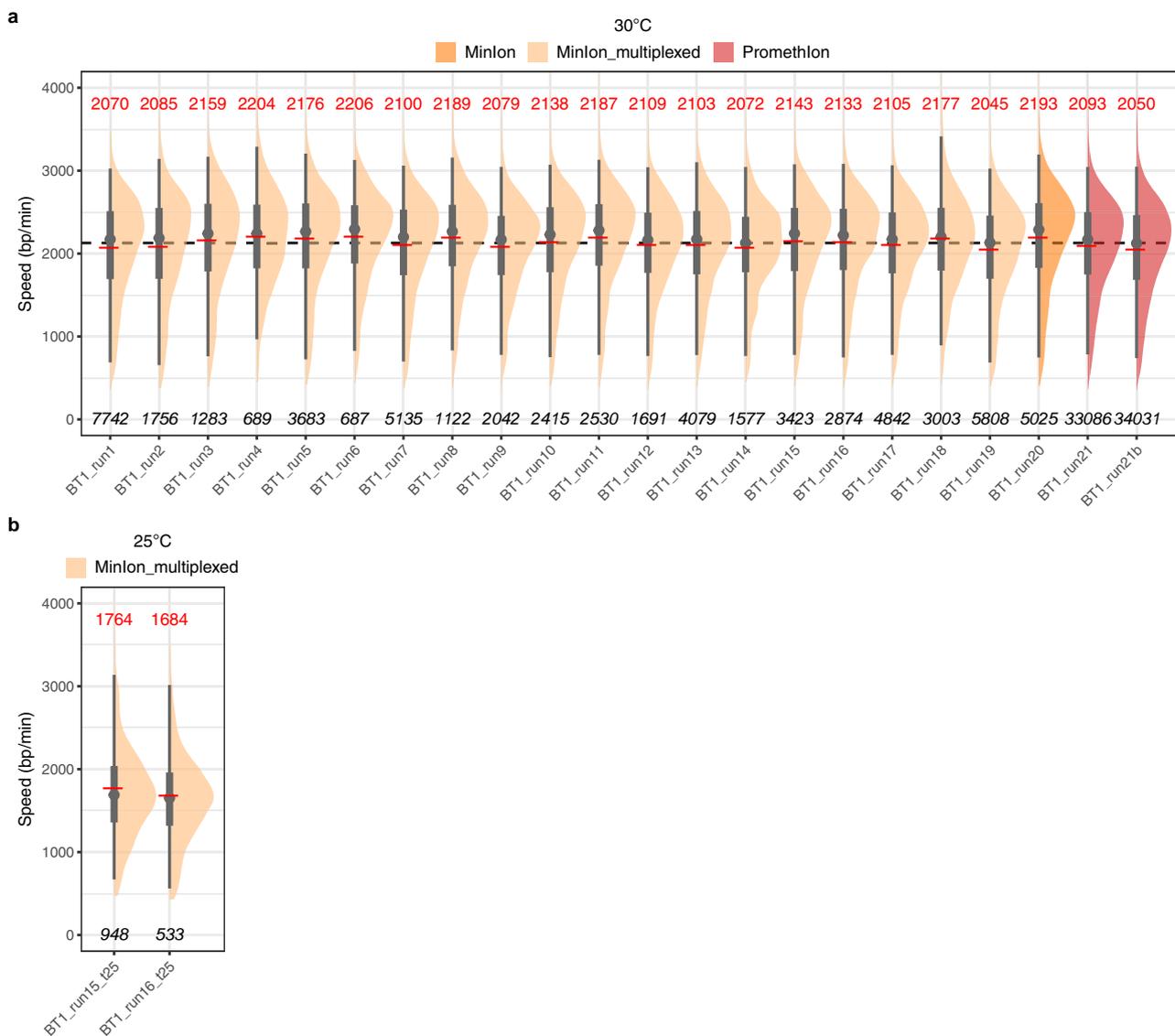
**Validation of fork speed measurement by NFS.** We performed multiple independent pulse-labelling experiments using the

aforementioned conditions (Fig. 2a). Thousands (tens of thousands) of individual measurements were typically collected in a single run using the ONT MinION (PromethION) device. All experiments yielded similar fork speed distributions and mean velocities between 2045 and 2206 bp/min (average of 2128 bp/min, Fig. 2a), emphasizing the reproducibility of our analysis. Above all, these values are in excellent agreement with previous estimates of  $\approx 2$  kb/min in *S. cerevisiae* cells grown at 30 °C<sup>23–26</sup>. NFS also found that the mean fork speed was 1.7 kb/min in cells grown at 25 °C, which again agreed very well with the average progression rate of 1.6 kb/min determined at this temperature in a preceding study<sup>12</sup> (Fig. 2b).

**Using simulated replication forks to evaluate NFS accuracy.** To further gauge NFS, we generated in silico reads with BrdU signals mimicking replication forks travelling at a defined velocity (‘Methods’) and tested NFS ability to recover these velocities. Simulated forks had similar amplitude range, signal dynamics and noise as those from authentic reads of 2-min, 100  $\mu$ M BrdU pulse-labelled BT1 DNA (Supplementary Fig. 6). In addition, they displayed the same speed distribution as that of real sequencing data to allow the comparison of our estimates after NFS processing to a known ground truth under the conditions of an actual experiment (‘Methods’). NFS measurements were extremely precise on simulated reads bearing one or multiple forks in the absence of signal noise (median error of 7 and 30 bp/min and interquartile range (IQR) of 123 and 186 bp/min for single and multiple forks, respectively; Fig. 3a). Importantly, although speed error distributions broadened when NFS was confronted to simulated forks with noise closely resembling experimental data (IQR of 371 and 438 bp/min for single and multiple forks, respectively; Fig. 3a), the median error remained remarkably low ( $-32$  and  $-11$  bp/min for single and multiple forks, respectively; Fig. 3a), demonstrating that the overall evaluation of fork velocity by NFS was extremely accurate. We noted that NFS tended to overestimate and underestimate slow and fast forks, respectively, and that the dispersion of NFS measurements increased with fork speed (Supplementary Fig. 7). However, the median speed error remained below 10% relative to theoretical velocities in the 1000 to 3000 bp/min range, which encompasses the vast majority of physiological fork rates (Supplementary Fig. 7).

**Estimating the true fork speed distribution in yeast.** The distribution of fork velocities determined by NFS in BT1 cells is an approximation of the genuine distribution of individual fork speeds in yeast owing to signal noise and measurement errors. Since we could characterize how NFS responded to fork velocities of known value thanks to simulated reads, we adopted a deconvolution strategy to predict the original fork speed distribution from our  $>125,000$  experimental measurements at 30 °C (‘Methods’). Interestingly, the main peak of our estimate of the true speed distribution consisted of values in the  $2486 \pm 150$  bp/min range that accounted for two-thirds of all fork velocities (Fig. 3b), suggesting a globally low dispersion of individual fork speeds in *S. cerevisiae*. Finally, to evaluate the deconvolution procedure, we generated 100,000 reads bearing forks simulated on the basis of the deconvoluted true fork speed distribution and analysed them with NFS; the obtained distribution closely approximated the experimental one (Fig. 3c), validating our approach.

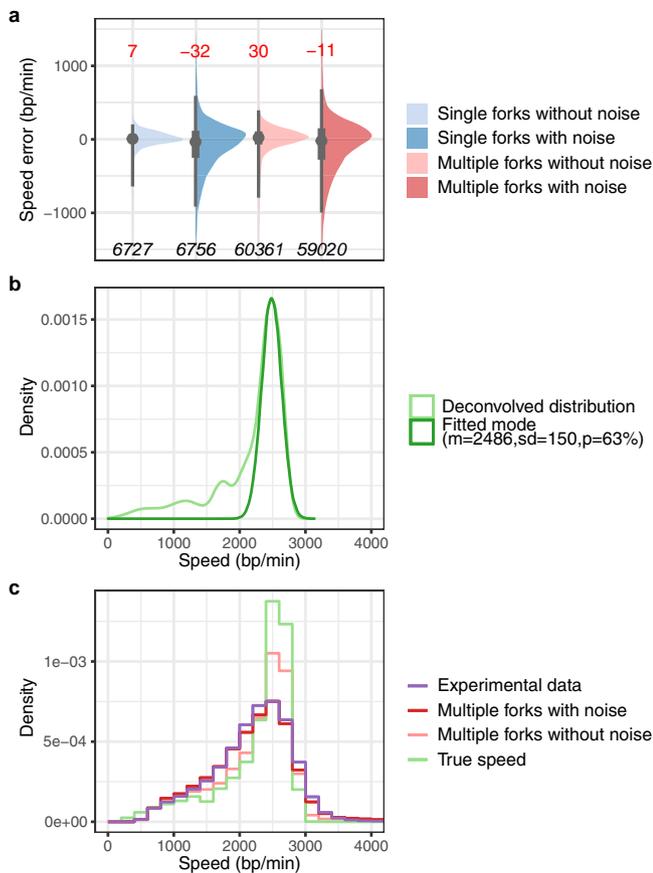
**NFS retrieves expected shifts in fork speed.** We next investigated if NFS was able to detect changes in replication fork velocity in conditions known to alter fork progression (Fig. 4). We first performed pulse-chase experiments in cells exposed to hydroxyurea (HU), a commonly used drug inducing replication fork



**Fig. 2 Measurement of replication fork speed in yeast by NFS. a, b** Replication fork speed in BT1 cells grown at 30 °C (**a**) or 25 °C (**b**). Half-eye plots of individual fork velocities detected by NFS on sequencing reads from MinION, multiplexed MinION or PromethION runs performed on pulse-labelled DNA from independent BT1 cell cultures (the two PromethION samples are technical replicates). Red line, mean speed, value indicated in red on top; grey dot, median speed; thick and thin grey vertical lines, 50 and 95% intervals, respectively; bottom, number of individual fork speed measurements; dotted line in **a**, average fork speed of all samples (2128 bp/min). The name of the sequencing run is indicated below each plot; detailed run information is presented in Supplementary Data 1.

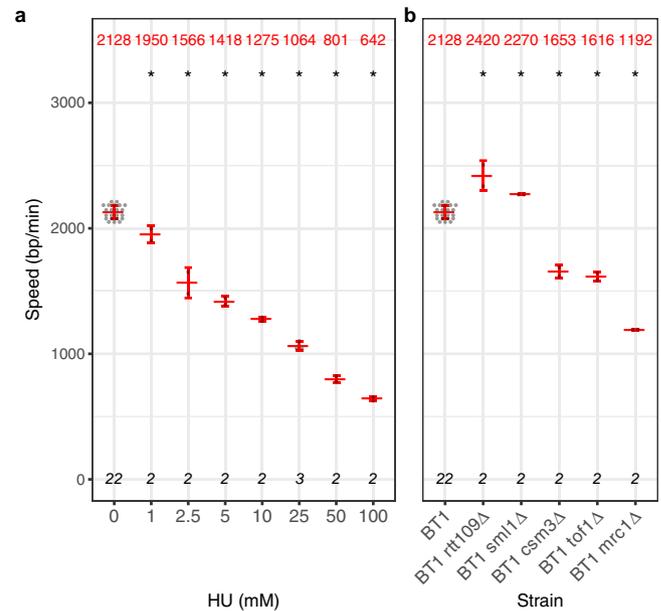
slowdown. As expected, we observed a gradual decrease in fork speed with increasing HU concentrations (Fig. 4a). We could detect fork slowdown induced by doses as low as 1 mM HU, indicating that NFS is a very sensitive tool to reveal replication stress. We then measured fork velocity in cells with mutations impacting replisome progression. We were able to recover the previously reported decrease in fork speed in the absence of Mrc1<sup>25,27–30</sup>, Tof1<sup>28,30</sup> and Csm3<sup>30</sup> (Fig. 4b). Interestingly, fork speed was equally reduced in *tof1Δ* and *csm3Δ* cells, consistent with the fact that these proteins form a complex at the front of the replisome<sup>31,32</sup>. Moreover, fork progression was affected to a lesser extent in those mutants than without Mrc1, as already described<sup>29,30,33</sup>. NFS also detected the acceleration of replication forks observed in the absence of Rtt109 acetyltransferase<sup>26</sup>, as well as in *smi1Δ* mutant exhibiting increased dNTP pools<sup>34</sup> (Fig. 4b). Altogether, these results demonstrate that NFS can precisely quantify fork speed variations both in physiological and perturbed conditions.

**Genome-wide mapping of replication fork progression.** NFS allowed us to generate a genome-wide map of replication fork progression based on individual fork velocities (Fig. 5 and Supplementary Fig. 8). In order to distinguish meaningful changes in local fork speed distribution, the experimental map was superimposed to a randomized map where fork locations remained unchanged but velocity values were shuffled. This showed that fork movement was largely uniform across the yeast genome, in agreement with the results obtained by Sekedat and colleagues<sup>12</sup>, although several loci clearly stood out. We next performed a multi-scale statistical analysis to precisely identify regions of significantly lower or higher fork speed than the bulk genome along with the underlying factors likely responsible for the observed alterations. We detected several regions of fork slowdown but only one locus, located between the *FLO10* and *NFT1* genes on chromosome XI, showed fork acceleration (Fig. 5 and Supplementary Fig. 9). Salient features inside slow regions narrowed down to 1 kb included centromeres, telomeres, tRNA



**Fig. 3 Evaluation of NFS accuracy and estimation of the true fork speed distribution in yeast using simulated replication forks.** **a** Half-eye plots showing the distribution of measurement errors made by NFS on simulated forks. Fork velocities were determined by NFS on simulated reads containing either a single or multiple forks of known speed, with or without noise, and differences between NFS measurements and true speeds were computed (see 'Methods'). Grey dot, median speed error, value indicated in red on top; thick and thin grey vertical lines, 50% and 95% intervals, respectively; bottom, number of measurements. **b** Deconvolved distribution of individual fork speeds in yeast (light green curve) and Gaussian distribution fitting its main peak (green curve). m, mean speed (bp/min); sd, standard deviation (bp/min); p, percentage (correction factor to fit the Gaussian distribution to the deconvolved distribution). See text and 'Methods' for details. **c** Histograms showing fork speed distribution on 100,000 reads with one or multiple forks simulated according to the deconvolved true fork speed distribution (True Speed, light green), fork speed distribution after NFS processing of these reads without (pink) or with (red) noise, and the experimental distribution of >125,000 individual fork velocities from BT1 cells at 30 °C (purple).

genes, the rDNA locus and the HML silent origin cluster (Fig. 5 and Supplementary Fig. 8), all known impediments to replication forks<sup>35,36</sup>. Targeted analysis of forks overlapping these DNA elements confirmed that centromeres and telomeres quite consistently exhibited a reduced fork speed (Fig. 6a and Supplementary Fig. 10) and that fork slowdown at the rDNA locus occurred in the vicinity of the replication fork barrier (RFB), as anticipated (Supplementary Fig. 11). Fork progression was globally slower at tRNA genes having a direction of transcription opposing that of forks (Fig. 6b), in agreement with tRNA genes being polar obstacles to replication<sup>35–37</sup>, although speed was markedly reduced only at a small subset of tRNA genes (Supplementary Fig. 12). We speculate that those exhibit a particularly strong binding of TFIIB transcription factor, considered

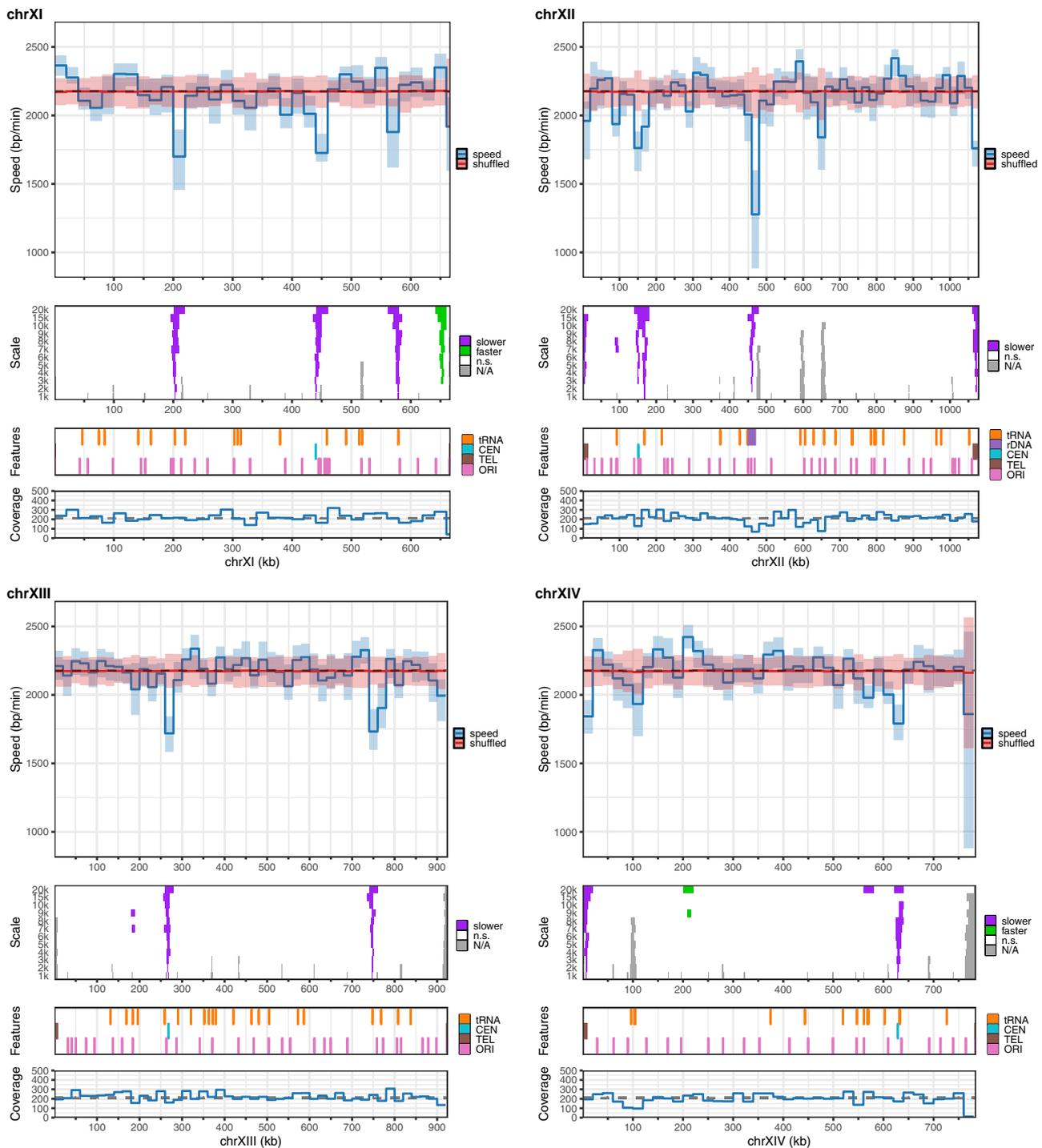


**Fig. 4 Measurement of replication fork speed in yeast by NFS in conditions of altered replisome progression.** **a** Replication fork speed in BT1 cells grown in the presence of increasing concentrations of hydroxyurea (HU). **b** Replication fork speed in mutant strains with altered fork progression. **a, b** Grey dot, mean fork speed of a sample; red centre line, average value of all mean fork speeds of the corresponding category, indicated in red on top; red whiskers, standard deviation; bottom, number of samples. Two-sided contrast comparisons between untreated BT1 cells and each HU concentration in **a** and between BT1 cells and each mutant in **b** are indicated by a star ( $p \leq 0.05$ ) or n.s. (not significant). Statistical analyses are detailed in Supplementary Table 2.

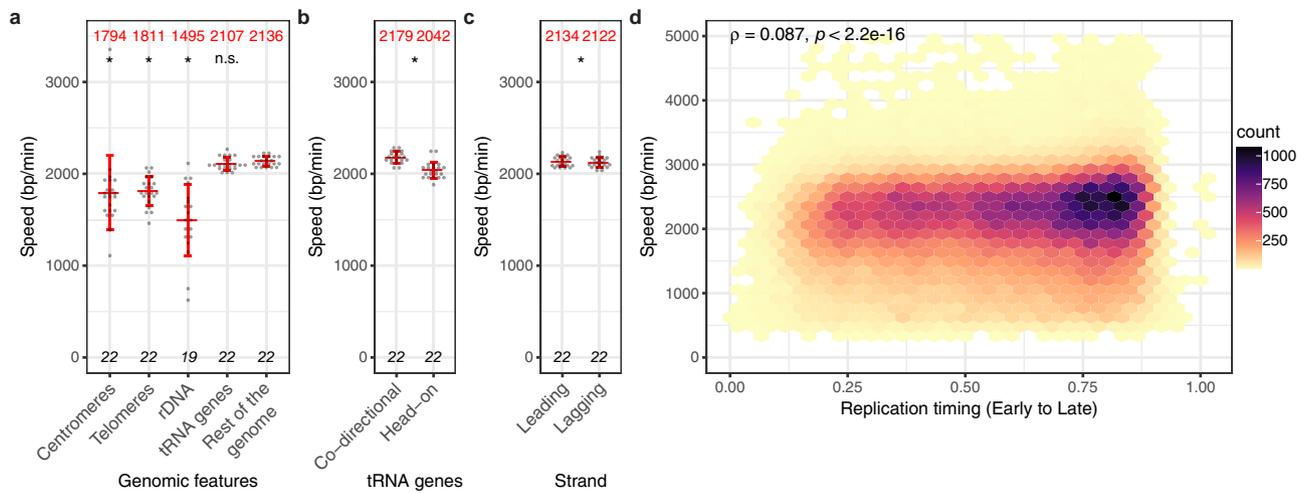
as the main impediment to fork movement at tRNA genes<sup>38</sup>. Thanks to the mapping of oriented tracks to Watson or Crick strands, we could distinguish fork progression on the leading and lagging strands. We found that DNA synthesis proceeds significantly faster on the leading strand (Fig. 6c), although the very weak difference in speed with the lagging strand (12 bp/min) is possibly devoid of biological significance. Finally, as it remains unclear whether fork velocity changes during S phase in eukaryotes<sup>2,3,8,39–41</sup>, we compared the speed of replication forks with the timing of the region they are replicating<sup>42</sup>, noting that forks appear to slightly accelerate during S phase in yeast (Fig. 6d).

## Discussion

Under conditions of BrdU incorporation that do not perturb DNA replication, NFS recovers former estimations of fork speed in yeast in a WT-like growing strain and in mutants known to have altered fork progression, with an excellent reproducibility between biological replicates. This validates our measurement procedure and emphasizes the robustness of our method. Moreover, NFS is remarkably efficient in detecting replication fork slowdown, from the slightest to major decelerations, making it a suitable tool to detect replication stress. NFS accuracy has been further verified on simulated forks of known speed, which also offered the opportunity to recreate the presumed true fork velocity distribution in *S. cerevisiae* from the experimental distribution that is inevitably altered by NFS measurement errors, however small they may be. Our results suggest that the majority of fork speeds are narrowly distributed in yeast (two-thirds within  $2500 \pm 150$  bp/min), in apparent contrast with the large dispersion of individual fork velocities observed by conventional SM



**Fig. 5 Replication fork progression map of yeast chromosomes.** Shown are chromosomes XI to XIV. Panels from top to bottom: (1) median of experimental fork speeds (blue line) with 98% confidence interval of the median (light blue) and median of reshuffled speeds (red line) with 98% confidence interval of the median (light red) computed in 20 kb windows, median fork speed in the whole genome) (dotted line, median fork speed in the whole genome); (2) results of Mann–Whitney–Wilcoxon tests with Holm correction (one-sided) performed along the chromosome to compare the speed distribution in a given window of a given width (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 and 20 kb) to the speed distribution on the whole genome (purple, regions of lower fork speed; green, regions of higher fork speed; white, n.s., not significant; statistical significance was set to  $p < 0.01$ ; grey, N/A, not applicable, regions with no fork); (3) position of selected genomic features (CEN, centromere; TEL, telomeres; ORI, known *S. cerevisiae* replication origins from ref. <sup>21</sup>); (4) coverage of individual replication fork velocities (dotted line, median coverage of the genome). N/A peaks in panel 2 and droughts in coverage in panel 4 essentially map to the position of active origins since fork pairs at initiation sites cannot be resolved by our analysis (see text for details).



**Fig. 6 Analysis of replication fork progression in the yeast genome.** **a** Replication fork speed at selected genomic features. Please note that “telomeres” correspond to subtelomeric sequences and do not comprise the terminal stretch of telomeric repeats. **b** Replication fork speed at tRNA genes with a direction of transcription in the same (Co-directional, 129 genes) or opposite (Head-on, 115 genes) orientation to the main direction of fork progression determined using BT1 RFD profile (see ‘Methods’ and Supplementary Fig. 3). **c** Replication fork speed on the leading and lagging strands. **a–c** Grey dot, mean speed of forks overlapping the corresponding DNA element (**a, b**) or progressing on the leading or lagging strand (**c**) in a sample; red centre line, average speed of all mean fork speeds of the corresponding category, indicated in red on top; red whiskers, standard deviation; bottom, number of samples ( $n_{\text{rDNA}} = 19$  since no fork overlapped the rDNA locus in 3 samples). Two-sided contrast comparisons between a given genomic feature and the rest of the genome in **a**, between co-directional and head-on tRNA genes in **b** and between the leading and lagging strands in **c** are indicated by a star ( $p \leq 0.05$ ) or n.s. (not significant). Statistical analyses are detailed in Supplementary Table 2. **d** Fork speed versus replication timing (data from ref. <sup>42</sup> normalized between 0 and 1 corresponding to the start and end of S phase, respectively) plotted as a 2D density plot using hexagonal bins. Coefficient ( $\rho$ ) and  $p$ -value of Spearman’s correlation test (two-sided) between fork speed and replication timing are indicated.

methods in eukaryotes<sup>2–4</sup>. However, it is important to note that although a 10-fold difference between the fastest and slowest forks is commonly found in standard DNA fibre analysis of a given cell line, most fork rates actually fall within less than twofold of the median velocity<sup>4</sup>. For instance, half of the forks travel within  $\pm 20\%$  of the 1.5 kb/min median speed in chicken DT40 cells<sup>9</sup>. Extreme values may reflect rare events or may be, at least in part, artefactual. For instance, exceptionally high velocities may originate from the fusion of forks emanating from adjacent origins that fired during the labelling pulse, which may go unnoticed even when using a robust analysis methodology<sup>4</sup>. In that regard, the use of a much shorter labelling pulse in our conditions compared to standard SM techniques minimizes the probability that BrdU tracks result from the activity of more than one fork.

In agreement with a low dispersion of fork speeds in *S. cerevisiae*, the mapping of individual fork velocities shows that the rate of fork progression does not significantly change across the yeast genome except at previously known fork pausing sites<sup>35,36</sup>, plus at a single locus situated between *FLO10* and *NFT1* genes where forks appear to move faster for reasons that remain to be determined. We have not attempted here to verify whether low fork speeds around pausing sites result from transient stalling rather than uniformly slow progression, but the former seems likely, especially since fork slowdown at the rDNA locus mapped to the ribosomal RFB pausing site. Our observations of a consistent velocity across the yeast genome support several earlier studies<sup>12,24</sup> and are in line with a very recent nanopore sequencing-based analysis of excised full-length replicons concluding that fork progression is remarkably homogeneous along *S. cerevisiae* chromosomes<sup>43</sup>, but they are at odds with those of Raghuraman and colleagues who interpreted variations in RT profile slopes along different parts of the genome as fluctuations in fork speed<sup>11</sup>. However, it has since been demonstrated that RT slopes depend not only on fork speed but also on the local

proportion of rightward and leftward forks in the cell population<sup>8,44,45</sup>. Our results suggest that RT profile slope variations observed in that pioneering study predominantly if not exclusively stem from variations in the proportion of rightward and leftward forks.

Determining fork speed is mandatory in circumstances known or speculated to influence replication dynamics. Existing methods include DNA fibre analysis<sup>2–4,46</sup>, hydrodynamic techniques<sup>2,3</sup>, slope computation from RT profiles (e.g., refs. <sup>8,11,25,26,47–50</sup>), mathematical modelling of DNA replication (e.g., refs. <sup>24,44</sup>) and time-resolved (i) dense-isotope substitution experiments (for instance, see ref. <sup>51</sup>), (ii) two-dimensional gels<sup>27</sup>, (iii) chromatin immunoprecipitation of replisome components<sup>12,23,26,27</sup> or (iv) isolation of BrdU- or 5-ethynyl-2'-deoxyuridine-labelled DNA<sup>52,53</sup>. All these approaches are demanding and difficult to implement, and none of them allows the mapping of individual fork velocities on entire genomes. Not only does NFS have this ability, but it also combines a straightforward protocol with a ready-to-use analytical pipeline (available on GitHub at <https://github.com/LacroixLaurent/NanoForkSpeed>) enabling an automated measurement of fork speed. NFS should advantageously replace DNA fibre analysis as the “gold standard” method to determine fork velocity as it surpasses it in simplicity, rapidity (estimation of fork speed in 6 days for NFS versus weeks or months for DNA combing or spreading especially when combined with FISH probe detection), number of measurements (thousands of values versus a few tens to a few hundreds) as well as in spatial (precision approaching the nucleotide resolution of sequencing versus precision limited to  $\approx 1$  kb by the resolution of optical microscopes) and temporal (fork speed averaged over 2 min versus typically 20 min) resolutions. Estimation of fork velocity by NFS is certainly applicable to BrdU pulse-labelled DNA from other eukaryotes, notably mammalian cells, although a higher sequencing depth will be required to obtain a comparable coverage. The rapid evolution of nanopore sequencing

technologies offering increased throughputs at reduced costs will contribute to overcome this limitation. Meanwhile, the simultaneous processing of multiplexed samples should facilitate the screening of knockouts libraries, particularly in yeast, to uncover new factors involved in replisome progression and broaden our understanding of eukaryotic DNA replication.

## Methods

**Yeast strains and growth conditions.** All strains used in this study are W303 derivatives and are listed in Supplementary Table 1. Standard yeast genetic techniques and media were used<sup>54</sup>. Cells were grown at 30 °C in YPD medium (MP Biomedicals #114001032) unless stated otherwise. BT1 strain was obtained by integrating at the *ura3-1* locus of W303 cells StuI-linearized pBL-hsvTK<sub>CO</sub>-hENT1<sub>CO</sub> plasmid (Supplementary Table 1), a modified version of the p306-BrdU-Inc vector<sup>17</sup> in which both the *hsvTK* and *hENT1* genes were replaced by versions that had been codon-optimized for expression in yeast. *hsvTK* was codon-optimized for yeast using JCat<sup>55</sup> and synthesized by GeneScript; codon-optimized *hENT1*<sup>19,56</sup> was a generous gift from E. Schwob (IGMM, Montpellier, France). Detailed cloning procedures are available upon request. The BT2 strain was obtained by reintroducing the WT *CDC21* allele encoding the thymidylate synthase into the MCM869 strain<sup>14</sup> by crossing. The BT3 strain was obtained by inserting the regular p306-BrdU-Inc plasmid<sup>17</sup> purchased from Addgene at the *ura3-1* locus of W303 cells.

**Growth curve and doubling time.** Yeasts were grown overnight in YPD (supplemented with 100 µM thymidine for MCM869 cells), diluted in fresh medium (supplemented with 100 µM thymidine for MCM869 cells) at an optical density at 600 nm ( $OD_{600}$ ) ≈ 0.1 and grown for 6 h.  $OD_{600}$  was measured every 2 h. Doubling times (*T*) were estimated on the basis of the growth curves based on the formula  $T = [\text{culture duration} * \log(2)] / [\log(\text{final } OD_{600}) - \log(\text{initial } OD_{600})]$ .

## Analysis of S phase progression by fluorescence-activated cell sorting (FACS).

Exponentially growing cells were synchronized in G1 by addition of 0.2 µM α-factor (Sigma #T6901) for 3 h then washed and resuspended in fresh, prewarmed YPD medium containing 50 µg.mL<sup>-1</sup> pronase (Millipore #53702) to release them into the cell cycle. In the experiment examining the impact of BrdU on S phase progression, BrdU was added 15 min after cell release. Aliquots were taken at regular time intervals and fixed in ethanol. Fixed cells were washed with 50 mM sodium citrate pH 7.4, incubated in sodium citrate buffer supplemented with 0.25 mg.mL<sup>-1</sup> RNase A for 1 h at 50 °C, added with 2 mg.mL<sup>-1</sup> proteinase K and incubated for one additional hour. DNA was counterstained overnight with 1 µM SYTOX Green (Invitrogen #S7020). Samples were analysed using a Beckman Coulter CytoFLEX LX flow cytometer. Data were collected using CytExpert v2.4.0.28 and analysed using FlowJo v10.7.1. Gating strategy is illustrated in Supplementary Fig. 13.

**Rad53 immunoblot analysis.** Total protein extracts were obtained by trichloroacetic acid (TCA) precipitation, separated by SDS-PAGE on a 7.5% gel and transferred to a nitrocellulose membrane. Rad53 immunoblot was performed with rabbit anti-Rad53 antibody at 1:10,000 (Abcam #104232, batch GR3353005-2), using HRP-conjugated anti-rabbit (Promega #W401B) at 1:5000 as secondary antibody. Detection was performed with SuperSignal West Pico (Thermo Scientific #34078) chemiluminescent reagents. ImageQuant LAS 4000 mini (GE Healthcare, software version 1.3) was used for imaging.

**Samples used for neural network training.** Neural network training was performed using nanopore-sequenced genomic DNA displaying variable BrdU substitution rates extracted from the thymidine-auxotroph MCM869 strain. For that purpose, MCM869 cells were grown overnight in synthetic complete medium (Dropout Base medium with Complete Supplement Mixture, MP Biomedicals #114025032 and #114500012) with 100 µM thymidine, washed twice to remove thymidine, transferred at  $OD_{600}$  ≈ 0.1 into fresh synthetic complete medium supplemented with various mixtures of BrdU and thymidine (0:100; 10:90; 20:80; 30:70; 40:60; 50:50; 60:40; 70:30; 80:20; 90:10 and 100:0) and grown for 24 h. Genomic DNA was isolated by zymolyase, RNase A, and proteinase K digestion using Genomic DNA Buffer Set (Qiagen #19060) and Qiagen Genomic-tips 20/G (Qiagen #10223) according to the manufacturer's instructions and subsequently subjected to nanopore sequencing.

**Mass spectrometry.** LC-MS/MS was performed on the 11 samples prepared for neural network training (see above) using a TSQ Quantiva triple quadrupole mass spectrometer (Thermo Scientific) coupled to an UltiMate 3000 XRS HPLC system (Dionex, Thermo Scientific) as described in ref. <sup>13</sup>. Samples were analysed in technical duplicates using the software TraceFinder (Thermo Scientific, version 5.1).

**Pulse-chase labelling.** In most experiments, BT1 cells and its derivatives as well as BT2 and BT3 strains were grown overnight in YPD, diluted in fresh medium at an  $OD_{600}$  ≈ 0.1 and pulsed after 3 doublings ( $OD_{600}$  ≈ 0.8) with 100 µM BrdU for 2 min followed by a 20 min incubation with 1 mM thymidine. For the experiment carried out with different BrdU concentrations, BT1 cells were pulsed in the same conditions as above with either 10, 25, 50, 100, 250, 500, 750 µM or 1 mM BrdU and chased with a ten-fold excess of thymidine. For experiments in the presence of HU, cells were grown for 2 doublings ( $OD_{600}$  ≈ 0.4) and treated for 1 h with HU prior to BrdU pulse-labelling. MCM869 cells were cultured overnight in YPD supplemented with 100 µM thymidine, diluted in fresh YPD with thymidine at an  $OD_{600}$  ≈ 0.1 and grown to  $OD_{600}$  ≈ 0.8 before being washed twice with YPD to remove thymidine, transferred to YPD for 30 min, pulsed with 100 µM BrdU for 2 min and chased with 1 mM thymidine for 45 min. Cells were pelleted after the thymidine chase, washed with water before DNA extraction as in ref. <sup>57</sup>.

**Library preparation and data acquisition.** All samples were sequenced using R9.4.1 chemistry flow cells from Oxford Nanopore Technology (ONT). MinION and GridION sequencing libraries were prepared using ligation sequencing kits SQK-LSK108 or SQK-LSK109 (ONT) in combination with ONT EXP-NBD103 Native Barcoding kit or ONT EXP-NBD104 Native Barcoding Expansion 1–12 pack in case of multiplexing according to ONT protocols with the modifications presented in ref. <sup>57</sup>. PromethION sequencing libraries were prepared using the ultra-long DNA sequencing kit SQK-ULK001 (ONT) according to ONT protocols. Data were acquired using MinKNOW (ONT, MinKNOW Core versions 1.14.1 to 4.4.13, Supplementary Data 1) with default parameters. Demultiplexing of bar-coded nanopore reads was performed as described in ref. <sup>57</sup>.

**Training of the BrdU basecalling model.** Our model was built in three successive steps. First, we modified the architecture of RepNano convolutional neural network<sup>13</sup> in order to obtain a nucleotide resolution for BrdU detection, creating RepNano v2. We next performed a first training of our model with ONT's Taiyaki from the Megalodon program (version 2.2.9 downloaded from <https://nanoporetech.github.io/megalodon/>, Guppy version 4.4.1) using RepNano v2 outputs, then trained our model a second time using the outputs of the first training and adding specific false positive BrdU signals to the training dataset in order to reduce background and false-positive signals.

**RepNano v2 architecture.** We used two convolutional layers with kernel size 7 and filter size 32 with no padding followed by a long short-term memory (LSTM) layer with 32 hidden units that output a vector of the same length as the input, then a convolutional layer with kernel size 1 and filter size 1 and finally an averaging layer outputting the BrdU percent of the segment. The initial length of the input vector was 112 so that after two convolutions without padding the final length was 100. The LSTM layer also produced a 100 × 32 vector and the last convolutional layer produced a 100 × 1 vector that was averaged and compared to the expected output of BrdU content via a logcosh loss. We extracted the vector from the last convolutional layer to basecall BrdU at the nucleotide level.

**RepNano v2 training.** RepNano v2 was trained using nanopore reads of the 11 genomic DNA samples with various BrdU substitution rates described above (BrdU contents measured by mass spectrometry of 0, 9.4, 16.6, 27.9, 35.1, 46.1, 54.8, 59, 72.6, 78.8 and 80.3%). For each of the 11 samples, composed of a mix of substituted and unsubstituted reads (corresponding to parental DNA), 400 reads were used for the learning. To separate substituted reads from unsubstituted ones, we first ran two training cycles with a neural network made of only three convolutional layers (no LSTM and no averaging layers) in order to avoid an overfitting of the data. At the end of the first cycle, the quality of the prediction was good enough to allow the discrimination of substituted from unsubstituted reads for the 40 to 100% BrdU samples. Reads were relabelled as substituted or unsubstituted based on these results, and a second training cycle was run. This allowed the separation of substituted from unsubstituted reads for the 20 and 30% BrdU samples. The final network (RepNano v2 architecture) was trained on this cleaner dataset of 400 reads per sample.

**Megalodon model training.** We performed a first training of our model with ONT's Taiyaki software on 400 reads from each of the 11 BrdU samples described above basecalled with RepNano v2 using a LSTM architecture referred to as mLstm\_cat\_mod\_flipflop with default parameters. Meanwhile, we selected 100 reads from the 0% BrdU sample where forks had been detected, corresponding to false positives (these were mainly found in the rDNA and at the positions of Ty elements). We also selected 100 reads in the same regions from the 100% BrdU sample not to introduce a bias in the neural network at these locations. These reads were added to the outputs of the first model to perform a second training. Finally, the Taiyaki model was converted into a Guppy model.

**BrdU basecalling and read mapping.** Basecalling and read mapping were performed from MinKNOW-generated fast5 files upon sequencing using Megalodon, that combines Guppy basecalling (using the model trained as described above to detect BrdU at the nucleotide level) and read mapping to the reference genome (see

basecalling\_sample.sh script on NanoForkSpeed GitHub repository). This step was carried out using a GPU-enhanced computer to allow fast processing. As the resulting BAM files could be quite big, a custom R<sup>58</sup> script (R version 4.0.3) named basecalling\_sample.r first split BAM files into 50,000 read subfiles. The parsing was then performed in R with the mega\_parsing homemade function. For each read, this function imported from the BAM file the read identity (read\_id), the DNA strand, the chromosome name, the start position and length of the read, the sequence and the Mm and Ml fields which contain the position and output score of detected BrdUs, respectively, using Samtools (version 1.10). The function then allocated a probability of being a BrdU at every position corresponding to a thymidine in the mapped sequence. All discontinuities (gaps) longer than 100 nucleotides introduced during the mapping step were recorded; this information was used after the fork detection procedure to filter out forks overlapping such gaps as discontinuities in the BrdU signal may interfere with fork speed measurement. Data were then smoothed using a combination of a rolling mean on 100 nucleotides and a Gaussian weighted rolling mean on 2.5 kb, with exclusion of the first and last 2.5 kb windows (which removed reads <5 kb from our analysis). This smoothing procedure was designed to allow the piecewise linear simplification (see below) without altering the shape of the signal too much. The output file contained both the raw and the smoothed signal as well as the mapping information (read\_id, chromosome name, start and end positions, strand and gap positions).

**BrdU basecaller comparison.** For each of the 11 genomic DNA samples with different BrdU substitution rates described above, a set of 8000 nanopore reads were basecalled either with Megalodon or with DNAscent v1<sup>15</sup>, DNAscent v2<sup>16</sup>, RepNano<sup>13</sup> transition matrices (RepNano\_TM) and RepNano convolutional neural network (RepNano\_CNN). RepNano\_TM and RepNano\_CNN were used as in ref. 13. For each basecaller, the BrdU content was computed in 1 kb bins for every read. The overall BrdU content of a given DNA sample corresponded to the mean of the BrdU content values per 1 kb window. Means of the 11 samples were plotted against BrdU contents measured by mass spectrometry (MS) in Supplementary Fig. 1a. The proportionality between basecaller and MS estimates was assessed using a linear regression and by computing the mean square error between the observed and ideal lines. Distributions of BrdU content values per 1 kb window for the reads from the 11 samples were represented as normalized densities using the geom\_freqpoly() function of the ggplot2 R package in Supplementary Fig. 1b–f. Although reads used for the analyses shown in Supplementary Fig. 1 came from a different sequencing run than reads used for the training of Megalodon, we cannot completely exclude that Megalodon's performance might benefit from having been trained on DNA originating from the same samples as those used to compare basecallers.

**Fork detection and orientation.** After pulse-labelling, BrdU signals corresponding to ongoing replication forks are visualized on nanopore reads as a steep ascending slope starting from a segment of null BrdU content followed by a shallower decreasing slope; the steep and shallow slopes reflect BrdU incorporation during the pulse and the chase, respectively, and this signal asymmetry allows the determination of the direction of fork progression<sup>13</sup>. Fork detection and orientation were performed with custom R scripts using a piecewise linear simplification approach deriving from the original FORK-seq manuscript<sup>13</sup>. Reads were first converted into a series of segments using the Ramer-Douglas-Peucker algorithm<sup>59,60</sup> (Hausdorff distance epsilon = 0.1, using DouglasPeuckerEpsilon function of the kmlShape R package). Only reads with 3 or more segments were kept as reads with less segments could not form a complete fork. Segments were classified into 4 categories using their slope and mean BrdU signal (B = flat segment with a background BrdU level, A = flat segment with a BrdU level above background, P = segment with a positive slope and N = segment with a negative slope). In order to set the background threshold, the distribution of the mean BrdU signal on 1 kb windows computed for the reads containing at least 3 segments of a given experiment was plotted; in every experiment, DNA replicated before the BrdU pulse (i.e., DNA with a theoretical null BrdU signal) was separated from the rest of the DNA by local minimum near 0.02, which was therefore set as the background threshold (b2a.thr parameter) for all samples.

We were specifically interested in capturing the segment corresponding to BrdU incorporation during the pulse, that is the section between the starting and ending points of the steep slope, in order to extract fork speed defined as the length of DNA replicated during the duration of the pulse (i.e., 2 min) divided by the labelling time. We thus focused on pulse segments for which we could determine the start and end positions. We determined two patterns to identify forks depending on their orientation on the nanopore reads. Rightward forks must be preceded by a B segment, then must contain one or several P segments that may be interrupted by A segments owing to noise and then at least one N segment; they must not be directly followed by a B segment as the BrdU level corresponding to the thymidine chase is above background. We then used a regular expression approach to recognize the “BP(P|A)\*N+” pattern, excluding forks for which the following segment was a B. Leftward forks were identified thanks to a symmetrical pattern “P+(N|A)\*NB”, excluding forks for which the preceding segment was a B. This prevented any overlap between forks detected on a given read (forks must start with a B segment and cannot end on a B segment) and excluded incomplete replication tracks as well as symmetrical signals due to pairs of forks initiated after

the start of the pulse or terminated before the end of the pulse, for which the actual labelling time could not be precisely estimated. Furthermore, in case of multiple replication signals on a single read detected as a succession of leftward and rightward forks, we could determine the position of individual initiation (termination) sites defined as the midpoints between diverging (converging) forks.

Our procedure was performed using the homemade NFSmaster function of the NFS\_function.r script. It creates an output file saved in the .rds format in which the data are organised as a list of 4 elements: (1) a list of tibble (specialised type of data.frame in the R tidyverse) containing (1.1) the reads filtered according to their length and the presence of 3 or more linearized segments and (1.2) the reads with detected forks; (2) a tibble of all the detected forks; (3) a tibble containing the detected initiation (Ini) and termination (Ter) events; (4) a table summary of different metrics of the experiment. Fork detection was performed on the split data corresponding to the basecalling of 50,000 reads and results were merged using the NFS\_merging function of the NFS\_function.r file. Merged files have the same organization with a slightly simplified summary table but do not contain reads without fork; they were used to produce the figures and data discussed in this manuscript.

**Fork speed analysis.** Fork data from every read and experiment were collected in a master data table containing read information (read\_id, chromosome name, read start and end positions, read strand), fork parameters (X0 = start of BrdU incorporation = position of the B/P and N/B transitions for rightward and leftward forks, respectively; X1 = start of the thymidine chase = position of the P(A|N) and P(A|N) transitions for rightward and leftward forks, respectively; speed in bp/min = fork speed averaged over the 2 min of the pulse = absolute value of (X1 - X0)/2; direction = Left or Right; type = leading or lagging; d.Y = BrdU signal amplitude at X1 position minus BrdU signal amplitude at X0 position) and experimental parameters (Exp = name of the experiment; B\_pulse = BrdU concentration in the medium during the pulse (in  $\mu$ M); t\_pulse = duration of the pulse (2 min in all experiments); T\_chase = thymidine concentration in the medium during the chase (in  $\mu$ M); temp = growing temperature (in  $^{\circ}$ C); strain = yeast strain used; mutant = wild-type (WT) or name of the inactivated gene; HU = hydroxyurea concentration in the medium (in mM)). Data were filtered using these parameters to group fork speeds according to the criteria presented in the figures of this manuscript. Detailed information for every sample sequenced in our study are presented in Supplementary Data 1.

Computing fork velocity from reads of increasing minimal length resulted in a negligible rise in the median fork speed estimated by NFS, confirming that a 2 min BrdU pulse duration was adequately suited with respect to read length to detect short as well as long tracks in our experiments; the median length of reads with forks is indicated for each sample in Supplementary Data 1.

**Genomic map of fork speed.** Forks detected in all BT1 biological replicates grown in standard conditions at 30  $^{\circ}$ C were converted into GenomicRanges using the GenomicRanges R package and reduced to their centre. The velocities of these forks were then binned into non-overlapping windows of different width (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 and 20 kb). The number of forks (coverage), the median speed and the 98% confidence interval of the median (using the MedianCI function of the DescTools R package) were computed for every window. In addition, a Mann-Whitney-Wilcoxon test was performed to compare the speed distribution in a given window to the speed distribution on the whole genome. *P*-values for slower and faster speed were corrected for the multiplicity of testing using the Holm method from the p.adjust function, taking into account the number of windows for all scales. Results were saved as bigwig files and used for plots in Fig. 5 and Supplementary Figs. 8, 9 and 11, setting statistical significance to *p* < 0.01. Bins with no fork were tagged N/A. In order to distinguish meaningful changes in local fork velocity distribution, the experimental map was superimposed to a randomized speed map. To generate this map, fork locations were kept unchanged but speed values were randomized 1000 times and binned into 20 kb (Fig. 5 and Supplementary Fig. 8) or 1 kb (Supplementary Figs. 9 and 11) windows. For each bin and each randomization, the median speed was computed, generating 1000 medians of randomized speeds per bin. The overall median and the 1st and 99th percentiles were extracted from these data for every bin and plotted as the median speed and 98% confidence interval of the median to build the shuffled speed map. In Fig. 5 and Supplementary Figs. 8 and 11, genomic feature coordinates are from the UCSC SGD\_other track and replication origins (ORs) are from ref. 21. Oriented genes in Supplementary Fig. 9 are from the Saccharomyces\_cerevisiae.R64-1-1.104.gtf file from Ensembl (see “Computational resources” below for further details).

**Replication fork directionality (RFD) computation.** RFD is calculated for a given position as the difference between the proportions of rightward- (R) and leftward- (L) moving forks (RFD = R - L). RFD profiles from MCM869 FORK-seq and OK-seq data were computed as in ref. 13. BT1 RFD profile from NFS data (i.e., X0 and X1 coordinates of oriented forks) was generated in a similar way with the simpleRFD function of the script helper\_function.r (this function produces 4 bigwig files corresponding to the RFD and the total, leftward and rightward forks coverage) using forks from all BT1 biological replicates grown in standard conditions

at 30 °C. RFD data were binned into non-overlapping 100 nucleotide windows to reduce the size of the plot files. The correlation table in Supplementary Fig. 3b reporting Spearman's pairwise correlation coefficients was produced using the `cor.rfd` function of the `helper_function.r` script and the `ggcorrplot` function of the `ggcorrplot` R package. The `cor.rfd` function rests on the base R `cor` function but works on a coverage type of data (RleList) and excludes positions where N/A are present in at least one of the RFD profiles for which the correlation is computed.

**Analysis of initiation and termination events.** Initiation (*ini*) and termination (*ter*) event positions, defined as the midpoints between diverging and converging forks, respectively, were extracted from all BT1 biological replicates grown in standard conditions at 30 °C and processed as in ref. <sup>13</sup>. Replication origins (ORIs) are from ref. <sup>21</sup>. Distance of each *ini* or *ter* to the centre of the nearest ORI was computed using the `distanceToNearest` function of the `GenomicRanges` R package. Empirical cumulative distribution functions were plotted using the `Ecdf` function of the `Hmisc` R package. Shuffled versions of *ini* and *ter* positions were produced as a control using the custom `shuffleGRen` function of the `helper_function.r` script, which randomizes `GenomicRanges` positions while conserving the number of events *per chromosome*.

**Simulation of BrdU incorporation during DNA replication.** We simulated reads containing a BrdU signal mimicking an elongating fork as well as reads bearing multiple forks. Simulations were performed using Python scripts (Python version 3.6).

**Simulation of BrdU level.** We modelled intracellular BrdU level  $b(t)$  as a function of time  $t$  as (i) zero before the BrdU pulse; (ii) an ascending section modelled by an exponential increase  $As(t) = M(1 - \exp(-\frac{t}{\tau_1}))$  where  $t$  is the time since the start of the pulse,  $M$  the saturating concentration of BrdU and  $\tau_1$  the characteristic time of BrdU intake; (iii) a decreasing section starting at the start of the chase at time  $t_{\text{pulse}}$  (also corresponding to the duration of the pulse) that was modelled by a similar exponential decrease from the value attained at  $t = t_{\text{pulse}}$ :

$Ds(t) = As(t_{\text{pulse}}) + (m - As(t_{\text{pulse}}))(1 - \exp(-\frac{t-t_{\text{pulse}}}{\tau_2}))$ , where  $m$  is the asymptotical BrdU concentration and  $\tau_2$  the characteristic time of BrdU outtake.

**Simulation of a single replication fork.** Given a fork speed  $v$  and a start position used as  $x = 0$ , BrdU time pulse shape was converted into a spatial BrdU incorporation pattern  $BrdU(x)$  using the simple relation:  $BrdU(x) = b(x/v)$ . 10,000 reads containing a single fork were simulated with or without noise (see below) for the analyses presented in Fig. 3a and Supplementary Fig. 7.

**Simulation of multiple replication forks.** To simulate multiple forks, we created a DNA segment of 300 kb in length. We then randomly positioned 6 origins on it to have on average an origin every 50 kb. The firing time of each origin was randomly chosen between 0 and 30 min. A fork speed randomly drawn from a given distribution (see below) was assigned to the whole segment, which was simulated only if firing times, origin positions and fork speed created a set of 12 forks. Next, the moment of the pulse was randomly chosen between the firing time of the earliest origin and the three-fifths of the firing time of the latest origin (this time interval was chosen because it allowed to obtain replication signals in most simulated reads). Finally, to simulate molecules of similar length as experimental reads, the DNA segment was sliced into fragments according to a truncated log-normal distribution (truncations at 5 and 300 kb) with a shape parameter of 0.5 and a scale of 350. 100,000 reads containing one or several forks were simulated with or without noise (see below) for the analyses in Fig. 3a and Supplementary Fig. 7.

**Choice of the simulation parameters.** A replication fork was modelled by two shifted exponentials, each having two parameters ( $M$ ,  $\tau_1$  and  $m$ ,  $\tau_2$  for the ascending and descending parts, respectively; see above). To simulate replication signals resembling those found on real sequencing reads, we used parameters coming from experimental data. To do so, we selected among the >125,000 individual forks detected by NFS in BT1 cells grown at 30 °C those for which the chase was long enough (>2.5 kb), fitted the ≈90,000 forks meeting this criterion using the model described in the 'Simulation of BrdU level' section and the speed measured by NFS, and extracted fork parameters. In order to retain potential correlations between parameters, all four parameters of a randomly chosen experimental fork were assigned to a simulated fork. Fork speed  $v$  was randomly chosen from the deconvolved true fork speed distribution (see below) except when indicated otherwise.

**Adding simulated experimental noise to the signal.** To take into account signal variability between reads, we added to the  $BrdU(x)$  signal a read-dependent offset value ( $O$ ) drawn from a log-normal distribution of shape 1.98, location  $-4.09e-06$  and scale 0.001 fitting the BrdU signal distribution of nanopore reads of DNA with no BrdU labelling.  $O$  was drawn again if its value was >0.2 or <1e-7.  $BrdU(x)+O$  signals reaching values >1 were truncated to 1. We then randomly assigned "Nan" (Not a number) values to ≈77% of the signal so that only thymidine (T) positions had information about the BrdU content. Finally, to mimic *Megalodon*'s output

signal, which is a probability of being a BrdU at each T site peaking at either 0 or 1, we assigned to each T position a value  $B$  drawn from a binomial law  $B(n,p)$  with parameters  $p = BrdU(x)+O$  and  $n$  randomly chosen in [1,2,3];  $B$  was subsequently divided by  $n$  to normalize between 0 and 1. Visual inspection of simulated versus experimental reads confirmed that they were virtually indistinguishable.

**Mean BrdU traces and signal noise analysis for experimental and simulated forks.** The experimental and simulated mean BrdU traces of oriented replication signals in Supplementary Fig. 6f, g were computed from ≈90,000 forks selected among the >125,000 detected by NFS in BT1 cells grown at 30 °C for which the chase was >2.5 kb (please note that leftward forks were reoriented in the rightward direction) and from 10,000 simulated reads containing a single fork, respectively,

with the BrdU signal being smoothed using a 100 bp running average to have information even at non-thymidine sites. In order to characterize signal fluctuations (i.e., signal noise) in experimental and simulated data, the autocorrelation of the mean BrdU signal smoothed at 100 bp minus the signal smoothed at 1000 bp was computed. Subtracting the signal smoothed at 1000 bp removed slow BrdU variations from each track to solely keep local fluctuations that define signal noise.

**Analysis of simulated data.** The smoothing process of simulated reads during the parsing procedure and fork detection with NFS was performed with the same settings as for experimental reads. Speed error was computed by subtracting, for each simulated replication fork, its true speed to the speed estimated by NFS. For simulated reads with multiple replication forks, all forks within a given read had the same speed. To estimate NFS error for different speed categories, true speeds from simulated reads were grouped into bins of 100 bp/min.

**Estimation of the true fork speed distribution in yeast from experimental measurements.** A measured fork speed is an approximation of the true fork speed as (i) BrdU incorporation and detection by nanopore sequencing are subject to noise, and (ii) the measurement by NFS is sensitive to the noise and amplitude of BrdU signals in a complex manner. We used simulated reads to estimate the "transfer function" between the true and the observed fork speed distributions allowing us to deconvolve the measurement errors from the observed fork speed distribution and in turn to estimate the true speed distribution underlying the observed speed distribution. We first built a library of 100,000 reads with known true speeds following a uniform distribution between 50 and 5000 bp/min. We processed each noisy BrdU incorporation profile with NFS and obtained, for each true speed, the response fork speed distribution of the experimental procedure. We then determined the weight of each true speed category required to recover the experimentally observed distribution of >125,000 fork velocities at 30 °C. The weights corresponded to the deconvolved true fork speed distribution. The computational details of this procedure are detailed in the following paragraph.

We considered 44 discretised true fork speed values  $v_k = k^*100$  bp/min for  $1 \leq k \leq 44$ . The 100,000 simulated reads following the uniform speed distribution were then affected to 44 classes if their true speed fell inside a window centred on  $v_k$  and of width 400 bp/min, i.e., reads belonged to more than one class. This strategy was adopted in order to have enough reads in each class. Then, for each class  $k$ , we collected fork speeds estimated by NFS processing, leading to NFS fork speed distribution in response to true speed  $v_k$  that was subsequently fitted with a five-component Gaussian mixture. We then determined the weight  $w_k$  so that the weighted sum of those 44 distributions best fitted the experimental fork speed distribution. The weights  $w_k$  are an estimate of the true fork speed distribution within the 44 classes. This procedure was repeated independently 5 times and the average of the 5 weight sets were used as the deconvolved true fork speed distribution. Finally, this discretised distribution was fitted using a Gaussian mixture with six components to create a continuous true fork speed distribution. To evaluate the deconvolution procedure, we generated a new set of 100,000 reads with multiple forks following the estimated true fork speed distribution and processed them with NFS. The obtained observed fork speed distribution closely resembled the experimental distribution of >125,000 fork velocities, validating our approach. Deconvolution was performed with custom Python scripts using a library from ref. <sup>61</sup>.

The mode of the estimated true speed distribution in Fig. 3b was determined using the `mlv1` function of the `modeest` R package. The major peak of the distribution was fitted with a normal distribution with a mean equal to this mode. Standard deviation and weight of the normal distribution were then adjusted to best fit the peak.

**Fork speed versus replication timing.** Replication timing (RT) data are from ref. <sup>42</sup>. The `liftOver` tool (<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer1/liftOver/>) was used to convert genomic positions from `sacCer1` to the `sacCer3` version of the yeast genome. Data were normalized between 0 and 1 (start and end of S phase, respectively). The mean RT was computed for each fork, then fork speed versus RT was plotted as a 2D density plot with hexagonal bins using the `geom_hex()` function of the `ggplot2` R package with default parameters. Spearman's correlation between fork speed and RT was computed using the `stat_cor()` function of the `ggpubr` R package.

**Statistical analysis.** The R environment v4.0.5 was used for all the analyses. Prior to analysis, fork speed values were averaged for every sample according to the classes of the studied factors, in order to (1) decrease the sensitivity of the tests, which otherwise detect fork speed differences of no practical significance when extremely large number of values are compared and (2) use as a unique source of variation the inter-experiment variations, which correspond to error variations, while fork speed variations have a biological origin. Data were fitted to a linear model that included fork speed as response variable, the factor of interest as predictive variable and the biological replicate/sequencing run as blocking factor. Two by two effect comparisons (two-sided contrast comparisons) were performed with the `emmeans()` function of the `emmeans` R package. When only two values were present in one group (Fig. 4a, b and Supplementary Fig. 12), a `limma` linear fitting and contrast analysis was performed with the `limma` R package. Statistical significance was set to  $p \leq 0.05$ . In each case, type I error was controlled by correcting the  $p$ -values according to the Benjamini & Hochberg method (“BH” option in the `p.adjust()` function of R). Speeds of forks overlapping two different features (e.g., a tRNA and a centromere) or two tRNAs were removed from the analyses presented in Fig. 6a, b and Supplementary Fig. 12 to respect the exclusive factor level requirement of statistical analysis. For the examination of speed at individual tRNAs (Supplementary Fig. 12), means above 5000 bp/min or showing a leverage of 1 in the fitted model were removed from the analysis as they strongly influenced the fitting. This concerned 11 values among the 2984 tRNA x sequencing run levels analysed. All samples corresponded to independent cell cultures except for the two BT1 PromethION runs, which were technical replicates (two sequencing libraries from the genomic DNA of one BT1 cell culture analysed on different PromethION flow cells); the “library plus flow cell” effect was deemed sufficient to consider these as distinct samples. Statistical analysis results are detailed in Supplementary Table 2.

**Visualisation tools.** Plots were made using custom R scripts. For half-eye plots, we used the `stat_slab` and `stat_pointinterval` functions of the `ggdist` R package. `stat_slab` outputs a density distribution of the experimental points; `stat_pointinterval` outputs the median and the 50% and 95% intervals of the data centred on the median.

**Reference genome.** The *sacCer3* genome release (S288C\_reference\_sequence\_R64-2-1) was used as the reference genome. The rDNA locus is composed of two ribosomal DNA units on chromosome XII (chrXII: 451,575–468,931) in this version. As rDNA reads containing more than two repeats tended to incorrectly map and to create artefactual BrdU signals, we added an artificial chromosome with 10 tandem rDNA repeats to the reference genome. The external rDNA repeats were flanked by 10 kb sequences located upstream and downstream of the original rDNA locus to allow the mapping of reads overlapping the first or last rDNA repeat of the rDNA array. In order to build this extra “chromosome”, 8 copies of the second rDNA unit (chrXII: 459,793–468,929 fragment) were inserted between the first and second unit of the original S288C genome, resulting in a 113-kb-long “chromosome” that we named rDNA-10R. Because the proper alignment of reads on tandem repeats the unit of which is of similar length to that of the reads requires specialised procedures, reads mapping to rDNA-10R, which accounted for the most part of rDNA reads, were subsequently excluded from our analyses.

**Computational resources.** Genomic feature coordinates (centromeres, telomeres, HML/HMR loci, rDNA and tRNA genes) were extracted from the UCSC SGD other track (<https://genome.ucsc.edu/cgi-bin/hgTables>) and saved as bed files (script `Script_YeastAnnotation.r`). Please note that “telomere” annotations in the SGD other track encompass both the terminal stretch of telomeric repeats (when present in the *sacCer3* genome) and subtelomeric sequences; moreover, since our smoothing procedure removes the first and last 2.5 kb of each read, telomeric repeats were excluded from our analyses. The UCSC SGD other track was also used to determine the direction of transcription of tRNA genes in Fig. 6b and Supplementary Fig. 12. In this figure, tRNA genes were subsequently categorized as co-directional (CD) or head-on (HO) with respect to replication if their transcription was in the same or opposite orientation to the main direction of fork progression, respectively, which was determined for each tRNA according to its mean RFD value (RFD > 0, forks travel mostly rightward; RFD < 0, forks travel mostly leftward). In Fig. 6a, b and Supplementary Figs. 10 and 12, forks were classified according to their overlap with either a group of features (centromeres, telomeres and tRNA genes) or individual features (rDNA locus and individual centromeres, telomeres and CD/HO-sorted tRNA genes) after fork conversion into `GenomicRanges`. The overlap was tested with the `OverlapsAny` function of the `GenomicRanges` R package (minimal overlap was one nucleotide). Yeast replication origins in Fig. 5 and Supplementary Figs. 3a, c, 8 and 11 are from ref. 21. RT data in Fig. 6d are from ref. 42 (data accessible from NCBI’s Gene Expression Omnibus repository, accession code GSM1036187, GSM1036187\_T7107\_normalised.wig.gz file). Oriented genes in Supplementary Fig. 9 are from the *Saccharomyces cerevisiae*.R64-1-1.104.gtf file downloaded from Ensembl ([http://ftp.ensembl.org/pub/release-104/gtf/saccharomyces\\_cerevisiae/](http://ftp.ensembl.org/pub/release-104/gtf/saccharomyces_cerevisiae/)). DNAscent v1 and v2 were downloaded from <https://github.com/MBoemo/DNAscent>.

**R packages.** R packages used in this study are `kmlShape` version 0.9.5 (<https://CRAN.R-project.org/package=kmlShape>); `DescTools` version 0.99.44 (<https://CRAN.R-project.org/web/packages/DescTools>); `RcppRoll` version 0.3.0 (<https://CRAN.R-project.org/package=RcppRoll>); `Hmisc` version 4.6-0 (<https://CRAN.R-project.org/package=Hmisc>); `tidyverse`<sup>62</sup>; `GenomicRanges`<sup>63</sup>; `rtracklayer`<sup>64</sup>; `BSgenome` version 1.56.0 (<https://bioconductor.org/packages/BSgenome>); `ggdist` version 3.0.1 (<https://mjskay.github.io/ggdist/>); `patchwork` version 1.1.1 (<https://CRAN.R-project.org/package=patchwork>); `ggplot2` version 3.3.5 (<https://CRAN.R-project.org/package=ggplot2>); `ggcorrplot` version 0.1.3 (<https://CRAN.R-project.org/package=ggcorrplot>); `ggpubr` version 0.4.0 (<https://CRAN.R-project.org/package=ggpubr>); `gridExtra` version 2.3 (<https://CRAN.R-project.org/package=gridExtra>); `modeest` version 2.4.0 (<https://CRAN.R-project.org/package=modeest>); `ggprism` version 1.0.3 (<https://CRAN.R-project.org/package=ggprism>); `ggrepel` version 0.9.1 (<https://CRAN.R-project.org/package=ggrepel>); `furrr` version 0.2.3 (<https://CRAN.R-project.org/package=furrr>); `devtools` version 2.4.2 (<https://CRAN.R-project.org/package=devtools>); `emmeans` version 1.5.5-1 (<https://CRAN.R-project.org/package=emmeans>); and `limma` version 3.46.0 (<https://bioconductor.org/packages/release/bioc/html/limma.html>).

**Genomic coordinates.** Coordinates are given according to the *sacCer3* yeast genome assembly.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Nanopore sequencing data generated in this study have been deposited in the ENA database under accession code PRJEB50302. Source data are available at <https://github.com/LacroixLaurent/NanoForkSpeed> and <https://doi.org/10.5281/zenodo.5958270>. Yeast genomic feature coordinates used in this study originate from UCSC SGD other track (<https://genome.ucsc.edu/cgi-bin/hgTables>), oriented genes are from Ensembl database (*Saccharomyces cerevisiae*.R64-1-1.104.gtf file available at [http://ftp.ensembl.org/pub/release-104/gtf/saccharomyces\\_cerevisiae/](http://ftp.ensembl.org/pub/release-104/gtf/saccharomyces_cerevisiae/)), replication origins are from ref. 21 and replication timing data are from ref. 42 (data accessible at NCBI GEO database, accession code GSM1036187). Source data are provided with this paper.

## Code availability

Megalodon-based BrdU basecaller, NFS software and associated R scripts can be accessed at <https://github.com/LacroixLaurent/NanoForkSpeed> and <https://doi.org/10.5281/zenodo.6535390>. Custom R scripts for statistical analyses can be accessed at [https://gitlab.pasteur.fr/gmillot/anova\\_contrasts/-/tree/v7.2.0](https://gitlab.pasteur.fr/gmillot/anova_contrasts/-/tree/v7.2.0). Python scripts are available at <https://github.com/organic-chemistry/simunano>.

Received: 3 March 2022; Accepted: 26 May 2022;

Published online: 08 June 2022

## References

- Gaillard, H., García-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
- Kapp, L. N. & Painter, R. B. in *International Review of Cytology* (eds Bourne, G. H., Danielli, J. F. & Jeon, K. W.) (Academic Press, 1982).
- Liapunova, N. A. in *International Review of Cytology* (eds Jeon, K. W. & Jarvik, J.) (Academic Press, 1994).
- Técher, H. et al. Replication dynamics: biases and robustness of DNA fiber analysis. *J. Mol. Biol.* **425**, 4845–4855 (2013).
- Yurov, Y. B. Rate of DNA replication fork movement within a single mammalian cell. *J. Mol. Biol.* **136**, 339–342 (1980).
- Anglana, M., Apiou, F., Bensimon, A. & Debatisse, M. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* **114**, 385–394 (2003).
- Letessier, A. et al. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**, 120–123 (2011).
- Guilbaud, G. et al. Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLOS Comput. Biol.* **7**, e1002322 (2011).
- Blin, M. et al. Transcription-dependent regulation of replication dynamics modulates genome stability. *Nat. Struct. Mol. Biol.* **26**, 58–66 (2019).
- Mendez-Bermudez, A. et al. Genome-wide control of heterochromatin replication by the telomere capping protein TRF2. *Mol. Cell* **70**, 449–461.e5 (2018).
- Raguraman, M. K. et al. Replication dynamics of the yeast genome. *Science* **294**, 115–121 (2001).

12. Sekedat, M. D. et al. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol. Syst. Biol.* **6**, 353 (2010).
13. Hennion, M. et al. FORK-seq: replication landscape of the *Saccharomyces cerevisiae* genome by nanopore sequencing. *Genome Biol.* **21**, 125 (2020).
14. Ma, E., Hyrien, O. & Goldar, A. Do replication forks control late origin firing in *Saccharomyces cerevisiae*? *Nucleic Acids Res.* **40**, 2010–2019 (2012).
15. Müller, C. A. et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat. Methods* **16**, 429–436 (2019).
16. Boemo, M. A. DNAscent v2: detecting replication forks in nanopore sequencing data with deep learning. *BMC Genomics* **22**, 430 (2021).
17. Viggiani, C. J. & Aparicio, O. M. New vectors for simplified construction of BrdU-Incorporating strains of *Saccharomyces cerevisiae*. *Yeast* **23**, 1045–1051 (2006).
18. Bianco, J. N. et al. Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods* **57**, 149–157 (2012).
19. Magiera, M. M., Gueydon, E. & Schwob, E. DNA replication and spindle checkpoints cooperate during S phase to delay mitosis and preserve genome integrity. *J. Cell Biol.* **204**, 165–175 (2014).
20. Vernis, L., Piskur, J. & Diffley, J. F. X. Reconstitution of an efficient thymidine salvage pathway in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **31**, e120 (2003).
21. Newman, T. J., Mamun, M. A., Nieduszynski, C. A. & Blow, J. J. Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts. *Nucleic Acids Res.* **41**, 9705–9718 (2013).
22. Lengronne, A., Pasero, P., Bensimon, A. & Schwob, E. Monitoring S phase progression globally and locally using BrdU incorporation in TK+ yeast strains. *Nucleic Acids Res.* **29**, 1433–1442 (2001).
23. Ohya, T. et al. The DNA polymerase domain of pole is required for rapid, efficient, and highly accurate chromosomal DNA replication, telomere length maintenance, and normal cell senescence in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **277**, 28099–28108 (2002).
24. Yang, S. C. H., Rhind, N. & Bechhoefer, J. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol. Syst. Biol.* **6**, 404 (2010).
25. Gispan, A., Carmi, M. & Barkai, N. Model-based analysis of DNA replication profiles: predicting replication fork velocity and initiation rate by profiling free-cycling cells. *Genome Res.* **27**, 310–319 (2017).
26. Frenkel, N., Jonas, F., Carmi, M., Yaakov, G. & Barkai, N. Rtt109 slows replication speed by histone N-terminal acetylation. *Genome Res.* **31**, 426–435 (2021).
27. Szyjka, S. J., Viggiani, C. J. & Aparicio, O. M. Mrc1 is required for normal progression of replication forks throughout chromatin in *S. cerevisiae*. *Mol. Cell* **19**, 691–697 (2005).
28. Tourrière, H., Versini, G., Cordon-Preciado, V., Alabert, C. & Pasero, P. Mrc1 and Top1 promote replication fork progression and recovery independently of Rad53. *Mol. Cell* **19**, 699–706 (2005).
29. Hodgson, B., Calzada, A. & Labib, K. Mrc1 and Top1 regulate DNA replication forks in different ways during normal S phase. *Mol. Biol. Cell* **18**, 3894–3902 (2007).
30. Yeeles, J. T. P., Janska, A., Early, A. & Diffley, J. F. X. How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol. Cell* **65**, 105–116 (2017).
31. Eickhoff, P. et al. Molecular basis for ATP-hydrolysis-driven DNA translocation by the CMG helicase of the eukaryotic replisome. *Cell Rep.* **28**, 2673–2688.e8 (2019).
32. Baretic, D. et al. Cryo-EM structure of the fork protection complex bound to CMG at a replication fork. *Mol. Cell* **78**, 926–940.e13 (2020).
33. Lewis, J. S. et al. Single-molecule visualization of *Saccharomyces cerevisiae* leading-strand synthesis reveals dynamic interaction between MTC and the replisome. *Proc. Natl Acad. Sci. USA* **114**, 10630 (2017).
34. Poli, J. et al. dNTP pools determine fork progression and origin usage under replication stress. *EMBO J.* **31**, 883–894 (2012).
35. Mirkin, E. V. & Mirkin, S. M. Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.* **71**, 13–35 (2007).
36. Gadaleta, M. C. & Noguchi, E. Regulation of DNA replication through natural impediments in the eukaryotic genome. *Genes* **8**, 98 (2017).
37. Osmundson, J. S., Kumar, J., Yeung, R. & Smith, D. J. Pif1-family helicases cooperatively suppress widespread replication-fork arrest at tRNA genes. *Nat. Struct. Mol. Biol.* **24**, 162–170 (2017).
38. Yeung, R. & Smith, D. J. Determinants of replication-fork pausing at tRNA genes in *Saccharomyces cerevisiae*. *Genetics* **214**, 825–838 (2020).
39. Rivin, C. J. & Fangman, W. L. Replication fork rate and origin activation during the S phase of *Saccharomyces cerevisiae*. *J. Cell Biol.* **85**, 108–115 (1980).
40. Malinsky, J. et al. The supply of exogenous deoxyribonucleotides accelerates the speed of the replication fork in early S-phase. *J. Cell Sci.* **114**, 747–750 (2001).
41. Takebayashi, S. I. et al. Regulation of replication at the R/G chromosomal band boundary and pericentromeric heterochromatin of mammalian cells. *Exp. Cell Res.* **304**, 162–174 (2005).
42. Müller, C. A. et al. The dynamics of genome replication using deep sequencing. *Nucleic Acids Res.* **42**, e3 (2014).
43. Claussin, C., Vazquez, J. & Whitehouse, I. Single-molecule mapping of replisome progression. *Mol. Cell* **82**, 1372–1382 (2022).
44. de Moura, A. P. S., Retkute, R., Hawkins, M. & Nieduszynski, C. A. Mathematical modelling of whole chromosome replication. *Nucleic Acids Res.* **38**, 5623–5633 (2010).
45. Baker, A. et al. Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput. Biol.* **8**, e1002443 (2012).
46. Norio, P. & Schildkraut, C. L. Visualization of DNA replication on individual Epstein-Barr virus episomes. *Science* **294**, 2361–2364 (2001).
47. Yabuki, N., Terashima, H. & Kitada, K. Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7**, 781–789 (2002).
48. Audit, B. et al. Multiscale analysis of genome-wide replication timing profiles using a wavelet-based signal-processing algorithm. *Nat. Protoc.* **8**, 98–110 (2013).
49. Zhao, P. A., Sasaki, T. & Gilbert, D. M. High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biol.* **21**, 76 (2020).
50. Ziane, R., Camasses, A. & Radman-Livaja, M. The asymmetric distribution of RNA polymerase II and nucleosomes on replicated daughter genomes is caused by differences in replication timing between the lagging and the leading strand. *Genome Res.* **32**, 337–356 (2022).
51. Reynolds, A. E., McCarroll, R. M., Newlon, C. S. & Fangman, W. L. Time of replication of ARS elements along yeast chromosome III. *Mol. Cell. Biol.* **9**, 4488–4494 (1989).
52. Peace, J. M., Villwock, S. K., Zeytounian, J. L., Gan, Y. & Aparicio, O. M. Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome Res.* **26**, 365–375 (2016).
53. Macheret, M. & Halazonetis, T. D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112–116 (2018).
54. Sherman, F. In *Methods in Enzymology* (eds Guthrie, C. & Fink, G. R.) (Academic Press, 2002).
55. Grote, A. et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**, W526–W531 (2005).
56. Talarek, N., Petit, J., Gueydon, E. & Schwob, E. In *DNA Replication: Methods and Protocols* (eds Vengrova, S. & Dalgaard, J.) (Springer New York, 2015).
57. Hennion, M., Theulot, B., Arbona, J. M., Audit, B. & Hyrien, O. In *Yeast Functional Genomics* (ed. Devaux, F.) (Humana Press, 2022).
58. R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2020).
59. Ramer, U. An iterative procedure for the polygonal approximation of plane curves. *Computer Graph. Image Process.* **1**, 244–256 (1972).
60. Douglas, D. H. & Peucker, T. K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: Int. J. Geographic Inf. Geovisualization* **10**, 112–122 (1973).
61. Schreiber, J. Pomegranate: fast and flexible probabilistic modeling in python. *J. Mach. Learn. Res.* **18**, 5992–5997 (2017).
62. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
63. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
64. Lawrence, M., Gentleman, R. & Carey, V. tracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

## Acknowledgements

The authors thank E. Schwob (IGMM, Montpellier, France) and Sophie Loeillet (Institut Curie, Paris, France) for kindly providing plasmids and strains, the Genoscope (Evry, France) for performing MinION and PromethION sequencing, Michael Mitter and Vienna BioCenter Metabolomics facility for mass spectrometry measurements as well as the IBENS IT platform and the BioClust computing cluster (Labex Memolife). The authors also thank Stevann Volant for his comments regarding statistical analyses, Erwan Sallard for his participation in strain construction, all members of the O.H. laboratory for helpful discussions and Nathalie Verin for proofreading the manuscript. This work was supported by the Ligue Nationale Contre le Cancer, the Association pour la Recherche sur le Cancer, the Agence Nationale de la Recherche [ANR-18-CE45-0002 to B.A. and ANR-19-CE12-0028 to O.H.], the Fondation pour la Recherche Médicale [FRM DE1201512344404 to O.H.] and the Cancéropôle Ile-de-France [PLBIO16-302 to O.H.]. This work was partly supported by the Genoscope, the Commissariat à l’Energie Atomique et aux Energies Alternatives (CEA) and France Génomique [ANR-10-INBS-09-08]. B.T. was supported by fellowships from the Ministère de l’Enseignement Supérieur et de la Recherche and the Fondation pour la Recherche Médicale [FRM FDT202106013030].

### Author contributions

B.T., B.L.T., J.P. and F.P. performed the experiments. L.L. developed the NanoForkSpeed pipeline. J.M.A. implemented BrdU detection in ONT's Megalodon program, simulated replication forks and performed the deconvolution analysis. B.T., C.C., S.E. and A.L. performed the nanopore sequencing. L.L., B.L.T., B.T., J.M.A., B.A., E.J., O.H. and M.H. analysed the data. G.A.M., L.L. and B.A. performed the statistical analyses. B.L.T. and L.L. wrote the manuscript, with inputs from the other authors. B.L.T., L.L. and O.H. designed the project and supervised the study.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-31012-0>.

**Correspondence** and requests for materials should be addressed to Laurent Lacroix, Olivier Hyrien or Benoît Le Tallec.

**Peer review information** *Nature Communications* thanks Susan Gerbi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2022

## X. Bibliographie

- Aparicio, J.G., VigAparicio, O.M., Stout, A.M., Bell, S.P., 1999. Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. *Proceedings of the National Academy of Sciences* 96, 9130–9135. <https://doi.org/10.1073/pnas.96.16.9130>
- Balakrishnan, L., Bambara, R.A., 2013. Okazaki fragment metabolism. *Cold Spring Harb Perspect Biol* 5, a010173. <https://doi.org/10.1101/cshperspect.a010173>
- Bell, S.P., Stillman, B., 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* 357, 128–134. <https://doi.org/10.1038/357128a0>
- Bianchi, A., Shore, D., 2007. Early Replication of Short Telomeres in Budding Yeast. *Cell* 128, 1051–1062. <https://doi.org/10.1016/j.cell.2007.01.041>
- Brewer, B.J., Fangman, W.L., 1988. A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell* 55, 637–643. [https://doi.org/10.1016/0092-8674\(88\)90222-x](https://doi.org/10.1016/0092-8674(88)90222-x)
- Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R., Thelander, L., 2003. Survival of DNA Damage in Yeast Directly Depends on Increased dNTP Levels Allowed by Relaxed Feedback Inhibition of Ribonucleotide Reductase. *Cell* 112, 391–401. [https://doi.org/10.1016/S0092-8674\(03\)00075-8](https://doi.org/10.1016/S0092-8674(03)00075-8)
- Cocker, J.H., Piatti, S., Santocanale, C., Nasmyth, K., Diffley, J.F., 1996. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* 379, 180–182. <https://doi.org/10.1038/379180a0>
- Cosgrove, A.J., Nieduszynski, C.A., Donaldson, A.D., 2002. Ku complex controls the replication time of DNA in telomere regions. *Genes Dev.* 16, 2485–2490. <https://doi.org/10.1101/gad.231602>
- Costa, A., Diffley, J.F.X., 2022. The Initiation of Eukaryotic DNA Replication. *Annu. Rev. Biochem.* 91, 107–131. <https://doi.org/10.1146/annurev-biochem-072321-110228>
- Coster, G., Diffley, J.F.X., 2017. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. *Science* 357, 314–318. <https://doi.org/10.1126/science.aan0063>
- Crampton, A., Chang, F., Pappas, D.L., Frisch, R.L., Weinreich, M., 2008. An ARS Element Inhibits DNA Replication through a SIR2-Dependent Mechanism. *Molecular Cell* 30, 156–166. <https://doi.org/10.1016/j.molcel.2008.02.019>
- Czajkowsky, D.M., Liu, J., Hamlin, J.L., Shao, Z., 2008. DNA Combing Reveals Intrinsic Temporal Disorder in the Replication of Yeast Chromosome VI. *Journal of Molecular Biology* 375, 12–19. <https://doi.org/10.1016/j.jmb.2007.10.046>

- Das, S.P., Borrmann, T., Liu, V.W.T., Yang, S.C.-H., Bechhoefer, J., Rhind, N., 2015. Replication timing is regulated by the number of MCMs loaded at origins. *Genome Res.* 25, 1886–1892. <https://doi.org/10.1101/gr.195305.115>
- Davé, A., Cooley, C., Garg, M., Bianchi, A., 2014. Protein Phosphatase 1 Recruitment by Rif1 Regulates DNA Replication Origin Firing by Counteracting DDK Activity. *Cell Reports* 7, 53–61. <https://doi.org/10.1016/j.celrep.2014.02.019>
- Deegan, T.D., Mukherjee, P.P., Fujisawa, R., Polo Rivera, C., Labib, K., 2020. CMG helicase disassembly is controlled by replication fork DNA, replisome components and a ubiquitin threshold. *eLife* 9, e60371. <https://doi.org/10.7554/eLife.60371>
- Dewar, J.M., Budzowska, M., Walter, J.C., 2015. The mechanism of DNA replication termination in vertebrates. *Nature* 525, 345–350. <https://doi.org/10.1038/nature14887>
- Dewar, J.M., Walter, J.C., 2017. Mechanisms of DNA replication termination. *Nat Rev Mol Cell Biol* 18, 507–516. <https://doi.org/10.1038/nrm.2017.42>
- Di Stefano, M., Di Giovanni, F., Pozharskaia, V., Gomar-Alba, M., Baù, D., Carey, L.B., Marti-Renom, M.A., Mendoza, M., 2020. Impact of Chromosome Fusions on 3D Genome Organization and Gene Expression in Budding Yeast. *Genetics* 214, 651–667. <https://doi.org/10.1534/genetics.119.302978>
- Diffley, J.F.X., Cocker, J.H., Dowell, S.J., Rowley, A., 1994. Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell* 78, 303–316. [https://doi.org/10.1016/0092-8674\(94\)90299-2](https://doi.org/10.1016/0092-8674(94)90299-2)
- Doughty, T.W., Yu, R., Chao, L.F.-I., Qin, Z., Siewers, V., Nielsen, J., 2021. A single chromosome strain of *S. cerevisiae* exhibits diminished ethanol metabolism and tolerance. *BMC Genomics* 22, 688. <https://doi.org/10.1186/s12864-021-07947-x>
- Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., Noble, W.S., 2010. A three-dimensional model of the yeast genome. *Nature* 465, 363–367. <https://doi.org/10.1038/nature08973>
- Dukaj, L., Rhind, N., 2021. The capacity of origins to load MCM establishes replication timing patterns. *PLOS Genet.* 17, e1009467. <https://doi.org/10.1371/journal.pgen.1009467>
- Eaton, M.L., Galani, K., Kang, S., Bell, S.P., MacAlpine, D.M., 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* 24, 748–753. <https://doi.org/10.1101/gad.1913210>

Ebrahimi, H., Donaldson, A.D., 2008. Release of yeast telomeres from the nuclear periphery is triggered by replication and maintained by suppression of Ku-mediated anchoring. *Genes Dev.* 22, 3363–3374. <https://doi.org/10.1101/gad.486208>

Ebrahimi, H., Robertson, E.D., Taddei, A., Gasser, S.M., Donaldson, A.D., Hiraga, S., 2010. Early initiation of a replication origin tethered at the nuclear periphery. *J Cell Sci* 123, 1015–1019. <https://doi.org/10.1242/jcs.060392>

Eser, U., Chandler-Brown, D., Ay, F., Straight, A.F., Duan, Z., Noble, W.S., Skotheim, J.M., 2017. Form and function of topologically associating genomic domains in budding yeast. *Proc Natl Acad Sci U S A* 114, E3061–E3070. <https://doi.org/10.1073/pnas.1612256114>

Evryn, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., Speck, C., 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences* 106, 20240–20245. <https://doi.org/10.1073/pnas.0911500106>

Fang, D., Lengronne, A., Shi, D., Forey, R., Skrzypczak, M., Ginalski, K., Yan, C., Wang, X., Cao, Q., Pasero, P., Lou, H., 2017. Dbf4 recruitment by forkhead transcription factors defines an upstream rate-limiting step in determining origin firing timing. *Genes Dev.* 31, 2405–2415. <https://doi.org/10.1101/gad.306571.117>

Ferguson, B.M., Fangman, W.L., 1992. A position effect on the time of replication origin activation in yeast. *Cell* 68, 333–339. [https://doi.org/10.1016/0092-8674\(92\)90474-Q](https://doi.org/10.1016/0092-8674(92)90474-Q)

Friedman, K.L., Brewer, B.J., Fangman, W.L., 1997. Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes to Cells* 2, 667–678. <https://doi.org/10.1046/j.1365-2443.1997.1520350.x>

Friedman, K.L., Diller, J.D., Ferguson, B.M., Nyland, S.V., Brewer, B.J., Fangman, W.L., 1996. Multiple determinants controlling activation of yeast replication origins late in S phase. *Genes Dev.* 10, 1595–1607. <https://doi.org/10.1101/gad.10.13.1595>

Guilliam, T.A., Yeeles, J.T.P., 2020. An updated perspective on the polymerase division of labor during eukaryotic DNA replication. *Critical Reviews in Biochemistry and Molecular Biology* 55, 469–481. <https://doi.org/10.1080/10409238.2020.1811630>

Hafner, L., Lezaja, A., Zhang, X., Lemmens, L., Shyian, M., Albert, B., Follonier, C., Nunes, J.M., Lopes, M., Shore, D., Mattarocci, S., 2018. Rif1 Binding and Control of Chromosome-Internal DNA Replication Origins Is Limited by Telomere Sequestration. *Cell Reports* 23, 983–992. <https://doi.org/10.1016/j.celrep.2018.03.113>

Hafner, L., Shore, D., Mattarocci, S., 2019. ChECing out Rif1 action in freely cycling cells. *Curr Genet* 65, 429–434. <https://doi.org/10.1007/s00294-018-0902-0>

Hardy, C.F., Dryga, O., Seematter, S., Pahl, P.M., Sclafani, R.A., 1997. *mcm5/cdc46-bob1* bypasses the requirement for the S phase activator Cdc7p. *Proc Natl Acad Sci U S A* 94, 3151–3155. <https://doi.org/10.1073/pnas.94.7.3151>

He, Y., Petrie, M.V., Zhang, H., Peace, J.M., Aparicio, O.M., 2022. Rpd3 regulates single-copy origins independently of the rDNA array by opposing Fkh1-mediated origin stimulation. *Proceedings of the National Academy of Sciences* 119, e2212134119. <https://doi.org/10.1073/pnas.2212134119>

Hediger, F., Neumann, F.R., Houwe, G.V., Dubrana, K., Gasser, S.M., 2002. Live Imaging of Telomeres:  $\gamma$ Ku and Sir Proteins Define Redundant Telomere-Anchoring Pathways in Yeast. *Current Biology* 12, 2076–2089. [https://doi.org/10.1016/S0960-9822\(02\)01338-6](https://doi.org/10.1016/S0960-9822(02)01338-6)

Heller, R.C., Kang, S., Lam, W.M., Chen, S., Chan, C.S., Bell, S.P., 2011. Eukaryotic Origin-Dependent DNA Replication In Vitro Reveals Sequential Action of DDK and S-CDK Kinases. *Cell* 146, 80–91. <https://doi.org/10.1016/j.cell.2011.06.012>

Hennion, M., Arbona, J.-M., Lacroix, L., Cruaud, C., Theulot, B., Tallec, B.L., Proux, F., Wu, X., Novikova, E., Engelen, S., Lemainque, A., Audit, B., Hyrien, O., 2020. FORK-seq: replication landscape of the *Saccharomyces cerevisiae* genome by nanopore sequencing. *Genome Biology* 21, 125. <https://doi.org/10.1186/s13059-020-02013-3>

Heun, P., Laroche, T., Shimada, K., Furrer, P., Gasser, S.M., 2001. Chromosome Dynamics in the Yeast Interphase Nucleus. *Science* 294, 2181–2186. <https://doi.org/10.1126/science.1065366>

Hiraga, S., Alvino, G.M., Chang, F., Lian, H., Sridhar, A., Kubota, T., Brewer, B.J., Weinreich, M., Raghuraman, M.K., Donaldson, A.D., 2014. Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Genes Dev.* 28, 372–383. <https://doi.org/10.1101/gad.231258.113>

Hiraga, S., Monerawela, C., Katou, Y., Shaw, S., Clark, K.R., Shirahige, K., Donaldson, A.D., 2018. Budding yeast Rif1 binds to replication origins and protects DNA at blocked replication forks. *EMBO Reports* 19. <https://doi.org/10.15252/embr.201846222>

Hiraga, S., Robertson, E.D., Donaldson, A.D., 2006. The Ctf18 RFC-like complex positions yeast telomeres but does not specify their replication time. *EMBO J* 25, 1505–1514. <https://doi.org/10.1038/sj.emboj.7601038>

Hoggard, T.A., Chang, F., Perry, K.R., Subramanian, S., Kenworthy, J., Chueng, J., Shor, E., Hyland, E.M., Boeke, J.D., Weinreich, M., Fox, C.A., 2018. Yeast heterochromatin regulators Sir2 and Sir3 act directly at euchromatic DNA replication origins. *PLoS Genet* 14, e1007418. <https://doi.org/10.1371/journal.pgen.1007418>

Jackson, A.L., Pahl, P.M., Harrison, K., Rosamond, J., Sclafani, R.A., 1993. Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. *Molecular and Cellular Biology* 13, 2899–2908. <https://doi.org/10.1128/mcb.13.5.2899-2908.1993>

Kamimura, Y., Tak, Y.-S., Sugino, A., Araki, H., 2001. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. *EMBO J.* 20, 2097–2107. <https://doi.org/10.1093/emboj/20.8.2097>

Knott, S.R.V., Peace, J.M., Ostrow, A.Z., Gan, Y., Rex, A.E., Viggiani, C.J., Tavaré, S., Aparicio, O.M., 2012. Forkhead transcription factors establish origin timing and long-range clustering in *S. cerevisiae*. *Cell* 148, 99–111. <https://doi.org/10.1016/j.cell.2011.12.012>

Knott, S.R.V., Viggiani, C.J., Tavaré, S., Aparicio, O.M., 2009. Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in *Saccharomyces cerevisiae*. *Genes Dev.* 23, 1077–1090. <https://doi.org/10.1101/gad.1784309>

Koç, A., Wheeler, L.J., Mathews, C.K., Merrill, G.F., 2004. Hydroxyurea Arrests DNA Replication by a Mechanism That Preserves Basal dNTP Pools \*. *Journal of Biological Chemistry* 279, 223–230. <https://doi.org/10.1074/jbc.M303952200>

Koç, A., Wheeler, L.J., Mathews, C.K., Merrill, G.F., 2003. Replication-independent MCB Gene Induction and Deoxyribonucleotide Accumulation at G1/S in *Saccharomyces cerevisiae* \*. *Journal of Biological Chemistry* 278, 9345–9352. <https://doi.org/10.1074/jbc.M213013200>

Labib, K., 2010. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev* 24, 1208–1219. <https://doi.org/10.1101/gad.1933010>

Lazar-Stefanita, L., Luo, J., Montagne, R., Thierry, A., Sun, X., Mercy, G., Mozziconacci, J., Koszul, R., Boeke, J.D., 2022. Karyotype engineering reveals spatio-temporal control of replication firing and gene contacts. *Cell Genomics* 2, 100163. <https://doi.org/10.1016/j.xgen.2022.100163>

Lazar-Stefanita, L., Scolari, V.F., Mercy, G., Muller, H., Guérin, T.M.M., Thierry, A., Mozziconacci, J., Koszul, R., 2017. Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO Journal* 36, 2684. <https://doi.org/10.15252/emboj.201797342>

- Lengronne, A., Pasero, P., 2020. Sir2 takes affirmative action to ensure equal opportunity in replication origin licensing. *Proceedings of the National Academy of Sciences* 117, 16723–16725. <https://doi.org/10.1073/pnas.2010001117>
- Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Lian, H.-Y., Robertson, E.D., Hiraga, S., Alvino, G.M., Collingwood, D., McCune, H.J., Sridhar, A., Brewer, B.J., Raghuraman, M.K., Donaldson, A.D., 2011. The effect of Ku on telomere replication time is mediated by telomere length but is independent of histone tail acetylation. *MBoC* 22, 1753–1765. <https://doi.org/10.1091/mbc.e10-06-0549>
- Lööke, M., Kristjuhan, K., Kristjuhan, A., 2017. Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques* 62, v. <https://doi.org/10.2144/000114497>
- Lööke, M., Kristjuhan, K., Värvi, S., Kristjuhan, A., 2013. Chromatin-dependent and -independent regulation of DNA replication origin activation in budding yeast. *EMBO Rep* 14, 191–198. <https://doi.org/10.1038/embor.2012.196>
- Lynch, K.L., Alvino, G.M., Kwan, E.X., Brewer, B.J., Raghuraman, M.K., 2019. The effects of manipulating levels of replication initiation factors on origin firing efficiency in yeast. *PLOS Genetics* 15, e1008430. <https://doi.org/10.1371/journal.pgen.1008430>
- Mantiero, D., Mackenzie, A., Donaldson, A., Zegerman, P., 2011. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *The EMBO Journal* 30, 4805–4814. <https://doi.org/10.1038/emboj.2011.404>
- Marahrens, Y., Stillman, B., 1992. A Yeast Chromosomal Origin of DNA Replication Defined by Multiple Functional Elements. *Science* 255, 817–823. <https://doi.org/10.1126/science.1536007>
- Maric, M., Maculins, T., De Piccoli, G., Labib, K., 2014. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* 346, 1253596. <https://doi.org/10.1126/science.1253596>
- Mattarocci, S., Shyian, M., Lemmens, L., Damay, P., Altintas, D.M., Shi, T., Bartholomew, C.R., Thomä, N.H., Hardy, C.F.J., Shore, D., 2014. Rif1 Controls DNA Replication Timing in Yeast through the PP1 Phosphatase Glc7. *Cell Reports* 7, 62–69. <https://doi.org/10.1016/j.celrep.2014.03.010>

McGuffee, S.R., Smith, D.J., Whitehouse, I., 2013. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell* 50, 123–135. <https://doi.org/10.1016/j.molcel.2013.03.004>

Muramatsu, S., Hirai, K., Tak, Y.-S., Kamimura, Y., Araki, H., 2010. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol  $\epsilon$ , and GINS in budding yeast. *Genes Dev* 24, 602–612. <https://doi.org/10.1101/gad.1883410>

Nathanailidou, P., Taraviras, S., Lygerou, Z., 2020. Chromatin and Nuclear Architecture: Shaping DNA Replication in 3D. *Trends in Genetics* 36, 967–980. <https://doi.org/10.1016/j.tig.2020.07.003>

Natsume, T., Müller, C.A., Katou, Y., Retkute, R., Gierliński, M., Araki, H., Blow, J.J., Shirahige, K., Nieduszynski, C.A., Tanaka, T.U., 2013. Kinetochores Coordinate Pericentromeric Cohesion and Early DNA Replication by Cdc7-Dbf4 Kinase Recruitment. *Mol Cell* 50, 661–674. <https://doi.org/10.1016/j.molcel.2013.05.011>

Newlon, C.S., Theis, J.F., 1993. The structure and function of yeast ARS elements. *Current Opinion in Genetics & Development* 3, 752–758. [https://doi.org/10.1016/S0959-437X\(05\)80094-2](https://doi.org/10.1016/S0959-437X(05)80094-2)

Ostrow, A.Z., Kalhor, R., Gan, Y., Villwock, S.K., Linke, C., Barberis, M., Chen, L., Aparicio, O.M., 2017. Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in *S. cerevisiae*. *Proc. Natl. Acad. Sci.* 114, E2411–E2419. <https://doi.org/10.1073/pnas.1612422114>

Ostrow, A.Z., Nellimoottil, T., Knott, S.R.V., Fox, C.A., Tavaré, S., Aparicio, O.M., 2014. Fkh1 and Fkh2 bind multiple chromosomal elements in the *S. cerevisiae* genome with distinct specificities and cell cycle dynamics. *PLoS One* 9, e87647. <https://doi.org/10.1371/journal.pone.0087647>

Pappas, D.L., Frisch, R., Weinreich, M., 2004. The NAD(+)-dependent Sir2p histone deacetylase is a negative regulator of chromosomal DNA replication. *Genes Dev* 18, 769–781. <https://doi.org/10.1101/gad.1173204>

Park, S., Patterson, E.E., Cobb, J., Audhya, A., Gartenberg, M.R., Fox, C.A., 2011. Palmitoylation controls the dynamics of budding-yeast heterochromatin via the telomere-binding protein Rif1. *Proceedings of the National Academy of Sciences* 108, 14572–14577. <https://doi.org/10.1073/pnas.1105262108>

Pasero, P., Bensimon, A., Schwob, E., 2002. Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. *Genes Dev* 16, 2479–2484. <https://doi.org/10.1101/gad.232902>

- Peace, J.M., Ter-Zakarian, A., Aparicio, O.M., 2014. Rif1 Regulates Initiation Timing of Late Replication Origins throughout the *S. cerevisiae* Genome. *PLoS ONE* 9, e98501. <https://doi.org/10.1371/journal.pone.0098501>
- Peace, J.M., Villwock, S.K., Zeytounian, J.L., Gan, Y., Aparicio, O.M., 2016. Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome Res.* 26, 365–375. <https://doi.org/10.1101/gr.196857.115>
- Pichugina, T., Sugawara, T., Kaykov, A., Schierding, W., Masuda, K., Uewaki, J., Grand, R.S., Allison, J.R., Martienssen, R.A., Nurse, P., Ueno, M., O'Sullivan, J.M., 2016. A diffusion model for the coordination of DNA replication in *Schizosaccharomyces pombe*. *Sci Rep* 6, 18757. <https://doi.org/10.1038/srep18757>
- Pohl, T.J., Brewer, B.J., Raghuraman, M.K., 2012. Functional Centromeres Determine the Activation Time of Pericentric Origins of DNA Replication in *Saccharomyces cerevisiae*. *PLOS Genetics* 8, e1002677. <https://doi.org/10.1371/journal.pgen.1002677>
- Raghuraman, M.K., Brewer, B.J., Fangman, W.L., 1997. Cell Cycle-Dependent Establishment of a Late Replication Program. *Science* 276, 806–809. <https://doi.org/10.1126/science.276.5313.806>
- Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., Fangman, W.L., 2001. Replication Dynamics of the Yeast Genome. *Science* 294, 115–121. <https://doi.org/10.1126/science.294.5540.115>
- Rao, H., Stillman, B., 1995. The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2224–2228. <https://doi.org/10.1073/pnas.92.6.2224>
- Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., Diffley, J.F.X., 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139, 719–730. <https://doi.org/10.1016/j.cell.2009.10.015>
- Rhind, N., n.d. DNA replication timing: Biochemical mechanisms and biological significance. *BioEssays* n/a, 2200097. <https://doi.org/10.1002/bies.202200097>
- Schwob, E., Nasmyth, K., 1993. CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in *Saccharomyces cerevisiae*. *Genes Dev.* 7, 1160–1175. <https://doi.org/10.1101/gad.7.7a.1160>

- Shao, Y., Lu, N., Wu, Z., Cai, C., Wang, S., Zhang, L.-L., Zhou, F., Xiao, S., Liu, L., Zeng, X., Zheng, H., Yang, C., Zhao, Z., Zhao, G., Zhou, J.-Q., Xue, X., Qin, Z., 2018. Creating a functional single-chromosome yeast. *Nature* 560, 331–335. <https://doi.org/10.1038/s41586-018-0382-x>
- Sheu, Y.-J., Stillman, B., 2010. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature* 463, 113–117. <https://doi.org/10.1038/nature08647>
- Sheu, Y.-J., Stillman, B., 2006. Cdc7-Dbf4 Phosphorylates MCM Proteins via a Docking Site-Mediated Mechanism to Promote S Phase Progression. *Molecular Cell* 24, 101–113. <https://doi.org/10.1016/j.molcel.2006.07.033>
- Siow, C.C., Nieduszynska, S.R., Müller, C.A., Nieduszynski, C.A., 2012. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Research* 40, D682–D686. <https://doi.org/10.1093/nar/gkr1091>
- Smith, O.K., Kim, R., Fu, H., Martin, M.M., Lin, C.M., Utani, K., Zhang, Y., Marks, A.B., Lalande, M., Chamberlain, S., Libbrecht, M.W., Bouhassira, E.E., Ryan, M.C., Noble, W.S., Aladjem, M.I., 2016. Distinct epigenetic features of differentiation-regulated replication origins. *Epigenetics Chromatin* 9, 18. <https://doi.org/10.1186/s13072-016-0067-3>
- Stevenson, J.B., Gottschling, D.E., 1999. Telomeric chromatin modulates replication timing near chromosome ends. *Genes Dev* 13, 146–151.
- Stinchcomb, D.T., Struhl, K., Davis, R.W., 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature* 282, 39–43. <https://doi.org/10.1038/282039a0>
- Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., Araki, H., 2003. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev* 17, 1153–1165. <https://doi.org/10.1101/gad.1065903>
- Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., Araki, H., 2011. Origin Association of Sld3, Sld7, and Cdc45 Proteins Is a Key Step for Determination of Origin-Firing Timing. *Current Biology* 21, 2055–2063. <https://doi.org/10.1016/j.cub.2011.11.038>
- Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., Araki, H., 2007. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* 445, 328–332. <https://doi.org/10.1038/nature05465>
- Tercero, J.A., Labib, K., Diffley, J.F.X., 2000. DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *The EMBO Journal* 19, 2082–2093. <https://doi.org/10.1093/emboj/19.9.2082>

- Teste, M.-A., Duquenne, M., François, J.M., Parrou, J.-L., 2009. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC Molecular Biol* 10, 99. <https://doi.org/10.1186/1471-2199-10-99>
- Theulot, B., Lacroix, L., Arbona, J.-M., Millot, G.A., Jean, E., Cruaud, C., Pellet, J., Proux, F., Hennion, M., Engelen, S., Lemainque, A., Audit, B., Hyrien, O., Le Tallec, B., 2022. Genome-wide mapping of individual replication fork velocities using nanopore sequencing. *Nat Commun* 13, 3295. <https://doi.org/10.1038/s41467-022-31012-0>
- Tjong, H., Gong, K., Chen, L., Alber, F., 2012. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.* 22, 1295–1305. <https://doi.org/10.1101/gr.129437.111>
- Umek, R.M., Kowalski, D., 1988. The ease of DNA unwinding as a determinant of initiation at yeast replication origins. *Cell* 52, 559–567. [https://doi.org/10.1016/0092-8674\(88\)90469-2](https://doi.org/10.1016/0092-8674(88)90469-2)
- Viggiani, C.J., Aparicio, O.M., 2006. New vectors for simplified construction of BrdU-Incorporating strains of *Saccharomyces cerevisiae*. *Yeast* 23, 1045–1051. <https://doi.org/10.1002/yea.1406>
- Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., Grunstein, M., 2002. Histone acetylation regulates the time of replication origin firing. *Mol. Cell* 10, 1223–1233. [https://doi.org/10.1016/s1097-2765\(02\)00702-5](https://doi.org/10.1016/s1097-2765(02)00702-5)
- Vujcic, M., Miller, C.A., Kowalski, D., 1999. Activation of Silent Replication Origins at Autonomously Replicating Sequence Elements near the HML Locus in Budding Yeast. *Molecular and Cellular Biology* 19, 6098–6109. <https://doi.org/10.1128/MCB.19.9.6098>
- Wilmes, G.M., Bell, S.P., 2002. The B2 element of the *Saccharomyces cerevisiae* ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. *Proc Natl Acad Sci U S A* 99, 101–106. <https://doi.org/10.1073/pnas.012578499>
- Wong, H., Marie-Nelly, H., Herbert, S., Carrivain, P., Blanc, H., Koszul, R., Fabre, E., Zimmer, C., 2012. A Predictive Computational Model of the Dynamic 3D Interphase Yeast Nucleus. *Current Biology* 22, 1881–1890. <https://doi.org/10.1016/j.cub.2012.07.069>
- Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H., Shirahige, K., 1997. The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2, 655–665. <https://doi.org/10.1046/j.1365-2443.1997.1530351.x>

- Yang, S.C.-H., Rhind, N., Bechhoefer, J., 2010. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol Syst Biol* 6, 404. <https://doi.org/10.1038/msb.2010.61>
- Yao, N.Y., O'Donnell, M.E., 2019. Machinery of DNA Replication, in: Roberts, G., Watts, A. (Eds.), *Encyclopedia of Biophysics*. Springer, Berlin, Heidelberg, pp. 1–10. [https://doi.org/10.1007/978-3-642-35943-9\\_435-1](https://doi.org/10.1007/978-3-642-35943-9_435-1)
- Yeeles, J.T.P., Deegan, T.D., Janska, A., Early, A., Diffley, J.F.X., 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* 519, 431–435. <https://doi.org/10.1038/nature14285>
- Yoshida, K., Bacal, J., Desmarais, D., Padioleau, I., Tsaponina, O., Chabes, A., Pantesco, V., Dubois, E., Parrinello, H., Skrzypczak, M., Ginalski, K., Lengronne, A., Pasero, P., 2014. The Histone Deacetylases Sir2 and Rpd3 Act on Ribosomal DNA to Control the Replication Program in Budding Yeast. *Molecular Cell* 54, 691–697. <https://doi.org/10.1016/j.molcel.2014.04.032>
- Zegerman, P., Diffley, J.F.X., 2007. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445, 281–285. <https://doi.org/10.1038/nature05432>
- Zhang, H., Petrie, M.V., He, Y., Peace, J.M., Chiolo, I.E., Aparicio, O.M., 2019. Dynamic relocalization of replication origins by Fkh1 requires execution of DDK function and Cdc45 loading at origins. *Elife* 8. <https://doi.org/10.7554/eLife.45512>

## RÉSUMÉ

---

Chez la levure, le génome s'organise selon la « configuration Rab1 » où les centromères se regroupent et s'attachent au corpuscule polaire, tandis que les télomères se regroupent sur l'enveloppe nucléaire. L'organisation tri-dimensionnelle du génome est supposée essentielle pour l'établissement et l'exécution du programme de réplication, bien qu'aucun lien clair entre ces deux mécanismes ne soit établi. Pour tenter d'élucider ce phénomène, nous nous sommes intéressés à la réplication d'une souche de levure à un chromosome, ne possédant plus qu'un unique centromère et deux télomères, ce qui élimine la configuration Rab1 et affecte le repliement tridimensionnel du génome (Shao et al, 2018). Nous avons utilisé NanoForkSpeed (NFS, Theulot et al, 2022), une méthode basée sur le séquençage Nanopore, qui permet de détecter, d'orienter, et de mesurer la vitesse de fourches de réplication après marquage au BrdU, pour comparer les profils de réplication de la souche à un chromosome et de son homologue sauvage. Nous avons mis en évidence l'inactivation d'une majorité d'origines juxta-centromériques, et un effet sur la direction des fourches de réplication et l'efficacité des origines au voisinage des fusions chromosomiques. En plus de ces effets attendus, nous avons également détecté des changements d'efficacité de plusieurs origines situées en d'autres points du génome, distants des télomères et centromères, suggérant un impact de l'organisation tridimensionnelle du génome sur l'activité des origines de réplication.

## MOTS CLÉS

---

Réplication de l'ADN, Organisation tridimensionnelle, Levures, Centromères, Télomères

## ABSTRACT

---

In yeast, centromeres cluster close to the spindle pole body and telomeres position away at the nuclear periphery, the so called "Rab1 configuration". It is often postulated that the three-dimensional organization of the genome is essential for the establishment and execution of its replication program. To address this question, we have investigated DNA replication in a single-chromosome yeast strain, in which all but two telomeres and one centromere have been eliminated, which strongly affects the 3D-folding of the genome and eliminates the Rab1 conformation (Shao et al, 2018). We have used NanoForkSpeed (NFS, Theulot et al, 2022), a nanopore sequencing method for detection, orientation and speed measurement of BrdU-labelled replication forks, to compare replication profiles in this strain and its wild-type counterpart. We observed the inactivation of some, but not all, origins next to deleted centromeres, and changes in fork direction and origin efficiency at chromosome fusions. In addition to these expected results, we also detected changes in origin efficiency at multiple sites through the genome, away from centromeres and telomeres, suggesting that perturbing the 3D folding of the genome can affect origin firing efficiency.

## KEYWORDS

---

DNA replication, tridimensional organization, Yeast, Centromeres, Telomeres