



HAL
open science

Spécialisation de Ku80c dans le couplage entre coupure et réparation de l'ADN lors des réarrangements programmés du génome chez *Paramecium tetraurelia*

Arthur Abello

► **To cite this version:**

Arthur Abello. Spécialisation de Ku80c dans le couplage entre coupure et réparation de l'ADN lors des réarrangements programmés du génome chez *Paramecium tetraurelia*. Biologie moléculaire. Université Paris Saclay (COMUE), 2019. Français. NNT : 2019SACLS083 . tel-02612222

HAL Id: tel-02612222

<https://theses.hal.science/tel-02612222>

Submitted on 19 May 2020

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.

Spécialisation de Ku80c dans le couplage entre coupure et réparation de l'ADN lors des réarrangements programmés du génomme chez *Paramecium tetraurelia*

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°577 Structure et Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat : Sciences de la vie et de la santé

Thèse présentée et soutenue à Gif-sur-Yvette, le 29 mars 2019, par

ARTHUR ABELLO

Composition du Jury :

Cécile Fairhead Professeur, Université Paris-Sud	Présidente
Jean-Pierre De Villartay Directeur de Recherche, INSERM	Rapporteur
Claude Bruand Directeur de Recherche, INRA	Rapporteur
Sébastien Britton Chargé de Recherche, IPBS	Examineur
Éric Meyer Directeur de Recherche, IBENS	Examineur
Mireille Bétermier Directrice de Recherche, I2BC	Directeur de thèse
Julien Bischerour Chargé de Recherche, I2BC	Co-directeur de thèse

REMERCIEMENTS

Je tiens tout d'abord à remercier Madame Mireille Bétermier et Monsieur Julien Bischerour, qui m'ont encadré tout au long de ma thèse. Mireille, pour avoir toujours été là, humainement comme scientifiquement. Ces quelques mots auront bien du mal à rendre hommage à ton exceptionnelle bonté qui m'aura été d'un grand secours au travers des nombreuses épreuves du doctorat. Le seul regret que je garderai est qu'à partir de maintenant, je serai forcément déçu de mes supérieurs car aucun ne pourra t'arriver à la cheville ! Julien, travailler à tes côtés m'a permis d'apprendre d'un des scientifiques les plus brillants que je connaisse. C'est grâce à tes nombreux conseils que j'ai pu avancer sur ce projet de recherche. Je dois cependant noter qu'il y a des progrès à faire du côté de Wii Tennis !

J'adresse tous mes remerciements à Messieurs Jean-Pierre de Villartay et Claude Bruand, qui m'ont fait l'honneur d'être rapporteurs de cette thèse. Je remercie également l'ensemble du jury avec qui j'ai pu aborder de très nombreux aspects de mes recherches lors de la soutenance. Je remercie ainsi Messieurs Sébastien Britton et Éric Meyer. Enfin, je remercie tout particulièrement Madame Cécile Fairhead d'avoir présidé le jury, lui permettant ainsi d'être présente à la conclusion de mon cursus universitaire, là où elle était également présente depuis mes premiers cours de génétique en licence.

Je remercie évidemment Madame Vinciane Régnier qui depuis mon premier passage dans l'équipe m'a guidé sur d'innombrables aspects et avec qui j'ai pu débattre de la potentielle mobilité de certain foci ! Ton soutien humain, scientifique et technique a été déterminant pour mon parcours au laboratoire. Je remercie également l'ensemble des membres de l'équipe des Réarrangements Programmés du Génome, présents et passés, Marc, Nathalie, Olivier, Linda, Eva et Loïc pour l'ensemble des questions quotidiennes dont je les accablais et du soutien moral quand un caryonide décidait de ne pas savoir passer l'autogamie. Je souhaite bon courage à Coralie et Mélanie pour les nombreuses injections qu'elles vont encore devoir affronter.

Je remercie Romain Le Bars avec qui, en plus des parties de basket, nous avons pu développer un outil d'automatisation d'analyse des images ; me sauvant de nombreuses heures et m'évitant ainsi des crampes à la main.

Pour m'avoir apporté un regard extérieur sur mes travaux tout au long de ma thèse, je remercie Jean-Baptiste Charbonnier et Karine Dubrana, membres de mon comité de thèse.

Une thèse ne serait pas une thèse sans avoir d'autres thésards à qui se confier sur nos malheurs. Pour cela, je remercie Vincent et l'ensemble des membres de YourI2BC qui ont pu alléger ces moments lors des ~~très sérieuses~~ *beer-sessions*.

Je remercie également le « couloir paramécie » pour tous ces instants partagés. Mes remerciements vont aussi vers les équipes paramécies parisiennes, dirigées par Éric Meyer et Sandra Duharcourt, pour les nombreuses discussions.

En s'éloignant du laboratoire, mes remerciements commenceront évidemment par Agnès, sans qui rien n'aurait été possible et qui a su être un pilier, me permettant d'avancer à chaque instant. Puis-je un jour te rendre la pareille, mon ADMV.

Je remercie enfin ma famille et mes Zinzins.

TABLE DES MATIÈRES

TABLE DES MATIÈRES	v
LISTE DES FIGURES	viii
A Introduction	1
I LES CASSURES DOUBLE-BRIN DE L'ADN	3
I.1 LA RECOMBINAISON HOMOLOGUE	4
I.1.1 Acteurs et mécanismes associés	6
I.1.2 La RH lors de CDB programmées : le cas de la méiose	9
I.1.3 Les voies dérivées de la RH	12
I.1.4 Les erreurs de la RH	12
I.2 LA VOIE DU <i>Non Homologous End Joining</i>	14
I.2.1 Mécanisme général	14
I.2.2 Les acteurs essentiels du NHEJ	16
I.2.2.1 Ku	16
I.2.2.2 DNA-PKcs	18
I.2.2.3 Le complexe de ligation	18
I.2.3 Les acteurs secondaires	22
I.2.4 Un rôle des ARN dans le NHEJ?	23
I.2.5 Conséquences du NHEJ sur les cassures accidentelles	24
I.2.6 Évolution de Ku	24
I.2.7 CDB programmées et NHEJ, exemple de la recombinaison V(D)J	26
I.2.8 Voie alternative de réparation des CDB	30
I.2.9 Choix de la voie de réparation des CDB	30
II LA PARAMÉCIE, MODÈLE D'ÉTUDE DU NHEJ LORS DE CDB PROGRAMMÉES	35
II.1 CARACTÉRISTIQUES GÉNÉRALES DE <i>Paramecium tetraurelia</i>	35
II.1.1 Cycles de vie de la paramécie	36
II.1.1.1 Cycle Végétatif	36
II.1.1.2 Cycles sexuels	37
II.1.2 Les duplications globales du génome	40
II.1.3 Organisation générale du génome de Paramécie	40
II.2 LES RÉARRANGEMENTS PROGRAMMÉS DU GÉNOME CHEZ LA PARAMÉCIE	41
II.2.1 Les réarrangements précis : élimination des IES	41
II.2.1.1 Clivage des IES et réparation des cassures introduites	43
II.2.1.2 Couplage entre coupure et réparation de l'ADN, le cas de Ku chez la paramécie	51
II.2.1.3 Reconnaissance des IES	56
II.2.2 Les réarrangements imprécis	60

II.2.3	Chez les autres ciliés	61
II.2.3.1	<i>Tetrahymena thermophila</i>	61
II.2.3.2	<i>Oxytricha trifallax</i>	62
III	PROJET DE THÈSE	65
B	Résultats	69
IV	LOCALISATION DES PROTÉINES PGM ET KU LORS DES RÉARRANGEMENTS PROGRAMMÉS	71
IV.1	LOCALISATION DE PGM	71
IV.1.1	Dynamique de la localisation nucléaire de Pgm	71
IV.1.2	Effet de l’extinction de Ku sur la localisation de Pgm	74
IV.1.2.1	Localisation nucléaire de Pgm en ARNi <i>KU</i>	74
IV.1.2.2	Quantité de protéine Pgm en ARNi <i>KU</i>	76
IV.1.2.3	Stabilité nucléaire de Pgm en ARNi <i>KU</i>	77
IV.1.2.4	Effet global d’une ARNi <i>KU</i> sur la rétention des IES	77
IV.1.2.5	Effets de l’inactivation d’autres gènes du NHEJ sur Pgm	79
IV.2	LOCALISATION DE KU80C	82
IV.3	ETUDE DE KU70 ENDOGÈNE PAR UN ANTICORPS α -KU70	86
IV.4	LES FOCI KU ET PGM, STRUCTURES FONCTIONNELLES?	87
IV.5	RÉSULTATS PRINCIPAUX DU CHAPITRE	91
V	SPÉCIALISATION FONCTIONNELLE DE KU80C	93
V.1	ANALYSE <i>in silico</i> DE L’ÉVOLUTION DES PROTÉINES KU80 DE CILIÉS	93
V.2	COMPLÉMENTATION FONCTIONNELLE D’UNE ARNi CONTRE <i>KU80c</i> PAR FLAG-KU80A	95
V.2.1	Création des transgènes nécessaires pour la complémentation	95
V.2.2	Flag-Ku80a ne complémente pas une déplétion de Ku80c	96
V.2.3	Flag-Ku80a provoque la rétention d’IES et donc une inhibition des coupures	98
V.2.4	Flag-Ku80c n’est pas détectable en IF sans pré-extraction	99
V.2.5	GFP-Ku80a se localise dans les ébauches mais n’y est pas stablement accroché	100
V.2.6	Pgm est localisé plus stablement dans les ébauches en présence de Flag-Ku80c qu’en présence de Flag-Ku80a	102
V.3	RECHERCHE D’UN SOUS-DOMAINES RESPONSABLE DE LA SPÉCIALISATION DE KU80C	104
V.3.1	Le domaine α/β de Ku80c est responsable de la spécialisation de la protéine pour les RPG	106
V.3.2	Le domaine central en tonneau β , piliers et pont est suffisant pour l’interaction avec Pgm	106
V.3.3	Colocalisation des foci Flag-Ku80ca et Pgm	108
V.3.4	Résultats principaux du chapitre	108
C	Discussion	111
VI	CARACTÉRISATION DU COUPLAGE ENTRE KU ET PGM	113
VI.1	L’IMPORTANCE DES INTERRELATIONS DE LOCALISATION	113
VI.1.1	Réinterprétation des résultats de MARMIGNON ET AL. (2014)	113
VI.1.2	La localisation nucléaire stable de Pgm dépend fortement de Ku	115

VI.1.3	DNA-PKcs permet-il le recyclage de Pgm ?	116
VI.1.4	Flag-Ku8oc ne dépend majoritairement pas de Pgm pour se localiser stablement dans l'ébauche	117
VI.1.5	Les foci Ku et Pgm	118
VII	SPÉCIALISATION DE KU8OC DANS LES RPG	119
VII.1	KU80A EST INCAPABLE DE REMPLACER KU8OC	119
VII.1.1	Le domaine α/β de Ku8oc est au cœur de la spécialisation	120
VII.2	DÉTAIL DE L'INTERACTION ENTRE KU8OC ET PGM	120
VII.2.1	L'interaction Ku/Pgm est-elle essentielle lors des RPG?	121
VII.3	MODÈLE DU COUPLAGE ENTRE COUPURE ET RÉPARATION DE L'ADN LORS DES RPG CHEZ <i>P. tetraurelia</i>	121
VIII	L'IMPORTANCE DU COUPLAGE	125
D	Matériels et Méthodes	129
E	Annexes	137
F	Article en préparation	149
	BIBLIOGRAPHIE	197

LISTE DES FIGURES

I.1	Dommmages de l'ADN et mécanismes de réparation associés	4
I.2	Mécanismes de la RH	5
I.3	Protéines de la RH	7
I.4	La RH en méiose	10
I.5	Stabilisation du chiasma par la cohésine	11
I.6	Les erreurs de la RH	13
I.7	Réparation d'une CDB par le NHEJ	15
I.8	Structure de Ku	17
I.9	Structure de DNA-PKcs	19
I.10	Structure de Lig4	20
I.11	Structure du filament XRCC4/XLF	21
I.12	Artemis, nucléase du NHEJ	22
I.13	Les polymérasés du NHEJ	23
I.14	Le chromothripsis	25
I.15	Diversité des Ku	25
I.16	La recombinaison V(D)J	27
I.17	Etapes de la recombinaison V(D)J	29
I.18	Le MMEJ, mécanisme alternatif de réparation des CDB	31
I.19	Choix de la voie de réparation de CDB	32
II.1	Paramécie	36
II.2	Cycle végétatif de paramécie	37
II.3	L'autogamie	38
II.4	La conjugaison	39
II.5	Les IES	41
II.6	Répartition des tailles d'IES et séquence aux bornes	42
II.7	Excision des IES	44
II.8	Pgm et les PgmL	46
II.9	Localisation nucléaire de Pgm et validation de la spécificité de l'anticorps	47
II.10	Délocalisation de Pgm en ARNi <i>PGML</i>	49
II.11	Les Ku de paramécie	52
II.12	Localisation nucléaire de Ku	53
II.13	Effets d'une ARNi <i>KU</i>	54
II.14	Rétention d'au moins 3 IES en ARNi <i>KU8oc</i>	55
II.15	Interaction entre Ku et Pgm	55
II.16	Modèle d'excision des IES	57
II.17	Scanning ARN	58
II.18	Modèle de reconnaissance des IES par Pgm	60
II.19	Les éliminations imprécises	60
II.20	Les RPG chez <i>Oxytricha trifallax</i>	63

III.1	Ku80a peut-il remplacer Ku80c?	66
IV.1	Taille des ébauches en conjugaison en fonction de l'avancement du cycle sexuel	73
IV.2	Dynamique de la localisation nucléaire de Pgm en conjugaison en fonction de l'avancement du cycle sexuel	73
IV.3	IF sans pré-extraction de Pgm au cours de l'autogamie en ARNi PGM ou KU	75
IV.4	WB de Pgm au cours de l'autogamie en ARNi PGM ou KU	76
IV.5	IF avec pré-extraction de Pgm au cours de l'autogamie en ARNi PGM ou KU	78
IV.6	Score de rétention des IES dans des ARNi KU80c	79
IV.7	IF avec pré-extraction de Pgm au cours de l'autogamie en ARNi DNA-PKcs ou LIG4	81
IV.8	Test d'anticorps α -Ku80c	83
IV.9	Transgène FLAG-KU80c et protéine associée théorique	83
IV.10	IF en pré-extraction de Flag-Ku80c au cours de l'autogamie en ARNi PGM, KU70 ou XRCC4	84
IV.11	WB de Ku70 au cours de l'autogamie dans différentes conditions d'ARNi .	86
IV.12	IF de Ku70 au cours de l'autogamie dans différentes conditions d'ARNi . .	88
IV.13	Disparition des foci Flag-Ku80c en ARNi PGM	89
IV.14	Colocalisation de Ku70 et Pgm en conjugaison et colocalisation de Ku70 et Flag-Ku80c en autogamie	90
IV.15	Résumé des résultats de localisation et stabilité nucléaire de Pgm et Flag-Ku80c en fonction des différentes ARNi	91
V.1	Datation des événements de WGD	93
V.2	Phylogénie des Ku80 chez les ciliés	94
V.3	Transgène FLAG-KU80c résistant à l'ARNi, transgène FLAG-KU80a et protéines associées théoriques	96
V.4	Complémentations de la déplétion de Ku80c par Flag-Ku80c ou Flag-Ku80a en fonction du CPGH	96
V.5	PCR autour d'IES lors de complémentations	98
V.6	IF α -Flag et α -Pgm, avec ou sans pré-extraction, sur des cellules injectées FLAG-KU80c, en déplétion du KU80c endogène	99
V.7	Localisation des constructions GFP-Ku80 avec ou sans pré-extraction dans des paramécies autogames à T5	101
V.8	Localisation des constructions Flag en pré-extraction et effet sur la stabilité de Pgm	103
V.9	Positions spécifiques de Ku80c marquées sur la structure	104
V.10	Transgènes FLAG-KU80ca et FLAG-KU80ac et protéines associées théoriques	105
V.11	Complémentations de la déplétion de Ku80c par Flag-Ku80ca ou Flag-Ku80ac en fonction du CPGH	105
V.12	Localisation des constructions Flag chimériques en pré-extraction et effet sur la stabilité de Pgm	107
V.13	Co-immunoprécipitation de sous-domaines de Ku80c avec Pgm	109
V.14	Colocalisation de Flag-Ku80ca avec Pgm en IF avec pré-extraction	110
VI.1	Localisation de Pgm-GFP en absence de Ku d'après MARMIGNON ET AL. (2014)	114
VII.1	Modèle du couplage entre coupure et réparation de l'ADN lors des RPG chez <i>P. tetraurelia</i>	122

VIII.1	Domaines de Tpb1	126
VIII.2	Formule de calcul du CPGH et emplacement des amorces pour l'amplification de <i>KU8oc</i> et <i>KU8oa</i>	132
VIII.3	Exemple d'organisation de fichiers avant analyse par la macro MicMacEpiV21	134
VIII.4	Exemple des cartes des ROI après analyse par la macro MicMacEpiV21 . .	135
VIII.5	Exemple de graphique de la fluorescence Pgm en fonction de la taille de l'ébauche en autogamie contrôlée après analyse par la macro MicMacEpiV21	136
6	Alignement des protéines Ku80 de <i>P. tetraurelia</i> d'après MARMIGNON ET AL. (2014)	139

Première partie

Introduction

Chapitre I

Les cassures double-brin de l'ADN

Origines et conséquences La molécule d'Acide DésoxyriboNucléique (ADN) est une des molécules biologiques les plus étudiées. Elle donne les plans de fabrication des protéines, outils de la cellule, et est transmise à la génération suivante. Elle est donc le support de l'information génétique. Cette macromolécule est composée de deux polymères de nucléotides. Chaque nucléotide est composé d'une base A, T, C ou G, d'un sucre et d'un à trois groupements phosphate. Les nucléotides sont associés au sein d'un polymère par une liaison phosphodiester dans le sens 5' vers 3'. Les deux polymères sont eux associés par des liaisons hydrogènes entre les bases, de manière anti-parallèle et formant ainsi les deux brins de la célèbre double-hélice (WATSON AND CRICK 1953). L'ADN peut former de très longues molécules (plusieurs centimètres voire mètres) et associé à des protéines forme les chromosomes. Les gènes sont les unités permettant la transcription des Acides RiboNucléiques (ARN), utilisés lors de la traduction pour la synthèse des protéines.

L'ADN est exposé à de nombreux stress (réplicatif, oxydatif, ionisant, ...) qui provoquent une grande variété de dommages accidentels (**Figure I.1**). L'ADN étant un élément essentiel de la cellule, ces dommages peuvent donc avoir de très graves conséquences en compromettant l'intégrité de l'information génétique. Des mécanismes de réparation ont donc évolué pour assurer la protection de l'ADN et se sont spécialisés dans certains types de dommages.

L'un de ces dommages est la Cassure Double-Brin (CDB), évènement où les deux brins de l'ADN subissent chacun la rupture d'une liaison phosphodiester. Cette cassure crée donc deux extrémités libres d'ADN, sensibles aux exonucléases. La cassure peut de

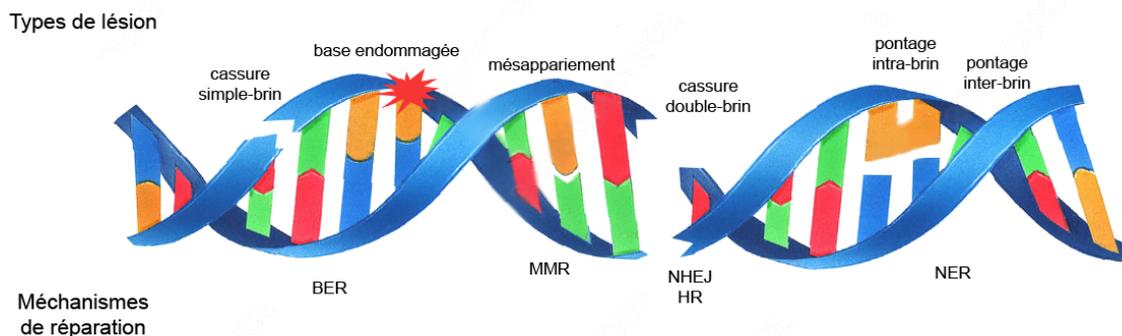


FIGURE I.1 – Dommages de l'ADN et mécanismes de réparation associés

Les cassures simple-brin et les bases endommagées sont réparées par le BER (*Base Excision Repair*). Les mésappariements par le MMR (*MisMatch Repair*). Les cassures double-brin par le NHEJ (*Non Homologous End Joining*) et le RH (*Recombinaison Homologue*). Les pontages intra et inter-brin par le NER (*Nucleotide Excision Repair*).

plus avoir lieu dans un gène ou une séquence régulatrice, perturbant la synthèse des protéines. Dans le cas où une réparation a lieu, des erreurs peuvent être introduites dans la séquence et avoir de possibles effets cancérogènes par exemple. Enfin, la cassure non réparée provoque une fragmentation du chromosome, pouvant aboutir à une mauvaise ségrégation chromosomique lors de la prochaine division cellulaire et ainsi provoquer une anomalie partielle de la ploïdie. Chez les bactéries, les chromosomes sont circulaires et une CDB non réparée provoque une linéarisation du chromosome létale pour la cellule. Il est donc essentiel de réparer le plus efficacement les CDB.

Les principaux mécanismes responsables de la réparation des CDB sont la Recombinaison Homologue (RH) et la Jonction d'Extrémités Non Homologues (NHEJ). Ils réparent les CDB accidentelles, mais également des CDB dites programmées. Différents exemples seront étudiés par la suite. De plus, le choix entre ces deux voies peut être régulé au cours du cycle cellulaire ou en fonction du type de cassure.

I.1 LA RECOMBINAISON HOMOLOGUE

La RH, très bien conservée dans l'ensemble du vivant, est un mécanisme de réparation généralement fidèle. Cette voie utilise l'information génétique présente sur un ADN homologue pour reconstituer le locus endommagé (**Figure I.2**).

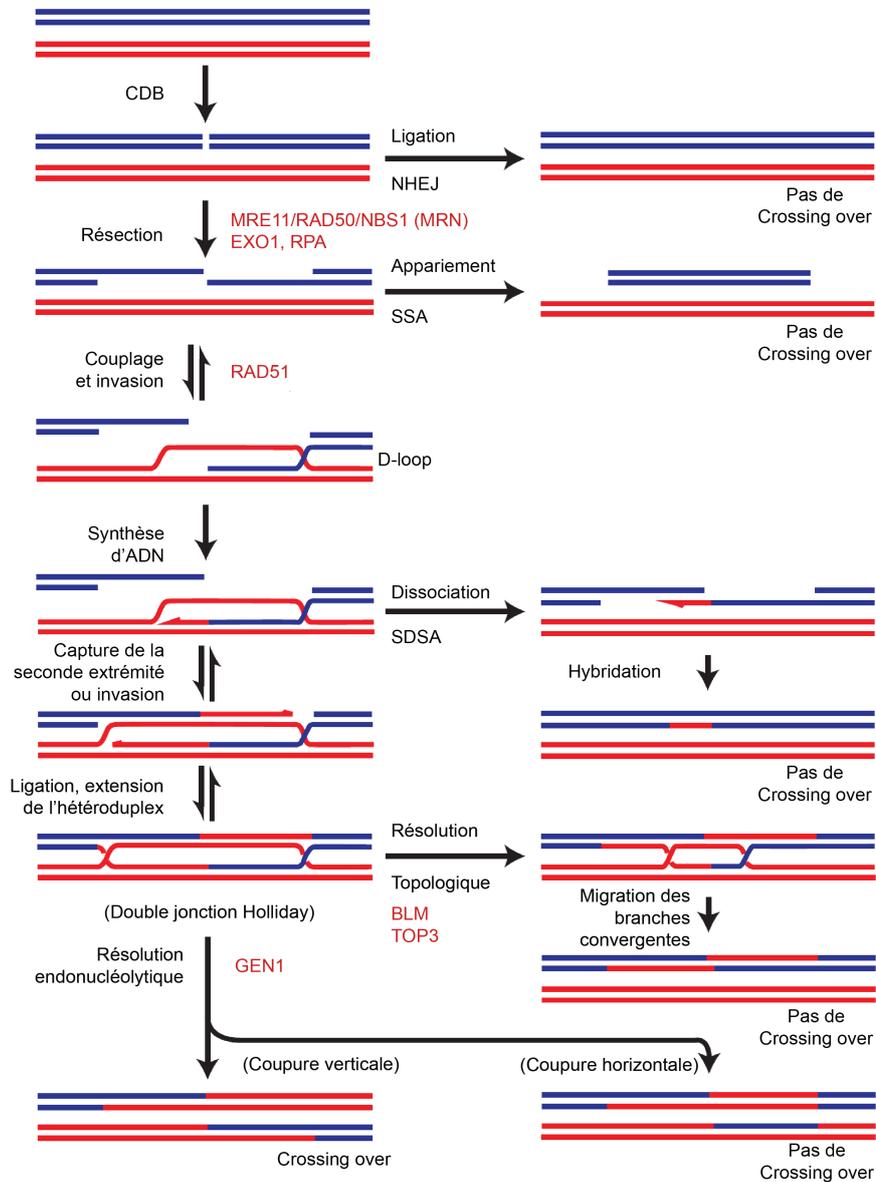


FIGURE I.2 – Mécanismes de la RH

Etats et conformations de l'ADN lors des différentes étapes de la RH. Des mécanismes de réparation tiers, tels que le NHEJ, le SSA ou le SDSA peuvent interférer avec la RH à différentes étapes. Ces mécanismes seront abordés plus loin. Adaptée de KOWALCZYKOWSKI (2015).

Pour ce faire, une résection d'ADN au niveau de la cassure se produit, dans le sens 5' vers 3'. Cette résection va produire un ADN simple brin (ADN sb) avec une extrémité 3'.

Cet ADN sb va ensuite envahir l'ADN homologue et s'associer à son brin complémentaire par homologie de séquence. L'ADN homologue est utilisé comme matrice pour resynthétiser la zone endommagée par la cassure, permettant ainsi de restaurer l'information génétique perdue. La seconde extrémité d'ADN est ensuite soit capturée par la D-loop soit réalise une invasion indépendante.

Une fois la polymérisation finie sur les deux extrémités, des structures intermédiaires présentes dans la voie classique, les jonctions de Holliday, sont résolues, complétant ainsi la réparation de la cassure. Les jonctions de Holliday peuvent être résolues de deux manières différentes : soit par une résolution topologique, nécessitant la migration des branches convergentes et l'action d'une topoisomérase soit par un clivage endonucléolytique introduisant une cassure de l'ADN. Un échange de bras de chromosome est possible entre les deux homologues suite à cette seconde résolution, résultant en un réarrangement chromosomique massif, communément appelé Crossing Over (CO) (SYMINGTON 2016).

Ce mécanisme de réparation est présent dans l'ensemble du vivant et les protéines sont très conservées, au moins au niveau de leur structure. Les protéines décrites ci-dessous sont les protéines humaines et ne représentent qu'une partie des protéines essentielles à la RH.

I.1.1 Acteurs et mécanismes associés

Le complexe MRN Ce complexe de trois protéines, Mre11, Rad50 et Nbs1 joue un rôle crucial pour la reconnaissance de la CDB et sa réparation. MRN est l'un des premiers complexes à s'associer à la CDB et entame la résection nécessaire à la RH (**Figure I.3**).

Mre11 possède une activité endonucléase et exonucléase 3'-5'. Des inhibiteurs spécifiques de chaque fonction ont pu être développés et ont mis en évidence les rôles distincts des deux activités nucléases (SHIBATA ET AL. 2014).

Rad50 est une protéine capable d'interaction avec l'ADN. Apparentée aux protéines SMC (cohésines), elle permet la juxtaposition des molécules d'ADN engagées dans la RH. En interagissant et en hydrolysant un ATP, Rad50 influence l'activité nucléase de Mre11.

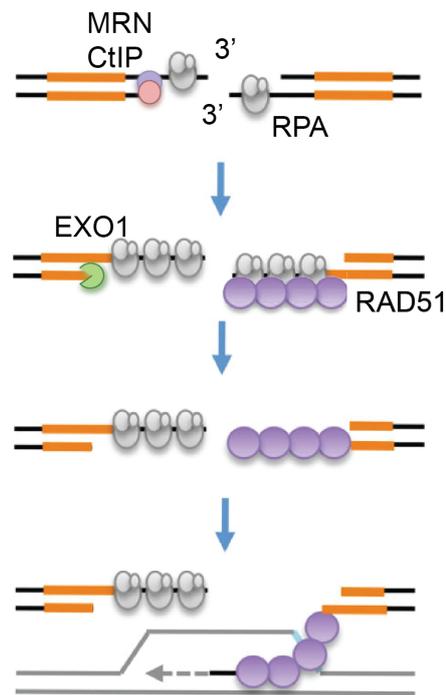


FIGURE I.3 – *Protéines de la RH*

Le complexe MRN/CtIP reconnaît la CDB et par une activité nucléase produit un ADN sb 3' sortant. L'exonucléase Exo1 prend le relai pour une résection étendue. L'ADN sb est d'abord protégé par RPA puis par Rad51 qui va permettre la recherche et l'invasion de l'ADN homologue. Adaptée de SYMINGTON (2016).

Nbs1 pourrait jouer un rôle de chaperonne. Elle permet la localisation nucléaire de Mre11 et présente des motifs d'interaction protéine/protéine.

L'activité exonucléase 3'-5' de Mre11 est à première vue étonnante, du fait de la résection 5'-3' observée *in vivo*. Au lieu de procéder à la maturation de la cassure directement sur l'extrémité, le complexe MRN peut réaliser une cassure simple-brin à distance du site de coupure, puis procéder à une dégradation de 3' vers 5', donc vers la cassure (voir **Figure I.4** pour exemple). Ce mécanisme permet à MRN de dégager les extrémités 3' d'ADN sb

CtIP La protéine CtIP a de multiples rôles. Elle interagit avec le complexe MRN et est un facteur dans l'activation de la résection en permettant de diriger la réparation de la CDB vers la RH (SARTORI ET AL. 2007). Des études *in vitro* ont dévoilé une activité endonucléase de CtIP (WANG ET AL. 2014, MAKHARASHVILI ET AL. 2014) mais ces travaux n'ont pas pu être reproduits par d'autres groupes (ANDRES ET AL. 2015). Un approfondissement du travail sur CtIP est donc nécessaire pour clarifier sa fonction biochimique.

Exo1 La résection créée par Mre11 n'est pas suffisante pour permettre l'invasion de l'ADN homologue par l'ADN sb car elle nécessite une résection d'au moins 500 nucléotides (MUNDIA ET AL. 2014). Une exonucléase 5'-3' avec une plus grande processivité entre donc en action. Il s'agit d'Exo1. Cette nucléase peut produire des extensions d'ADN sb longues de plusieurs kb (HODGSON ET AL. 2011).

RPA L'ADN sb ainsi créé est pris en charge par RPA. Cette protéine permet de protéger cette forme fragile de l'ADN et évite la formation de structure secondaire. En effet, un ADN sb a la capacité de former des structures secondaires s'il possède au moins deux séquences complémentaires. RPA permet donc de garder une structure linéaire, quelle que soit la séquence de l'ADN sb. RPA joue également un rôle essentiel de signalisation et de recrutement des protéines de réparation (BHAT AND CORTEZ 2018).

Rad51 RPA est ensuite remplacé par Rad51, une autre protéine permettant la protection de l'ADN sb, mais qui permet de plus la recherche du duplexe d'ADN homologue et une fois localisé, son invasion. La structure à trois brins ainsi formée, appelée *D-loop*, va

permettre l'initiation de la réplication. La seconde extrémité double brin de la cassure aura subi les mêmes maturations que la première et aura soit envahi de manière indépendante sa région homologue, soit aura été capturée par la *D-loop* (capture secondaire) (**Figure I.2**). Les polymérisations et ligations achevées, deux jonctions de Holliday sont créées. Ces jonctions lient très fortement les deux ADN double-brin (ADN db) entre eux. La RH utilise deux mécanismes pour résoudre ces jonctions :

- Un mécanisme de résolution topologique, nécessitant l'hélicase Blm et l'introduction d'une CDB par une topoisomérase de type II, Top3. Cette résolution permet d'empêcher l'apparition de CO.

- Un mécanisme de résolution nucléolytique, nécessitant une résolvasse, Gen1, qui va résoudre verticalement ou horizontalement la jonction. Une résolution horizontale va, comme pour la résolution topologique, garder la continuité des chromosomes, alors qu'une résolution verticale va provoquer l'échange complet des séquences entre les deux chromosomes, soit un CO.

La RH est donc un mécanisme de réparation des CDB essentiel à l'ensemble du domaine du vivant qui permet de préserver l'intégrité de l'ADN et ainsi d'assurer une bonne transmission de l'information génétique.

I.1.2 La RH lors de CDB programmées : le cas de la méiose

Malgré la dangerosité des CDB pour la cellule, certains mécanismes requièrent l'introduction de ces coupures. C'est le cas de la méiose.

Lors d'une division mitotique, les deux chromatides sœurs d'un chromosome sont séparées. Pour pouvoir les ségréger dans deux cellules différentes, il faut au préalable les associer fortement l'une à l'autre. Ainsi, les mécanismes mitotiques pourront identifier un chromosome à deux chromatides et donc les séparer sans erreur. C'est le centromère qui joue ce rôle de lien solide entre chromatides.

A l'inverse de la mitose, la première division de méiose consiste à séparer les chromosomes homologues. Ils ne possèdent pas entre eux le lien qu'est le centromère, mais la nécessité de les faire migrer sans erreur est tout aussi présente. Le mécanisme qui a été sélectionné est celui de l'introduction de CDB par Spo11 (DE MASSY 2013).

Spo11 est une endonucléase apparentée aux topoisomérases de type II. Elle introduit

des CDB lors de la prophase de méiose I et reste covalamment associée à l'ADN. Les CDB sont prises en charge par MRN et CtIP. Les activités endonucléase et exonucléase 3'-5' de Mre11 permettent de libérer Spo11 de la cassure et de procéder à la maturation de l'ADN (**Figure I.4**) (NEALE ET AL. 2005).

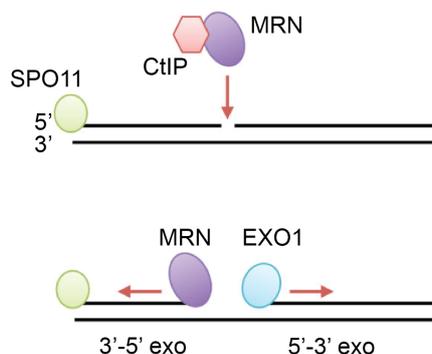


FIGURE I.4 – *La RH en méiose*

Spo11 introduit une CDB lors de la méiose et se retrouve lié covalamment à l'ADN, bloquant la réparation. Mre11 au sein du complexe MRN/CtIP introduit une cassure simple-brin puis dégrade de 3' vers 5' l'ADN, libérant le fragment lié à Spo11. Exo1 dégrade l'ADN de 5' vers 3' afin de compléter la maturation de l'ADN pour la RH. Adaptée de SYMINGTON (2016).

La cassure est ensuite réparée par la RH classique avec l'intervention d'acteurs spécifiques. En plus de la recombinaise Rad51, la protéine Dmc1 est exprimée spécifiquement pendant la méiose (CRICKARD AND GREENE 2018). Dmc1 est une recombinaise, homologue de Rad51. Les deux protéines proviennent d'une duplication de gène datant de l'apparition des eucaryotes (LIN ET AL. 2006). Dmc1 s'est depuis spécialisée dans la RH lors de la méiose et Rad51 intervient dans la voie de RH générale. Une des différences entre ces deux protéines est la capacité de Dmc1 à tolérer plus de mésappariements lors de la recherche de brin matrice que Rad51. Cette différence permet donc à Dmc1 d'autoriser la RH entre les chromatides de deux chromosomes, malgré les possibles différences alléliques.

La résolution de la double jonction de Holliday qui s'en suit peut provoquer un CO comme décrit précédemment. Malgré la résolution, les chromosomes homologues sont maintenus ensemble grâce à la présence de cohésines entre les chromatides sœurs et de CO entre chromosomes homologues (**Figure I.5**). Les CO sont donc essentiels pour assurer la ségrégation correcte des chromosomes homologues en méiose I.

Le nombre de CO est régulé par chromosome afin de limiter leur nombre. Ce phéno-

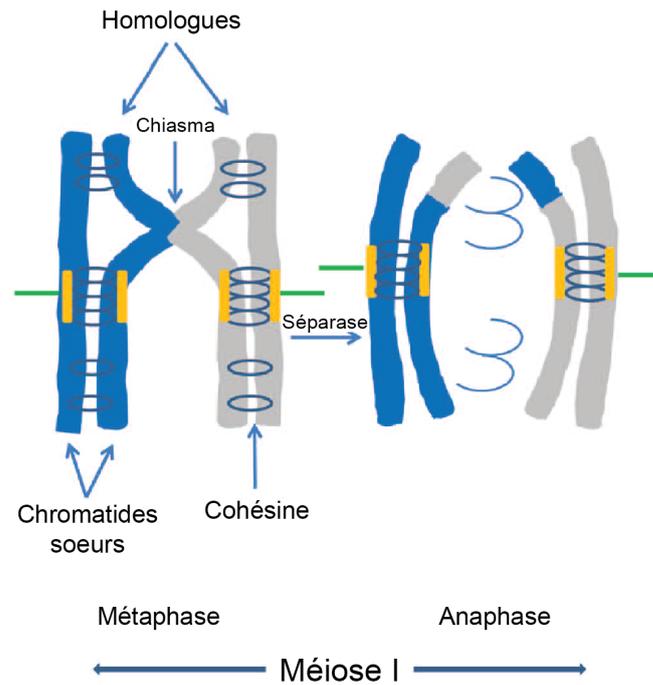


FIGURE I.5 – *Stabilisation du chiasma par la cohésine*

En métaphase de méiose I, la réparation des CDB par la RH peut provoquer des CO. Les chromosomes homologues réarrangés sont bloqués l'un avec l'autre du fait de la forte association des chromatides sœurs via la cohésine. La zone de croisement entre les deux chromosomes est appelée chiasma. Cette association entre chromosomes homologues permet une bonne orientation du fuseau méiotique. En anaphase, une séparase détruit le lien cohésine entre les chromatides sœurs permettant la résolution du chiasma et la migration des chromosomes. Adaptée de MCKEE ET AL. (2012).

mène est appelé *crossover interference* (HILLERS 2004). Les mécanismes moléculaires impliqués sont encore aujourd'hui peu connus. En revanche, ce phénomène peut être ajouté aux activités de Spo11 et de Dmc1 dans la caractérisation d'une spécialisation d'un mécanisme de réparation des CDB dans un évènement cellulaire programmé et essentiel.

Ce ne sera qu'après la séparation des bras des chromatides sœurs, via une séparase rompant le lien de cohésine entre elles, que les chromosomes réarrangés pourront être séparés en anaphase (BUONOMO ET AL. 2000).

I.1.3 Les voies dérivées de la RH

A partir de certaines étapes de la RH, il est possible de diverger vers d'autres mécanismes de réparation (LI AND HEYER 2008) (**Figure I.2**).

Après la résection initiale, s'il existe une homologie entre les deux extensions d'ADN sb de part et d'autre de la CDB et que l'invasion d'un duplexe homologue n'a pas eu lieu, il est possible qu'ils se religuent directement entre eux. Ce mécanisme est le *Single Strand Annealing* ou SSA. Il provoque donc une délétion chromosomique autour de la CDB.

Pour le *Synthesis-Dependent Strand Annealing* ou le SDSA, il y a une dissociation rapide de la *D-loop* après une courte réplication sur l'ADN homologue. L'ADN nouvellement formé s'apparie avec l'autre extrémité d'ADN et les espaces sont comblés par des polymérase. Ce mécanisme empêche la formation d'une double jonction de Holliday et donc d'un CO.

I.1.4 Les erreurs de la RH

La RH est classiquement réputée pour permettre une réparation fidèle, de par son mécanisme d'action utilisant une séquence homologue. C'est cependant ce mécanisme qui peut mener à de nombreuses erreurs de réparation. En effet, la matrice peut ne pas correspondre au locus d'origine. Un CO peut également provoquer des réarrangements chromosomiques. Les différents exemples déjà observés *in vivo* sont décrits en **Figure I.6** (GUIROUILH-BARBAT ET AL. 2014).

Un mécanisme de réparation se doit d'être le plus fidèle possible pour être efficace. Ces différents exemples montrent donc les lacunes de la RH. Ce manque de fidélité, potentiellement délétère, peut également être source de diversité génétique. Un génome

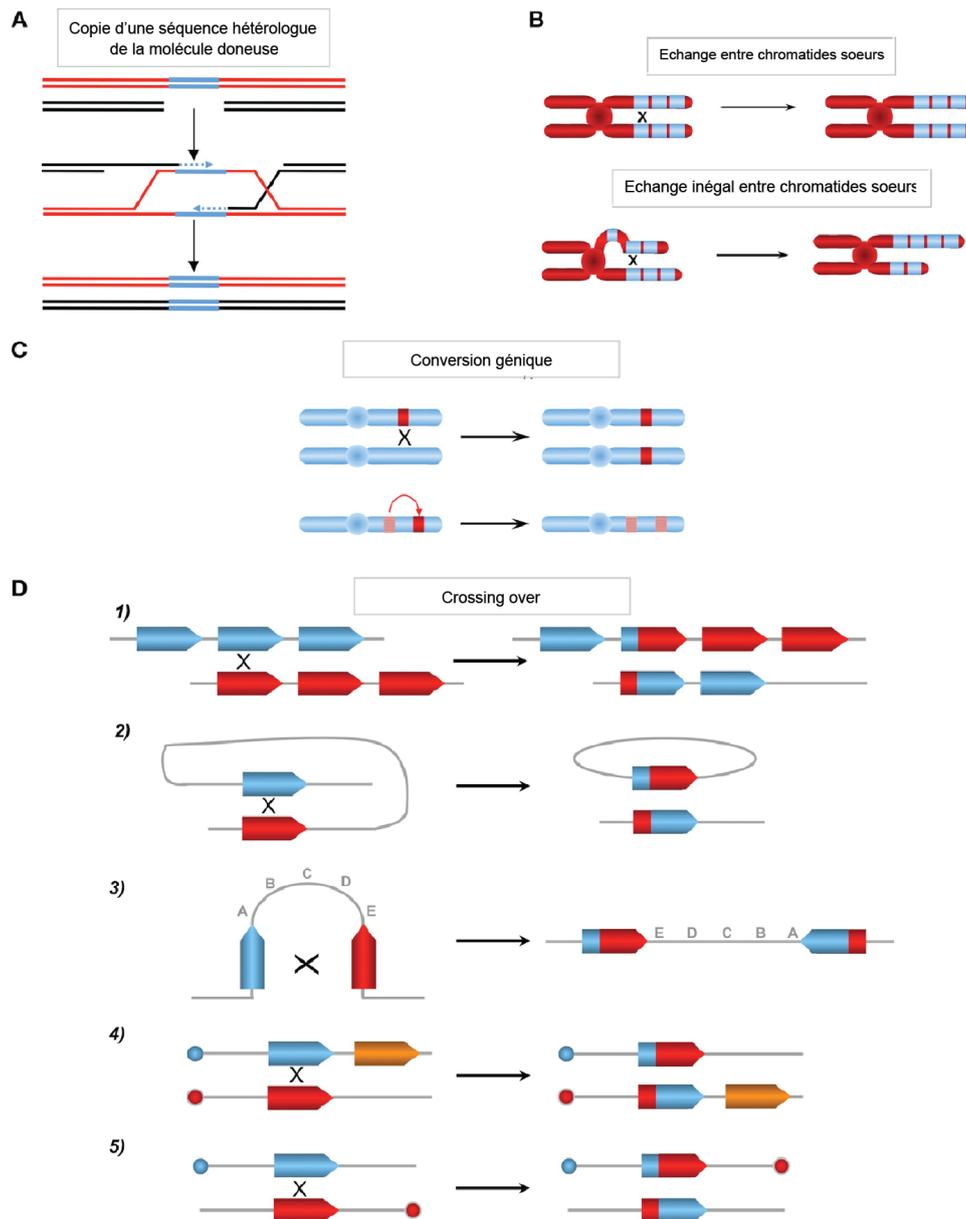


FIGURE I.6 – Les erreurs de la RH

(A) Copie d'une séquence absente de la molécule cassée. La séquence noire ne possédant pas la séquence bleue mais ayant des homologies avec la séquence rouge, la RH va provoquer l'insertion de la séquence bleue et donc une modification du chromosome noir. (B) En présence de séquences répétées, une erreur sur le choix de la séquence homologue peut provoquer des amplifications ou pertes d'ADN. (C) Perte d'une hétérozygotie suite à un échange non réciproque ou introduction d'un codon stop à partir d'un pseudogène homologue. (D) Différentes possibilités de réarrangements suite à un CO entre séquences répétées. 1) CO inégal entre séquences homologues, provoquant une modification du nombre de répétitions pour les deux molécules. 2) Fragmentation et circularisation d'un chromosome suite au CO entre deux séquences répétées dans la même orientation. 3) Inversion de l'orientation suite au CO entre deux séquences répétées d'orientations opposées. 4) CO entre deux chromosomes, provoquant une translocation (mécanisme méiotique). 5) CO entre deux séquences d'orientation inversée, provoquant un chromosome dicentrique et un acrocentrique. Adaptée de GUIROUILH-BARBAT ET AL. (2014)

parfaitement stable ne permet pas à l'organisme de s'adapter à une situation nouvelle et risquerait de provoquer son extinction. L'ensemble des mutations et réarrangements sont donc un moteur essentiel à l'évolution et à la survie des espèces. Il n'est donc pas péjoratif de considérer la RH comme potentielle source d'erreurs.

I.2 LA VOIE DU *Non Homologous End Joining*

Le NHEJ est l'autre grand mécanisme de réparation des CDB. Il procède à la réparation en reliant directement les deux extrémités d'ADN. Contrairement à la RH, la réparation n'est pas basée sur une homologie de séquence et la réparation peut provoquer de nombreux changements de l'information génétique. Cette absence de contrôle permet cependant de faire du NHEJ un mécanisme de réparation plus souple et plus rapide que la RH.

I.2.1 Mécanisme général

Le NHEJ prend en charge les extrémités d'ADN introduites par la CDB avec l'hétérodimère Ku₇₀/Ku₈₀ (Ku) (**Figure I.7**). Ce complexe permet de protéger les extrémités de la dégradation et de recruter les protéines de la voie de réparation. Une CDB pouvant présenter différentes géométries d'extrémités, il est parfois nécessaire de procéder à une maturation des extrémités afin de les rendre compatibles pour la ligation. Des polymérasés et des nucléasés entrent donc en jeu lors de ce processus de maturation. Dans le cas où il n'y a pas besoin de maturation, tel que pour les extrémités franches, la ligation peut directement avoir lieu.

Pour les extrémités avec ADN sb, le NHEJ peut se servir d'une microhomologie pour stabiliser le complexe avant la réparation. Quelques nucléotides peuvent donc être enlevés des extrémités pour permettre un tel appariement. L'ADN réparé qui en résultera aura donc subi une courte délétion. Des polymérasés peuvent de plus ajouter des nucléotides sans séquence matrice, et donc créer l'addition d'une séquence *de novo*. Le NHEJ est donc un mécanisme de réparation ayant un fort potentiel d'introduction d'erreurs, potentiel qui a pu être utilisé par la cellule lors de la recombinaison V(D)J que nous aborderons plus tard. A l'inverse, le NHEJ peut également se révéler très précis lors des Réarrangements Programmés du Génome (RPG) chez la paramécie, également abordés plus tard.

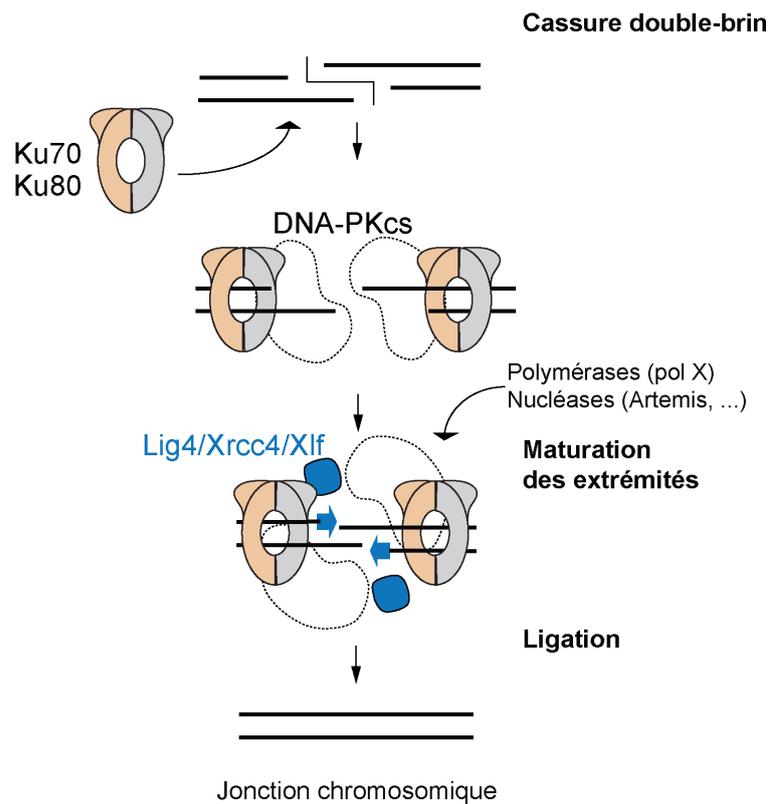


FIGURE I.7 – Réparation d'une CDB par le NHEJ

Ku reconnaît la CDB et recrute DNA-PKcs. Ensemble, ils permettent le recrutement de l'ensemble de la voie NHEJ incluant des polymérase et des nucléases potentiellement nécessaire pour la maturation de la cassure. Le complexe de ligation composé de Lig4, Xrcc4 et Xlf procède à la ligation une fois que les extrémités sont compatibles.

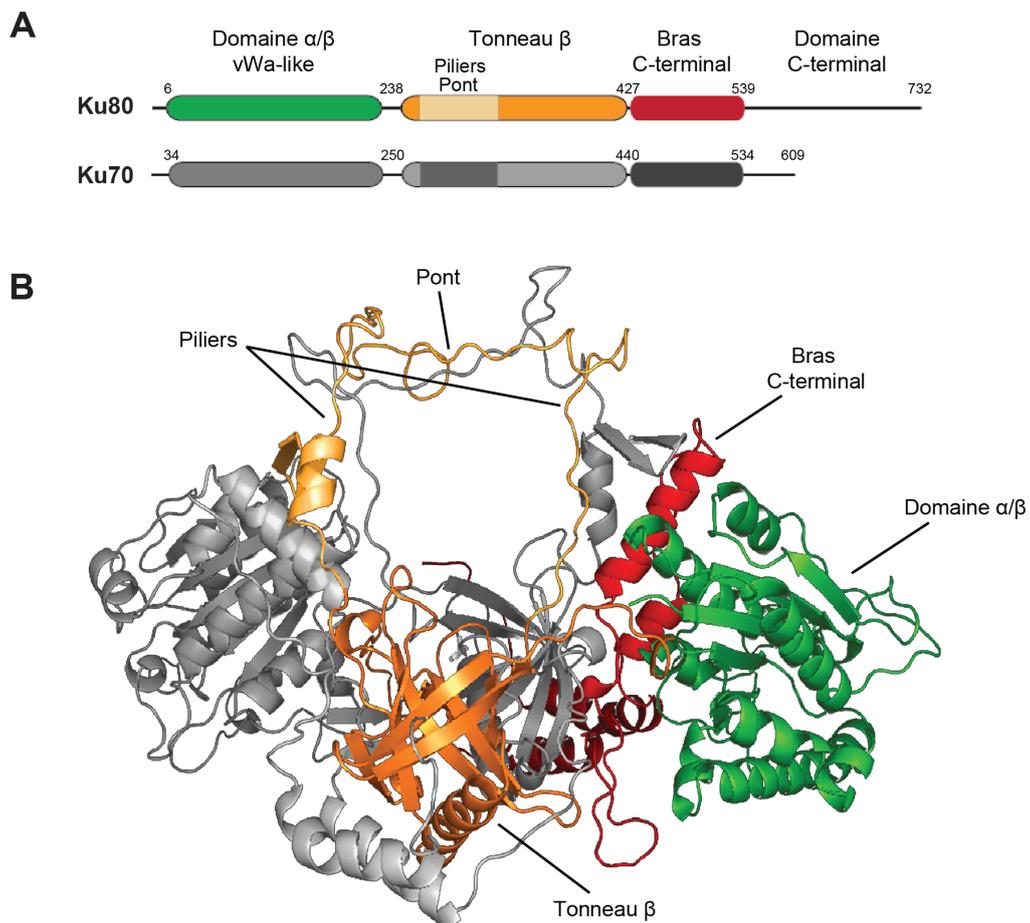
I.2.2 Les acteurs essentiels du NHEJ

I.2.2.1 Ku

Ku est l'acteur agissant le plus tôt dans la voie NHEJ. Il sert à la fois de marqueur de la CDB pour les voies de signalisation, de protecteur des extrémités et de recruteur de protéines de réparation. Ku est pour la première fois décrit en 1981 chez un patient atteint de myopathies (MIMORI ET AL. 1981). Le patient, dont le nom commence par Ku, produit des auto-anticorps dirigés contre une protéine nucléaire. Deux sous-unités seront ensuite décrites pour cette protéine, une de 70 kDa et une autre de 80 kDa, donnant donc Ku70 et Ku80. La structure du complexe humain a été déterminée par cristallographie aux rayons X (WALKER ET AL. 2001) (**Figure I.8**).

Les deux sous-unités, bien qu'ayant des séquences différentes, possèdent une structure similaire. Elles sont composées en N-terminal d'un domaine appartenant à la famille des facteurs de Von Willebrand de type A, ou vWA. Ce domaine est présent dans de nombreuses protéines aux fonctions diverses. Il est composé d'hélices α et d'un feuillet β et est donc également appelé domaine α/β , ce qui sera le cas pour la suite du manuscrit. Ce domaine permet l'interaction avec différents partenaires.

Le domaine central des protéines, ou domaine cœur, est composé d'un tonneau β permettant l'hétérodimérisation entre Ku70 et 80 et d'une structure en anneau, composée de deux "piliers" et d'un "pont". Cette structure, chargée positivement sur la face interne, permet une forte affinité de Ku pour l'ADN db, dont le squelette est chargé négativement. Ku n'a donc pas de spécificité de séquence pour son interaction avec l'ADN via ce domaine. La structure en anneau n'est pas capable de s'ouvrir, comme le font les hélicases ou les topoïsomérasés pour s'associer à l'ADN. Ku ne peut donc se charger sur l'ADN avec son anneau qu'à partir d'une extrémité. Cette rigidité structurelle de l'anneau pose la question du devenir de Ku après la réparation. Il semble compliqué d'imaginer que Ku puisse être recyclé pour une nouvelle réparation. Piégé sur l'ADN reliqué, sa libération doit forcément passer soit par un clivage protéolytique de l'anneau soit par un dépliage complet via la déstructuration du tonneau β . De récentes études montrent le rôle essentiel de l'ubiquitination de Ku pour sa libération du site de réparation (BROWN ET AL. 2015,

FIGURE I.8 – *Structure de Ku*

(A) Représentation linéaire des domaines de Ku70 et Ku80 et coordonnées en acide aminé. (B) Structure de Ku d'après la cristallographie de WALKER ET AL. (2001). La boucle formé par le tonneau β , les piliers et le pont permet l'interaction avec le squelette de la double-hélice d'ADN. Le domaine C-terminal de Ku80 n'est pas présent dans la structure.

ISHIDA ET AL. 2017). Les études montrent également que Ku est dégradé par le protéasome mais ne montrent pas directement comment Ku est enlevé de l'ADN.

Vient ensuite le bras C-terminal, porteur des séquences d'import nucléaire de Ku70 (KOIKE ET AL. 1999b) et Ku80 (KOIKE ET AL. 1999a). Le domaine C-terminal présente la plus grande différence entre Ku70 et 80. Ku70 possède un domaine SAP ayant des propriétés d'interactions avec l'ADN (CHOU ET AL. 1992, ARAVIND AND KOONIN 2000). Ku80 de son côté possède une extension flexible (non présente dans le cristal) avec à son extrémité C-terminale un domaine d'interaction avec DNA-PKcs.

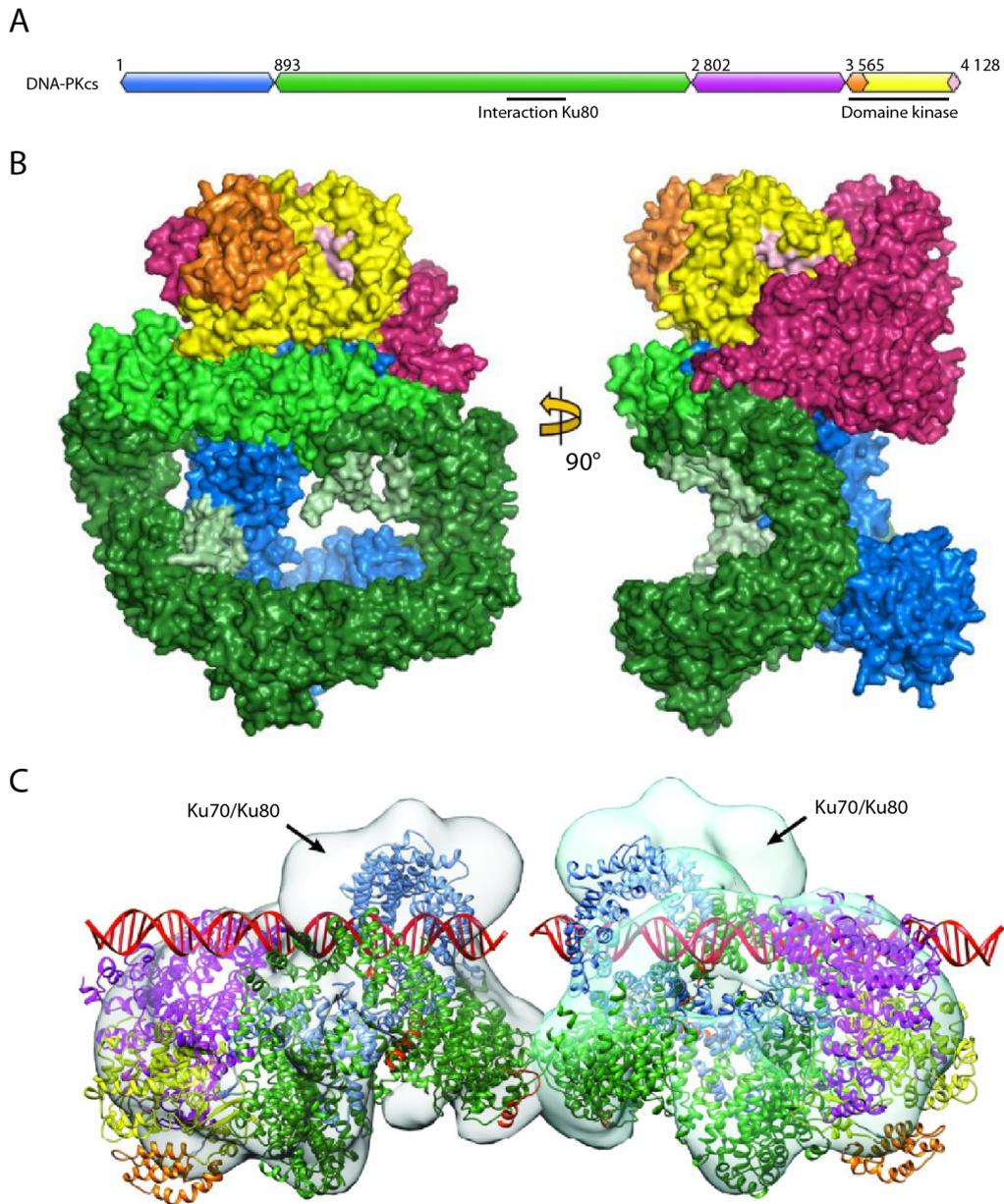
I.2.2.2 DNA-PKcs

DNA-PKcs est la sous-unité catalytique du complexe DNA-PK, pour protéine kinase dépendante de l'ADN (pour revue JETTE AND LEES-MILLER (2015)) (**Figure I.9**). Ce complexe est composé de DNA-PKcs et de Ku70/80. DNA-PKcs est une protéine kinase de type sérine/thréonine. Très grande (350 kDa), elle interagit avec Ku une fois que celui-ci s'est fixé à l'ADN. La formation de ce complexe permet un changement de conformation de DNA-PKcs et stimule de 5 à 10 fois son activité kinase qui a de nombreuses cibles, y compris elle-même. On retrouve parmi ses cibles Ku70 et 80. L'activité kinase de DNAPK-cs est essentielle au bon déroulement du NHEJ. La seule exception est celle des extrémités franches d'ADN, où DNA-PKcs et autres protéines de maturation (nucléases/polymérase) ne sont pas impliquées.

I.2.2.3 Le complexe de ligation

Lig4 La ligase 4 est la ligase spécialisée du NHEJ. Elle est ATP-dépendante et est la seule ligase à pouvoir être recrutée par Ku. Lig4 interagit avec Ku grâce à ses domaines BRCT. Elle permet la religation des extrémités d'ADN et complète ainsi la réparation (**Figure I.10**). La stœchiométrie de Lig4 lors de la réparation n'est pas encore clairement établie. Des modèles proposent une action conjointe de deux Lig4, réparant chacune un brin d'ADN mais il est possible d'une seule Lig4 soit présente à la jonction.

XRCC4 et XLF XRCC4 est un partenaire essentiel de Lig4. Il forme un homodimère qui peut interagir avec Lig4 via son domaine en hélice (**Figure I.10C**). XLF, pour *XRCC4 Like*

FIGURE I.9 – *Structure de DNA-PKcs*

(A) Représentation linéaire des domaines de DNA-PKcs et coordonnées en acide aminé. (B) Structure de DNA-PKcs d'après la cristallographie de SIBANDA ET AL. (2017). (C) Modèle structural d'un dimère de complexes DNA-PK/ADN. L'arrangement du dimère peut faciliter l'alignement des extrémités d'ADN et ainsi promouvoir la réparation. Adaptée de SIBANDA ET AL. (2017).

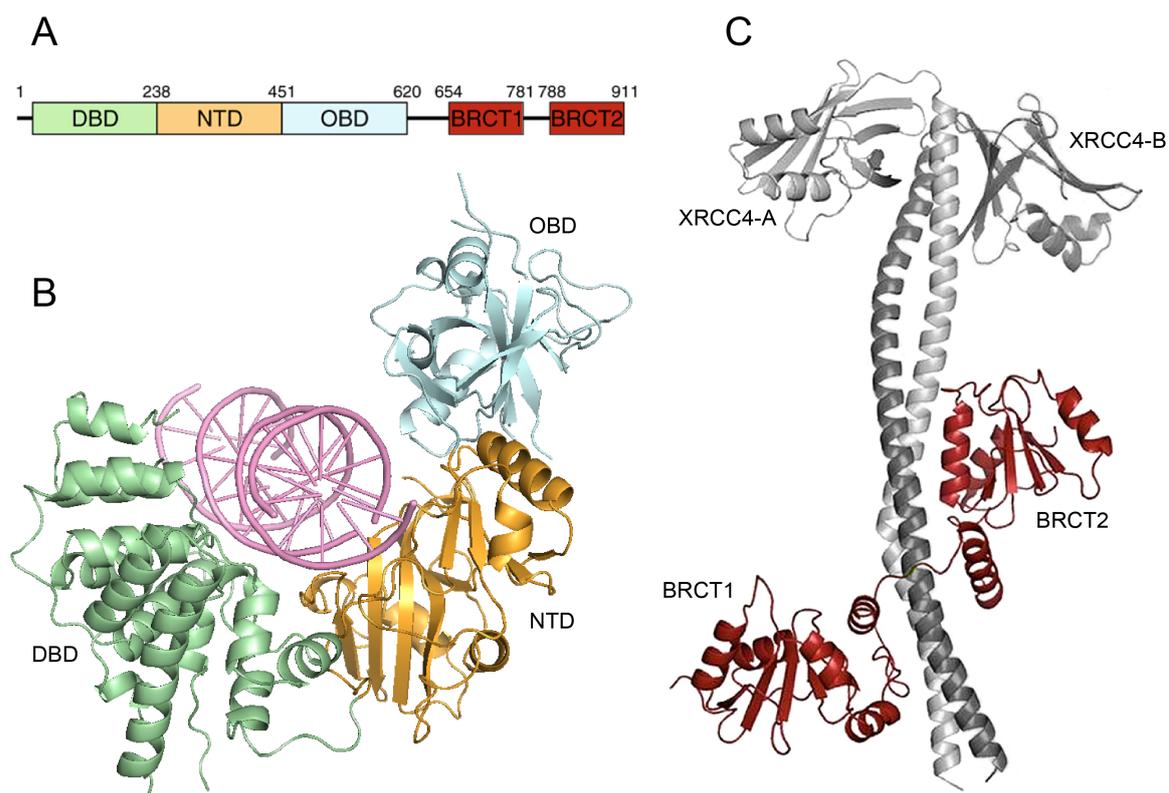


FIGURE I.10 – Structure de Lig4

(A) Représentation linéaire des domaines de Lig4 et coordonnées en acide aminé. DBD : DNA Binding Domain ; NTD : NucleotidylTransferase Domain ; OBD : Oligonucleotide/oligosaccharide-Binding fold Domain. (B) Structure du domaine catalytique de Lig4 associé à un ADN avec une coupure sb (rose) d'après la cristallographie de KAMINSKI ET AL. (2018). (C) Structure des domaines BRCT et interaction avec un homodimère XRCC4 d'après la cristallographie de WU ET AL. (2009).

Factor (connu également sous le nom de Cernunnos), est une protéine proche de XRCC4 mais n'interagissant pas avec Lig4 (BUCK ET AL. 2006, AHNESORG ET AL. 2006). Son rôle était mystérieux jusqu'à ce que la structure du complexe XRCC4/XLF soit découverte (ROPARS ET AL. 2011, HAMMEL ET AL. 2011, WU ET AL. 2011) (**Figure I.11**). Les homodimères de XRCC4 et XLF peuvent interagir entre eux et former un hétéro-oligomère, pouvant aller jusqu'à la formation d'un filament. De plus, deux hétéro-oligomères peuvent interagir entre eux via les domaines en hélice.

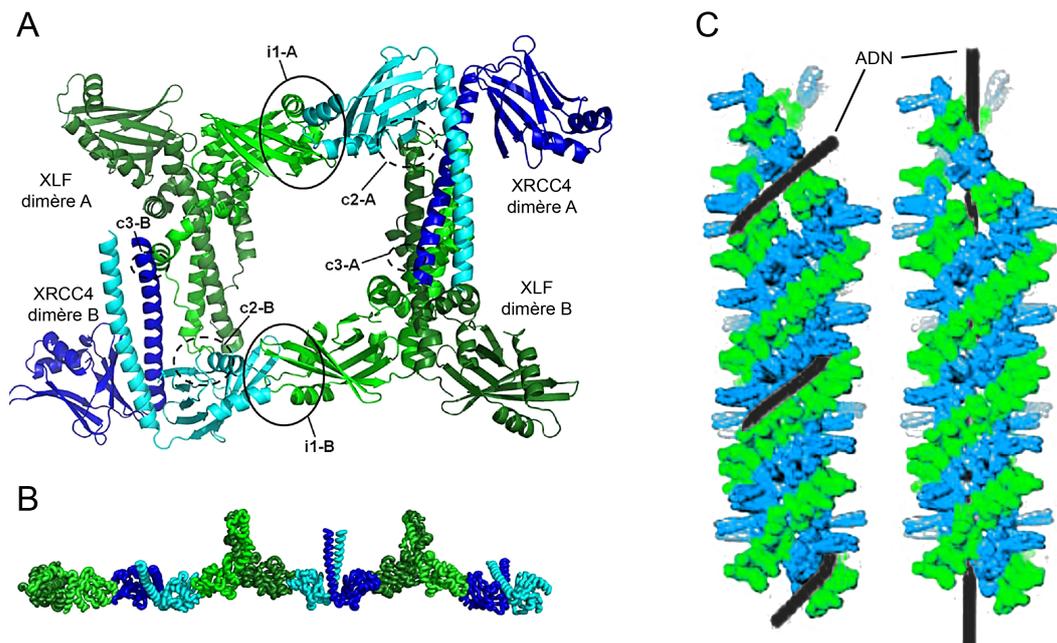


FIGURE I.11 – Structure du filament XRCC4/XLF

(A) Tetra-dimère de XRCC4 et XLF montrant les capacités d'interaction multiples de ces protéines. (B) Filament de XRCC4/XLF. (C) Modèles d'interaction entre le filament et l'ADN. L'ADN passe soit par la périphérie du filament soit par le centre. (A) et (B) adaptées de ROPARS ET AL. (2011). (C) adaptée de MAHANEY ET AL. (2013)

Ce filament peut interagir avec l'ADN et les modèles actuels proposent que c'est grâce à ce filament que les deux extrémités d'ADN sont maintenues à proximité au cours du processus de réparation, réduisant ainsi le risque de translocations. Deux modèles sont proposés pour l'interaction du filament avec l'ADN (**Figure I.11C**). Un paralogue de XRCC4 et XLF, PAXX, a été décrit dans OCHI ET AL. (2015), XING ET AL. (2015), CRAXTON ET AL. (2015) comme stabilisant le complexe de réparation, grâce à son interaction avec Ku70, mais n'étant pas essentiel.

I.2.3 Les acteurs secondaires

Les acteurs principaux suffisent à réparer une cassure avec des extrémités compatibles. Cependant, il est rare que ce soit le cas. Dès qu'une maturation de l'ADN est nécessaire, ces protéines ne sont plus suffisantes. Différentes nucléases et polymérase rentrent donc en action.

Artemis La protéine Artemis est une nucléase interagissant fortement avec DNA-PKcs (MOSHOUS ET AL. 2001) **figure I.12**. Lors de l'activation de DNA-PKcs, Artemis est elle aussi activée par phosphorylation par DNA-PKcs. Elle gagne ainsi la capacité de couper les extrémités d'ADN simple ou double-brin, avec une double activité endo et exonucléase. Elle peut ainsi procéder à la maturation de structures telles que les repliements en épingle de l'ADN, l'élimination de nucléotides aux extrémités et autres structures pouvant provenir d'un dommage de l'ADN (MA ET AL. 2002, PANNUNZIO ET AL. 2018).

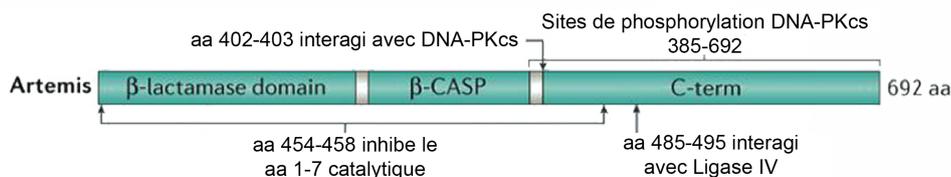


FIGURE I.12 – *Artemis*, nucléase du NHEJ

Domaines de Artémis. Adaptée de CHANG ET AL. (2017).

Dans certains types de cassures, d'autres nucléases peuvent être utilisées telles que l'aprataxine ou bien APLF. De manière générale, les résections d'ADN de 5' vers 3' lors d'une réparation par la voie du NHEJ sont inférieures à 10 nucléotides.

Les polymérase L'ensemble des polymérase utilisées dans le NHEJ sont des polymérase de la famille X. La famille X est composé de quatre polymérase : Pol β , Pol λ , Pol μ et la terminal transferase (Tdt) (YAMTICH AND SWEASY 2010). Elles sont toutes utilisées dans le NHEJ sauf Pol β , qui est spécifique du mécanisme de réparation BER **figure I.13**. De fait, Pol β est la seule polymérase de la famille X à ne pas posséder un domaine BRCT, essentiel pour interagir avec la structure Ku-ADN.

Les polymérase de la famille X ne possèdent pas de capacité de relecture et sont beaucoup moins fidèles que les polymérase utilisées pour la réplique de l'ADN. Leur

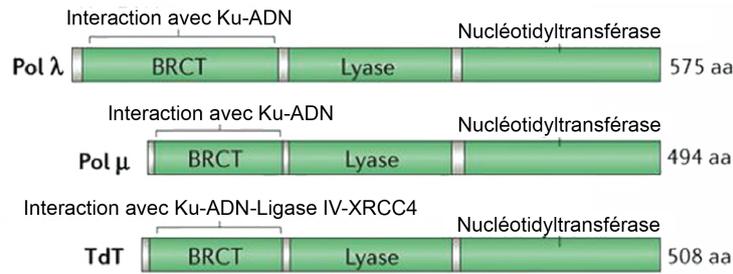


FIGURE I.13 – Les polymérases du NHEJ

Domaines des trois polymérases du NHEJ. ELles possèdent toutes le domaine BRCT permettant l'interaction avec Ku-ADN. Adaptée de CHANG ET AL. (2017).

manque global de fidélité peut être expliqué par la nécessité d'être capable de travailler avec une matrice d'ADN potentiellement discontinue ou endommagée, matrice qu'une polymérase classique ne peut exploiter.

Les polymérases de la famille X ont récemment été décrites comme capable d'intégrer des ribonucléotides à la place de désoxyribonucléotides. Il a été proposé que cette incorporation augmente l'efficacité *in vivo* de la réparation (MOON ET AL. 2017, PRYOR ET AL. 2018). Cette capacité fait exception dans le monde des ADN polymérases.

I.2.4 Un rôle des ARN dans le NHEJ ?

Jusqu'à récemment, la voie du NHEJ était décrite comme exclusivement protéique à l'exception de phosphorylations de ribonucléoprotéines par DNA-PKcs (pour revue HALEY ET AL. (2009)). De récentes études ont montré l'implication d'un long ARN non codant, LINP1, dans le NHEJ (ZHANG ET AL. 2016, WANG ET AL. 2018). LINP1, pour *LncRNA In Non-homologous end joining Pathway 1*, interagit avec Ku80 et DNA-PKcs et augmente l'efficacité et la vitesse de réparation des CDB *in vivo*. LINP1 pourrait donc avoir un rôle de stabilisateur du complexe DNA-PK ou d'une autre fonction dans le NHEJ, augmentant son efficacité.

Il a également été montré par BRITTON ET AL. (2013) qu'un traitement à la RNase permettait d'extraire du noyau les protéines Ku ou XRCC4 non engagées sur une CDB. Cela implique qu'une large fraction des protéines est associée à la chromatine via l'ARN. Cette association permettrait une proximité physique constante des facteurs du NHEJ avec l'ADN et donc une plus grande efficacité dans le traitement des CDB.

I.2.5 Conséquences du NHEJ sur les cassures accidentelles

Le NHEJ est un mécanisme qui ne recherche pas d'homologie. Il reconnaît des extrémités et les recolle. Ces propriétés font du NHEJ un mécanisme pouvant être source d'erreurs et de modification du génome.

L'erreur classique du NHEJ est la micro-délétion. Les bases endommagées et la maturation nécessaire des extrémités entraînent la perte de quelques nucléotides. Cette "petite" erreur peut s'avérer très délétère si elle se situe dans une séquence codante et change le cadre de lecture.

Les choses se compliquent pour le NHEJ dès que le nombre de CDB augmente. Avec une seule cassure, le locus sera forcément reconstitué, avec ou sans micro-délétion. Avec plusieurs cassures concomitantes, il existe un risque d'associer ensemble différents loci, provoquant ainsi des translocations (GHEZRAOUI ET AL. 2014). Ces translocations peuvent être nocives dans le cas par exemple de fusion entre deux chromosomes, créant ainsi un chromosome à deux centromères.

Il existe un phénomène de fragmentation étendu du chromosome, appelé chromothripsis. Il consiste en une apparition massive de CDB sur un chromosome dont la réparation provoque de nombreux réarrangements chromosomiques et de possibles perte de matériel génétique (PELLESTOR ET AL. 2014) (**Figure I.14**). Ce phénomène peut avoir plusieurs origines : radiations, érosion télomérique, apoptose abortive, etc. Le NHEJ pourrait jouer un rôle important dans la réparation du chromothripsis.

I.2.6 Évolution de Ku

Ku70 et Ku80 sont présents chez l'ensemble des eucaryotes. On les retrouve strictement sous forme d'hétérodimère. Chez les procaryotes, la présence du NHEJ n'est pas systématique. En effet, de nombreuses espèces, telle que la célèbre *Escherichia coli*, ne possèdent pas de protéines Ku. Celles qui en possèdent présentent un unique gène, produisant un homodimère Ku "simplifié" car dépourvu de domaine N-terminal α/β et de domaine SAP (**Figure I.15**). Les Ku eucaryotes étaient donc certainement à leur origine également des homodimères, qui se sont dupliqués puis spécialisés au cours de l'évolution dans différentes sous-fonctions en gagnant de nouveaux domaines (PITCHER ET AL. 2007).

Il existe à ce jour une exception au dogme du Ku unique chez les procaryotes. Dans la

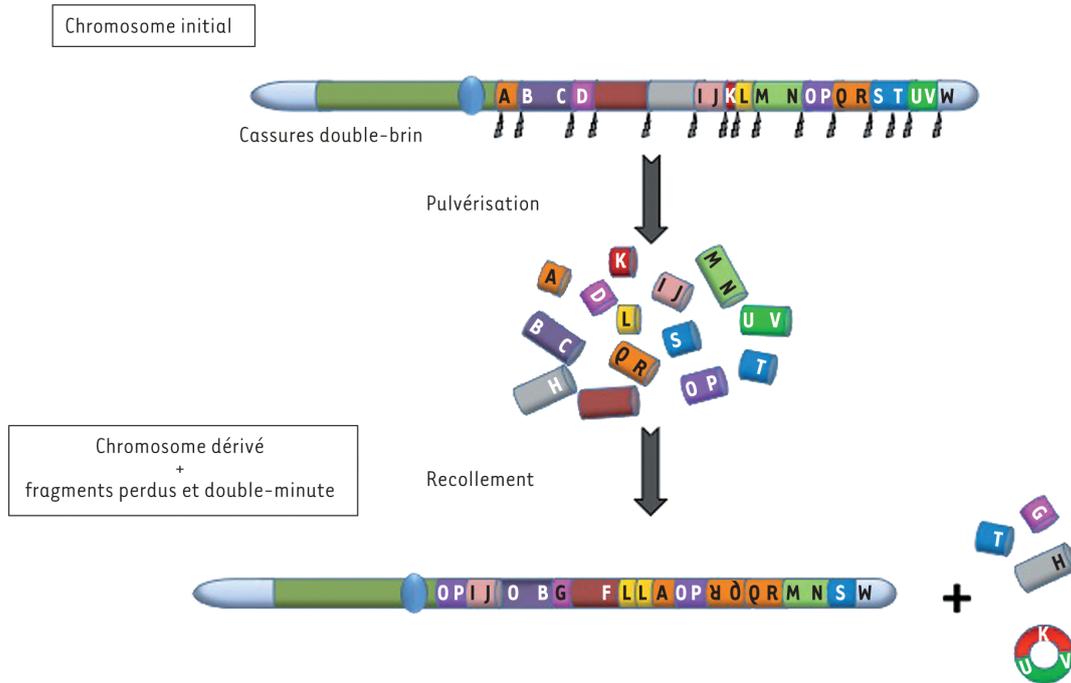


FIGURE I.14 – *Le chromothripsis*

Fragmentation extrême d'un chromosome, provoquant de nombreux réarrangements et pertes d'ADN (PELLESTOR ET AL. 2014).

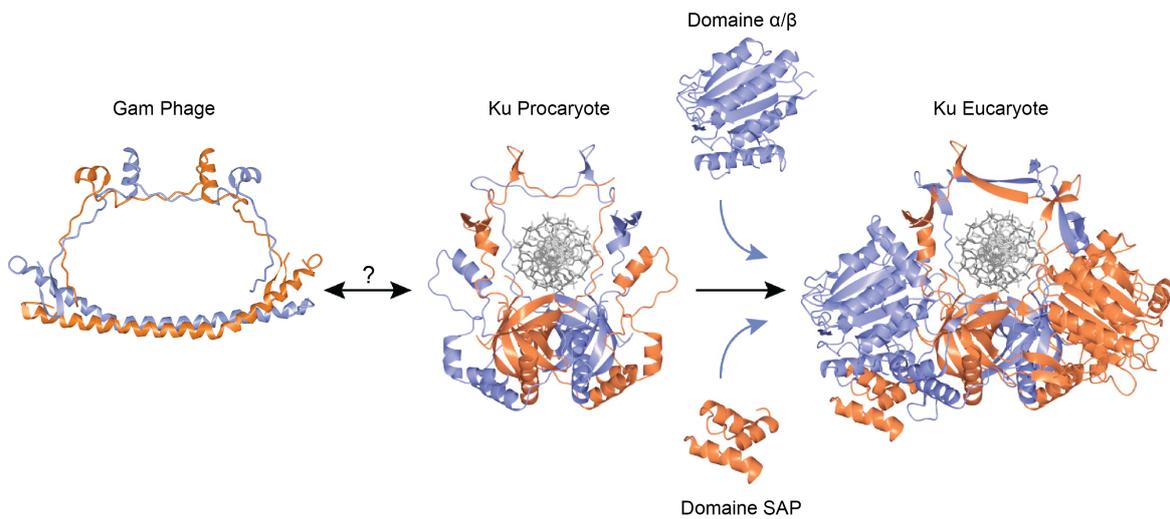


FIGURE I.15 – *Diversité des Ku*

Structure du Gam de phage, du Ku procaryote et du Ku eucaryote. Adaptée de PITCHER ET AL. (2007).

classe des alpha-protéobactéries, on retrouve au moins quatre versions de Ku, étudiées en particulier chez *Sinorhizobium meliloti* par DUPUY ET AL. (2018) et nommées Ku₁, 2, 3 et 4. Là où Ku₂ semble être le Ku procaryote classique et Ku₁ une version en pseudogenisation, Ku₃ et 4 semblent former une voie du NHEJ parallèle au sein du même organisme. Il est de plus proposé que Ku₃ et 4 forment un hétérodimère. Ce serait le premier exemple d'hétérodimère Ku chez les procaryotes, ouvrant la voie à une spécialisation des sous-unités, comme ça a été le cas pour les protéines eucaryotes.

L'exemple de la protéine Gam du phage Mu est également intéressant. Cette protéine, formant un homodimère (**Figure I.15**), protège les extrémités double-brin de l'ADN viral. Elle possède des homologies de séquences avec Ku mais présente une structure sensiblement différente. Enfin, il a été récemment montré qu'elle inhibait les exonucléases, se liait à une ligase d'*E. coli* et promouvait *in vivo* chez *E. coli* une réparation de l'ADN similaire au NHEJ (BHATTACHARYYA ET AL. 2018). Il y a donc pu avoir soit une convergence évolutive entre Gam et Ku soit un transfert horizontal de gène à l'occasion d'une infection du phage Mu. *E. coli* n'ayant pas de NHEJ, la présence de ce système viral a pu être conservé et favoriser les symbioses entre ces deux organismes en protégeant le génome bactérien.

I.2.7 CDB programmées et NHEJ, exemple de la recombinaison V(D)J

Les CDB étant très toxiques si non réparées, il peut être considéré dangereux pour l'organisme d'en introduire volontairement. Cependant, elles sont régulièrement utilisées dans différents processus cellulaires. Cette introduction volontaire de CDB vient donc avec la nécessité pour l'organisme d'être prêt à effectuer une réparation efficace. Un exemple de cassures programmées utilisant la RH a déjà été décrit, avec Spo11, lors de la méiose. La recombinaison V(D)J, quant à elle est un exemple parfait de l'utilisation du NHEJ, de ses qualités et de ses défauts, dans un processus d'introduction de CDB volontaires et de leur réparation (LITTLE ET AL. (2015) pour revue).

Le système immunitaire repose sur la détection de particules exogènes par les cellules lymphocytaires grâce aux anticorps. La spécificité et la diversité de détection des anticorps sont assurées par le mécanisme de recombinaison V(D)J. Un anticorps est composé de deux chaînes lourdes et deux chaînes légères. La région de reconnaissance de l'antigène est elle-même composée de segments V, D et J pour la chaîne lourde et V et J pour la

chaîne légère. Les loci germinaux de ces deux sous-unités sont composés de nombreuses versions différentes de ces segments. Les différents réarrangements ayant lieu lors de la différenciation des lymphocytes provoquent l'excision des segments à l'exception d'une version de chaque segment, choisie de manière stochastique, donnant un exon fonctionnel (**Figure I.16**).

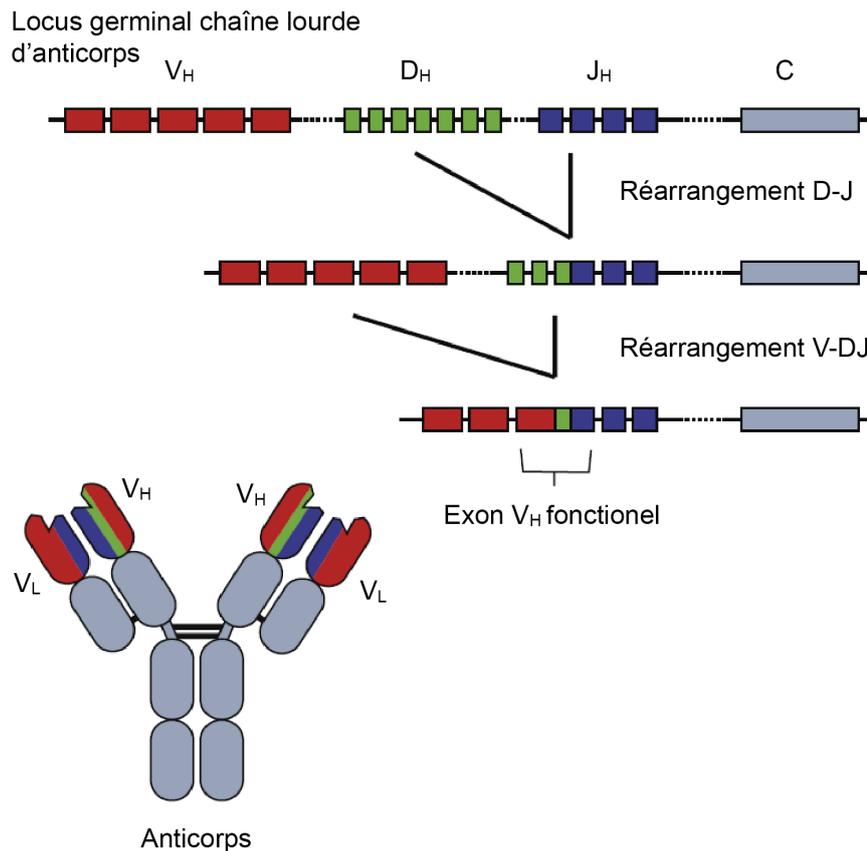


FIGURE I.16 – *La recombinaison V(D)J*

Recombinaison VDJ du locus germinale de la chaîne lourde d'anticorps. H : *Heavy*, L : *Light*, V : *Variable*, D : *Diversité*, J : *Joining*, C : *partie constante*. Adaptée de LITTLE ET AL. (2015).

L'excision des segments éliminés dépend de la transposase domestiquée RAG1. Les transposons sont des séquences d'ADN qui ont la capacité de se déplacer ou de se répliquer dans un génome à l'aide de la transposase codée par le transposon lui-même (HICKMAN AND DYDA (2015) pour revue). Les transposases les plus étudiées possèdent un domaine catalytique apparenté aux RNases H contenant trois acides aminés catalytiques DD(D/E). Une transposase domestiquée est une transposase ayant perdu la capacité d'exciser son transposon mais conservée par la cellule hôte pour ses fonctions (endonucléase,

fixation à l'ADN). Elle intègre ainsi des processus cellulaires et devient essentielle pour l'organisme (VOLFF 2006, JANGAM ET AL. 2017). Différents exemples de domestication ont été décrits. SETMAR, une transposase domestiquée *mariner* partiellement active est impliquée dans la réparation des CDB chez les primates (LIU ET AL. 2007, KIM ET AL. 2014). $\alpha 3$, dérivée d'un transposon hAT permet, avec Kat1, le changement de type sexuel de la levure *Kluyveromyces lactis* (BARSOUM ET AL. 2010, RAJAEI ET AL. 2014).

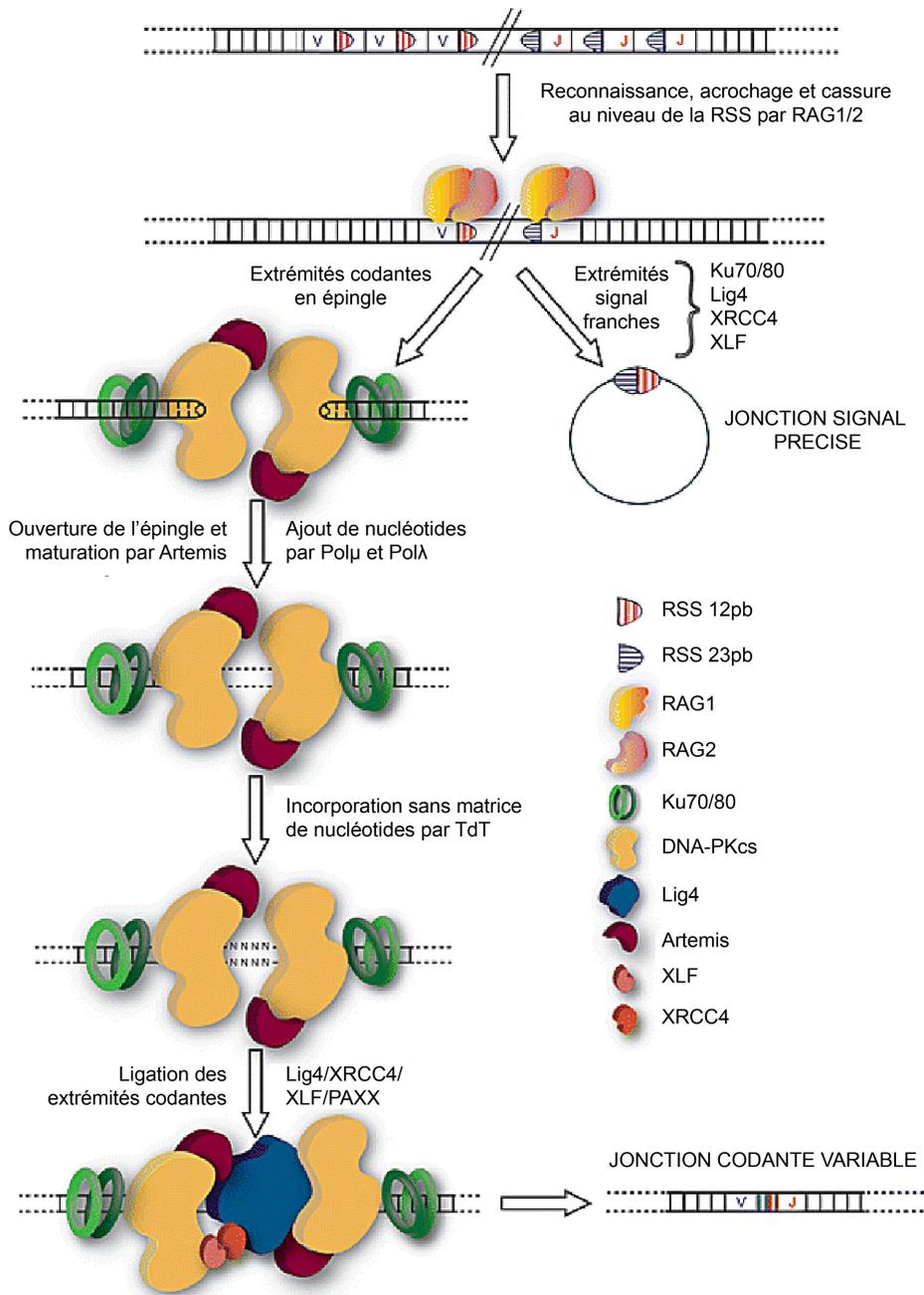
RAG1 dérive des transposases de la famille Transib (KAPITONOV AND JURKA 2005, HUANG ET AL. 2016) et a été domestiquée pour effectuer les RPG nécessaires à la synthèse des anticorps.

RAG1, au côté de RAG2, partenaire essentiel mais non catalytique, se fixe sur les *Recombination Signal Sequences* (RSS), séquences flanquant les segments V, D ou J, provoque la formation d'une synapse et la coupure de l'ADN, avec la formation de structure en épingle sur les extrémités codantes (**Figure I.17**).

La maturation et la ligation sont effectuées par le NHEJ. Artemis permet l'ouverture de l'épingle à une position variable. L'action des polymérases, et notamment TdT, conduit à l'ajout de nucléotides. TdT réalise une incorporation sans matrice puis le complexe de ligation réalise la jonction chromosomique.

Le mécanisme d'ouverture de l'épingle à une position variée associée à l'action de TdT produit une zone de jonction très variable. Cette variabilité est essentielle pour la production d'un haut niveau de diversité de reconnaissance d'antigènes par les anticorps. Les mécanismes imprécis du NHEJ ont donc été utilisés par la cellule comme moteurs de diversité.

Bien que recherchant l'introduction d'erreurs, la recombinaison V(D)J doit s'assurer de la religation des CDB introduites. Des études se sont donc penchées sur une possible interdépendance entre le mécanisme de coupure et celui de réparation, ce qui permettrait de diriger le NHEJ sur la coupure avant même qu'elle n'ait lieu. Les travaux de RAVAL ET AL. (2008) démontrent que Ku est capable d'interagir avec RAG1. Cependant, il est connu depuis longtemps que RAG1 et RAG2 sont suffisants pour l'introduction des cassures (MCBLANE ET AL. 1995). D'autre part, RAG2 a été montré comme important dans la stabilisation de la synapse entre les deux extrémités d'ADN (LESCALE ET AL. 2016). Lors de la recombinaison V(D)J, il y a donc une forme de couplage entre cassure et réparation,

FIGURE I.17 – *Etapes de la recombinaison V(D)J*

Le complexe RAG₁/RAG₂ reconnaît et lie les RSS 12 et 23pb. Ensuite, une CDB est introduite par RAG₁ produisant des extrémités franches sur la séquence excisée et la formation d'épingles à cheveux sur la séquence codante. Ku se fixe aux extrémités et la séquence excisée est réparée précisément et circularisée par le NHEJ à l'aide de Ku, Lig4 XRCC₄ et XLF. Concernant la séquence codante, Artemis procède à l'ouverture de l'épingle à des positions variables, les espaces sont comblés par Polμ et Polλ et TdT ajoute des nucléotides sans aide de matrice. Enfin, la ligation est effectuée par Lig4/XRCC₄/XLF et PAXX. Adaptée de MALU ET AL. (2012).

ce couplage se limitant à un "guidage" du NHEJ vers la cassure et à une stabilisation de la synapse.

I.2.8 Voie alternative de réparation des CDB

Il existe un autre mécanisme de réparation des CDB, moins étudié. Il s'agit de l'*alternative End-Joining*, ou alt-EJ, aussi appelé MMEJ pour *Micro-Homology End Joining*. Ce mécanisme de réparation est indépendant de Ku (**Figure I.18**). La difficulté dans l'étude du MMEJ réside dans le fait que le substrat utilisé et le produit de réparation ressemblent à ceux du NHEJ, mécanisme très efficace. Cette voie alternative est donc étudiée avec des souches Δ KU ou Δ LIG4, afin d'être en condition de déplétion du NHEJ. Ce mécanisme dépend de PARP1 et d'une résection de l'ADN, provoquée par le complexe MRN et CtIP. La résection permet de découvrir des régions de micro-homologie (3-10 nucléotides), favorisant la formation d'une synapse entre les deux extrémités d'ADN et donc leur ligation. La polymérase θ , ou Pol θ , est spécifique du mécanisme. Enfin la ligation est assurée par Lig3 ou Lig1. La nature des maturations de l'ADN nécessaires provoque donc forcément une délétion de séquence de part et d'autre de la CDB.

I.2.9 Choix de la voie de réparation des CDB

La présence concomitante de l'ensemble de ces mécanismes de réparation engendre une bataille sur les extrémités d'ADN. En compétition les unes avec les autres, le choix de la voie est grandement déterminé par la vitesse d'accession à l'ADN. Ku est une protéine ayant une très forte affinité pour les extrémités d'ADN et très abondante dans le noyau. Elle est donc la première à s'associer à la cassure et déclenche majoritairement le NHEJ (**Figure I.19**).

Différents processus peuvent cependant ralentir le NHEJ et donc favoriser la RH. Lors des phases S et G2 du cycle cellulaire, les chromosomes possèdent deux chromatides. L'utilisation d'une chromatide sœur comme matrice par la RH garantit une réparation parfaite. Différents mécanismes ont évolué pour activer la RH lors de ces phases du cycle, mais également pour ralentir le NHEJ, tels que l'inhibiteur CYREN, pour *Cell cycle Regulator of NHEJ* (ARNOULT ET AL. 2017). CYREN interagit avec Ku70/80 et provoque une inhibition du NHEJ en phase S/G2. En parallèle, la RH est activement inhibée par 53BP1

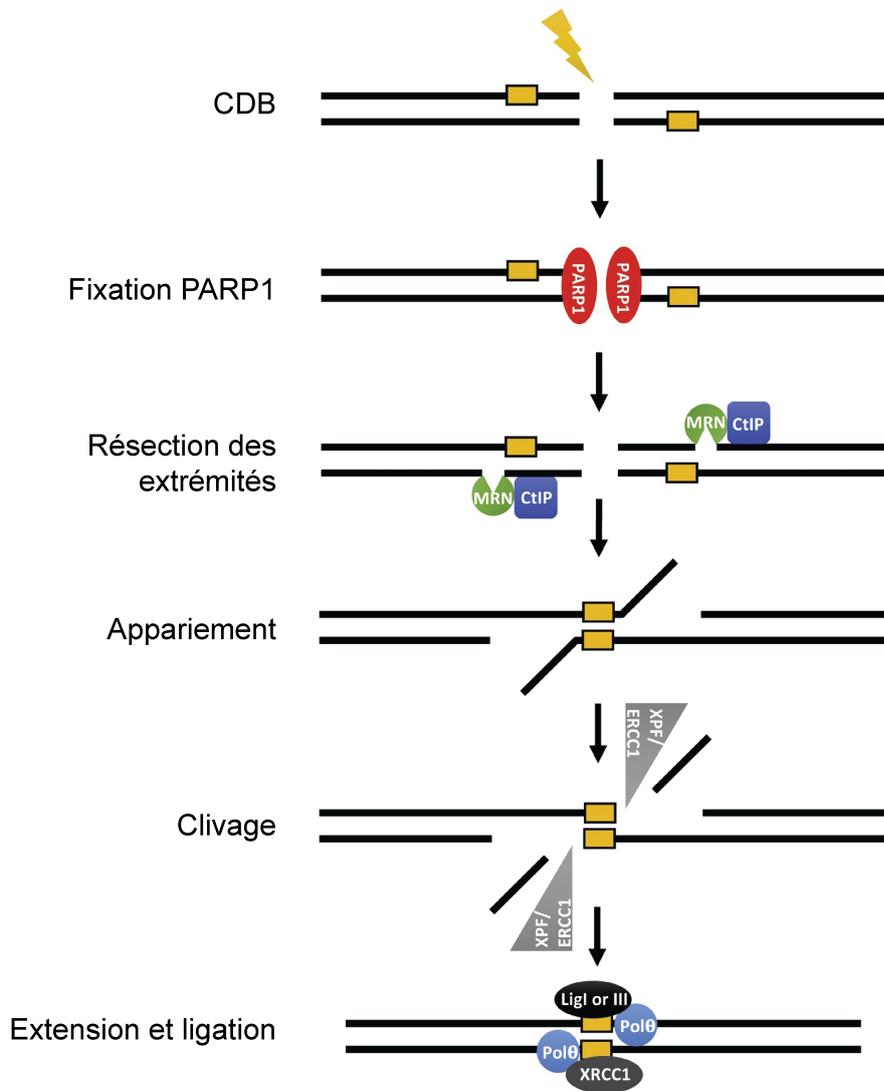


FIGURE I.18 – *Le MMEJ, mécanisme alternatif de réparation des CDB*

PARP1 se fixe aux extrémités d'ADN et participe au recrutement des protéines de résection (CtIP et le complexe MRN) pour découvrir les micro-homologies (rectangle jaune). L'appariement des micro-homologies provoque la formation d'extrémités non homologues. Ces extrémités sont clivées par XPF/ERCC1 avant que Polθ comble l'espace et que LigI/III ligue les extrémités. Adaptée de SEOL ET AL. (2018).

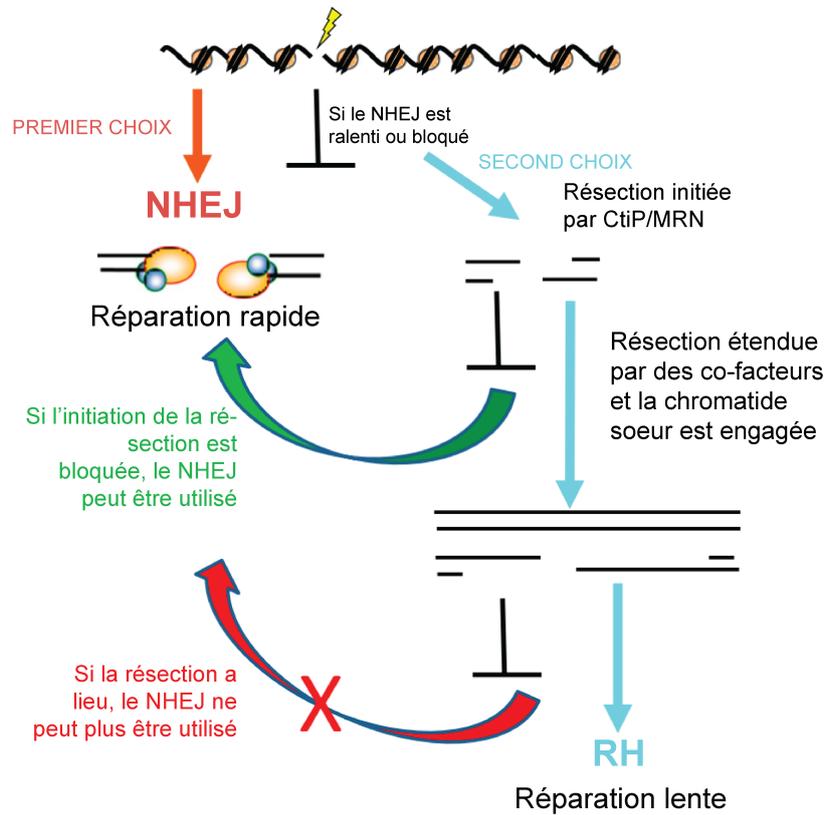


FIGURE I.19 – Choix de la voie de réparation de CDB

Le NHEJ est le mécanisme de premier choix pour la réparation d'une CDB. Si le NHEJ est ralenti ou bloqué, le second choix se porte sur la RH. Tant que la résection initiée par CtIP et MRN est de faible ampleur, il est possible de retourner au NHEJ. Une résection étendue engage la réparation de la CDB vers la RH. Adaptée de KAKAROUGKAS AND JEGGO (2014).

en phase G₁ (Daley et al. 2014 pour revue). Cette protéine empêche l'introduction des premières étapes de résection du complexe MRN et favorise donc le NHEJ. Lors des phases S/G₂, la protéine BRCA₁ inhibe 53BP₁, permettant ainsi le début de la résection par MRN.

Il arrive également qu'une protéine Ku se fixe à un fragment d'ADN que le NHEJ ne peut réparer, comme c'est le cas avec les extrémités uniques d'ADN. On retrouve ces structures sur des fourches de réplifications arrêtées. Dans cette situation, Ku est dégagé de l'extrémité par les activités exonucléase de Mre₁₁ et endonucléase de CtIP (CHANUT ET AL. 2016), ouvrant ainsi la voie à une possible réparation par la RH.

La structure de l'ADN est également importante dans le choix de la voie. Si de simples extrémités db sont compatibles avec le NHEJ et la RH, dès lors que la résection par la RH s'étend, le NHEJ perd ses capacités de réparation et la cassure s'engage vers une réparation par la RH. Les extrémités peuvent aussi être obstruées, comme c'est le cas avec Spo₁₁. Liée covalamment à l'ADN, elle empêche Ku de se fixer et favorise ainsi l'action de Mre₁₁, nécessaire pour l'introduction des CO essentiels à la méiose.

Comme vu au cours de ce premier chapitre, l'étude et la compréhension des mécanismes de réparation de l'ADN ont largement bénéficié de l'analyse de CDB programmées. Reproductibles, elles fournissent un outil précieux dans l'analyse des régulations et des partenaires impliqués. Les ciliés sont des organismes procédant à des RPG massifs. Ils ont été et sont encore des organismes modèles pour l'étude des CDB programmées.

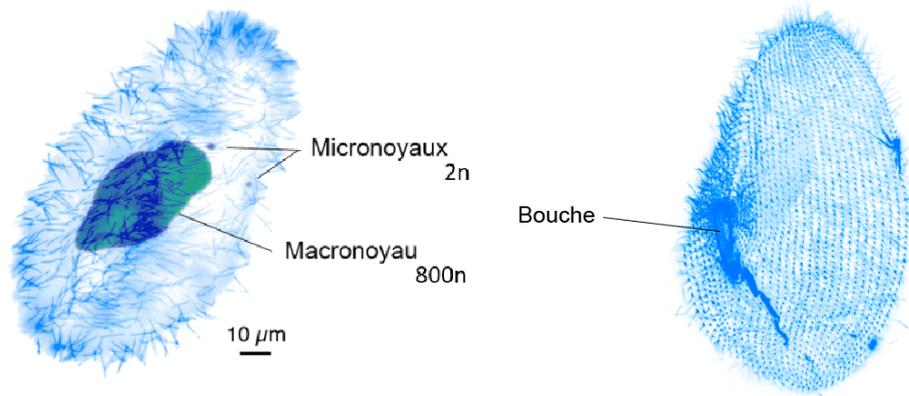
Chapitre II

La paramécie, modèle d'étude du NHEJ lors de CDB programmées

II.1 CARACTÉRISTIQUES GÉNÉRALES DE *Paramecium tetraurelia*

Le genre paramécie regroupe les organismes appartenant à l'embranchement des ciliés. On y retrouve différentes espèces telles que *Paramecium tetraurelia*. La paramécie est un organisme eucaryote unicellulaire recouvert de cils sensoriels et vibratoires lui permettant de se déplacer dans son milieu naturel (lac, rivière, ...). De très grande taille pour une cellule (plus de 100 μm^2 de longueur), elle se nourrit de micro-organismes tels que des bactéries, des algues ou des levures. Elle possède une structure appelée bouche, où elle réalise une endocytose formant ainsi des vacuoles digestives (**Figure II.1**).

Comme tous les ciliés, elle présente un dimorphisme nucléaire. La paramécie possède deux noyaux diploïdes, appelés Micronoyaux (Mic) et un noyau hautement polyploïde, appelé Macronoyau (Mac) avec une ploïdie d'environ 800 copies par locus. Les Mic sont les noyaux dits germinaux, car ils ne servent pas à l'expression des protéines et sont transmis au cours des processus sexuels. Le Mac est le noyau dit somatique, car il sert à l'expression de l'ensemble des protéines nécessaires à la vie de la paramécie. Il est perdu à chaque cycle sexuel et est remplacé par un nouveau noyau issu des Mic au cours de processus impliquant des RPG. Ces processus nécessitent l'introduction de CDB et leur réparation par le NHEJ.

FIGURE II.1 – *Paramécie*

Photos de paramécie montrant sa structure, son dimorphisme nucléaire et sa bouche. Image de gauche, Janine BEISSON, communication personnelle. Image de droite, Pierrick LE BORGNE, communication personnelle.

II.1.1 Cycles de vie de la paramécie

II.1.1.1 Cycle Végétatif

En laboratoire, la paramécie réalise une division végétative toutes les 5-6 heures environ à 27°C (**Figure II.2**). (1) Lors de cette division, les Mic réalisent une mitose en conservant leur enveloppe nucléaire intacte. Le Mac lui se divise de manière amitotique. L'ADN est répliqué mais les chromosomes du Mac ne sont pas ségrégués de manière régulée. Ils ne possèdent pas de centromères et ne se condensent pas au moment de la division. Le Mac s'allonge puis se scinde en deux, séparant de manière aléatoire les copies de chaque chromosome. Ce mode de division nucléaire pourrait conduire à une perte de matériel génétique en perdant l'ensemble des copies d'un gène dans un des noyaux, mais la forte ploïdie permet un maintien global de l'ensemble des chromosomes dans les deux nouveaux Mac. (2) La division cellulaire se termine par la scission entre les deux cellules filles.

Les paramécies peuvent se diviser végétativement jusqu'à 200 fois, avec une diminution progressive de la vitesse de croissance jusqu'à l'arrêt des divisions. Le mécanisme de scission du Mac pourrait être à l'origine de ce phénomène, où un déséquilibre chromosomique finirait finalement par arriver après un grand nombre de divisions. Il y a également un raccourcissement des chromosomes dû à des cassures (GILLEY AND BLACKBURN 1994). Ce mécanisme serait donc un équilibre entre une division facilitée, où la cellule n'a pas à

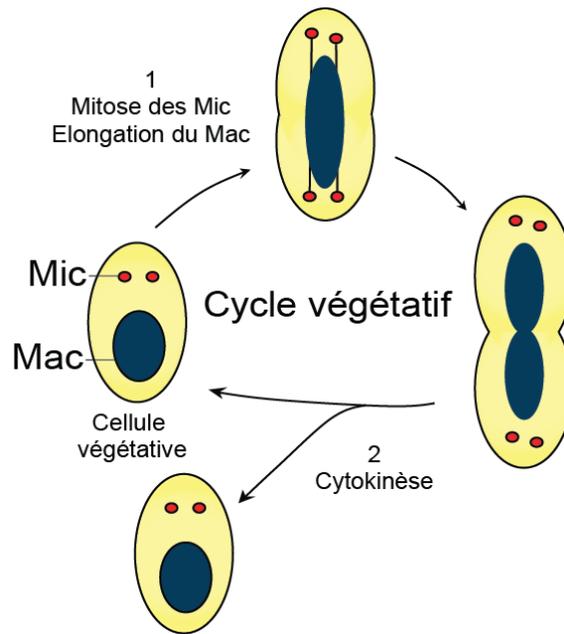


FIGURE II.2 – Cycle végétatif de paramécie

Division végétative d'une paramécie. Le cycle dure environ 5-6h dans des conditions standard de culture au laboratoire.

condenser, aligner, et ségréger finement chacun de ses très nombreux chromosomes, mais aux dépens, sur le long terme, de l'intégrité du Mac.

II.1.1.2 Cycles sexuels

Un mécanisme permettant de "remettre les compteurs à zéro" pour le Mac a évolué chez les ciliés. A chaque nouveau cycle sexuel, un nouveau Mac est formé à partir des Mic. Il existe deux types de cycles sexuels : l'autogamie, qui est une auto-fécondation et la conjugaison, où deux paramécies de types sexuels différents échangent des noyaux gamétiques.

II.1.1.2.1 L'autogamie

L'autogamie peut être déclenchée en laboratoire. Des paramécies "vieilles" d'une vingtaine de divisions végétatives et placées en carence alimentaire (milieu dépourvu de bactérie) commenceront le processus d'autogamie (Figure II.3).

(1) Le cycle commence par une méiose des deux Mic (sans rompre l'enveloppe nucléaire), produisant ainsi 8 noyaux haploïdes. (2) Par un processus inconnu, un des noyaux est sélectionné et subit une mitose donnant deux noyaux gamétiques haploïdes rigoureux-

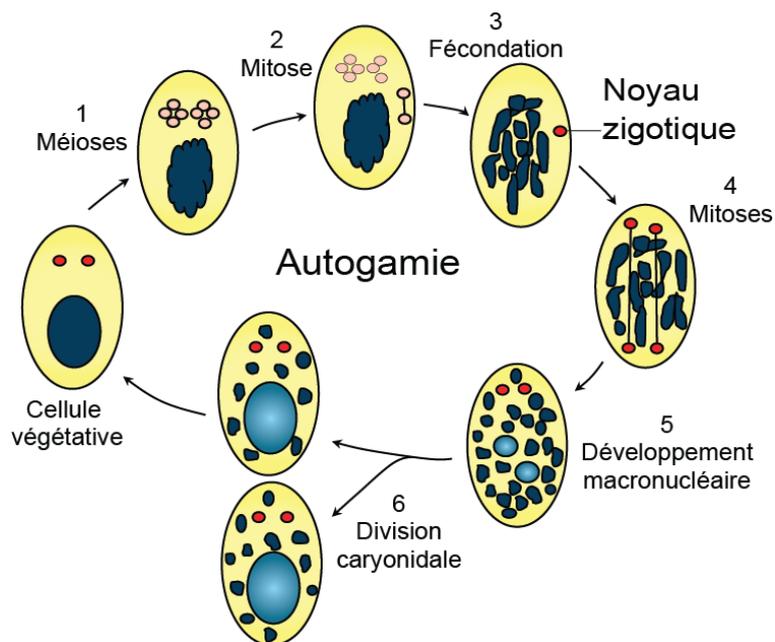


FIGURE II.3 – L'autogamie

Reproduction sexuelle de la paramécie par auto-fécondation.

sement identiques. Les sept autres noyaux sont détruits. (3) Les deux noyaux gamétiques sont ensuite fusionnés, menant ainsi à une étape d'auto-fécondation où le noyau zygotique diploïde créé est 100% homozygote. (4) Le noyau zygotique subit deux mitoses successives. Deux des noyaux produits deviennent les nouveaux Mic et les deux autres sont destinés à devenir les nouveaux Mac. (5) Le développement de ces noyaux diploïdes en Mac contenant 800 copies du génome nécessite de nombreuses réplifications de l'ADN et des réarrangements de celui-ci.

(6) En parallèle des réplifications massives des nouveaux Mac, la cellule réalise une division cellulaire, appelé division caryonidale. La division caryonidale consiste en une mitose des nouveaux Mic, en une ségrégation des deux nouveaux Mac et en une scission des deux nouvelles cellules filles. Cette division est appelée caryonidale car elle va former deux caryonides. Un caryonide est donc une cellule issue d'un événement de développement macronucléaire. Une population caryonidale représente l'ensemble des cellules issues de divisions végétatives à partir d'un caryonide. Une population caryonidale est donc apparentée à une population clonale.

Tout au long du processus d'autogamie, de la méiose des Mic jusqu'au-delà de la division caryonidale, l'ancien Mac se détruit. Cette destruction débute par un débobinage en

parallèle de la méiose puis se poursuit par une fragmentation. L'ensemble des fragments, une trentaine, est ensuite progressivement dégradé. Les quelques fragments pouvant rester après la division caryonidale sont perdus par dilution lors des divisions végétatives. Le cycle complet prend environ 24h en condition standard de culture au laboratoire.

II.1.1.2.2 La conjugaison

La conjugaison est l'autre mode de reproduction sexuelle possible. Contrairement à l'autogamie, il se déclenche avec des cellules "jeunes" qui deviennent "réactives" en condition de carence alimentaire. Il existe deux types sexuels chez la paramécie, le type sexuel O et le E (ou respectivement mt7 et mt8 pour *mating type* chez *P. tetraurelia*). (1) Les deux cellules s'associent au niveau de leur bouche et déclenchent la méiose des Mic (**Figure II.4**).

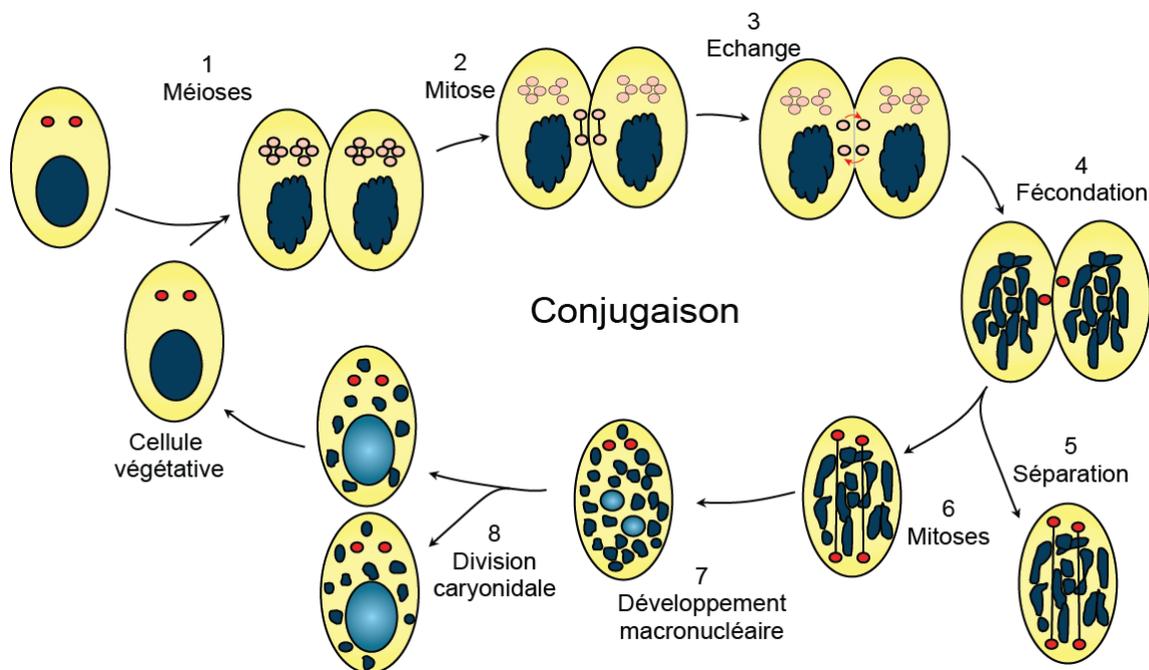


FIGURE II.4 – La conjugaison

Reproduction sexuelle de la paramécie par échange de noyau gamétique avec une paramécie de type sexuel différent.

La grande différence entre l'autogamie et la conjugaison est qu'il y a échange de noyaux gamétiques entre les deux cellules par un pont cytoplasmique formé au niveau des bouches (3). Il y a ensuite fusion entre les deux noyaux, formant un noyau zygotique diploïde non nécessairement homozygote si les deux parents possèdent des génotypes

différents. Les cellules se séparent ensuite et le reste du cycle se poursuit de manière similaire à l'autogamie.

II.1.2 Les duplications globales du génome

Il est arrivé de nombreuses fois au cours de l'évolution qu'un doublement du matériel génétique ait été conservé chez certaines espèces. On appelle ce phénomène une duplication globale du génome, ou *Whole Genome Duplication* (WGD). Bien que majoritairement létales, ces WGD, quand elles sont maintenues, peuvent être de véritables accélérateurs évolutifs. En effet, lors d'une duplication, la redondance des gènes permet la dérive évolutive d'une des versions. Majoritairement, une des copies est perdue (par pseudogénéisation ou par délétion), mais il peut également arriver qu'une des copies se spécialise pour une nouvelle fonction. Dans certain cas également, les copies sont maintenues car la protéine associée au gène nécessite une forte expression. Les gènes paralogues qui sont issus d'une WGD sont appelés ohnologues.

Le séquençage du génome de *P. tetraurelia* a révélé qu'elle a subi 3 WGD au cours de son évolution (AURY ET AL. 2006). Elles sont appelées WGD récente, intermédiaire et ancienne. Parmi les 40 000 gènes, seuls 32% sont présents en une seule copie. La polypléidie du Mac de la paramécie a pu participer à sa résistance pour les WGD. L'expression des gènes se faisant dans les Mac, il suffit "simplement" de supprimer une étape de réplication de l'ADN au moment du développement du Mac pour arriver à la pléidie antérieure à la WGD.

II.1.3 Organisation générale du génome de Paramécie

Le génome macronucléaire de paramécie est compact avec 78% de séquences codantes, des régions intergéniques très courtes (352 bases en moyenne) et des petits introns (25 bases en moyenne) (AURY ET AL. 2006). Elle possède cependant beaucoup d'introns par rapport aux autres unicellulaires eucaryotes mais semble incapable de réaliser de l'épissage alternatif. Les séquences régulatrices de l'expression des gènes ne sont pas définies, mais l'utilisation des régions intergéniques d'un gène donné dans un transgène permet généralement de reproduire le profil d'expression endogène.

Concernant la nomenclature appliquée à la paramécie, les gènes sont notés en majus-

cules et italiques (*PGM*). Les noms des protéines ont la première lettre majuscule et ne sont pas en italiques (*Pgm*). Et enfin pour différencier les ohnologues entre eux, une lettre de l'alphabet est ajoutée en minuscule à la fin, même pour les gènes (*PGML4a*, *Pgml4a*).

La paramécie est l'un des rares contre-exemples à l'universalité du code génétique. Seul le codon UGA est un codon STOP. Les codons UAG et UAA codent pour une glutamine (CARON AND MEYER 1985, PREER ET AL. 1985).

II.2 LES RÉARRANGEMENTS PROGRAMMÉS DU GÉNOME CHEZ LA PARAMÉCIE

La complexité du génome des Mic est d'environ 100 Mb (GUÉRIN ET AL. 2017) alors que celle du Mac est de 72 Mb (AURY ET AL. 2006). Lors du développement du Mac, il y a donc élimination d'environ 28 Mb d'ADN germinale. Il existe deux types de réarrangements, les réarrangements précis et imprécis.

II.2.1 Les réarrangements précis : élimination des IES

Les IES, pour *Internal Eliminated Sequences*, sont des séquences éliminées précisément lors du développement du Mac. On en dénombre 45 000 par génome haploïde que l'on retrouve en moyenne tous les 1,6 kb et qui représentent 3 Mb d'ADN éliminé (ARNAIZ ET AL. 2012). On les retrouve dans les régions intergéniques mais également dans les gènes : 47% des gènes sont interrompus par au moins une IES. Leur excision précise est donc essentielle à l'expression de protéines fonctionnelles (Figure II.5).

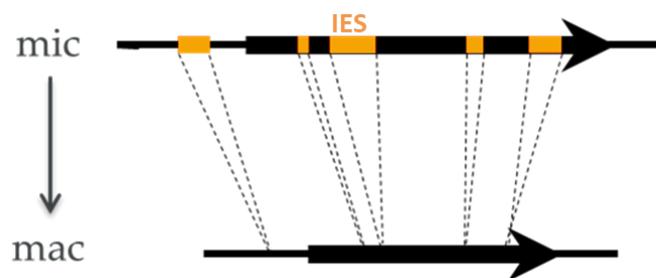


FIGURE II.5 – Les IES

Différence entre le génome des Mic et celui des Mac. Le chromosome est en noir, le gène est une flèche et les IES sont en orange. Les IES sont éliminées précisément du génome Mac lors des RPG.

Les IES peuvent avoir des tailles très variables, allant de quelques dizaines à plusieurs milliers de bases. Cependant, la répartition des tailles n'est pas aléatoire (**Figure II.6A**).

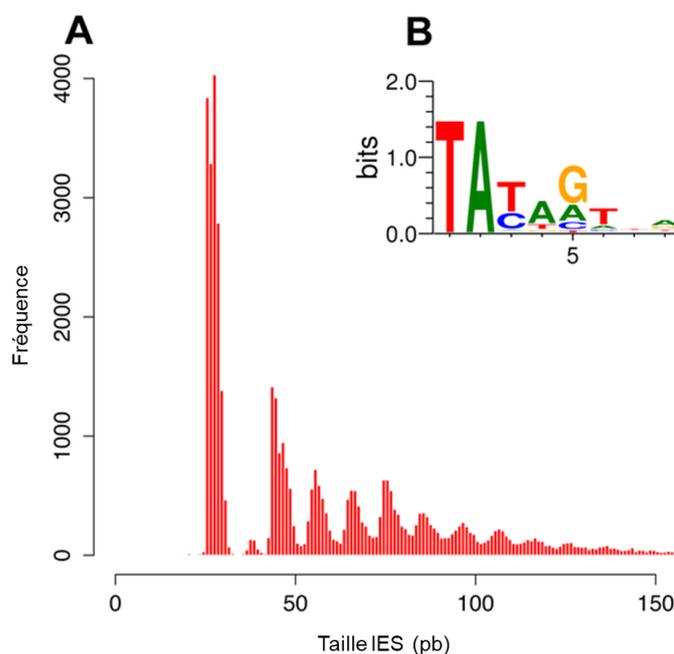


FIGURE II.6 – Répartition des tailles d'IES et séquence aux bornes

(A) Fréquence d'IES en fonction de leur taille. (B) Séquence consensus aux bornes des IES. Le TA est essentiel à l'excision pour l'ensemble des IES, la suite du consensus est moins stricte. Il se rapproche néanmoins du consensus des transposons Tc1/mariner (TACAGTKS avec K=G/T et S=C/G) (ARNAIZ ET AL. 2012).

Plusieurs caractéristiques très intéressantes sont à noter concernant la répartition des tailles d'IES. On observe une périodicité des pics de 10 pb, correspondant à un tour d'hélice de l'ADN. De plus, on ne trouve pas d'IES plus petite que 26 pb et très peu d'IES entre 32 et 45 pb. L'ensemble de ces éléments indique une contrainte topologique lors de l'excision des IES

De plus, 93% des IES sont plus petites que 150 pb et 35% ont une taille de 26-30 pb. L'ensemble des IES sont bornées à chacune de leurs extrémités par un di-nucléotide TA. Il existe un faible consensus pour les bases adjacentes au TA, ressemblant au consensus des extrémités des éléments transposables de la famille Tc1/mariner (**Figure II.6B**). Certaines IES parmi les plus grandes (plus de 1,5 kb) portent des gènes présentant une homologie avec ceux des transposases de la famille Tc1/mariner.

Ces éléments donnent des indications sur l'origine et l'évolution des IES : Les IES proviendraient d'une ou plusieurs invasions massives du génome, au moins en partie

par des transposons Tc₁/mariner. Les séquences auraient ensuite dégénéré et se seraient progressivement raccourcies, jusqu'à atteindre la taille minimale de 26 pb contrainte par le complexe d'excision. Les IES les plus courtes seraient donc les plus anciennes et les plus longues celles intégrées le plus récemment.

Il aurait pu donc être logique de soupçonner une transposase de type Tc₁/mariner comme étant responsable de l'excision des IES. Cependant, le mécanisme d'excision des Tc₁/mariner laisse les deux TA sur le chromosome. Or, la cicatrice retrouvée dans le génome du Mac après excision d'une IES est un unique di-nucléotide TA. Une transposase Tc₁/mariner ne peut donc pas être responsable de l'excision des IES.

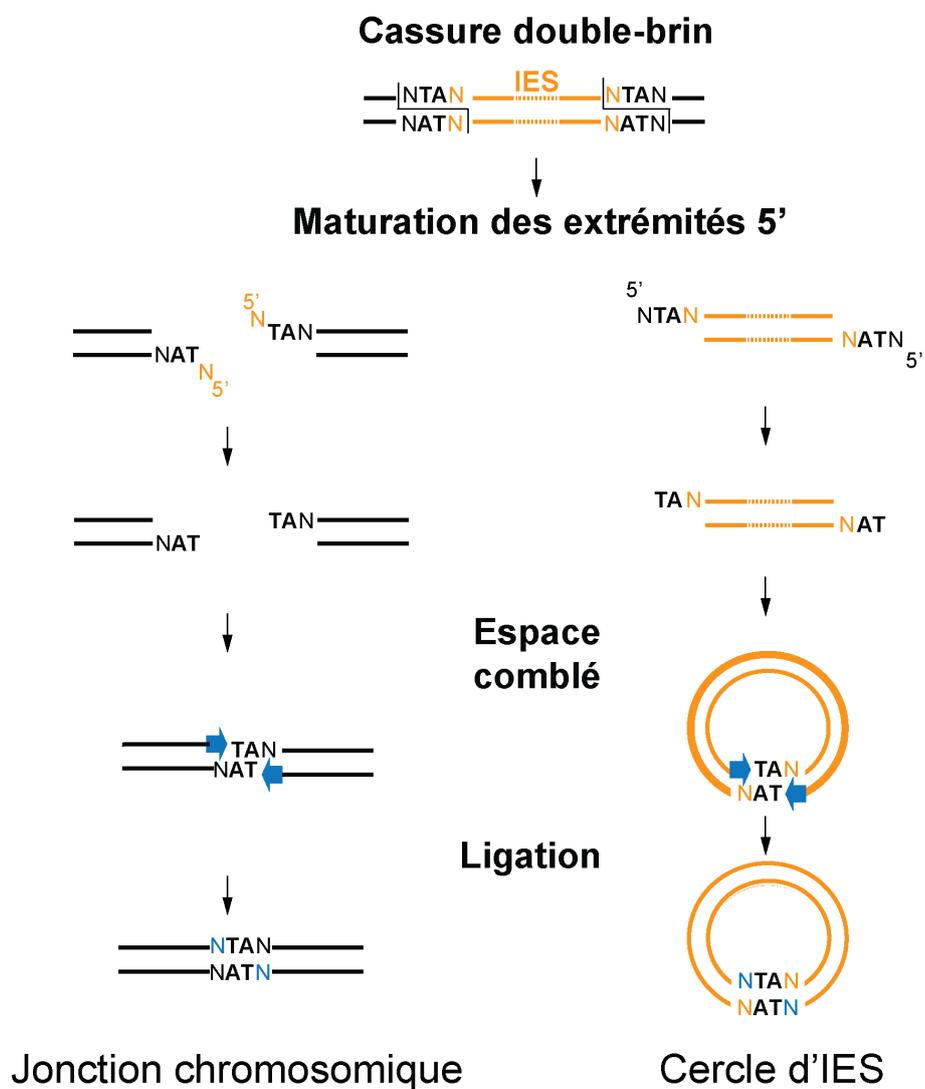
II.2.1.1 Clivage des IES et réparation des cassures introduites

Mécanismes moléculaires L'étude des mécanismes moléculaires de la coupure a pu montrer que les IES étaient clivées à chacune des extrémités, en laissant 4 bases sortantes en 5', centrées sur le TA (GRATIAS AND BÉTERMIER 2003) (**Figure II.7**).

Le nucléotide en 5' est ensuite éliminé et la synapse se forme grâce à la micro-homologie entre les deux TA. Cette synapse permet le comblement de l'espace et enfin la ligation. Ces mécanismes moléculaires ont lieu à la fois sur le chromosome et sur l'IES. En effet, si sa taille le permet, elle peut être circularisée et être détectée par PCR au cours des RPG (BÉTERMIER ET AL. 2000). Les IES trop courtes pour être circularisées peuvent être concatémérisées avec d'autres IES excisées puis circularisées (ALLEN ET AL. 2017).

Les acteurs du clivage La connaissance de la géométrie de la coupure ainsi que la disponibilité d'une première version annotée du génome (AURY ET AL. 2006) ont donné les moyens de procéder à la recherche de l'endonucléase responsable par une approche de gène candidat.

PiggyMac, ou Pgm, est une protéine dérivant de la transposase PiggyBac. Les transposases PiggyBac ont pour substrat des séquences TTAA aux bornes de leur transposon et laissent une cicatrice TTAA précise dans le chromosome une fois le transposon excisé de son site donneur (ELICK ET AL. 1996). Le gène *PGM* est présent en une seule copie dans le génome de la paramécie et n'est pas dans une structure transposable. Comme RAG₁ pour le V(D)J, Pgm a donc été domestiquée par la cellule (BAUDRY ET AL. 2009)

FIGURE II.7 – *Excision des IES*

Modèle de l'excision des IES. Deux CDB sont introduites aux bornes de l'IES, avec quatre nucléotides sortant en 5', centré sur le TA. La base en 5' est clivée, le chromosome apparié par micro-homologie, l'espace est comblé puis le chromosome reliqué. La réparation se produit aussi sur les IES, pouvant donc les circulariser.

Des expériences d'ARN interférence dirigées contre *PGM* pendant l'autogamie ont montré une inhibition complète des RPG, avec rétention de l'ensemble des IES dans le génome du nouveau Mac (ARNAIZ ET AL. 2012, GUÉRIN ET AL. 2017). Les réplifications successives de l'ADN ne sont pas impactées pendant le développement du nouveau Mac mais la descendance n'est pas viable, la moitié des gènes étant interrompus par au moins une IES et n'étant donc pas capables d'exprimer une protéine fonctionnelle. Une ARNi contre *PGM* provoque donc la mortalité des cellules post-autogames en conséquence de l'absence d'introduction des CDB aux bornes des IES.

PiggyBac est une transposase possédant un domaine catalytique avec trois acides aminés Aspartate (D) essentiels pour l'activité endonucléase (MITRA ET AL. 2008) et un domaine C-terminal en doigt de zinc également essentiel pour son activité (MORELLET ET AL. 2018) (**Figure II.8A**).

Pgm possède le domaine catalytique, avec les trois D et un domaine en doigt de zinc variant (DUBOIS ET AL. 2017). Il possède également des domaines N et C-terminaux supplémentaires. L'extension C-terminale est prédite pour être un *coiled-coil*.

Des expériences de complémentation fonctionnelle ont permis de vérifier le rôle de certains domaines de Pgm (DUBOIS ET AL. 2017). Les expériences de complémentation utilisaient l'expression de différents mutants de Pgm lors des RPG en absence du Pgm endogène. En particulier, une triple mutation des D catalytiques en Alanine (A) inhibe complètement l'introduction des CDB aux bornes des IES. Cette expérience a permis de préciser le rôle de Pgm. Bien qu'essentielle aux RPG, il n'était pas sûr qu'elle soit directement en charge de la coupure de l'ADN. Cette nouvelle donnée laisse donc peu de doutes sur le rôle catalytique direct de Pgm lors de l'introduction des cassures.

Cette même étude a pu démontrer des capacités de multimérisation de Pgm avec lui-même, comme attendu pour une protéine de type transposase.

Différentes approches ont été utilisées pour déterminer la localisation cellulaire de Pgm lors des RPG. La première fut de créer des transgènes *GFP-PGM* (BAUDRY ET AL. 2009, DUBOIS ET AL. 2017) et la plus récente a été de développer un anticorps dirigé spécifiquement contre la protéine endogène (DUBOIS ET AL. 2017) (**Figure II.9**). Ces deux méthodes ont révélé que Pgm se localise spécifiquement dans les Mac en développement, également appelés ébauches.

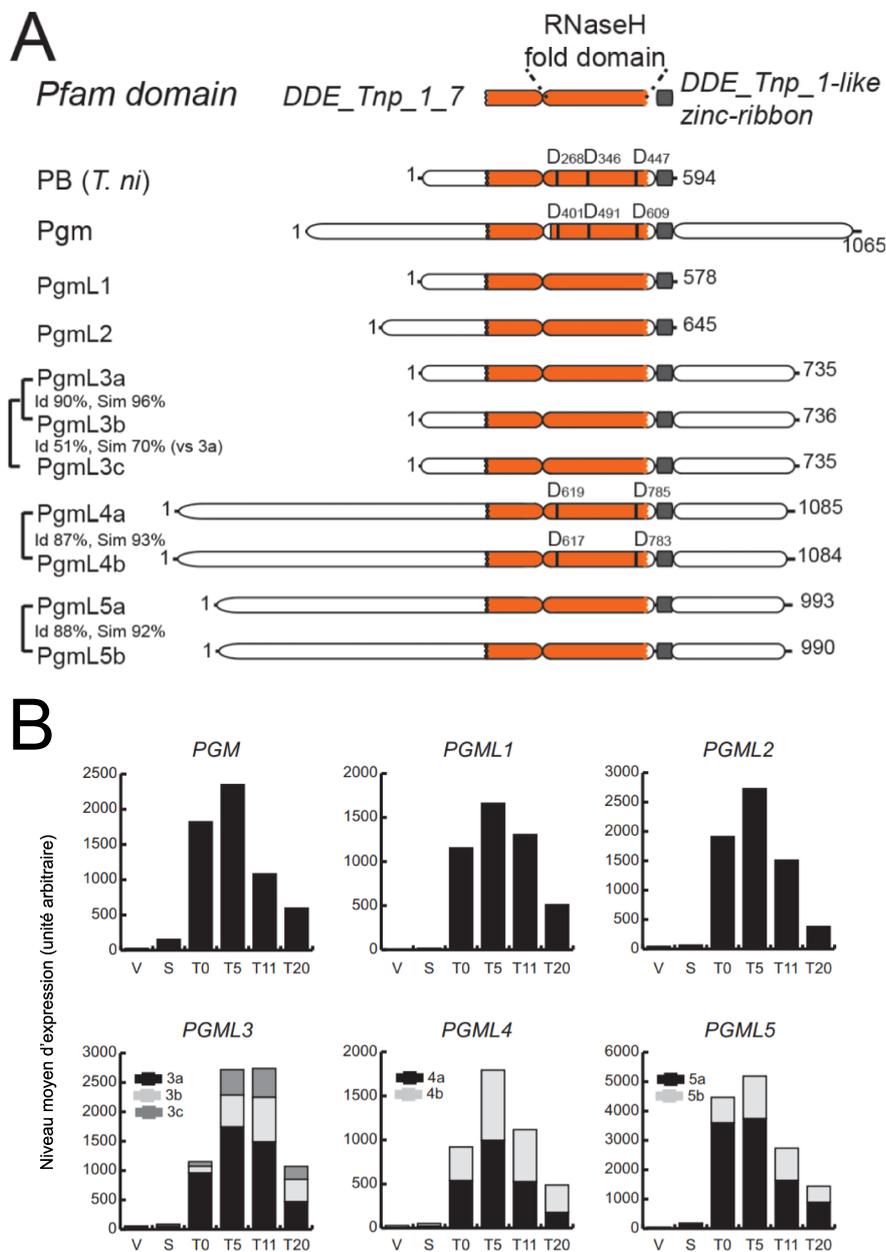


FIGURE II.8 – *Pgm* et les *PgmL*

(A) Domaines de PiggyBac (PB), *Pgm* et des *PgmL*. La position des éventuels acides aminés D catalytiques est affichée. (B) Niveau d'expression des ARNm de *PGM* et des *PGML* par RNAseq, en unité arbitraire. V = cellule végétatives, S = stade méiotique, T = temps en heure depuis le début de l'autogamie. Adaptée de BISCHEROUR ET AL. (2018).

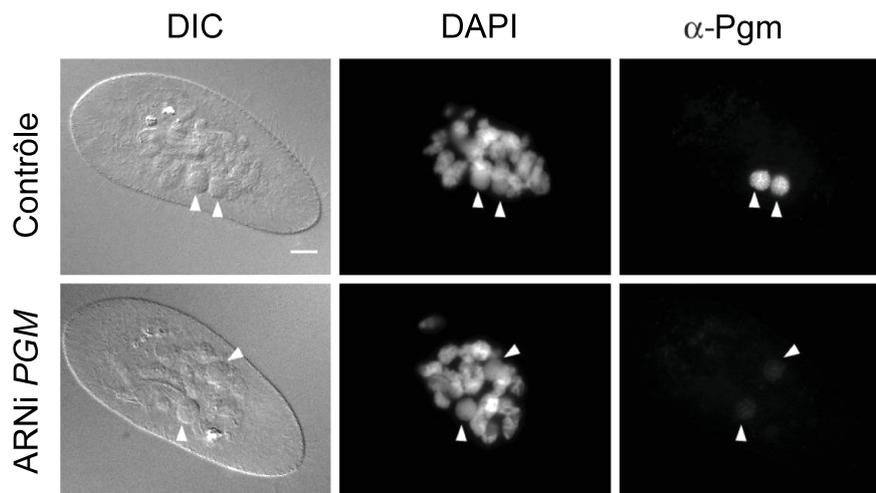


FIGURE II.9 – Localisation nucléaire de Pgm et validation de la spécificité de l'anticorps

Immunomarquage de Pgm dans de jeunes cellules autogames en ARNi contrôle ou PGM. Les ébauches sont indiquées par les flèches blanches. L'absence de signal dans l'ARNi contre PGM indique que l'anticorps reconnaît spécifiquement Pgm. La barre d'échelle est de 10 μm . Adaptée de BISCHEROUR ET AL. (2018).

Une recherche par modèle de Markov caché (comparaison de prédiction de structure + homologie de séquence) à partir de Pgm dans le protéome de paramécie a permis de détecter une série de protéines homologues (BISCHEROUR ET AL. 2018). Ces protéines ont été nommées PiggyMac-Like, ou PgmL (**Figure II.8A**). On retrouve ainsi 9 protéines, réparties en 5 groupes (en raison des WGD). Comme Pgm, elles sont toutes spécifiquement exprimées lors des RPG (**Figure II.8B**). Elles ne possèdent aucun des acides aminés catalytiques, sauf PgmL4a et b qui en possèdent deux sur trois. La régulation de leur expression indique un rôle potentiel dans les RPG mais l'absence d'acide aminé catalytique semble exclure un rôle direct.

Des expériences de déplétion par ARNi ont été effectuées sur l'ensemble des *PGML*. Chacun des 5 groupes de PgmL est essentiel pour le bon déroulement de l'autogamie. Différentes approches de microscopie (anticorps spécifiques et/ou fusions protéiques avec étiquette Flag) démontrent que l'ensemble des PgmL se localisent dans les ébauches. Des analyses bio-chimiques montrent également que chaque groupe de PgmL peut former des complexes *in vitro* avec Pgm. Enfin, des séquençages du génome macronucléaire post-RPG en condition d'ARNi contre les *PGML* montrent un effet global de l'absence de ces protéines sur les RPG. L'ensemble des IES sont spécifiquement retenues, indiquant une

absence d'introduction des CDB. PgmL₁ et 3 semblent faire légèrement exception, avec respectivement 17% et 8% des IES excisées efficacement mais avec un taux d'erreur élevé dans le positionnement des bornes d'excision (BISCHEROUR ET AL. 2018).

Suspectant un effet sur Pgm lorsque le phénotype de rétention se montre global, la localisation nucléaire de Pgm a été observée en fonction des différentes ARNi contre les *PGML* (**Figure II.10**).

Alors qu'aucune des PgmL n'est nécessaire à la stabilité protéique de Pgm (**Figure II.10B**), elles sont toutes nécessaires à sa localisation nucléaire (**Figure II.10A et C**). En absence de Pgm stable dans l'ébauche, il est difficile d'imaginer de possibles introductions de CDB.

La déplétion du signal Pgm nucléaire en ARNi contre *PGML1* est intermédiaire. Ce résultat est en corrélation avec celui du séquençage génomique. Quelques IES restent excisées, donc il doit rester du Pgm nucléaire au moins partiellement actif. Le résultat pour PgmL₃ soulève plus de questions. Malgré une absence nucléaire visible de Pgm, 8% des IES sont encore excisées. Une manière d'expliquer cela serait que l'on se trouve peut-être à la limite du seuil de détection de l'immunofluorescence. Une petite fraction de molécules Pgm présentes mais non détectées pourrait suffire à l'excision de ces quelques IES.

L'excision des IES nécessite donc un ensemble de protéines dérivant de transposases PiggyBac. Elles ont été domestiquées et se sont spécialisées pour les RPG, en étant capables entre autre de reconnaître une séquence différente de la séquence consensus du transposon d'origine, ne portant pas de motif conservé évident en dehors du TA. Le complexe d'excision serait composé d'au moins six sous-unités dont une catalytiquement active, Pgm. Chacune des sous-unités permet au complexe de se localiser stablement dans les noyaux en développement. Enfin, bien qu'indispensables, les sous-unités PgmL₁ et 3 seraient les plus accessoires.

Les acteurs de la réparation Les RPG ont lieu en parallèle de l'amplification du génome. Ainsi, l'excision des IES démarre alors que l'ébauche est déjà à une ploïdie autour de 16-32n (BÉTERMIER ET AL. 2000). Cette ploïdie augmente mécaniquement le nombre d'IES à exciser et donc d'événements de coupure et de réparation. On peut donc estimer qu'il y

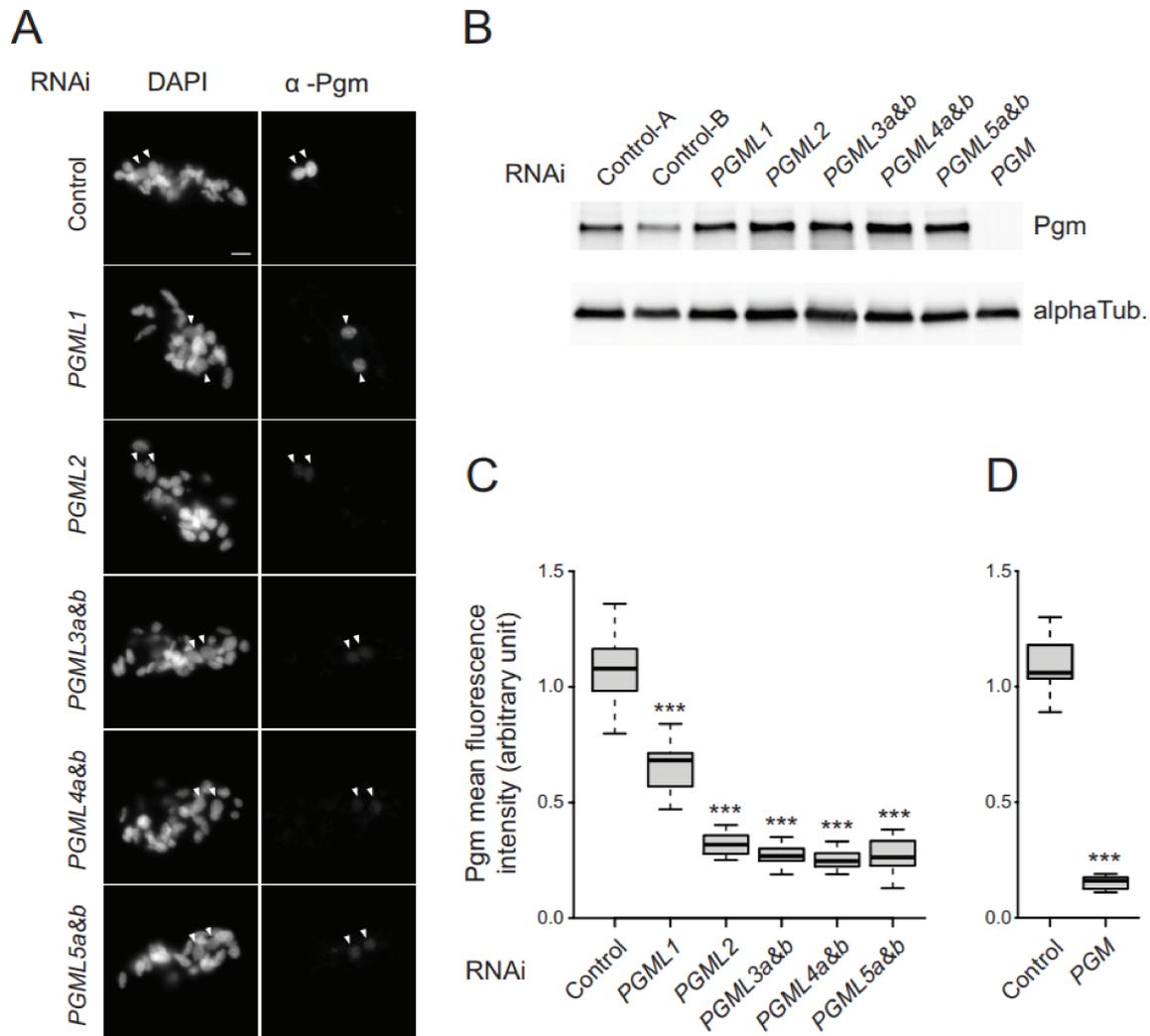


FIGURE II.10 – Délocalisation de Pgm en ARNi PGML

Expression et localisation de Pgm en ARNi PGML. (A) Immunomarquage de Pgm dans de jeunes cellules autogames en situation contrôle ou ARNi contre les différentes PGML. Les ébauches sont indiquées par les flèches blanches. La barre d'échelle est de 10 μ m. (B) Western-Blot (WB) anti-Pgm montrant la stabilité du niveau de Pgm lors d'ARNi contre les PGML. (C) Représentation en *boxplot* de l'intensité de fluorescence de Pgm en ARNi PGML dans des ébauches. (D) Représentation en *boxplot* de l'intensité de fluorescence de Pgm en ARNi PGM dans des ébauches. Les *** indiquent que $p < 0,001$ par rapport au contrôle dans un test statistique de Mann-Whitney. Adaptée de BISCHEROUR ET AL. (2018).

a plus d'un million d'événements distincts. Dans ces conditions extrêmes, il est crucial d'avoir un mécanisme de réparation très efficace et rapide, tout en étant extrêmement fidèle du fait de la présence d'IES dans les gènes.

Les travaux d'Aurélié Kapusta, ancienne doctorante de l'équipe, ont cherché à déterminer quelle voie de réparation pouvait être utilisée (KAPUSTA ET AL. 2011). Les protéines Lig4 et Xrcc4 ont pu être identifiées par recherche d'homologie avec les protéines humaines. La paramécie possède deux orthologues de Lig4, Lig4a et b et une copie unique de Xrcc4. Les deux gènes *LIG4* sont identiques à 94% entre eux. Il n'a donc pas été possible de réaliser des ARNi strictement spécifiques d'un seul des deux gènes. Les gènes *LIG4* et *XRCC4* sont exprimés fortement lors des RPG. Des expériences d'ARNi contre *LIG4* et *XRCC4* ont conduit à un blocage des RPG et à une létalité pour la descendance sexuelle. Contrairement à la déplétion du complexe d'excision qui provoque la rétention des IES, une déplétion de Lig4 ou Xrcc4 provoque une accumulation de CDB non réparées et une inhibition de l'amplification du génome dans les nouveaux Mac. Une analyse de la composition des extrémités d'ADN a permis de déterminer que la base 5' sortante était toujours clivée mais que le comblement de l'espace vide au niveau de la jonction était inhibé. Lig4 et Xrcc4 participent donc au recrutement ou à l'activation de la polymérase à ADN nécessaire pour le comblement de l'espace, avant de procéder à la ligation du chromosome.

En absence de Lig4 ou Xrcc4, non seulement les jonctions chromosomiques n'ont plus lieu mais la circularisation des IES est également inhibée. Ces travaux ont mis en évidence le rôle crucial de ces protéines spécifiques du NHEJ, et donc du NHEJ lui-même dans les RPG chez la paramécie.

L'excision des IES un exemple extraordinaire de NHEJ précis. La ligation est précise au nucléotide près, sans introduction d'erreurs et cela de manière reproductible sur plus d'un million d'événements, ayant lieu dans une courte période de temps et à proximité directe les uns des autres. La paramécie est donc un formidable modèle pour l'étude du NHEJ.

II.2.1.2 Couplage entre coupure et réparation de l'ADN, le cas de Ku chez la paramécie

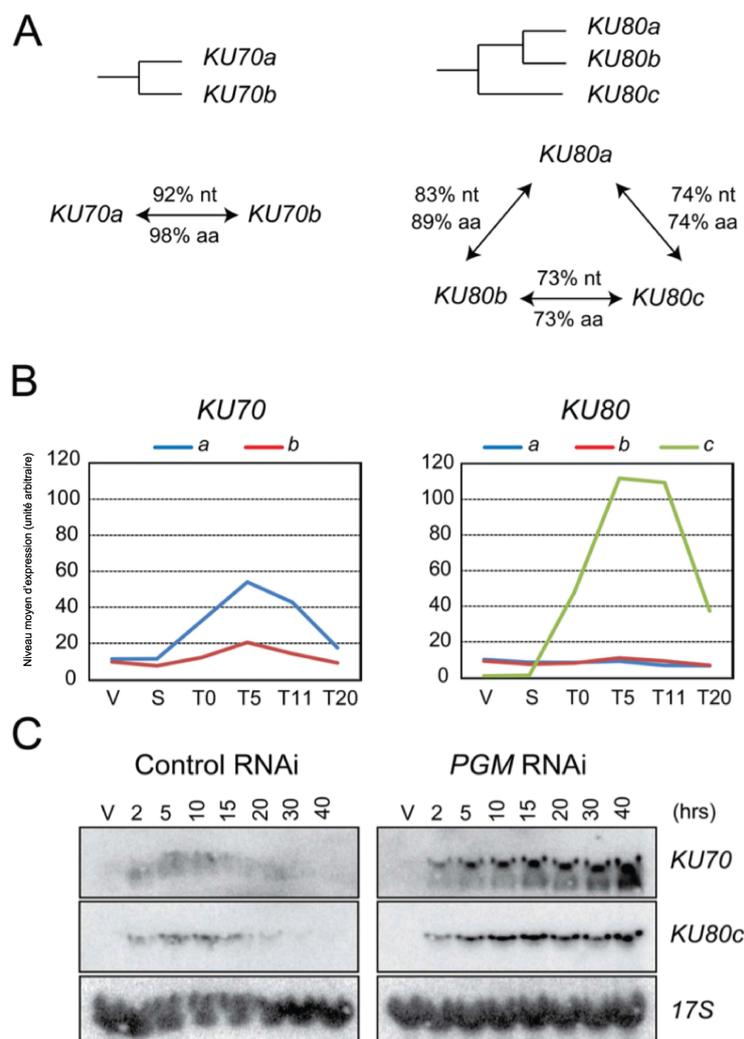
Les Ku de paramécie Dans la poursuite de l'étude du NHEJ chez la paramécie, les travaux d'Antoine Marmignon, ancien doctorant dans l'équipe également, l'ont dirigé vers les protéines Ku de la paramécie. Des recherches par homologie avec les protéines humaines ont permis d'identifier cinq protéines (MARMIGNON ET AL. 2014). Deux orthologues Ku70 et trois Ku80. Ku70a et Ku70b sont extrêmement proches l'une de l'autre, avec 92% d'identité sur la séquence génique et 98% sur la séquence protéique (**Figure II.11A**).

Les trois copies de Ku80 sont un peu plus divergentes. Ku80a et Ku80b sont relativement proches, avec 83% d'identité sur la séquence génique et 89% sur la séquence protéique alors que Ku80c diverge un peu plus avec 74% d'identité génique et protéique avec Ku80a et 73% d'identité génique et protéique avec Ku80b. Ku70a et b sont exprimées en croissance végétative et Ku70a particulièrement est sur-exprimée lors de l'autogamie (**Figure II.11B**). Ku80a et b sont exprimées de manière constitutive alors que Ku80c n'est pas exprimée dans des cellules végétatives et est fortement exprimée lors de l'autogamie.

Lors d'une ARNi PGM, les gènes *KU70* et *KU80c* sont sur-exprimés aux temps tardifs, c'est à dire après la fin des RPG en situation contrôle (**Figure II.11C**). Cette dérégulation a été observée depuis pour d'autres gènes comme les *PGML* (BISCHEROUR ET AL. 2018) et pourrait, plus généralement, être le signe qu'un gène est impliqué dans les RPG.

Localisation nucléaire de Ku Des constructions *GFP-KU70a* et *GFP-KU80c* ont été utilisées pour déterminer la localisation des protéines Ku70a et Ku80c (**Figure II.12**). Les deux protéines se localisent dans les ébauches. De plus, leur localisation nucléaire n'est pas uniforme. Des points de concentration, ou foci, se forment, particulièrement visibles avec GFP-Ku70.

Effets des ARNi KU Des expériences d'ARNi contre l'ensemble des gènes *KU* ont été réalisées (**Figure II.13**). *KU70a* et *KU70b* étant trop proches, les deux gènes sont impactés de la même manière par l'ARNi. L'ARNi *KU70* a conduit à une mortalité de la descendance sexuelle. L'ARNi *KU80a* n'a pas donné de phénotype particulier, tout comme l'ARNi *KU80b* ainsi qu'une double ARNi contre *KU80a* et *KU80b*. L'ARNi *KU80c* a elle

FIGURE II.11 – Les *Ku* de paramécie

(A) Diagramme des relations de WGD entre les gènes *KU70* ou *KU80* chez *P. tetraurelia*. Le % d'identité entre les gènes (nt) ou les protéines (aa) sont indiqués le long des flèches connectant les gènes. (B) Profils de transcription exprimés en unité arbitraire des gènes *KU70* et *KU80* durant une autogamie, déterminés par RNA-seq haut-débit. V : végétative; S : stade méiotique; T₀ : 50% des cellules ont l'ancien Mac fragmenté; les T suivants sont les temps en heures après T₀. (C) Détection des ARNm de *KU70* et *KU80c* durant l'autogamie par hybridation de northern blot en ARNi contrôle ou *PGM*. Les temps en heures sont basés sur le T₀. Adaptée de MARMIGNON ET AL. (2014).

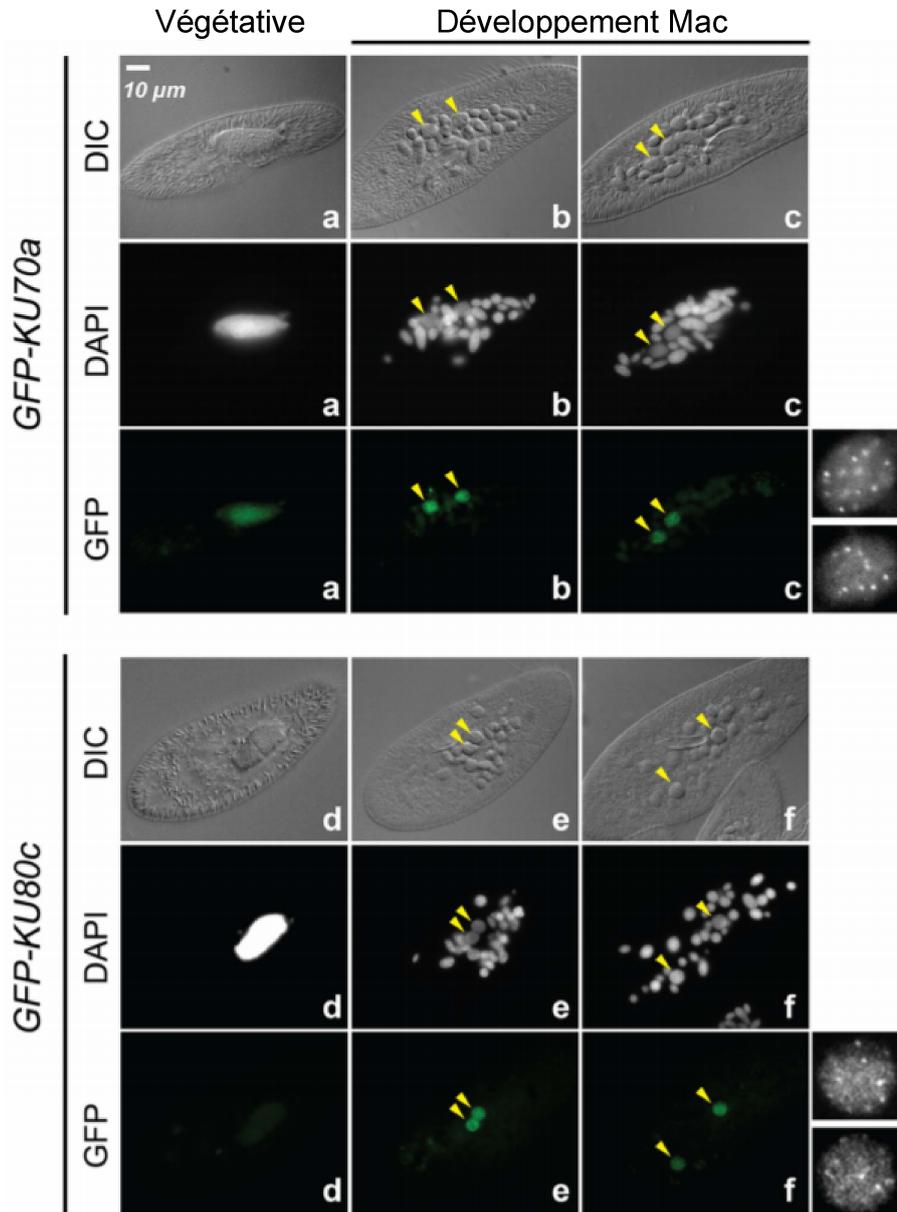


FIGURE II.12 – Localisation nucléaire de Ku

Les cellules ont été micro-injectées avec les transgènes de fusion exprimant GFP-KU70a (a, b, c) et GFP-Ku80c (d, e, f) sous le contrôle de leur signaux de transcription respectifs. Les ébauches sont indiquées par les flèches jaunes, les autres noyaux sont les fragments de l'ancien Mac. Adaptée de MARMIGNON ET AL. (2014).

présenté un fort phénotype de mortalité de la descendance sexuelle. Ceci indique donc que Ku80c joue un rôle spécifique pendant les RPG.

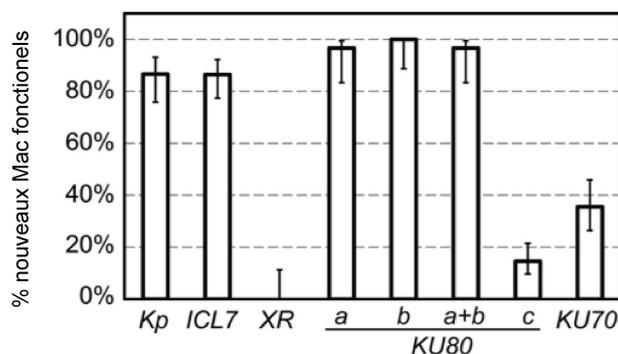


FIGURE II.13 – Effets d'une ARNi KU

Pourcentage de cellules survivantes avec un nouveau Mac fonctionnel après une autogamie sous différentes conditions d'ARNi. *Kp* et *ICL7* : contrôles positifs; *XR* : ARNi contre *XRCC4* (contrôle négatif); ARNi contre *KU80a*, *b*, *a+b*, *c* ou *KU70*. Adaptée de MARMIGNON ET AL. (2014).

Effets moléculaires d'une ARNi KU Les protéines Ku70 et Ku80c sont donc essentielles lors de l'autogamie et sont localisées dans les ébauches. Il était attendu à ce que leur rôle dans les RPG soit directement connecté à celui de Lig4 et Xrcc4 : la réparation des CDB. Le phénotype moléculaire associé à un tel rôle est l'introduction des coupures par Pgm et les PgmL suivi d'une accumulation de cassures, due à l'inhibition du NHEJ. L'analyse de l'ADN de paramécie en autogamie montre qu'en ARNi *KU80c* les trois IES testées sont retenues dans le génome (**Figure II.14**).

De plus, aucune extrémité d'ADN n'a pu être détectée aux bornes de ces IES et l'amplification du nouveau Mac n'a pas été inhibée. Ces résultats diffèrent grandement de ceux obtenus avec une ARNi *LIG4* et sont comparables à ceux obtenus en ARNi *PGM*.

En absence de Ku, Pgm est donc inactivée. Ku a donc un rôle antérieur à la prise en charge des CDB en étant essentiel à l'introduction des cassures. Il y a donc ici un couplage strict entre le mécanisme d'introduction des CDB et celui de réparation de l'ADN. Les deux mécanismes auraient co-évolué pour s'assurer qu'une réparation est possible avant d'introduire de graves dommages à l'ADN.

Interaction entre Ku et Pgm Dans le cadre de l'étude de ce couplage, l'interaction entre les protéines Ku et Pgm a été testée par co-immunoprécipitation en utilisant des protéines

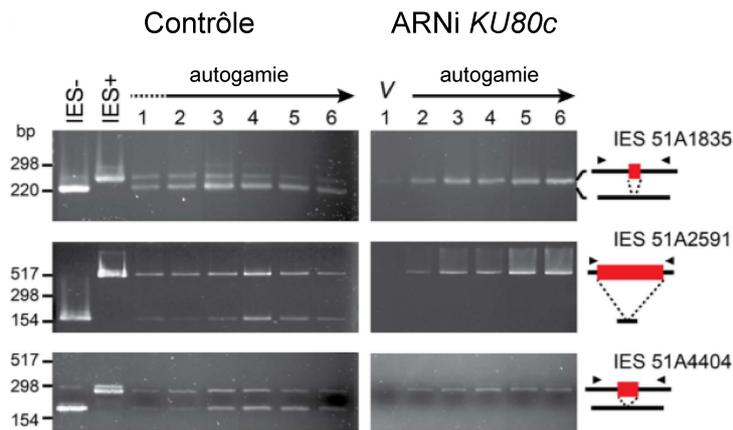


FIGURE II.14 – Rétention d'au moins 3 IES en ARNi KU80c

Détection par PCR de jonctions *de novo* d'excision d'IES pendant l'autogamie, en situation contrôle ou en ARNi KU80c. Adaptée de MARMIGNON ET AL. (2014).

produites dans des cellules d'insectes (Figure II.15). Que ce soit seuls ou en hétérodimère, Ku70a et Ku80c co-précipitent avec Pgm.

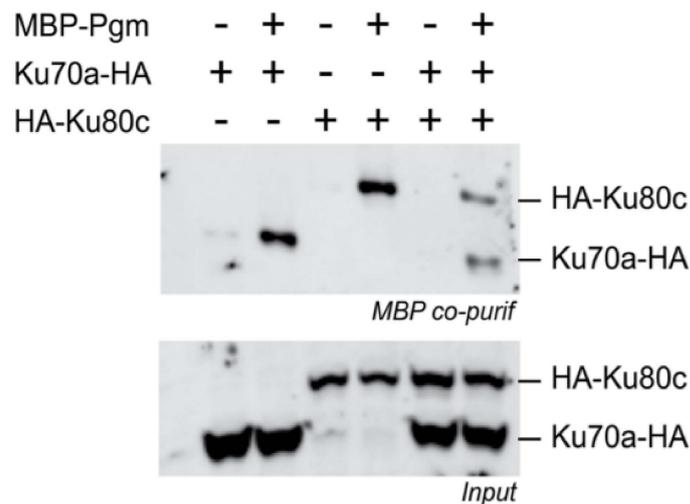


FIGURE II.15 – Interaction entre Ku et Pgm

Co-précipitation de HA-Ku80c et Ku70a-HA avec MBP-Pgm exprimées en cellules d'insectes, révélées par WB en utilisant un anticorps anti-HA. Adaptée de MARMIGNON ET AL. (2014).

Modèle de l'excision des IES L'ensemble de ces résultats ont permis de définir un modèle pour l'excision des IES (Figure II.16). L'hétérodimère Ku70/Ku80c interagit avec Pgm, permettant d'activer le complexe d'excision composé de Pgm et des PgmL et donc d'introduire les CDB aux bornes de l'IES. L'hétérodimère Ku est ensuite directement

chargé sur l'ADN et permet une protection rapide et efficace des extrémités cassées. Le reste du NHEJ est ensuite recruté : DNA-PKcs, une ou plusieurs nucléases responsables de la maturation des extrémités (encore non identifiée(s)) et enfin Lig4 et Xrcc4 permettant l'activation d'une ou plusieurs polymérases pour combler l'espace suivi de la ligation des fragments chromosomiques par Lig4.

Des travaux préliminaires de Mireille Bétermier ont mis en évidence des protéines homologues de polymérases de la famille X ayant un rôle dans les RPG. De plus, des données non publiées mettent en évidence des protéines homologues de XLF chez la paramécie, ayant également un rôle dans les RPG.

II.2.1.3 Reconnaissance des IES

Rôle des ARN non codants Une grande problématique posée par l'excision des IES est leur reconnaissance par le complexe d'excision. En effet, la seule séquence strictement essentielle pour la définition d'une IES est la présence de di-nucléotides TA à ses bornes. Or, le génome de paramécie est composé à 73% de nucléotides A/T et est donc rempli de di-nucléotides TA. Malgré cela, Pgm parvient à introduire une coupure précise. De nombreuses équipes ont donc cherché s'il existait un mécanisme capable de définir une IES autrement que par ses TA. L'ensemble des résultats a permis de définir le modèle du *scanning* par des petits ARN scan (scnRNA) (**Figure II.17**).

Ce modèle décrit l'utilisation de petits ARN essentiels à l'excision d'une partie des IES. L'ensemble du génome macronucléaire est transcrit en longs ARN non codants (ncRNA maternels) de manière constitutive (pendant la croissance végétative et le cycle sexuel). Spécifiquement lors de la méiose des Mic, le génome micronucléaire est transcrit, produisant de long ARN double-brins couvrant l'ensemble du génome germinale non réarrangé. Ces ARNdb sont clivés par les protéines Dcl2/Dcl3 en petits ARNdb de 25 nucléotides, les scnRNA (LEPÈRE ET AL. 2008). Pris en charge sous forme simple-brin par les protéines Ptiwio1 et Ptiwio9, ils sont exportés vers le Mac maternel (BOUHOUCHE ET AL. 2011). Là, les scnRNA pouvant s'associer aux ncRNA maternels sont dégradés ou séquestrés. Seul les scnRNA n'ayant pas trouvé de séquence complémentaire sont conservés, définissant ainsi les loci spécifiques du Mic absents de l'ancien Mac. Ce processus de *scanning* est dépendant des protéines Nowa1 et 2 (NOWACKI ET AL. 2005) et de l'hélicase Ptmb220 (NO-

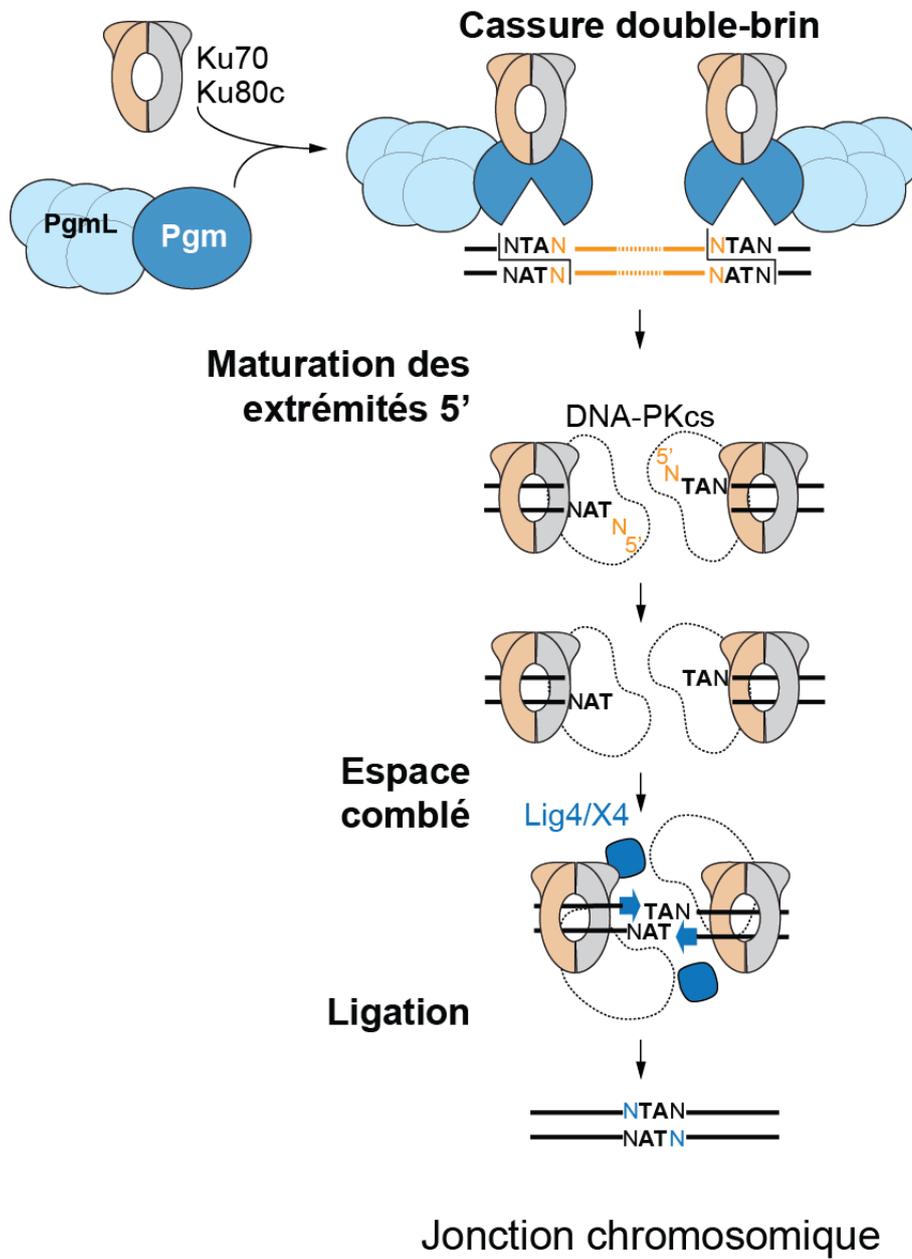
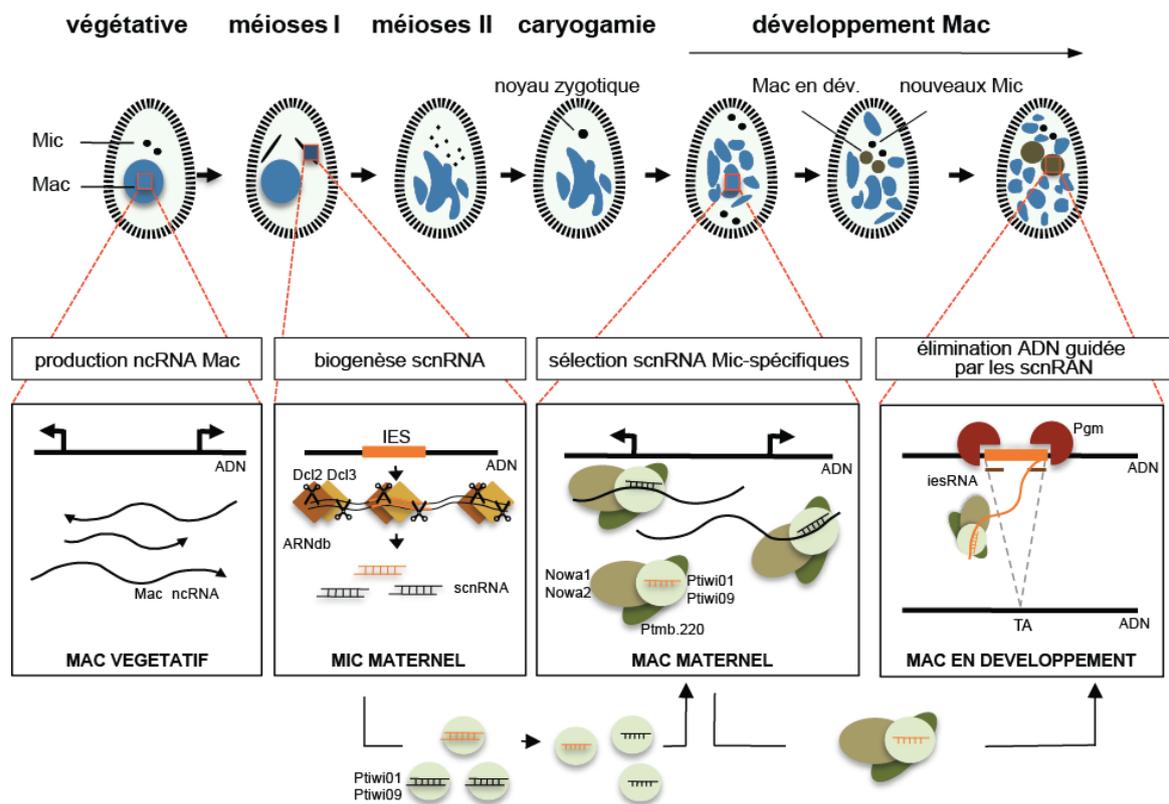


FIGURE II.16 – Modèle d'excision des IES

Couplage extrême entre le mécanisme de coupure et celui de réparation. Ku est essentiel à l'activation de la cassure, ce qui assure une prise en charge efficace par le NHEJ.

FIGURE II.17 – *Scanning ARN*

Modèle des scnRNA. Une comparaison entre le génome du Mac maternel et des Mic est effectuée en utilisant des ARN non codants (scnRNA). Les scnRNA Mic spécifiques sont transportés dans les ébauches et permettent le marquage des régions à éliminer. Adaptée de BETERMIER AND DUHARCOURT (2014).

WAK ET AL. 2011). Les scnRNA Mic spécifiques sont ensuite exportés vers les ébauches et guident Pgm sur les séquences à exciser, reproduisant ainsi le profil de réarrangement de l'ancien Mac. Ce guidage pourrait se faire via l'appariement des scnRNA avec des transcrits naissants non codants produits dans les Mac en développement grâce à un facteur de transcription TFIIS spécialisé (MALISZEWSKA-OLEJNICZAK ET AL. 2015).

Un mécanisme complémentaire des scnRNA utilise des ARN transcrits à partir des IES circularisées (seules ou en concatémères d'IES) (ALLEN ET AL. 2017). Dcl5 les clive en fragments de 23 à 31 nucléotides (SANDOVAL ET AL. 2014, HOEHENER ET AL. 2018). Ils sont ensuite pris en charge par Ptiwi10 et Ptiwi11 où ils vont pouvoir continuer à marquer les IES encore non excisées, permettant ainsi une amplification du signal d'excision (FURRER ET AL. 2017).

Cependant, des ARNi contre les *DCL2/3* ou *DCL5* ne provoquent que la rétention d'environ 20% d'IES au total. Une étude encore en révision (SWART ET AL. 2017) suggère que la triple déplétion de *DCL2/3/5* pourrait affecter l'excision de 50% des IES. Il y a donc certainement différents mécanismes encore inconnus responsables de l'identification des IES restantes à éliminer.

La reconnaissance par Pgm Comment les scnRNA et autres mécanismes indiquent-ils à Pgm la localisation précise des IES? Les données et modèles actuels indiquent que des marques épigénétiques spécifiques sont déposées sur les IES. Ainsi, la protéine Ezl1, potentielle lysine histone méthyltransférase, est essentielle pour les modifications H3K9me3 et H3K27me3 et pour l'excision de 70% des IES (LHULLIER-AKAKPO ET AL. 2014). Des données non publiées de Marc Guérineau montrent la capacité du domaine en doigt de zinc de Pgm d'interagir avec l'histone H3, notamment le peptide N-terminal de H3 (1-19) mais pas H3 (16-35). Pgm directement ou des partenaires du complexe d'excision pourraient interagir avec ces marques, et d'autres encore inconnues, permettant de positionner le complexe d'excision sur les régions à éliminer. Les scnRNA ou les iesRNA pourraient également avoir un rôle dans le guidage précis de Pgm pour l'identification de la borne de l'IES, une grande partie des IES étant plus courtes que l'ADN enroulé autour d'un nucléosome (**Figure II.18**). Des études sont également en cours pour évaluer le rôle de la méthylation de bases de l'ADN dans la reconnaissance des IES.

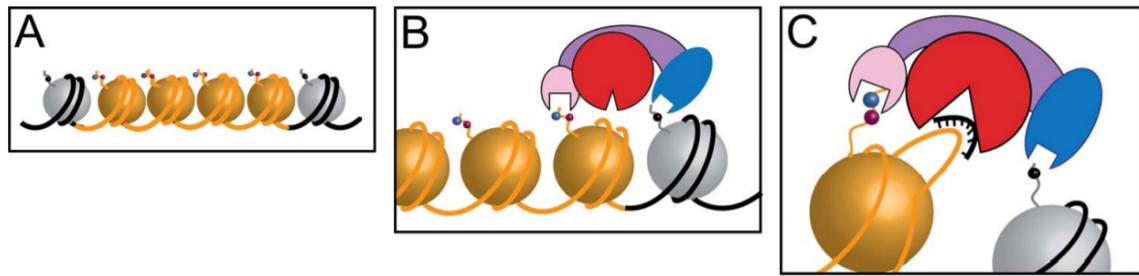


FIGURE II.18 – Modèle de reconnaissance des IES par Pgm

(A) L'IES (en jaune) est associée à des histones porteuses de modifications épigénétiques. (B) Pgm (en rouge) ou ses partenaires reconnaissent ces modifications et ciblent ainsi l'IES. (C) Les scnRNA ou les iesRNA participent à la reconnaissance précise de l'extrémité de l'IES. Adaptée de BETERMIER AND DUHARCOURT (2014).

II.2.2 Les réarrangements imprécis

L'autre catégorie de réarrangements est celle des réarrangements imprécis. Ils représentent 25 Mb d'ADN éliminé, essentiellement dans des régions inter-géniques. Ils consistent en l'élimination de transposons et de séquences répétées lors des RPG. L'élimination est dite imprécise car les bornes des régions éliminées peuvent être variables et la cassure est soit réparée par religation soit télomérisée, fragmentant ainsi les chromosomes (**Figure II.19**). Il n'est pas exclu qu'il y ait une excision à des bornes précises, mais que ce soit la réparation qui apporte l'imprécision.

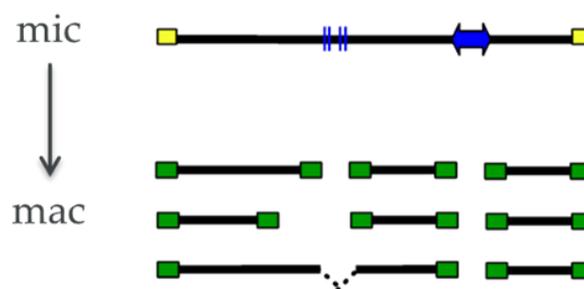


FIGURE II.19 – Les éliminations imprécises

Élimination des transposons et séquences répétées lors des RPG. Ils donnent lieu à des fragmentations des chromosomes, avec néo-télomérisations et religations variables.

L'élimination imprécise est dépendante de Pgm (BAUDRY ET AL. 2009, GUÉRIN ET AL.

2017) et le mécanisme de réparation impliqué, quand elles sont réparées, n'est pas connu. En effet, la nature imprécise de ces réarrangements rend leur étude extrêmement compliquée.

II.2.3 Chez les autres ciliés

L'ensemble des ciliés possèdent un dimorphisme nucléaire et effectuent des RPG. L'étude de ces différents ciliés permet d'enrichir les connaissances sur les RPG, leur diversité et les différents mécanismes associés. L'un de ces organismes, particulièrement étudié, est le cilié *Tetrahymena thermophila*

II.2.3.1 *Tetrahymena thermophila*

Tetrahymena possède un génome Mic de 157 Mb avec 5 chromosomes et est diploïde. Le génome du Mac est de 103 Mb avec 181 chromosomes et une ploïdie de 45. Il y a donc élimination d'environ 34% de l'ADN germinale et fragmentation des chromosomes. Il existe deux types d'IES chez *Tetrahymena*. Le premier est le majoritaire et comprend environ 6 000 IES (YAO ET AL. 2014). Ces IES sont toutes intergéniques et dépendantes du mécanisme des scnRNA, responsable du dépôt des marques H₃K₉me₃ et H₃K₂₇me₃ au niveau des IES pour leur élimination. L'endonucléase responsable de leur excision est Tpb2. Comme Pgm, c'est une transposase domestiquée de la famille des transposases PiggyBac. Elle possède la triade DDD catalytique et un domaine C-terminal en doigt de zinc qui permet l'interaction avec les marques d'hétérochromatine.

Comme pour la paramécie, le NHEJ est responsable de la réparation des CDB aux bornes des IES. Cependant, une grande différence entre ces deux organismes est que Ku n'est pas nécessaire pour l'étape d'introduction des cassures (LIN ET AL. 2012). Cette absence de couplage entre coupure et réparation pourrait être due à la nature intergénique des IES impliquées.

L'autre type d'IES semble confirmer cette hypothèse. Au nombre de 12, elles interrompent des exons et leur excision précise est essentielle pour l'assemblage de gènes fonctionnels. Une étude récente a identifié les protéines essentielles à leur excision, Tpb1 et Tpb6 (CHENG ET AL. 2016). Homologues de Tpb2, elles sont également des transposases domestiquées de la famille des transposases PiggyBac. Ce qui les rend uniques dans le

monde des transposases est le fait qu'elles soient des protéines de fusion entre une transposase Piggybac et le domaine cœur d'une protéine Ku80. Là où la paramécie a évolué vers un couplage extrême entre coupure et réparation, *Tetrahymena* est potentiellement allée un cran plus loin en sélectionnant directement une protéine chimérique pouvant couper et directement réparer le dommage. Malheureusement, les études jusqu'ici n'ont pas cherché à répondre aux questions liées au couplage. Quelles fonctions une protéine Tpb1/6 déléetée de son domaine Ku80 conserve-t-elle? Y-a-t-il un découplage, avec cassures mais absence de réparation? La protéine Ku80 pleine taille prend-t-elle le relais? Y-a-t-il des erreurs introduites par ce découplage? Ku70 interagit-elle directement avec le domaine Ku80 de Tpb1/6, formant ainsi un potentiel hétérodimère Ku actif? Ces questions seront je l'espère étudiées dans le futur.

II.2.3.2 *Oxytricha trifallax*

Oxytricha est un cilié faisant partie de la sous-classe des spirotriches. Plus éloigné évolutivement de la paramécie et de *Tetrahymena*, ses RPG sont de nature différente (YERLICI AND LANDWEBER 2014).

Avec un génome Mic de 500 Mb et un génome Mac de 50 Mb, *Oxytricha* élimine 90% de son ADN germinale. Les RPG chez *Oxytricha* sont extrêmes, de par la quantité d'ADN éliminé mais également par l'ampleur du réarrangement des séquences codantes. En effet, les ORF sont fortement mélangées dans le génome Mic (**Figure II.20**).

Un mécanisme inverse des scnRNA existe chez *Oxytricha*. De longs ARN non codants sont produits dans le Mac maternel et indiquent l'organisation des séquences Mac. Il y a également des petits ARN (piRNA) produits à partir du Mac parental qui ciblent les séquences à conserver. Ces deux éléments permettent l'obtention d'ORF fonctionnelles.

L'autre grande particularité des RPG chez *Oxytricha* est la création de nano-chromosomes, regroupant à chaque fois une à trois ORF maximum. Le Mac est ainsi divisé en 15 600 chromosomes et chaque chromosome est répliqué de nombreuses fois, amenant la ploïdie finale à environ 2 000.

Les acteurs des RPG sont encore peu connus chez *Oxytricha* mais des protéines de type transposase Tc1/mariner ont été décrites comme essentielles pour les RPG (NOWACKI ET AL. 2009).

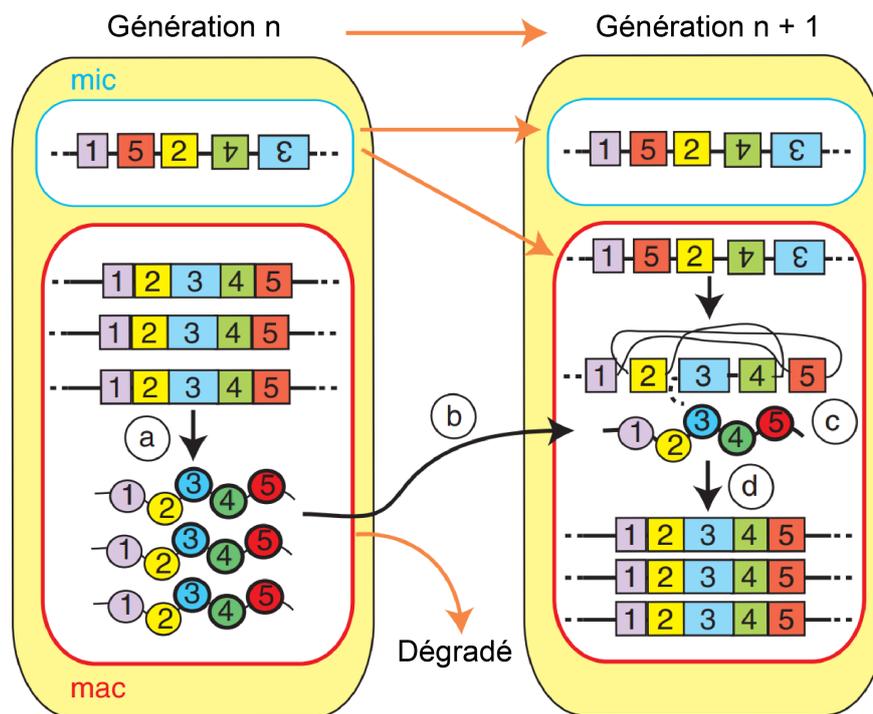


FIGURE II.20 – Les RPG chez *Oxytricha trifallax*

Les éléments des ORF sont représentés en carrés numérotés, selon leur ordre fonctionnel. Les séquences entre ces éléments sont les IES et sont représentées par un trait noir. Lors des RPG, (a) des ncRNA scannent le génome Mac maternel, (b) sont transférés vers le Mac en développement, (c) marquent les séquences à retenir et l'organisation de l'ORF et (d) permettent l'élimination des IES et la réorganisation de l'ORF. Adaptée de CHALKER ET AL. (2013).

Chapitre III

Projet de thèse

De façon totalement inattendue, les travaux d'Antoine Marmignon ont pu déterminer que la présence de Ku lors des RPG est essentielle à l'introduction de cassures de l'ADN par Pgm. Les premiers résultats obtenus étaient cependant limités à l'étude moléculaire de trois IES particulières. Ces premiers travaux ont également montré que Ku localise spécifiquement dans les Mac en développement et qu'il interagit avec Pgm. Cependant, le mécanisme d'activation conduisant à l'introduction des coupures restait encore à déterminer. C'est donc à cette question que je me suis intéressé pendant ma thèse.

Caractérisation du couplage entre Pgm et Ku Dans un premier temps, j'ai regardé de façon quantitative l'interdépendance des protéines Pgm et Ku vis-à-vis de leur localisation nucléaire en utilisant des approches d'immunofluorescence et de microscopie. Mes résultats démontrent que Pgm est dépendant de Ku pour sa localisation stable dans les ébauches. Des expériences de séquençage haut-débit confirment une rétention de toutes les IES dans le génome lors d'une déplétion de Ku80c.

En revanche, la localisation nucléaire de Ku80c n'est que partiellement affectée lors d'une déplétion de Pgm, indiquant que Ku80c n'est pas dépendant de la présence de CDB pour sa localisation stable dans les ébauches.

Etude de la spécialisation fonctionnelle de Ku80c pour les RPG Les gènes *KU80* chez la paramécie ont été dupliqués lors de deux WGD et fournissent un modèle pour l'étude de la spécialisation fonctionnelle de protéines suite à des événements de duplications. Les ARNi menées par Antoine Marmignon montrent que seul l'ARNi de *KU80c* conduit à un

blocage des RPG, alors que des ARNi contre *KU80a*, *KU80b* ou les deux ne provoquent aucun phénotype. Deux hypothèses peuvent être formulées : soit la particularité du gène *KU80c* est simplement son profil d'expression spécifique des RPG, soit c'est la protéine Ku80c elle-même qui a acquis une fonction particulière par rapport à Ku80a et b.

Lors d'une ARNi *KU80c*, il est possible que les faibles niveaux d'expression de *KU80a* et *b* ne permettent pas de compléter l'absence de la protéine Ku80c (**Figure III.1B**).

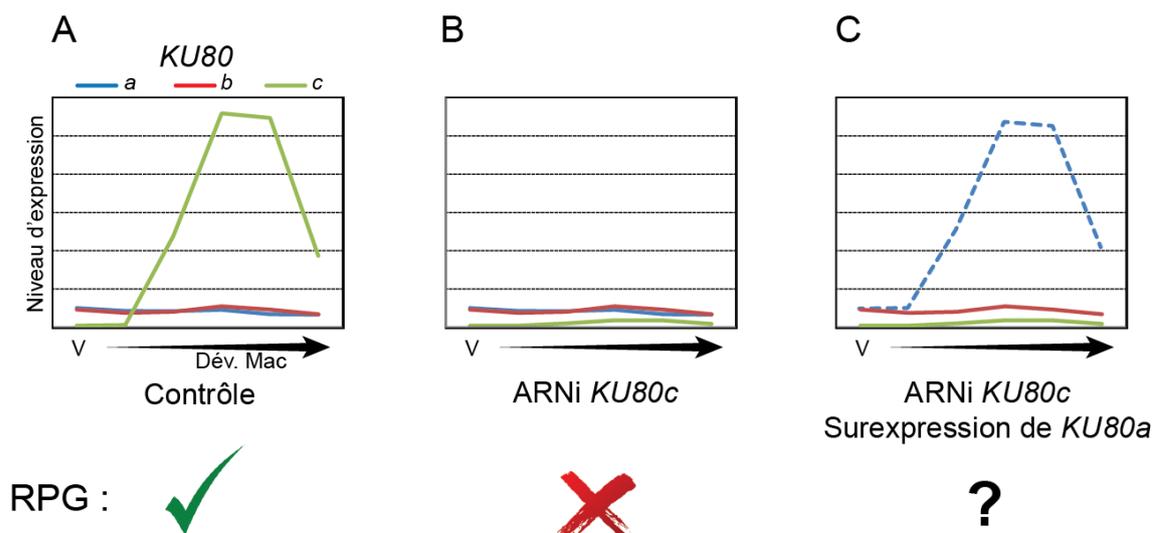


FIGURE III.1 – *Ku80a* peut-il remplacer *Ku80c* ?

A) Niveau d'expression des trois gènes *KU80* lors du développement macronucléaire en situation contrôle. B) ARNi *KU80c* provoquant la perte d'expression de *Ku80c*. Les quantités de *Ku80a* et *b* sont probablement faibles du fait du faible niveau d'expression constitutif. Les RPG n'ont pas lieu. C) ARNi *KU80c* couplée à une surexpression de *KU80a*, permettant de retrouver une quantité sauvage d'une protéine *Ku80* lors des RPG.

L'absence de complémentation pourrait ainsi être simplement due à un effet dose, la quantité globale de *Ku80a* et *b* dans la cellule étant probablement très faible dans une ARNi *KU80c*. J'ai donc cherché à surexprimer *Ku80a* lors des RPG, en condition d'ARNi *KU80c* afin de vérifier si une protéine *Ku80a* partageait les mêmes capacités que *Ku80c* (**Figure III.1C**). Mes résultats indiquent que *Ku80a* surexprimée ne remplace pas une absence de *Ku80c*, les RPG n'ayant pas lieu. La protéine *Ku80c* s'est donc bien spécialisée dans les RPG. Grâce à des constructions de protéines chimériques, j'ai démontré que le domaine α/β de *Ku80c* était le principal responsable de cette spécialisation.

L'ensemble des résultats obtenus pendant ma thèse, enrichis d'expériences complémentaires réalisées par l'équipe, ont été incorporés dans un manuscrit dont une première

version est disponible en annexe. Ma discussion générale s'appuiera sur l'ensemble des résultats disponibles.

Deuxième partie

Résultats

Chapitre IV

Localisation des protéines Pgm et Ku lors des réarrangements programmés

IV.1 LOCALISATION DE PGM

IV.1.1 Dynamique de la localisation nucléaire de Pgm

La méthode de suivi du développement Mac utilisée en routine dans l'équipe consistait à observer le grossissement des ébauches au microscope sur des cellules fixées et colorées au DAPI (DUBOIS ET AL. 2017). Les ébauches sont classées en trois classes : précoces - intermédiaires - tardives, définies par l'avancement du cycle sexuel et par la taille moyenne des nouveaux Mac. En effet, les réplifications massives du génome et donc, l'accroissement de la taille des ébauches, sont concomitantes avec les RPG.

Afin d'étayer l'ensemble des phénotypes associés à une ARNi *KU*, l'analyse en ImmuFluorescence (IF) de Pgm est un outil puissant. Cependant, la localisation nucléaire des protéines des RPG est dynamique. Pgm-GFP était décrite comme nucléaire dans des ébauches précoces mais absente dans les tardives (DUBOIS ET AL. 2017). Ces catégorisations sont suffisantes pour des effets forts et très contrastés, mais manquent de précision pour les effets plus subtils, néanmoins tout aussi intéressants, d'où la nécessité de développer des méthodes de quantification de la taille des ébauches et du signal d'IF. L'autogamie peut être déclenchée sur de larges populations par carence alimentaire. En conditions optimales, la synchronie de la population autogame est d'environ 6 heures, c'est à dire que la première cellule à avoir déclenché le processus peut avoir 6 heures d'avance sur la

dernière. Cela produit donc une population hétérogène. Ainsi, il est possible d'avoir en IF, dans une même population, des cellules marquées ou non par Pgm dans les ébauches. Cette hétérogénéité empêche donc une comparaison fine des intensités de fluorescence entre différentes conditions.

Afin de contourner ce problème, nous avons cherché à vérifier s'il était possible de lier précisément la taille des ébauches présentant un marquage Pgm à l'état d'avancement des RPG. Ainsi, nous pourrions comparer l'effet en IF de différentes conditions d'ARNi en comparant des ébauches de tailles identiques.

Des expériences sur des cellules en conjugaison ont été réalisées par Nathalie Mathy (AI CNRS dans l'équipe). La conjugaison présente l'avantage sur l'autogamie d'être beaucoup plus synchrone. Cependant, elle ne permet pas de travailler facilement sur de larges population car elle implique généralement un tri manuel de chaque cellule. Afin d'avoir des données plus précises sur l'évolution de la taille des ébauches au cours des RPG, des cellules en conjugaison ont été fixées à différents temps, de 8 à 20 heures après le début de la conjugaison et la taille des ébauches a été mesurée grâce à un module d'analyse écrit par Romain Le Bars de la plateforme d'imagerie de l'I2BC (détails en matériels et méthodes) (**Figure IV.1**).

Cette expérience montre qu'une ébauche a une taille comprise entre 30 et 75 μm^2 lors de l'excision des IES. Il était ensuite important de vérifier si Pgm était localisée dans les ébauches correspondant à ces tailles en utilisant l'anticorps α -Pgm (**Figure IV.2**). La dynamique de la localisation nucléaire de Pgm correspond aux tailles d'ébauches lors de l'excision des IES (30-75 μm^2). En autogamie, la localisation nucléaire de Pgm a été décrite comme organisée en foci, sur un fond nucléoplasmique granuleux par DUBOIS ET AL. (2017), observation reproduite ici plus finement grâce à l'utilisation de cellules en conjugaison, fournissant une plus grande synchronie.

Note technique Le protocole de fixation utilisé dans les figures IV.1 et IV.2 est un protocole optimisé pour la détection du Pgm nucléaire (DUBOIS ET AL. 2017). Il inclut une étape de perméabilisation par du Triton (détergent) antérieure à la fixation des cellules. Ce protocole, en plus de permettre une observation optimale de Pgm, provoque l'extraction des

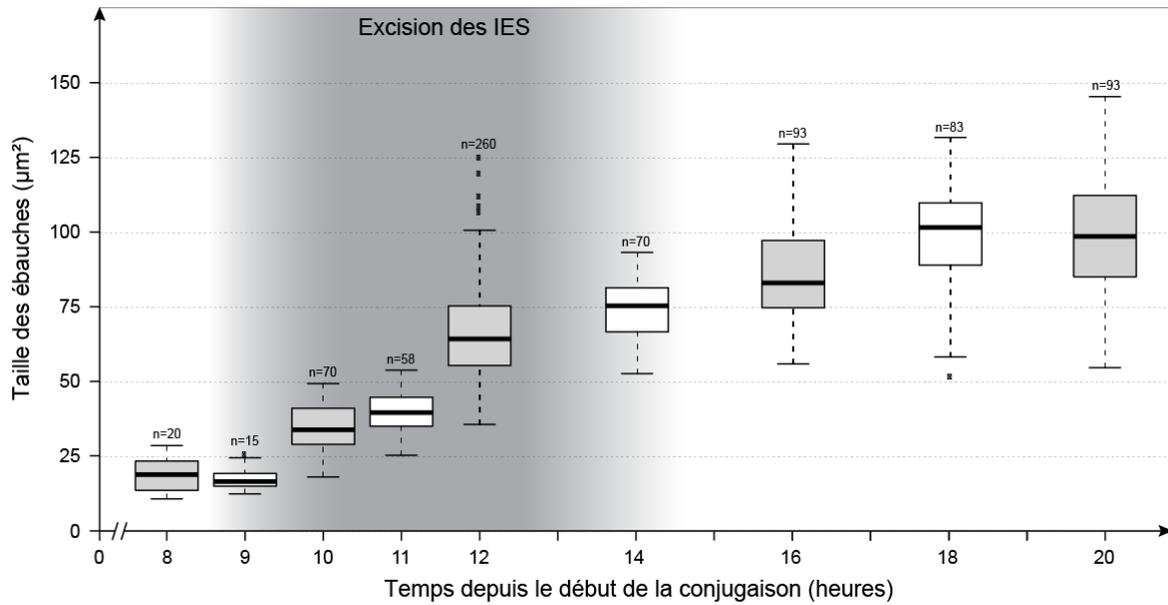


FIGURE IV.1 – Taille des ébauches en conjugaison en fonction de l'avancement du cycle sexuel

Représentation en boîtes à moustaches des tailles d'ébauches en fonction du temps en heure depuis le début de la conjugaison, observées au microscope. Les fixations ont été réalisées en condition de pré-extraction. La taille de l'échantillon pour chaque boîte à moustache est représentée par la valeur n. Le moment correspondant à l'excision des IES, représenté en gris, a été déterminé par GRATIAS AND BÉTERMIER (2001). Le moment du début et de la fin des réarrangement n'est pas encore défini très précisément, d'où la représentation en gradient. Expérience de Nathalie Mathy.

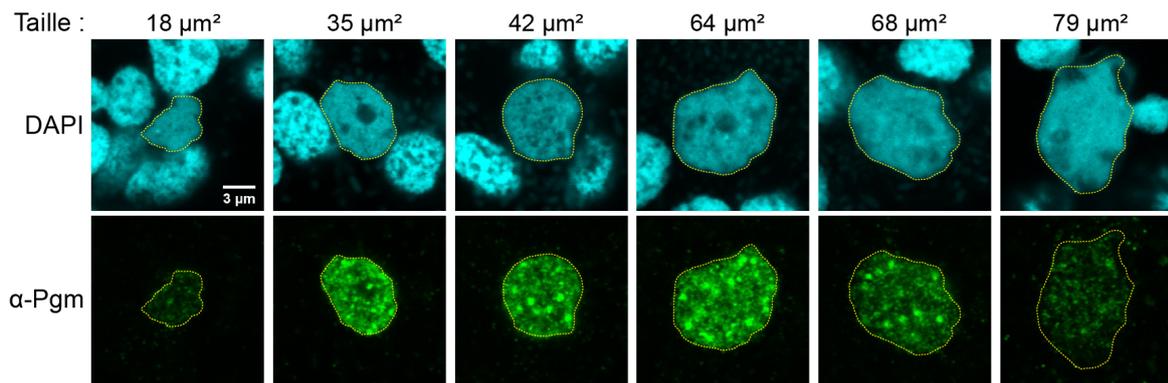


FIGURE IV.2 – Dynamique de la localisation nucléaire de Pgm en conjugaison en fonction de l'avancement du cycle sexuel

Imagerie confocale en un plan d'IF α -Pgm sur des cellules en conjugaison après un traitement de pré-extraction. Les ébauches sont entourées de pointillés jaunes et ont été triées par taille en μm^2 . Les noyaux marqués au DAPI mais non marqués par Pgm sont les fragments de l'ancien Mac. Expérience réalisée conjointement avec Nathalie Mathy.

protéines labiles (MARTINI ET AL. 1998) et met donc en évidence les protéines solidement accrochées à une structure, ici le noyau et donc probablement la chromatine.

Un protocole sans pré-perméabilisation et donc sans pré-extraction a également été utilisé dans différentes expériences de ce manuscrit et dans BISCHEROUR ET AL. (2018). Il permet de mettre en évidence l'ensemble des protéines, labiles ou non. La comparaison des résultats obtenus entre les protocoles avec ou sans pré-extraction permet donc de conclure sur la force d'association de la protéine étudiée avec le noyau.

Pgm est fortement associée aux ébauches excisant les IES Ces deux premières expériences nous permettent de dire qu'un marquage nucléaire de Pgm en protocole de pré-extraction correspond à des ébauches de 35 à 68 μm^2 de section (**Figure IV.2**) qui est la gamme de taille des ébauches pendant la fenêtre de temps où se produit l'excision des IES (**Figure IV.1**). Dans les expériences suivantes, les ébauches présentant un marquage Pgm en condition contrôle définissent la gamme de tailles d'ébauches utilisée pour comparer les intensités de fluorescence dans des conditions d'ARNi différentes. Cette méthode d'analyse a été développée conjointement avec Vinciane Régnier.

IV.1.2 Effet de l'extinction de Ku sur la localisation de Pgm

IV.1.2.1 Localisation nucléaire de Pgm en ARNi KU

En utilisant cette nouvelle méthode d'analyse, j'ai cherché à déterminer si la localisation nucléaire de Pgm pouvait être affectée par les ARNi KU. Une absence d'activité de Pgm en ARNi KU pourrait ainsi être expliquée par une délocalisation complète.

Les cellules ont subi une ARNi contre *KU70*, *KU80c* ou *PGM* en contrôle négatif ou bien n'ont pas subi d'ARNi (contrôle positif) et l'autogamie a été provoquée, toujours dans les conditions d'ARNi. Les cellules ont été récoltées à T5 (rappel : T0 est le moment où 50% des cellules présentent un Mac fragmenté. Une récolte 5 heures après cet instant permet d'avoir un grand nombre de cellules en cours de RPG). Elles ont été fixées avec le protocole sans pré-extraction. Une IF α -Pgm couplée à l'analyse de nombreuses images de microscopie à épifluorescence a permis la quantification du signal Pgm dans ces différentes conditions d'ARNi (**Figure IV.3**).

En situation contrôle, les ébauches présentant un signal Pgm positif s'étalent de 55 μm^2

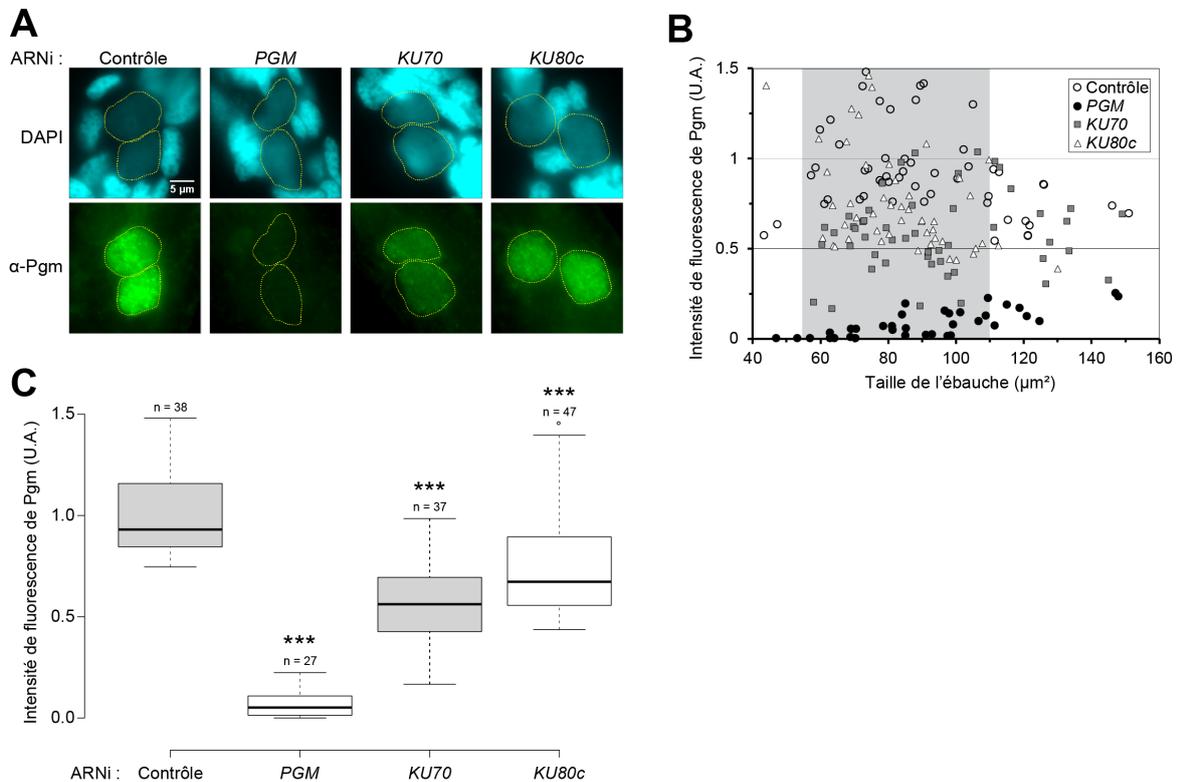


FIGURE IV.3 – IF sans pré-extraction de Pgm au cours de l'autogamie en ARNi PGM ou KU

(A) Localisation de Pgm dans des cellules autogames à T5 en fonction de l'ARNi. Microscopie à épifluorescence. Les ébauches sont indiquées par les pointillés. (B) Intensité de fluorescence de Pgm dans les ébauches en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm². Un bruit de fond correspondant au signal total de cellules végétatives en fluorescence verte a été soustrait à l'ensemble des données. En condition contrôle, Pgm présente un pic de localisation nucléaire entre 55 et 110 μm², définissant ainsi la zone d'intérêt pour la comparaison entre les différentes ARNi (zone grise). (C) *Boxplot* de l'intensité de fluorescence de Pgm des ébauches comprises entre 55 et 110 μm². Le n représente le nombre d'ébauches par condition. Signal moyen de Pgm : Contrôle 100%, PGM 6%, KU70 57%, KU80c 75%. *** pour p<0.001 dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle. Dans cette expérience il y a eu 100% de descendance sexuelle fonctionnelle pour la condition contrôle et 0% pour les ARNi PGM, KU70 et KU80c

à 110 μm^2 environ (**Figure IV.3B**). Cette différence par rapport à la **figure IV.2** s'explique par l'utilisation du protocole sans pré-extraction. Chaque protocole agit différemment sur les structures cellulaires et peut provoquer des différences de taille apparente des noyaux. L'information de la gamme de taille d'ébauches Pgm positives en condition contrôle permet donc de s'affranchir de la variation de taille intrinsèque au protocole utilisé.

En comparant l'ensemble des ébauches se situant entre 55 μm^2 et 110 μm^2 (**Figure IV.3C**), on observe une disparition presque totale du signal Pgm dans une ARNi *PGM* par rapport au contrôle avec 6% restant en moyenne. Dans les ARNi *KU*, Pgm est toujours localisée dans les ébauches en développement mais en moindre quantité par rapport au contrôle. On détecte un signal moyen de 57% de Pgm dans l'ARNi *KU70* et 75% dans l'ARNi *KU80c*.

La présence nucléaire de Pgm est donc partiellement dépendante des protéines Ku. Cependant, cette diminution du signal pourrait être une diminution de l'expression ou de la stabilité de Pgm.

IV.1.2.2 Quantité de protéine Pgm en ARNi *KU*

Afin de tester le niveau d'expression de la protéine Pgm en absence des protéines Ku au cours de l'autogamie, des WB ont été réalisés sur des cellules autogames récoltées à T5 provenant de l'expérience présentée en **figure IV.3** (**Figure IV.4**).

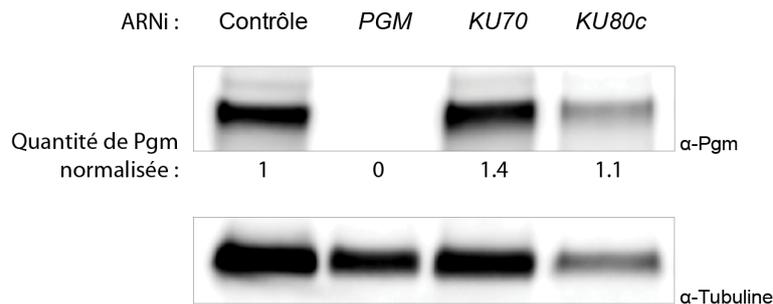


FIGURE IV.4 – WB de Pgm au cours de l'autogamie en ARNi *PGM* ou *KU*

Extraits protéiques totaux de cellules à T5, migrés par SDS-PAGE et révélés en WB par l'anticorps α -Pgm. La membrane a ensuite été deshybridée puis révélée par un α -Tubuline pour permettre la normalisation de la quantité de Pgm dans les différents extraits. Pourcentage de cellules avec un nouveau Mac fonctionnel par condition d'ARNi : contrôle 100%, *PGM*, *KU70* et *KU80c* 0%.

En condition d'ARNi *KU70* et *KU80c*, la quantité globale de protéine Pgm est stable par rapport au contrôle alors que Pgm n'est plus du tout détectée lors d'une ARNi *PGM*.

Pgm est donc toujours autant présent dans les cellules déplétées de Ku. La diminution du signal nucléaire de Pgm dans la **figure IV.3** en ARNi *KU* (25 à 40% de diminution) est donc uniquement liée à une délocalisation nucléaire partielle de Pgm. Différentes hypothèses permettent d'expliquer cette délocalisation :

1) Ku est impliqué dans l'import nucléaire de 25 à 40% de Pgm.

2) En ARNi *KU*, Pgm est au moins partiellement inactif (MARMIGNON ET AL. 2014).

Cette inactivation pourrait réduire la stabilité nucléaire de Pgm, entraînant une baisse globale de la quantité de protéine dans les ébauches.

IV.1.2.3 Stabilité nucléaire de Pgm en ARNi *KU*

L'expérience sans pré-extraction a montré qu'une large fraction de Pgm était nucléaire en absence de Ku, malgré les résultats d'Antoine Marmignon indiquant qu'en absence de Ku, Pgm n'est pas apte à procéder à l'excision des IES étudiées. Les propriétés du Pgm nucléaire en absence de Ku sont donc potentiellement différentes de celle du Pgm nucléaire en présence de Ku. J'ai testé ces propriétés en réalisant des expériences d'ARNi *KU* couplées à une observation de Pgm dans des cellules fixées avec le protocole de pré-extraction (**Figure IV.5**).

Ces expériences permettent d'observer une forte diminution du signal Pgm dans les ARNi *PGM*, *KU70* et *KU80c*, avec respectivement 18, 30 et 20% du signal restant. En pré-extraction, le niveau du signal Pgm en ARNi *KU* est donc proche du niveau du contrôle négatif, c'est à dire de l'ARNi *PGM*. Cela signifie qu'en absence de Ku, l'accrochage nucléaire de Pgm est presque complètement aboli. Ku est donc probablement essentiel pour que Pgm soit solidement associé à la chromatine et actif. Le mécanisme d'activation de Pgm par Ku pourrait donc être de permettre l'accrochage de Pgm à la chromatine et déclenchant ainsi les CDB.

IV.1.2.4 Effet global d'une ARNi *KU* sur la rétention des IES

Les quelques IES testées dans MARMIGNON ET AL. (2014) ont toutes montré une rétention en ARNi *KU* et les extrémités db aux bornes de ces IES n'ont pas pu être détectées en utilisant des approches moléculaires. L'analyse de ces données et de celles obtenues en **figure IV.5** semblent indiquer une inactivation complète de Pgm dans les RPG en ARNi

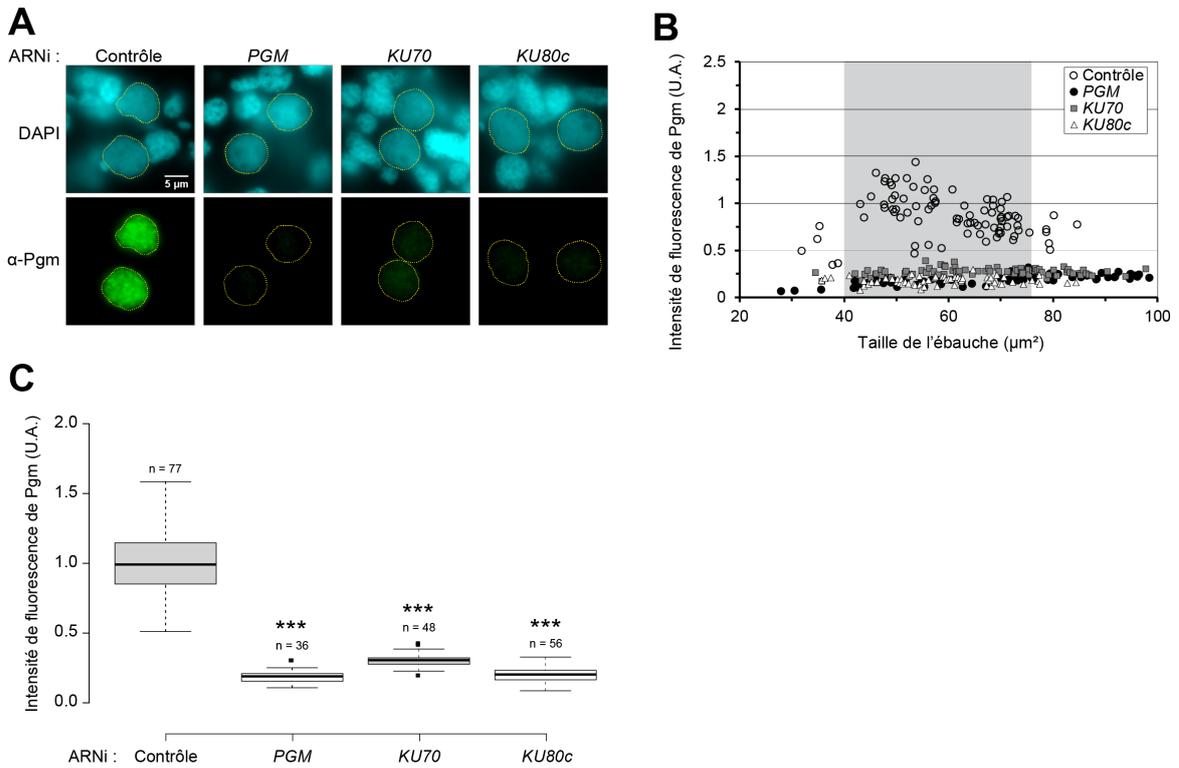


FIGURE IV.5 – IF avec pré-extraction de Pgm au cours de l'autogamie en ARNi PGM ou KU

(A) Localisation de Pgm dans des cellules autogames à T5 en fonction de l'ARNi. Microscopie à épifluorescence. (B) Intensité de fluorescence de Pgm dans les ébauches en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm². Un bruit de fond correspondant au signal de la lame en fluorescence verte a été soustrait à l'ensemble des données (zone sans cellule). En condition contrôle, Pgm présente un pic de localisation nucléaire entre 40 et 75 μm², définissant ainsi la zone d'intérêt pour la comparaison entre les différentes ARNi (zone grise). (C) *Boxplot* de l'intensité de fluorescence de Pgm des ébauches comprises entre 40 et 75 μm². Le n représente le nombre d'ébauches par condition. Signal moyen de Pgm : Contrôle 100%, PGM 18%, KU70 30%, KU80c 20%. *** pour p<0.001 dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle. Dans cette expérience il y a eu 100% de descendance sexuelle fonctionnelle pour la condition contrôle et 0% pour les ARNi PGM, KU70 et KU80c.

KU du fait de la déstabilisation de la localisation nucléaire. Une telle inactivation doit conduire à la rétention de l'ensemble des IES. Mireille Bétermier a cherché à vérifier cela en procédant au séquençage du génome de paramécies autogames tardives en ARNi *KU*. Ces données de séquençage ont été analysées par Olivier Arnaiz (IE CNRS dans l'équipe) qui a pu déterminer le nombre d'IES retenues significativement par rapport à une population de paramécies autogames tardives en situation contrôle (**Figure IV.6**). La distribution du score de rétention dans les deux réplica de l'ARNi *KU80c* montre une très forte inhibition de l'excision des IES sur l'ensemble du génome. Les différences entre les deux réplica ainsi que la présence d'IES non significativement retenues peuvent être expliquées par des contaminations de fragments d'ancien Mac.

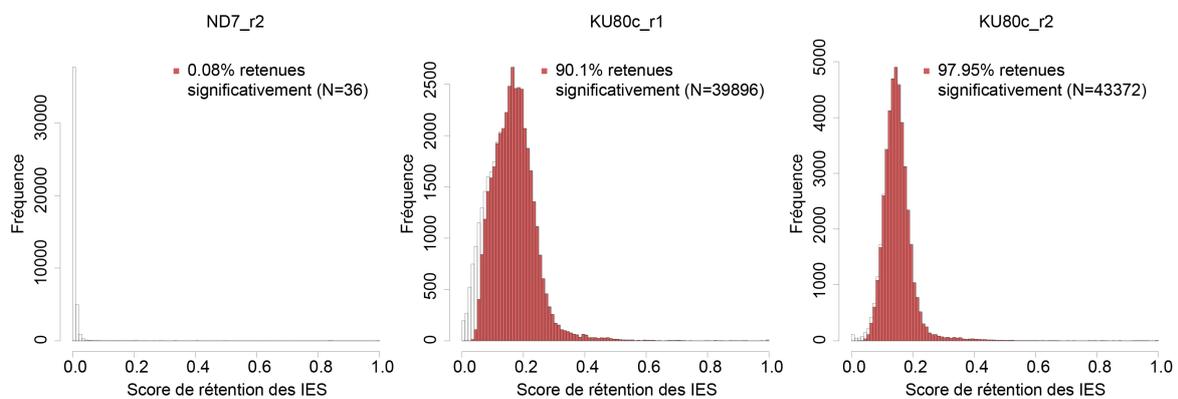


FIGURE IV.6 – Score de rétention des IES dans des ARNi *KU80c*

Distribution des scores de rétention d'IES dans des ARNi *ND7* (contrôle) ou *KU80c*. Les barres grises représentent la distribution de toutes les IES en fonction de leur score de rétention, de 0 à 1 (par des pas de 0,025). La variation du nombre d'IES non significativement retenues peut être expliquée par la variation du degré de contamination par les fragments de l'ancien Mac dans la préparation d'ADN.

Jusqu'à 97,95% des IES sont significativement retenues. Pgm est donc très probablement complètement inactif en absence de *Ku80c*. La présence de *Ku* est donc strictement nécessaire pour l'excision de toutes les IES.

IV.1.2.5 Effets de l'inactivation d'autres gènes du NHEJ sur Pgm

La nouvelle méthode d'analyse de la fluorescence des ébauches en fonction de leur taille a donc permis une meilleure caractérisation des effets d'une ARNi *KU* sur Pgm. J'ai utilisé cette nouvelle méthode pour tester les effets d'une ARNi *DNA-PKcs* ou *LIG4* sur la stabilité nucléaire de Pgm.

Des expériences ont été menées sur le rôle de DNA-PKcs lors des RPG chez la paramécie par Sophie Malinsky (communication personnelle). DNA-PKcs est essentiel pour les RPG mais son rôle ne semble pas se limiter à la réparation des cassures. Les résultats moléculaires montrent en effet une ambiguïté, empêchant de comprendre précisément le rôle de DNA-PKcs dans les RPG. En effet, l'analyse moléculaire d'IES révèle un retard d'excision de certaines IES et un retard de réparation pour d'autres. Il est proposé que DNA-PKcs puisse être requise pour une excision optimale des IES. Elle pourrait contribuer comme Ku, à l'activation de la coupure par Pgm. Elle pourrait également être impliquée dans la libération du complexe d'excision par phosphorylation. Un complexe non libéré empêcherait une réparation optimale et ne serait pas disponible pour procéder à de nouvelles excisions. Avec cette hypothèse, on s'attend à observer une accumulation de Pgm fortement associés aux ébauches en absence de DNA-PKcs. J'ai réalisé l'expérience sur des paramécies autogames ayant subi une ARNi *DNA-PKcs* et j'ai préparé les cellules avec le protocole de pré-extraction. Une IF dirigée contre Pgm a ensuite été réalisée et les intensités de fluorescence ont été analysées en fonction des conditions d'ARNi (**Figure IV.7**).

On observe une forte accumulation de Pgm en ARNi *DNAPKcs*, avec 182% du signal moyen de Pgm par rapport au contrôle. La stabilité nucléaire de Pgm a donc été augmentée en absence de DNA-PKcs. Ce résultat semble confirmer l'hypothèse d'un rôle de DNA-PKcs dans la libération de Pgm des sites de coupure.

Dans un contexte ARNi *LIG4*, les cassures sont introduites par Pgm mais ne sont pas réparées (KAPUSTA ET AL. 2011). On s'attend donc à ce que la stabilité nucléaire de Pgm ne soit pas impactée par une ARNi *LIG4* (**Figure IV.7**).

lors d'une ARNi *LIG4*, on observe une diminution du signal Pgm, avec 41% restant par rapport au contrôle. Ce résultat inattendu pourrait être expliqué par une inhibition de l'amplification du génome en ARNi *LIG4*. Un défaut d'amplification est observée avec le marquage DAPI en déplétion de Lig4 (KAPUSTA ET AL. 2011). Une absence d'amplification implique une diminution du nombre d'IES à exciser pour Pgm et donc une diminution de sites de fixation de Pgm.

Ces deux résultats pour *DNAPKcs* et *Lig4* devront être confirmés, étant donné que l'expérience n'a été réalisée qu'une seule fois et que la taille de l'échantillon analysé est

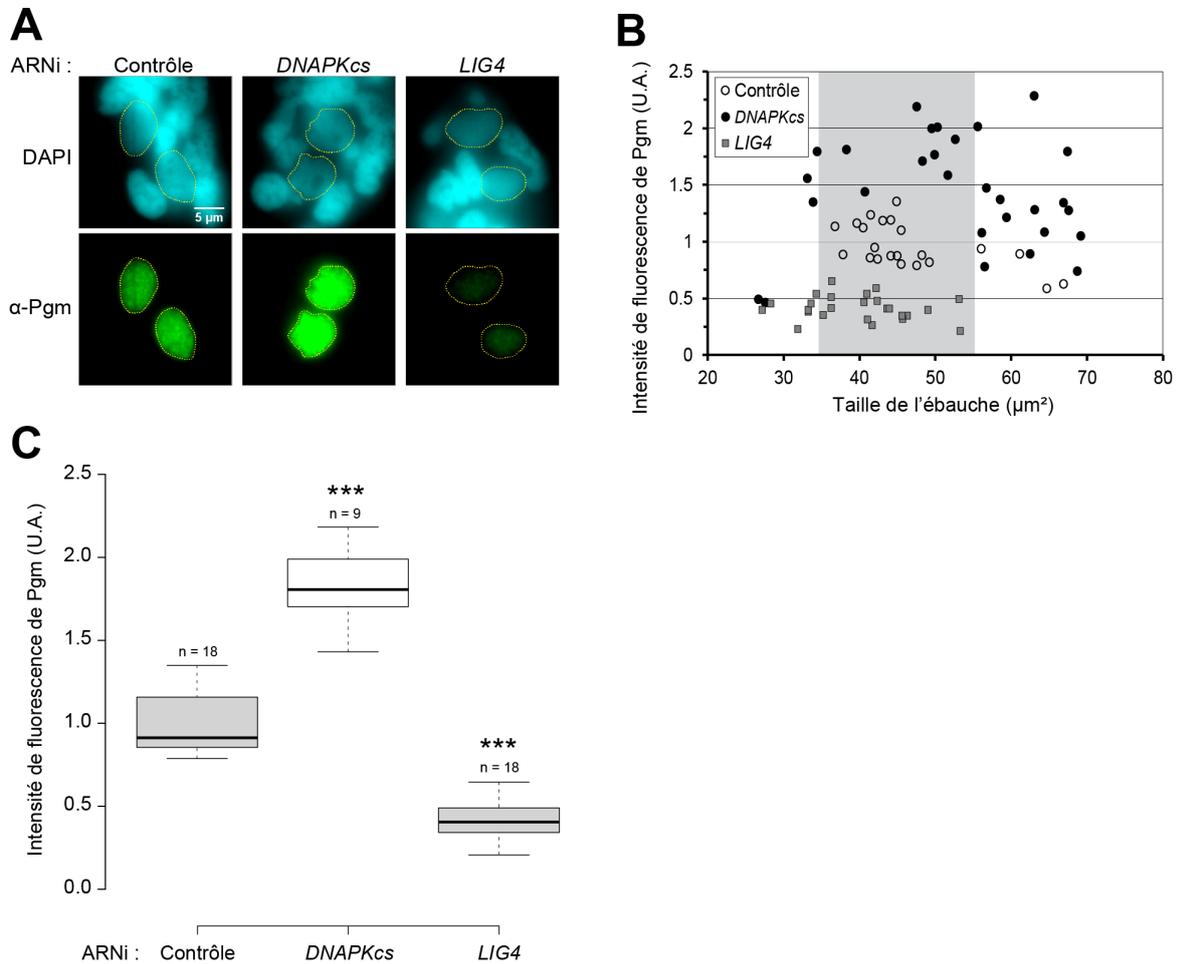


FIGURE IV.7 – IF avec pré-extraction de Pgm au cours de l'autogamie en ARNi DNA-PKcs ou LIG4

(A) Localisation de Pgm dans des cellules autogames à T5 en fonction de l'ARNi. Microscopie à épifluorescence. (B) Intensité de fluorescence de Pgm dans les ébauches en unité arbitraire en fonction de la taille de l'ébauche mesurée en µm². Un bruit de fond correspondant au signal total de cellules végétatives en fluorescence verte a été soustrait à l'ensemble des données. La zone d'intérêt a été choisie entre 35 et 55 µm² (zone grise). (C) Boxplot de l'intensité de fluorescence de Pgm des ébauches comprises entre 35 et 55 µm². Le n représente le nombre d'ébauches par condition. Signal moyen de Pgm : Contrôle 100%, DNAPKcs 186%, LIG4 41%. *** pour p<0.001 dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle. Dans cette expérience il y a eu 100% de descendance sexuelle fonctionnelle pour la condition contrôle et 0% pour les ARNi DNAPKcs et LIG4

faible. De plus, des quantifications du niveau de protéines Pgm par WB seront nécessaire corrélés les observations d'IF avec la quantité de protéine Pgm présente dans la cellule.

IV.2 LOCALISATION DE KU80C

L'analyse des effets de la déplétion de Ku sur la localisation de Pgm a permis d'observer la capacité de Pgm à localiser dans les ébauches en absence de Ku mais il n'y était pas stablement associé à la chromatine. Cette analyse nous a donc permis d'améliorer notre compréhension du mécanisme de couplage entre Ku et Pgm. Une analyse inverse, c'est à dire de l'effet d'une déplétion de Pgm sur la localisation de Ku, était donc susceptible de nous apporter également des informations sur l'interdépendance des protéines quant à leur localisation.

Essais de développement d'anticorps dirigés contre Ku80c Des transgènes exprimant des protéines de fusion GFP-Ku70a ou GFP-Ku80c ont été introduits par micro-injection dans des Mac de paramécies végétatives et les protéines ont été localisées spécifiquement dans les ébauches (MARMIGNON ET AL. 2014). Dans ces expériences, l'observation du signal GFP a été compliquée par la difficulté d'avoir un signal homogène entre les cellules d'une population transformée, particulièrement pour GFP-Ku80c. Ces problèmes techniques étant probablement liées soit à la micro-injection, soit à l'étiquette GFP, nous avons cherché à développer des anticorps spécifiques de Ku80c, permettant une observation de la protéine endogène. De précédents essais de créations d'anticorps s'étant révélés infructueux, nous avons tenté de nouvelles immunisations de lapin purifiées par affinité contre le peptide immunogène avec deux nouvelles régions de Ku80c qui nous ont donné deux lots : α -Ku80c-1 et α -Ku80c-2. Les deux ont été testés en IF et ont montré des résultats identiques (seul α -Ku80c-1 est montré ici) **Figure IV.8**. Dans la condition contrôle, l'anticorps marque l'ensemble des noyaux faiblement (ébauches + fragments) et aucune différence n'est visible avec l'expérience d'ARNi *KU80c*. Cela signifie que l'anticorps n'est pas spécifique à Ku80c et n'est donc pas utilisable tel quel en IF.

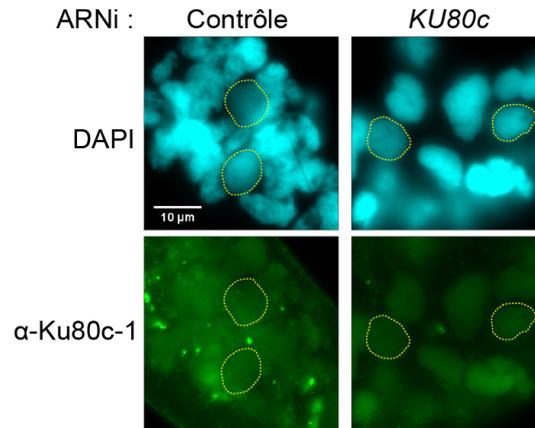


FIGURE IV.8 – Test d'anticorps α -Ku80c

IF avec l' α -Ku80c-1 en protocole de pré-extraction sur des cellules autogames à T5 en fonction de l'ARNi. Microscopie à épifluorescence. Dans cette expérience il y a eu 100% de descendance sexuelle fonctionnelle pour la condition contrôle et 0% pour l'ARNi *KU80c*

Utilisation d'un transgène *FLAG-KU80c* et effets d'ARNi *PGM*, *KU70* ou *XRCC4* sur la stabilité de la localisation de Flag-Ku80c Pour pouvoir malgré ces problèmes techniques étudier la protéine Ku80c, nous avons décidé de retourner à des expériences de micro-injections utilisant la même construction que celle décrite dans MARMIGNON ET AL. (2014) en changeant l'étiquette GFP par une étiquette Flag. Ce transgène porte les séquences régulatrices endogène du *KU80c* et permet donc une induction pendant les RPG.

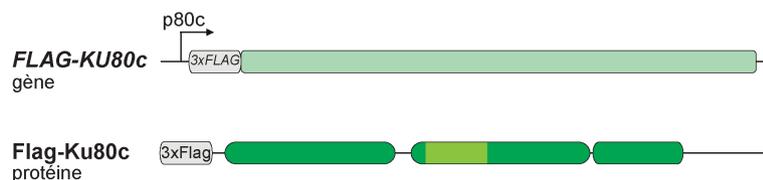


FIGURE IV.9 – Transgène *FLAG-KU80c* et protéine associée théorique

Le transgène est sous le contrôle du promoteur endogène de *KU80c*. L'étiquette fusionnée est trois Flag successifs (séquence dans le matériels et méthodes).

La transformation se fait par micro-injection de l'ADN transgénique, directement dans le Mac de paramécies végétatives. Les transgènes sont protégés par l'addition de télomères à leurs extrémités. Ils peuvent également former des concatémères avant cette étape de télomérisation. Ils sont maintenus lors des divisions végétatives comme des mini-chromosomes se répliquant de manière autonome (GILLEY ET AL. 1988, BOURGAIN AND KATINKA 1991, KATINKA AND BOURGAIN 1992). Les transgènes sont conservés et exprimés

jusqu'à la destruction de l'ancien Mac, provoquant ainsi la perte du transgène dans la nouvelle génération sexuelle.

Suite à la transformation de paramécies par micro-injection, nous avons pu observer en IF (en protocole de pré-extraction) la protéine Flag-Ku80c dans des paramécies autogames. Nous avons réalisé différentes conditions d'ARNi pour chercher un possible effet de Pgm, Ku70 ou Xrcc4 sur la stabilité de la localisation nucléaire de Flag-Ku80c (**Figure IV.10**).

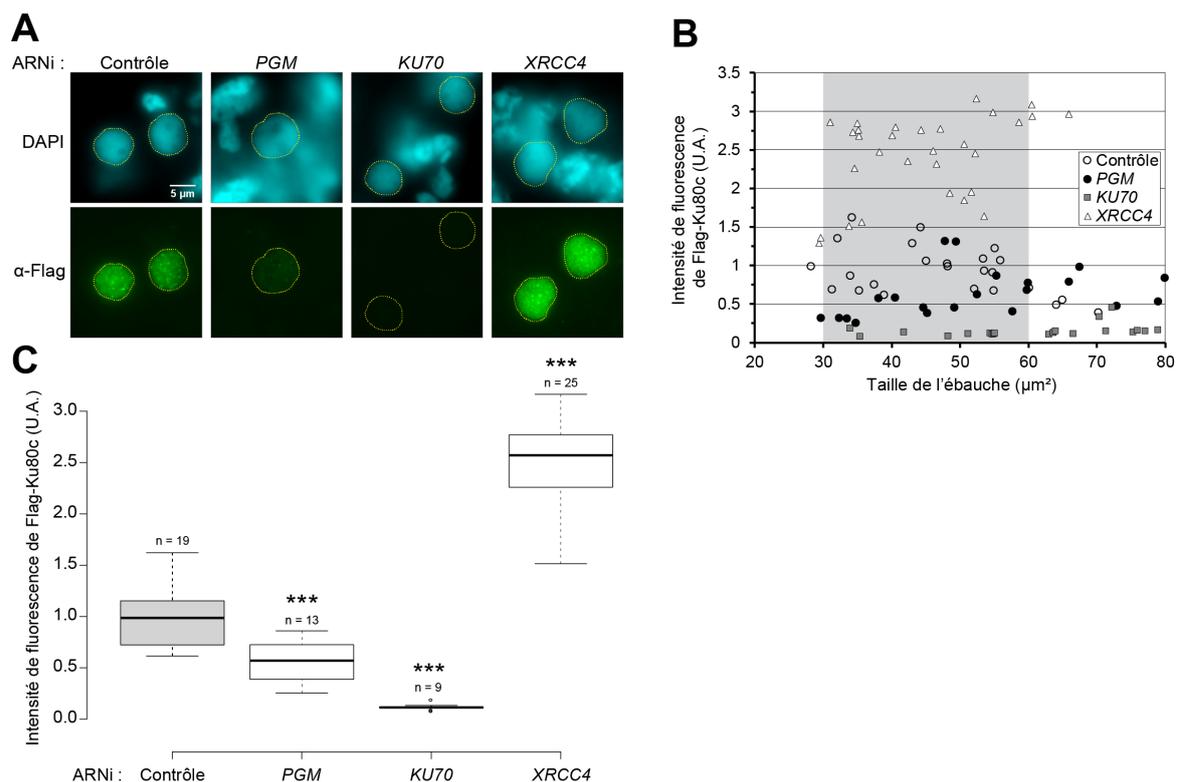


FIGURE IV.10 – IF en pré-extraction de Flag-Ku80c au cours de l'autogamie en ARNi PGM, KU70 ou XRCC4

(A) Localisation de Flag-Ku80c dans des cellules autogames à T5 en fonction de l'ARNi. Microscopie à épifluorescence. (B) Intensité de fluorescence de Flag-Ku80c dans les ébauches en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm^2 . Un bruit de fond correspondant au signal de la lame en fluorescence verte a été soustrait à l'ensemble des données (zone sans cellule). En condition contrôle, Flag-Ku80c présente un pic de localisation nucléaire entre 30 et 60 μm^2 , définissant ainsi la zone d'intérêt pour la comparaison entre les différentes ARNi (zone grise). (C) *Boxplot* de l'intensité de fluorescence de Flag-Ku80c des ébauches comprises entre 30 et 60 μm^2 . Le n représente le nombre d'ébauches par condition. Signal moyen de Flag-Ku80c : Contrôle 100%, PGM 62%, KU70 12%, XRCC4 245% (Signal partiellement saturant, induisant donc une sous-estimation de la valeur réelle). *** pour $p < 0.001$ dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle. Dans cette expérience il y a eu 100% de descendance sexuelle fonctionnelle pour la condition contrôle et 0% pour les ARNi PGM, KU70 et XRCC4.

En condition contrôle, Flag-Ku80c a une localisation nucléaire spécifique des ébauches.

Le signal nucléaire présente un profil en foci, tout comme le profil qui avait été observé avec GFP-Ku8oc (MARMIGNON ET AL. 2014) et ressemblant au profil de Pgm. Ce profil est difficilement visible ici et sera étudié plus loin dans des expériences utilisant de la microscopie confocale.

Pour une ARNi *PGM*, nous pouvions attendre une forte diminution du signal Flag-Ku8oc due à l'absence de CDB et donc de substrat permettant une forte stabilité de l'interaction Ku avec l'ADN dans le noyau. En ARNi *PGM*, nous observons que Flag-Ku8oc baisse mais pas de manière drastique, avec 62% du signal restant par rapport à la condition contrôle. La diminution peut être attribuée à l'absence de cassure mais les 62% du signal restant sont surprenants. Cela signifie qu'en dépit de l'absence de cassure, Flag-Ku8oc reste majoritairement stablement associé au noyau. Ce résultat très important montre que Flag-Ku8oc interagit probablement avec la chromatine, indépendamment de Pgm. Cette capacité d'interaction pourrait être un élément important dans la mécanique de couplage entre coupure et réparation, potentiellement en guidant Pgm.

Ku70 est le partenaire essentiel de Ku80. Une déplétion de Ku70 est donc sensée déstabiliser Flag-Ku8oc ou au moins l'inactiver. Dans une ARNi *KU70*, on observe une quasi disparition du signal Flag-Ku8oc, avec 12% du signal restant par rapport au contrôle. Ce résultat prévisible confirme le lien étroit entre les deux sous-unités. En absence de Ku70, Flag-Ku8oc est donc soit dégradé, soit sa localisation stable est perturbée. Une étude par WB de Flag-Ku8oc en ARNi *KU70* serait nécessaire pour trancher entre ces deux hypothèses.

En ARNi *XRCC4*, les CDB sont accumulées, de la même manière que dans un ARNi *LIG4* (KAPUSTA ET AL. 2011). La stabilité nucléaire de Ku doit donc être fortement augmentée. Dans la condition ARNi *XRCC4*, une nette augmentation du signal de Flag-Ku8oc est observée, avec au moins 245% de fluorescence par rapport au contrôle (saturation partielle de certaines acquisitions). Ce résultat confirme l'augmentation de la stabilité de Ku lorsque les CDB s'accumulent dans le noyau.

Une étude par WB de la quantité de Flag-Ku8oc dans ces différentes conditions d'ARNi serait nécessaire pour vérifier que la quantité de protéine est stable. De plus, il serait nécessaire de répéter l'expérience d'IF afin de consolider et d'approfondir les résultats présentés. En effet, elle n'a été réalisée qu'une seule fois, la taille de l'échantillon

de données est plutôt faible (13 ébauches quantifiées dans l'ARNi *PGM*, 9 dans l'ARNi *KU70*) et enfin, elle a été réalisée en présence de la protéine Ku80c endogène. Ce dernier point peut poser problème car il pourrait y avoir une compétition entre la protéine endogène et la protéine Flag-Ku80c pour leur localisation nucléaire et ainsi biaiser les résultats. Il faudrait recommencer l'expérience en étant en plus dans chaque condition en ARNi *KU80c*. Le transgène *FLAG-KU80c* devra donc être de plus résistant à l'ARNi.

IV.3 ETUDE DE KU70 ENDOGÈNE PAR UN ANTICORPS α -KU70

Dans la même logique que pour Ku80c, le laboratoire a cherché à développer des anticorps dirigés contre Ku70. Un anticorps, l' α -Ku70B particulièrement prometteur nous encouragea à réaliser différentes expériences de WB et d'IF.

Stabilité de la protéine Ku70 en ARNi *PGM* En prenant des extraits protéiques totaux de paramécies à T5 de l'autogamie, ayant subi différentes conditions d'ARNi, nous avons réalisé un WB contre Ku70 avec l' α -Ku70B (**Figure IV.11**).

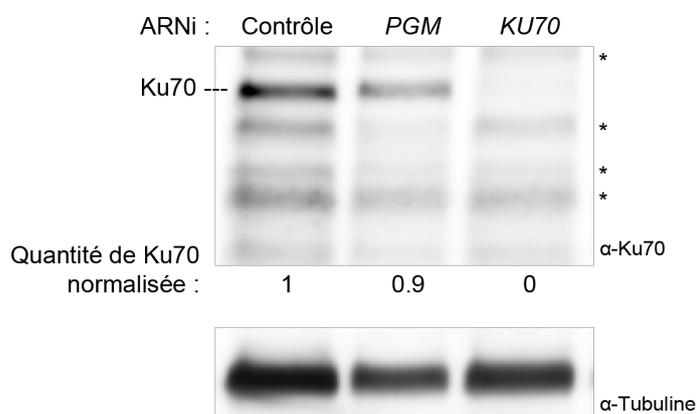


FIGURE IV.11 – WB de Ku70 au cours de l'autogamie dans différentes conditions d'ARNi

Extraits protéiques totaux de cellules à T5, migrés en SDS-PAGE et révélés en WB par l'anticorps α -Ku70. La membrane a ensuite été deshybridée puis révélée par l' α -Tubuline pour permettre la normalisation de la quantité de Ku70 dans les différents extraits. Pourcentage de cellules avec un nouveau Mac fonctionnel par condition d'ARNi : contrôle 100%, *PGM* et *KU70* 0%.

En condition contrôle, plusieurs bandes sont visibles, dont une majoritaire à la taille attendue de Ku70. Cette bande disparaît totalement en ARNi *KU70*, alors que les autres bandes signalées par un * persistent. Cela signifie que l'anticorps nous permet de détec-

ter la protéine Ku70 endogène en WB. Il détecte également de nombreuses bandes non spécifiques.

De manière similaire à l'observation de la stabilité de Pgm en absence de Ku, j'ai voulu tester la stabilité de Ku70 en absence de Pgm. En ARNi *PGM*, la bande correspondant à Ku70 est toujours présente et la normalisation à l'aide de la tubuline révèle que la quantité de Ku70 est stable, ce qui indique donc que Pgm n'est pas nécessaire pour la stabilité de la protéine Ku70.

Non spécificité de l' α -Ku70 en IF L'anticorps a ensuite été testé en IF avec le protocole de pré-extraction sur des cellules autogames à T5 (**Figure IV.12**).

En condition contrôle, la fluorescence est spécifique des ébauches et présente un motif en foci. Cette observation est très proche de ce qui avait été observé avec la fusion GFP-Ku70a (MARMIGNON ET AL. 2014) et semblait donc très encourageante sur la qualité de l' α -Ku70. Malheureusement, en ARNi *KU70*, la fluorescence est encore très présente, à hauteur de 79% par rapport à la condition contrôle. La quantité de protéine Ku70 dans cette condition avait déjà pu être déterminée par WB dans la **figure IV.11**, qui montrait que la protéine Ku70 n'était plus détectable dans une ARNi *KU70*. Le signal d'IF fluorescent qui est détecté ici ne correspond donc pas à la protéine Ku70 mais à une autre protéine, inconnue et localisée spécifiquement et stablement dans les ébauches. Cela confirme les données de WB et signifie que l'anticorps dirigé contre Ku70 n'est pas spécifique, et empêche donc l'utilisation de cet anticorps en IF.

De la même manière que pour Ku80c ce problème a été contourné en utilisant une fusion Flag-Ku70. Ces travaux ont été réalisés récemment par Julien Bischerour et les résultats ne sont pas présentés dans ce manuscrit.

IV.4 LES FOCI KU ET PGM, STRUCTURES FONCTIONNELLES ?

La présence de foci pour Pgm (DUBOIS ET AL. 2017) et Ku (MARMIGNON ET AL. 2014) questionne. Lors d'une ARNi *PGM*, le profil en foci de Flag-Ku80c est perdu et le signal devient globalement homogène (**Figure IV.13**).

La disparition de ces foci indique qu'ils sont dépendants de Pgm et donc probablement des RPG. Ils représentent donc potentiellement une structure fonctionnelle : soit

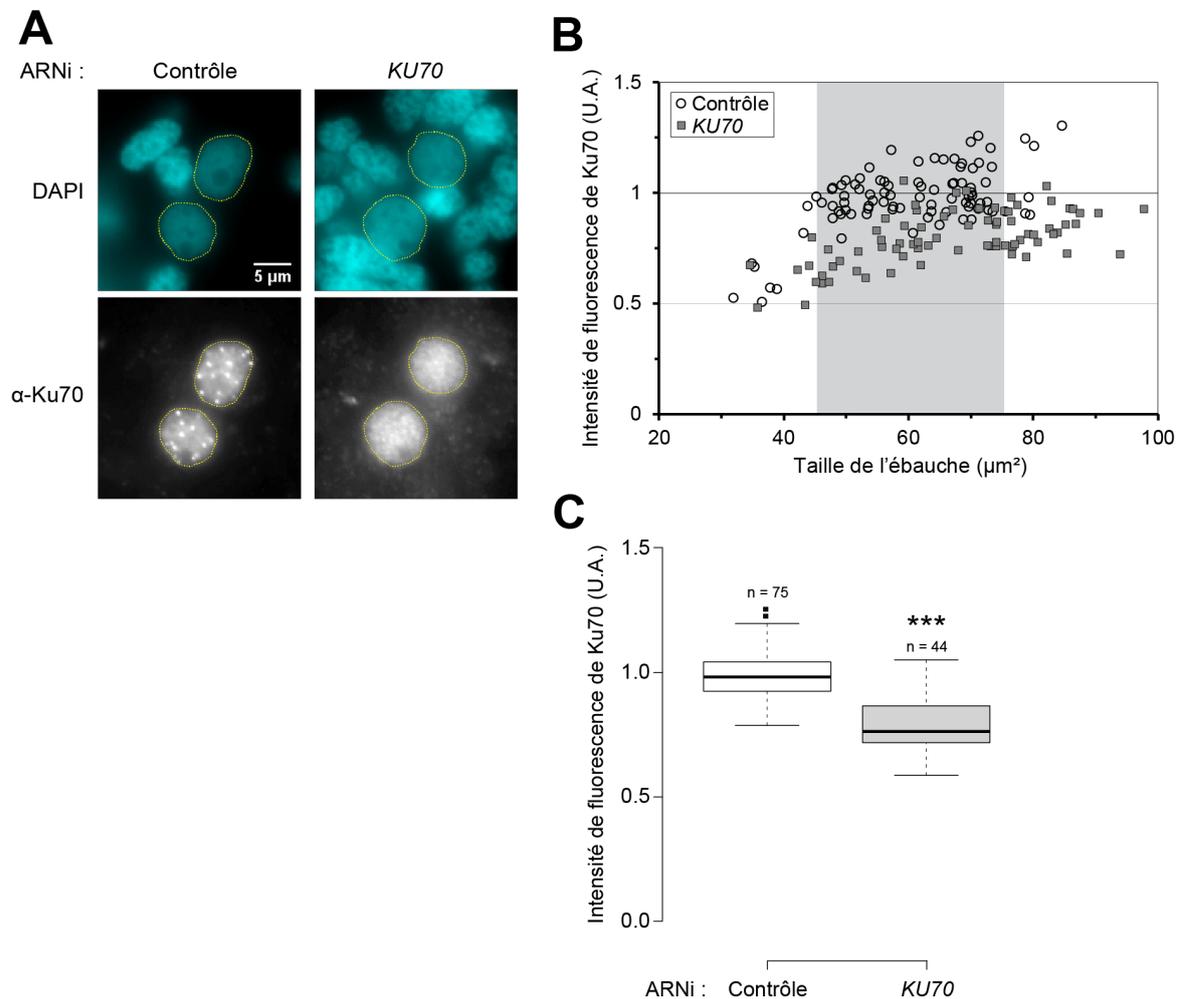


FIGURE IV.12 – IF de *Ku70* au cours de l'autogamie dans différentes conditions d'ARNi

(A) IF anti-Ku70 dans des cellules autogames à T5 en ARNi contrôle ou *KU70*. Microscopie à épifluorescence. (B) Intensité de fluorescence de Ku70 dans les ébauches en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm^2 . Un bruit de fond correspondant au signal de la lame en fluorescence verte a été soustrait à l'ensemble des données (zone sans cellule). En condition contrôle, Ku70 présente un pic de localisation nucléaire entre 45 et 75 μm^2 , définissant ainsi la zone d'intérêt pour la comparaison entre les différents ARNi (zone grise). (C) *Boxplot* de l'intensité de fluorescence de Ku70 des ébauches comprises entre 45 et 75 μm^2 . Le n représente le nombre d'ébauches par condition. Signal moyen de Ku70 : Contrôle 100%, *KU70* 79. *** pour $p < 0.001$ dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle. Dans cette expérience il y a eu 100% de descendance sexuelle fonctionnelle pour la condition contrôle et 0% pour l'ARNi *KU70*.

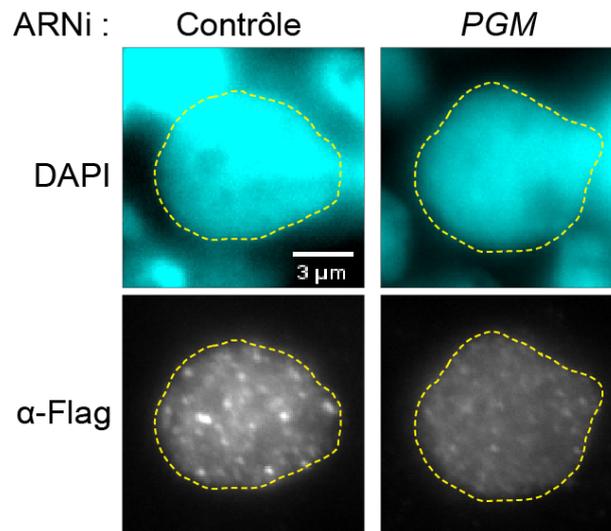


FIGURE IV.13 – Disparition des foci *Flag-Ku8oc* en ARNi PGM

IF de *Flag-Ku8oc* dans des cellules autogames à T5 en ARNi contrôle ou PGM présentant deux profils de localisation. Microscopie à épifluorescence. Dans cette expérience il y a eu 100% de descendance sexuelle fonctionnelle pour la condition contrôle et 0% pour l'ARNi PGM.

un lieu où les cassures se produisent, soit un lieu où les réparations ont lieu, soit les deux. L'accumulation de protéines Ku dans une telle zone pourrait être expliquée par une concentration d'extrémités double-brin d'ADN. Nous avons également cherché à savoir si les foci *Flag-Ku8oc* colocalisaient avec les foci Pgm. Malheureusement, des difficultés techniques nous ont empêché d'obtenir ce résultat directement. En effet, le marquage Pgm est très sensible aux conditions techniques de fixation et le profil en foci peut être facilement perdu. Nous avons cependant pu obtenir un résultat indirect, impliquant l' α -Ku70 non spécifique (**Figure IV.14**).

Bien qu'il soit impossible de dire que Ku70 est bien la protéine révélée par cet anticorps en IF, l'utilisation d' α -Ku70 révèle un profil en foci. Les foci de l' α -Ku70 colocalisent en grande majorité avec ceux de Pgm (**Figure IV.14A**) et ils colocalisent également avec ceux de *Flag-Ku8oc* (**Figure IV.14B**). Il est donc très probable que les protéines Pgm et Ku8oc fassent partie de cette même structure en foci, dont la nature exacte et la fonction restent à ce jour inconnues. Des expériences complémentaires de localisation avec d'autres partenaires des RPG permettront de mieux déterminer la nature de ces structures.

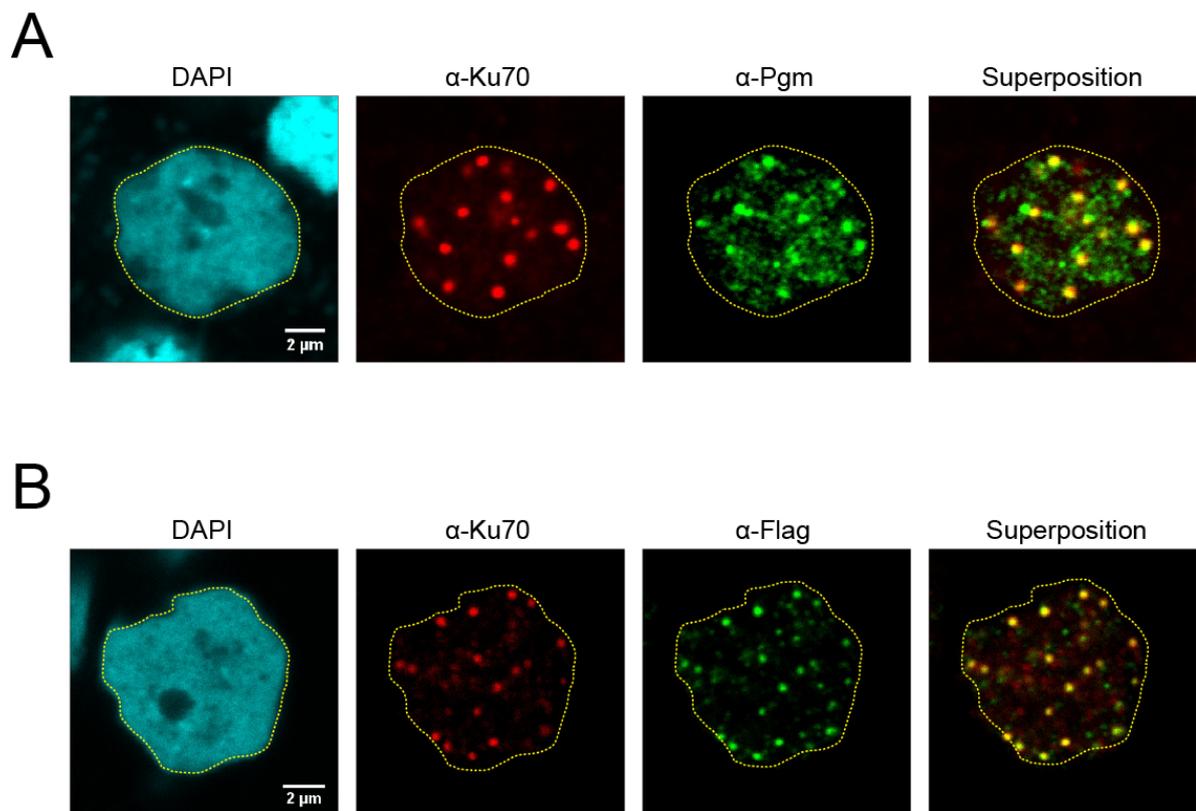


FIGURE IV.14 – Colocalisation de Ku70 et Pgm en conjugaison et colocalisation de Ku70 et Flag-Ku80c en autogamie

(A) IF α -Ku70 et α -Pgm sur des cellules en conjugaison. Microscopie confocale. (B) IF α -Ku70 et α -Flag sur des cellules autogames transformées par FLAG-KU80c. Microscopie confocale.

IV.5 RÉSULTATS PRINCIPAUX DU CHAPITRE

La stabilité de la localisation nucléaire de Pgm est abolie en absence de Ku. De plus, une déplétion de Ku provoque une rétention globale des IES. A l'inverse, en absence de Pgm, la majorité de Flag-Ku80c reste stablement associé aux ébauches. Ku80c interagit donc probablement avec la chromatine, indépendamment de Pgm. Cette capacité d'interaction pourrait être un élément important dans la mécanique de couplage entre coupure et réparation, potentiellement en guidant Pgm.

Fluorescence Pgm		ARNi					
		Contrôle	<i>PGM</i>	<i>KU70</i>	<i>KU80c</i>	<i>DNAPK-cs</i>	<i>LIG4</i>
Sans pré-extraction		= (x1)	↓↓↓ (x0,06)	↓ (x0,57)	↓ (x0,75)	Non testé	Non testé
Avec pré-extraction		= (x1)	↓↓↓ (x0,18)	↓↓ (x0,30)	↓↓↓ (x0,20)	↗ (x1,86)	↓↓ (x0,41)

Fluorescence Flag-Ku80c		ARNi			
		Contrôle	<i>PGM</i>	<i>KU70</i>	<i>XRCC4</i>
Avec pré-extraction		= (x1)	↓ (x0,62)	↓↓↓ (x0,12)	↗↗ (x2,45)

FIGURE IV.15 – Résumé des résultats de localisation et stabilité nucléaire de Pgm et Flag-Ku80c en fonction des différentes ARNi

Chapitre V

Spécialisation fonctionnelle de Ku80c

V.1 ANALYSE *in silico* DE L'ÉVOLUTION DES PROTÉINES KU80 DE CILIÉS

Paramecium tetraurelia possède trois gènes *KU80*, issus des deux dernières WGD. Ces deux WGD ont eu lieu avant la spéciation des différentes espèces composant le complexe des *aurelia* (Figure V.1).

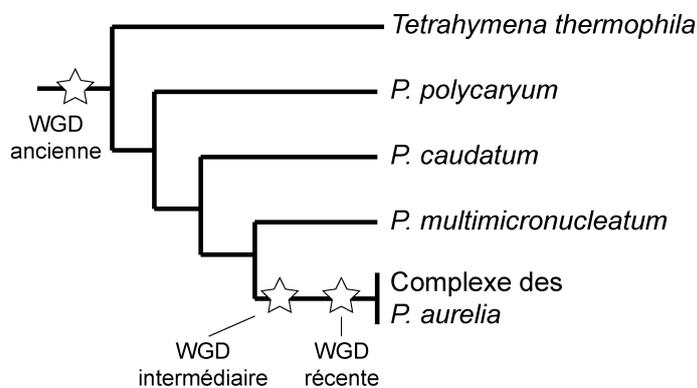


FIGURE V.1 – Datation des événements de WGD

Emplacement des WGD dans la phylogénie des paramécies et de *T. thermophila* (AURY ET AL. 2006, McGRATH ET AL. 2014).

Chez *P. tetraurelia*, seul l'ARNi de *KU80c* conduit à un blocage des RPG, alors que des ARNi contre *KU80a*, *KU80b* ou les deux ne provoquent aucun phénotype. Une recherche par homologie de séquence a permis de révéler tout un ensemble de protéines Ku80 chez les espèces de paramécies dont le génome a été séquencé récemment et de construire un arbre phylogénétique avec ces séquences (Figure V.2).

Cet arbre phylogénétique révèle que l'ensemble des paramécies du complexe *aurelia*

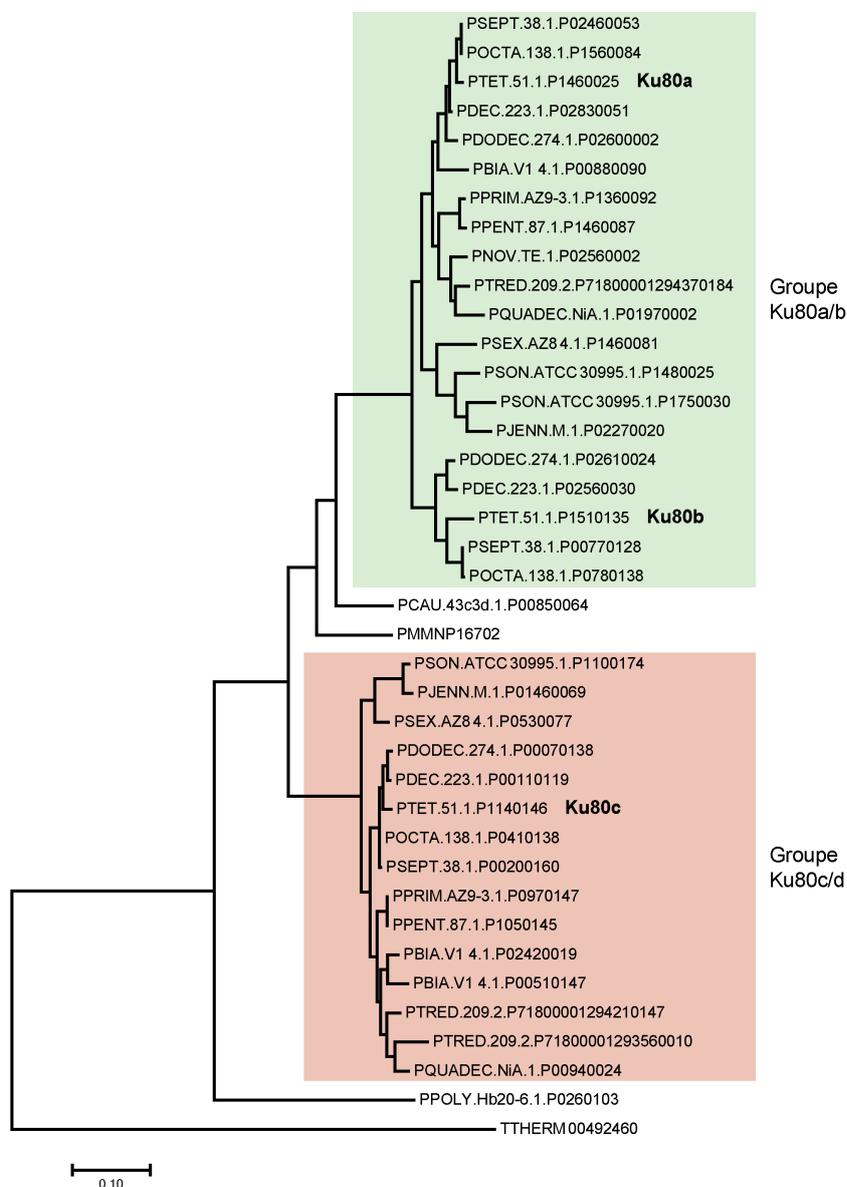


FIGURE V.2 – Phylogénie des *Ku80* chez les ciliés

Les séquences ont été obtenues par homologie de séquence protéique de la base de données paramciumdb (ARNAIZ AND SPERLING 2011, GOUT ET AL. 2019). L'arbre phylogénétique a été construit en utilisant la méthode de *Maximum Likelihood* basée sur le modèle *JTT matrix* (JONES ET AL. 1992). Les données relatives à *P. septaurelia* (PSEPT.38) sont à prendre avec précaution. En effet, il pourrait s'agir d'une souche de *P. octaurelia* mal caractérisée (Linda Sperling, communication personnelle).

possèdent au moins un membre du groupe Ku8oa/b et au moins un membre du groupe Ku8oc/d. Cette présence systématique des deux versions de Ku8o semble indiquer un rôle important et spécifique de chacune d'entre elles.

P. caudatum, *multimicronucleatum polycarium* et *T. thermophila* ne possèdent qu'une version de Ku8o. De manière intéressante, les versions Ku8o de *P. caudatum* et *multimicronucleatum* sont intermédiaires entre Ku8oa et Ku8oc, donnant ainsi un indice supplémentaire sur une possible spécialisation de Ku8oc pour les RPG.

V.2 COMPLÉMENTATION FONCTIONNELLE D'UNE ARNi CONTRE *KU8oc* PAR FLAG-KU8OA

Afin de tester la spécialisation fonctionnelle de Ku8oc, j'ai mis en place un système de complémentation d'une ARNi contre *KU8oc* endogène par des transgènes exprimant soit Ku8oc soit Ku8oa.

V.2.1 Création des transgènes nécessaires pour la complémentation

Lors d'une ARNi *KU8oc*, le niveau d'expression de *KU8oa* sous contrôle de ses signaux de transcription endogènes pourrait ne pas être suffisant pour compléter l'absence de Ku8oc. Il est donc nécessaire de surexprimer *KU8oa* lors des RPG pour tester sa capacité à compléter l'absence de Ku8oc. Il est possible de réaliser cela en utilisant des transgènes portant les séquences régulatrices de *KU8oc* et permettant l'expression de *FLAG-KU8oc* ou de *FLAG-KU8oa* lors des RPG (**Figure V.3**). De plus, pour permettre l'expression de *FLAG-KU8oc* en situation de déplétion du Ku8oc endogène, la séquence nucléotidique de *FLAG-KU8oc* a été changée sur la zone visée par l'ARNi. Les changements de nucléotides n'impactent pas la séquence protéique grâce à la redondance du code génétique. L'expression de cette version Flag-Ku8oc en absence du Ku8oc endogène permettra de vérifier que l'étiquette Flag ne perturbe pas le déroulement des RPG.

La quantité de transgènes introduits dans le Mac par micro-injection n'est pas finement régulée et peut avoir une incidence sur la quantité de protéine exprimée. Pour chaque clone transformé, le niveau d'injection est mesuré par PCR quantitative (qPCR) et est exprimé en Copies Par Génome Haploïde (CPGH).

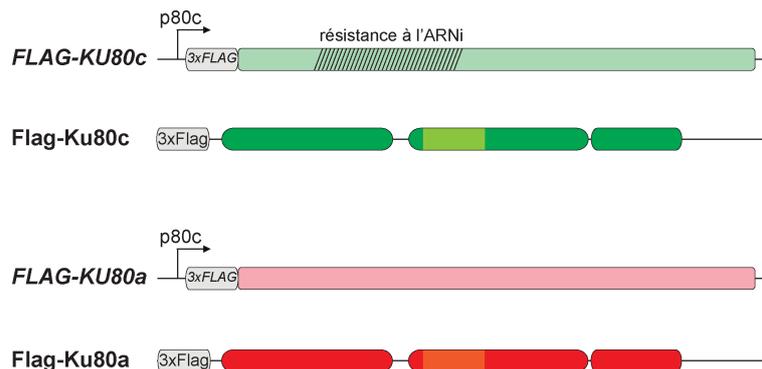


FIGURE V.3 – Transgène *FLAG-KU80c* résistant à l'ARNi, transgène *FLAG-KU80a* et protéines associées théoriques

Les transgènes sont sous le contrôle du promoteur endogène de *KU80c*. L'étiquette fusionnée est trois Flag successifs. La zone barrée représente les modifications de la séquence du transgène rendant l'ARNm résistant à l'ARNi, sans modifier la séquence de la protéine.

V.2.2 Flag-Ku80a ne complète pas une déplétion de Ku80c

Des injections de *FLAG-KU80a* ou *FLAG-KU80c* ont été réalisées sur de nombreux clones afin d'obtenir une gamme variée de CPGH. Pour chacun des CPGH, un test de survie de la descendance post-autogame a été réalisé, soit en condition contrôle, soit en ARNi *KU80c* (Figure V.4).

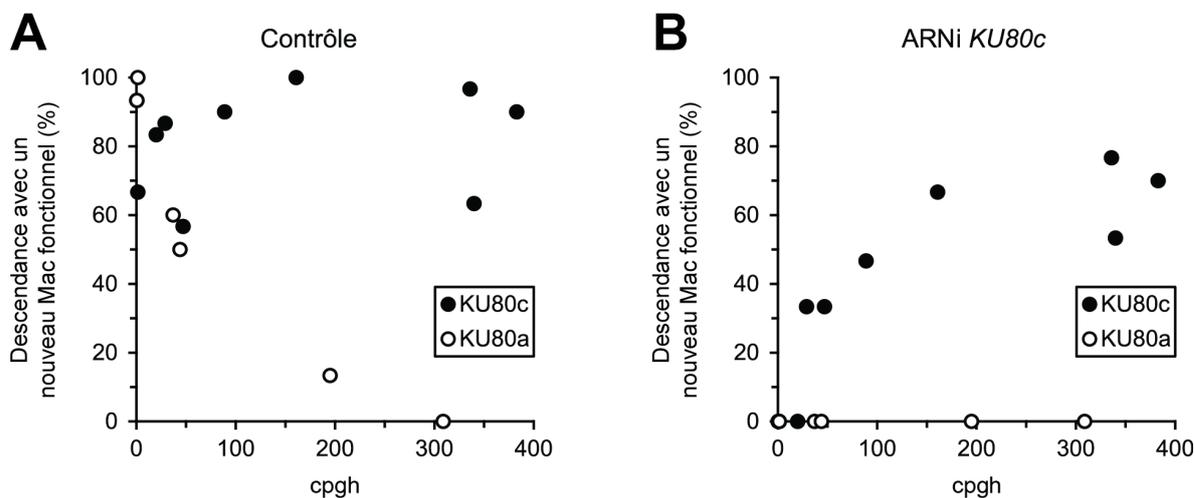


FIGURE V.4 – Complémentations de la déplétion de *Ku80c* par *Flag-Ku80c* ou *Flag-Ku80a* en fonction du CPGH

Pourcentage de descendance sexuelle viable en fonction du niveau d'injection dans des cellules complémentées par *FLAG-KU80c* ou *FLAG-KU80a*.

La condition contrôle permet de vérifier que le transgène n'a pas un effet toxique sur les RPG et l'ARNi *KU80c* permet de tester les capacités de complémentation du transgène.

Complémentation par Flag-Ku80c Pour *Flag-KU80c* en condition contrôle, on observe entre 60 et 100% de survie par clone entre 0 et 400 CPGH (**Figure V.4A**). À ces niveaux d'injection, *Flag-KU80c* n'est donc pas toxique. En ARNi contre *KU80c*, Flag-Ku80c permet une complémentation de l'absence de la protéine endogène, avec une restauration de la survie comprise entre 50 et 80% à partir de 100 CPGH (**Figure V.4B**). Flag-Ku80c est donc au moins partiellement fonctionnelle. La petite baisse entre la situation contrôle et l'ARNi *KU80c* pourrait être expliquée soit par la présence du Flag qui pourrait perturber l'activité de la protéine ou bien par le fait qu'il s'agit d'un transgène, pouvant entraîner une expression différentielle de la protéine par rapport au locus endogène. Il existe également un phénomène où des ARN aberrants peuvent être transcrits à partir du transgène et déclencher une ARNi, surtout à haut CPGH (GARNIER ET AL. 2004).

Bien que ce système de transgène avec une étiquette ne soit pas parfait, il permet donc d'obtenir une bonne complémentation de l'absence du Ku80c endogène. Il permet donc de tester l'effet de la surexpression de Ku80a lors de l'autogamie.

Non complémentation par Flag-Ku80a En condition contrôle, Flag-Ku80a est toxique à 50% à 50 CPGH et est fortement toxique au-delà de 200 CPGH (**Figure V.4A**). De plus, aucune complémentation de l'absence du Ku80c endogène n'est observée, quel que soit le CPGH (**Figure V.4B**).

Flag-Ku80a n'est donc pas capable de compléter l'absence de Ku80c, indiquant que Ku80c et Ku80a ne sont pas équivalents. Nous pouvons donc conclure que Ku80c s'est spécialisé pour les RPG. De plus Flag-Ku80a est toxique en situation contrôle. Cette toxicité pourrait provenir d'une compétition entre le Ku80c endogène et la fusion Flag-Ku80a pour l'interaction avec des partenaires essentiels, tels que Ku70 ou Pgm par exemple.

Le CPGH indique la quantité de transgène présent dans le clone mais n'est pas une mesure directe de la quantité de protéine. Des analyses par WB devront être effectuées sur de futures expériences pour être sûr de comparer des situations de complémentation à niveaux de protéines similaires entre Flag-Ku80c et Flag-Ku80a.

V.2.3 Flag-Ku80a provoque la rétention d'IES et donc une inhibition des coupures

Ces résultats permettent de proposer que Ku80a ne peut pas remplir le rôle de Ku80c lors de l'autogamie. Cependant, Ku80c intervient à différents niveaux des RPG (introduction des cassures puis réparation) et cette expérience ne permet pas de tester le ou les rôles que Ku80a est incapable d'effectuer par rapport à Ku80c. Pour cela, il est possible de regarder par PCR la présence ou non d'IES dans le génome de paramécies autogamies tardives (**Figure V.5**).

Les PCR sont effectuées autour de sites d'excision d'IES. Si le génome a été réarrangé, le produit de PCR sera petit et si l'IES a été maintenue dans le génome, sa taille s'ajoutera au produit, ce qui l'allongera. Une migration par électrophorèse permet ensuite de séparer les deux produits, avec ou sans IES.

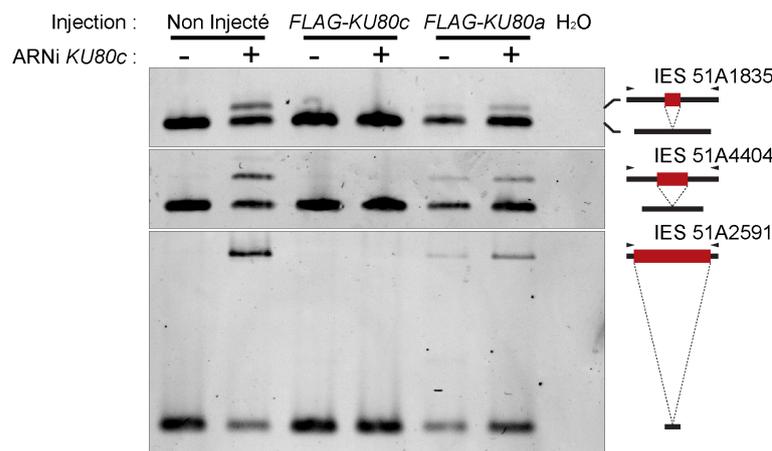


FIGURE V.5 – PCR autour d'IES lors de complémentations

Produits de PCR autour de sites d'excisions d'IES sur des paramécies autogames tardives. En situation contrôle positif (non injectée ARNi KU80c -), les RPG sont finis et ne sont détectées que des formes où l'IES a été excisée. En contrôle négatif (non injectée ARNi KU80c +) les RPG ont été bloqués et les IES sont retenues dans le génome. Les formes IES- visibles proviennent des fragments de l'ancien Mac. CPGH clone *FLAG-KU80c* : 1585. CPGH clone *FLAG-KU80a* : 882. Taux de survie de la descendance sexuelle : non injectée ARNi- 75%; non injectée ARNi+ 0%; *FLAG-KU80c* ARNi- 96%; *FLAG-KU80c* ARNi+ 96%; *FLAG-KU80a* ARNi- 0%; *FLAG-KU80a* ARNi+ 0%.

Dans la situation contrôle sans ARNi, les formes avec IES ne sont pas observables, indiquant le bon déroulement des RPG. En ARNi contre *KU80c*, les formes avec IES apparaissent, confirmant ainsi l'absence d'introduction des CDB.

En complémentation avec Flag-Ku80c (ARNi *KU80c* +), aucune forme avec IES n'est observable indiquant donc la bonne excision de ces IES. En complémentation avec Flag-

Ku80a (ARNi *KU80c* +), les produits de PCR avec IES sont observables, les IES sont donc retenues dans le génome. Cela montre qu'en complémentation par Flag-Ku80a, les CDB sur au moins les trois IES testées ici ne sont pas introduites efficacement.

La toxicité de Flag-Ku80a en absence d'ARNi s'accompagne d'une faible rétention des IES, expliquant le phénotype létal de cette condition : les CDB ne sont pas efficacement introduites, renforçant l'hypothèse de la compétition avec le Ku80c endogène.

V.2.4 Flag-Ku80c n'est pas détectable en IF sans pré-extraction

Parmi les caractéristiques principales de Ku80c, on retrouve sa capacité à se localiser stablement dans les ébauches et celle de permettre à Pgm de lui-même se localiser stablement dans les ébauches. Pour déterminer si Ku80a partage ces caractéristiques ou si elles sont spécifiques à Ku80c, des expériences d'IF sur des cellules transformées en condition de déplétion du *KU80c* endogène ont été réalisées, avec ou sans pré-extraction (Figure V.6).

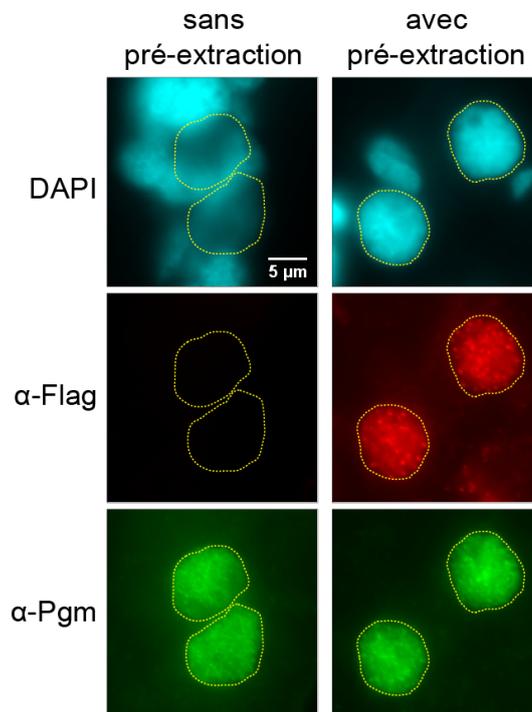


FIGURE V.6 – IF α -Flag et α -Pgm, avec ou sans pré-extraction, sur des cellules injectées FLAG-KU80c, en déplétion du KU80c endogène

IF dans des cellules autogames à T5 en ARNi *KU80c* injectées avec FLAG-KU80c. Utilisation du protocole sans pré-extraction ou de celui avec sur la même récolte de cellules. Microscopie à épifluorescence.

Malheureusement en conditions de fixation sans pré-extraction, aucune fluorescence nucléaire Flag n'est visible, alors que dans les même condition, Pgm est observable facilement et que les mêmes cellules ont été traitées en parallèle au protocole avec pré-extraction, permettant de révéler un signal nucléaire fort pour Flag-Ku80c. Ce résultat montre donc qu'en IF sans pré-extraction, Flag-Ku80 n'est pas observable, probablement dû à un problème d'accès de l'anticorps pour l'antigène Flag. Ce problème technique nous empêche donc d'observer la capacité de Flag-Ku80c à se localiser (stablement ou non) dans les ébauches.

V.2.5 GFP-Ku80a se localise dans les ébauches mais n'y est pas stablement accroché

Pour malgré tout réussir à répondre à la question importante de la localisation nucléaire différentielle de Ku80a et Ku80c dans les ébauches, nous sommes ponctuellement retournés aux transgènes *GFP-KU80c* et *a*, permettant d'éviter tout problème d'accession à un antigène. Le transgène *GFP-KU80c* est celui utilisé dans MARMIGNON ET AL. (2014) et le transgène *GFP-KU80a* disponible au laboratoire est identique à *FLAG-KU80a* à l'exception du changement d'étiquette. Malheureusement, le transgène *GFP-KU80c* n'est pas résistant à l'ARNi contre le *KU80c* endogène. Ce transgène ne permet donc pas de vérifier la complémentation lors d'un test de survie. Seule la localisation des protéines a été observée (en présence du *Ku80c* endogène) (**Figure V.7**).

Sans pré-extraction (**Figure V.7A-C**), GFP-Ku80c et GFP-Ku80a sont fortement visibles dans les ébauches, avec un signal moyen de 100% pour GFP-Ku80c et 140% pour GFP-Ku80a (les différences sont dues à la différence de protéine totale provoquée par la différence de CPGH - 549 CPGH pour l'injectée GFP-Ku80c ; 850 CPGH pour l'injectée GFP-Ku80a). L'incapacité de Ku80a à activer les RPG pouvait laisser supposer une incapacité à rentrer dans les ébauches. Sa présence est donc surprenante.

Avec pré-extraction (**Figure V.7D-F**), GFP-Ku80c présente un signal similaire à Flag-Ku80c : une localisation nucléaire spécifique des ébauches avec présence de foci. GFP-Ku80a présente lui un comportement complètement différent : la quasi-totalité du signal est perdu dans ces conditions de fixation avec seulement 20% de signal par rapport à GFP-Ku80c dans les mêmes conditions.

GFP-Ku80a se localise donc dans les ébauches mais ne s'y associe pas fortement

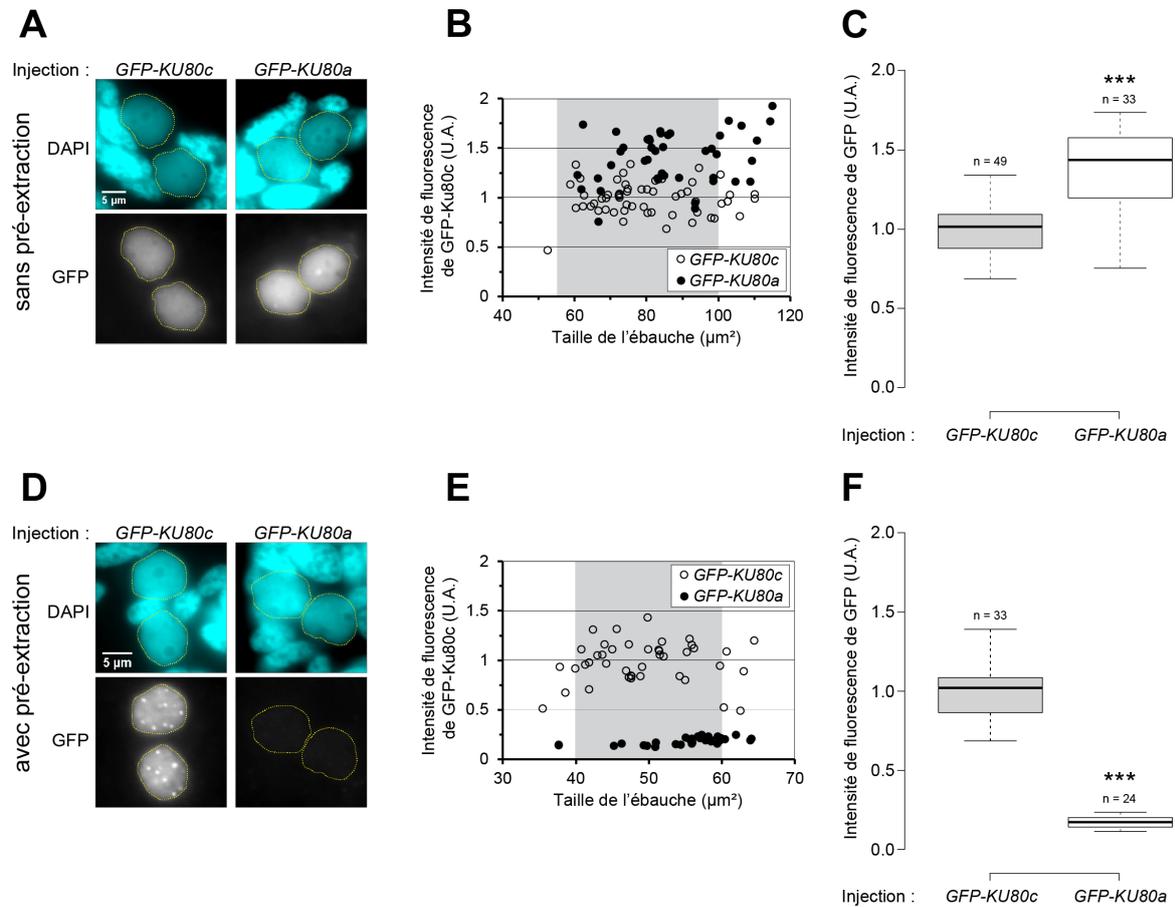


FIGURE V.7 – Localisation des constructions GFP-Ku80 avec ou sans pré-extraction dans des paramécies autogames à T5

(A) Fluorescence GFP dans des cellules autogames à T5 en protocole sans pré-extraction. Microscopie à épifluorescence. (B) Intensité de fluorescence de la GFP dans les ébauches en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm^2 . Un bruit de fond correspondant au signal de la lame en fluorescence verte a été soustrait à l'ensemble des données (zone sans cellule). La zone d'intérêt pour la comparaison des deux transgènes est définie entre 55 et 100 μm^2 (zone grise). (C) Boîte à moustaches de l'intensité de fluorescence de la GFP des ébauches comprises entre 55 et 100 μm^2 . Le n représente le nombre d'ébauches par conditions. Signal moyen de la GFP : GFP-KU80c 100%, GFP-KU80a 140%. (D) Fluorescence GFP dans des cellules autogames à T5 en protocole avec pré-extraction. Microscopie à épifluorescence. (E) Intensité de fluorescence de la GFP dans les ébauches en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm^2 . Un bruit de fond correspondant au signal de la lame en fluorescence verte a été soustrait à l'ensemble des données (zone sans cellule). La zone d'intérêt pour la comparaison des deux transgènes est définie entre 40 et 60 μm^2 (zone grise). (F) Boîte à moustaches de l'intensité de fluorescence de la GFP des ébauches comprises entre 40 et 60 μm^2 . Le n représente le nombre d'ébauches par conditions. Signal moyen de la GFP : GFP-KU80c 100%, GFP-KU80a 20%. *** pour $p < 0.001$ dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle.

comme le fait GFP-Ku80c. Il faut cependant noter que dans ces conditions, on s'attend à avoir des RPG normaux pour le clone exprimant GFP-KU80c, et donc la présence de très nombreuses CDB, alors que les RPG sont très probablement inhibés pour le clone exprimant GFP-Ku80a. Cette différence pourrait-elle expliquer l'absence de localisation nucléaire stable de GFP-Ku80a ?

Pour rappel, l'expérience présentée en **figure IV.10** montrait que Flag-KU80c, en protocole de pré-extraction, était capable de se localiser dans les ébauches en ARNi PGM, malgré l'absence de CDB. Sans CDB, Ku80c est stable dans l'ébauche alors que Ku80a est complètement délocalisé. Ku80c doit donc interagir avec un composant des ébauches essentiel pour les RPG, interaction impossible pour Ku80a.

V.2.6 Pgm est localisé plus stablement dans les ébauches en présence de Flag-Ku80c qu'en présence de Flag-Ku80a

Ku80a n'étant ni capable de se localiser stablement dans les ébauches ni d'activer les RPG, il était proposé qu'il ne soit pas capable non plus d'assurer la localisation nucléaire stable de Pgm. Pour tester cela, des IF en pré-extraction ont été réalisées sur des cellules autogames en ARNi *KU80c* et transformées par *FLAG-KU80c* ou *FLAG-KU80a* (**Figure V.8**).

En IF anti-Flag, les résultats confirment les données de la **figure V.7D** : seul Ku80c est fortement présent dans les ébauches et forme des foci alors que Ku80a n'est pas ou très peu présent dans les ébauches (**Figure V.8A-C**).

Concernant Pgm, une complémentation par Flag-Ku80c permet une stabilité nucléaire accrue de la protéine par rapport à une complémentation par Flag-Ku80a (**Figure V.8A,D,E**).

L'ensemble de ces expériences permettent donc de dire que Ku80c s'est spécialisé dans le couplage entre coupure et réparation de l'ADN lors des RPG, en étant essentiel à la localisation nucléaire stable de Pgm et à son activation.

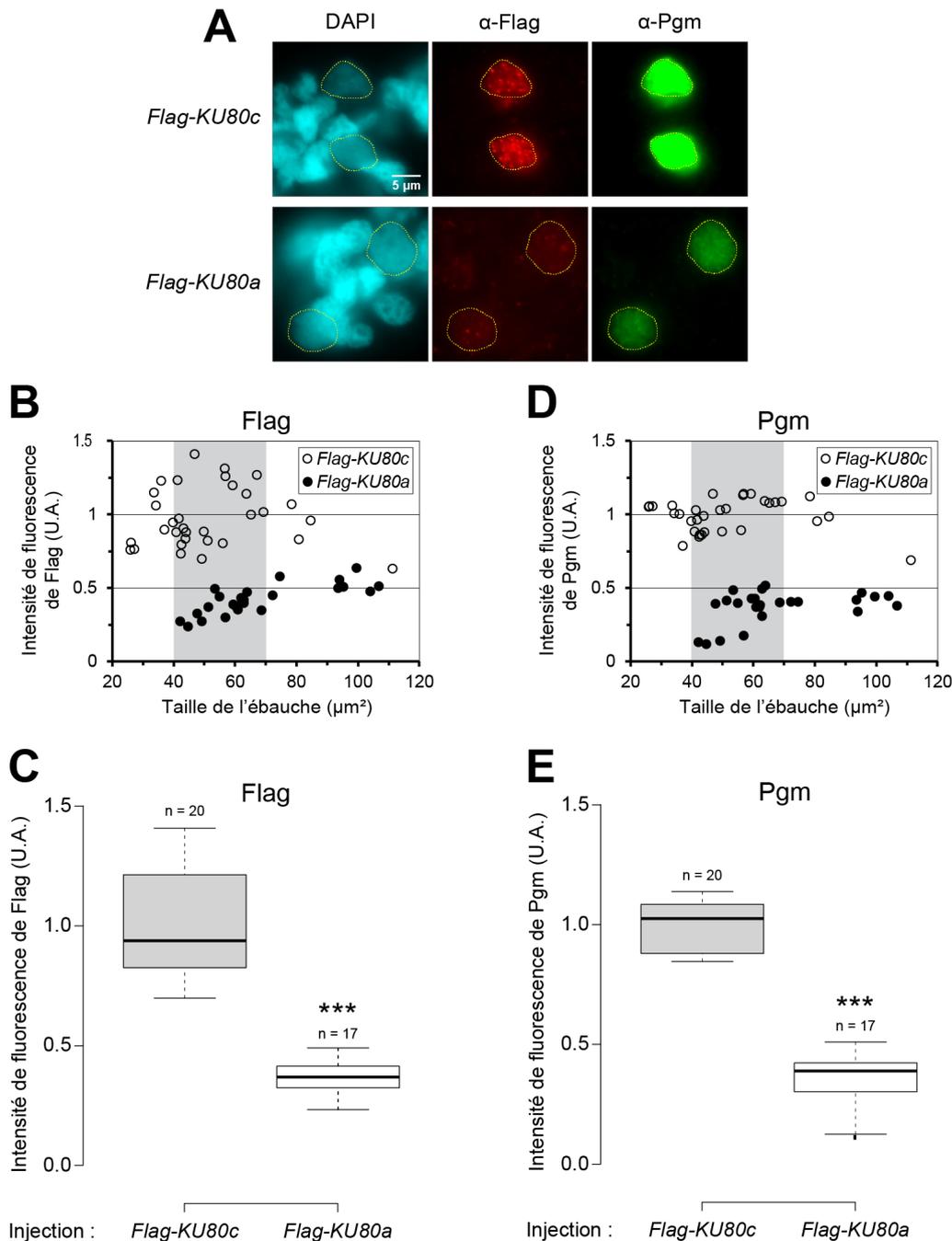


FIGURE V.8 – Localisation des constructions Flag en pré-extraction et effet sur la stabilité de Pgm

(A) IF de Pgm, Flag-Ku80c et α dans des cellules autogames à T5 en déplétion du Ku80c endogène. Microscopie à épifluorescence. (B) Intensité de fluorescence dans l'ébauche de Flag-Ku80c ou Flag-Ku80a en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm^2 . Un bruit de fond correspondant au signal de la lame en fluorescence verte a été soustrait à l'ensemble des données (zone sans cellule). La zone d'intérêt a été choisie entre 40 et 70 μm^2 (zone grise). (C) Boxplot de l'intensité de fluorescence du Flag des ébauches comprises entre 40 et 70 μm^2 . Le n représente le nombre d'ébauches par conditions. (D) Intensité de fluorescence dans l'ébauche de Pgm en unité arbitraire en fonction de la taille de l'ébauche mesurée en μm^2 . La zone d'intérêt a été choisie entre 40 et 70 μm^2 (zone grise). (E) Boxplot de l'intensité de fluorescence du Pgm des ébauches comprises entre 40 et 70 μm^2 . Le n représente le nombre d'ébauches par conditions. *** pour $p < 0.001$ dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle. CPGH clone *FLAG-KU80c* : 161. CPGH clone *FLAG-KU80a* : 881. Taux de survie de la descendance sexuelle : *FLAG-KU80c* 67%; *FLAG-KU80a* ARNi- 0%.

V.3 RECHERCHE D'UN SOUS-DOMAINE RESPONSABLE DE LA SPÉCIALISATION DE KU80C

La spécialisation de Ku80c a dû passer par des modifications de sa séquence protéique. Afin de vérifier s'il existait une région de Ku80c où les modifications se concentraient, les différences entre Ku80c et Ku80a/b ont été représentées sur la structure de l'hétérodimère Ku (**Figure V.9**) (l'alignement des séquences est présent en annexe).

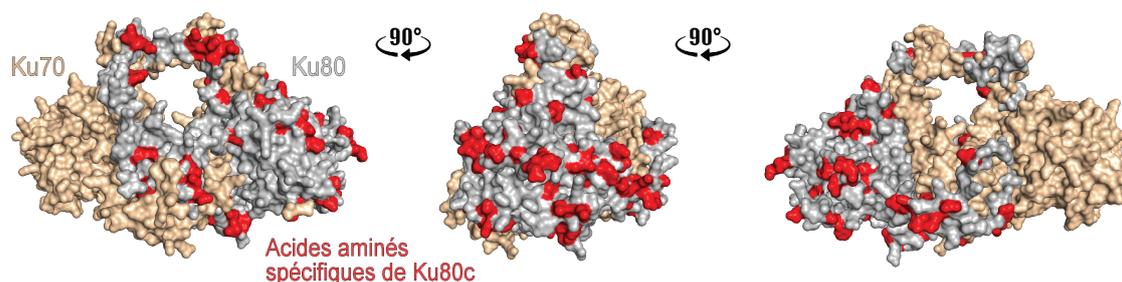


FIGURE V.9 – Positions spécifiques de Ku80c marquées sur la structure

Prédiction de structure de l'hétérodimère Ku70a/Ku80c de paramécie basée sur la structure 1JEQ de l'hétérodimère humain en utilisant l'outil web phyre (KELLEY ET AL. 2015). Les acides aminés spécifique de Ku80c non similaires par rapport à Ku80a et b sont représentés en rouge.

Cette représentation montre que les différences sont réparties sur l'ensemble de la structure, ne permettant donc pas directement de soupçonner une sous-partie de la protéine.

Afin d'identifier s'il existe une sous-région permettant la spécialisation de Ku80c pour les RPG, des constructions chimériques de Ku80c et Ku80a ont été réalisées. Les chimères possèdent le domaine α/β d'une protéine (1-241) et le reste de la séquence de l'autre protéine (242-737 pour Ku80c et 242-739 pour Ku80a) formant ainsi les chimères Ku80ca (α/β de Ku80c et corps de Ku80a) et Ku80ac (inverse) (**Figure V.10**). Comme pour les expériences précédentes, les transgènes sont sous le contrôle des séquences régulatrices de *KU80c* et l'étiquette *FLAG* est fusionnée en 5' afin de suivre les protéines en IF. Les deux chimères sont résistantes à l'ARNi contre *KU80c*.

Ces transgènes ont été micro-injectés dans des paramécies afin de tester leurs capacités de complémentarité et de localisation.

Des tests de survie sur une gamme de CPGH (de 0 à 400 pour *FLAG-KU80ca* et de 0 à 200 pour *FLAG-KU80ac*) ont été réalisés avec ces transgènes (**Figure V.11**). En condition

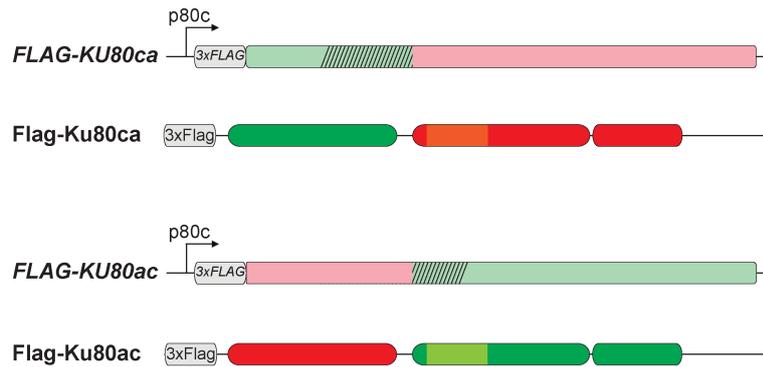
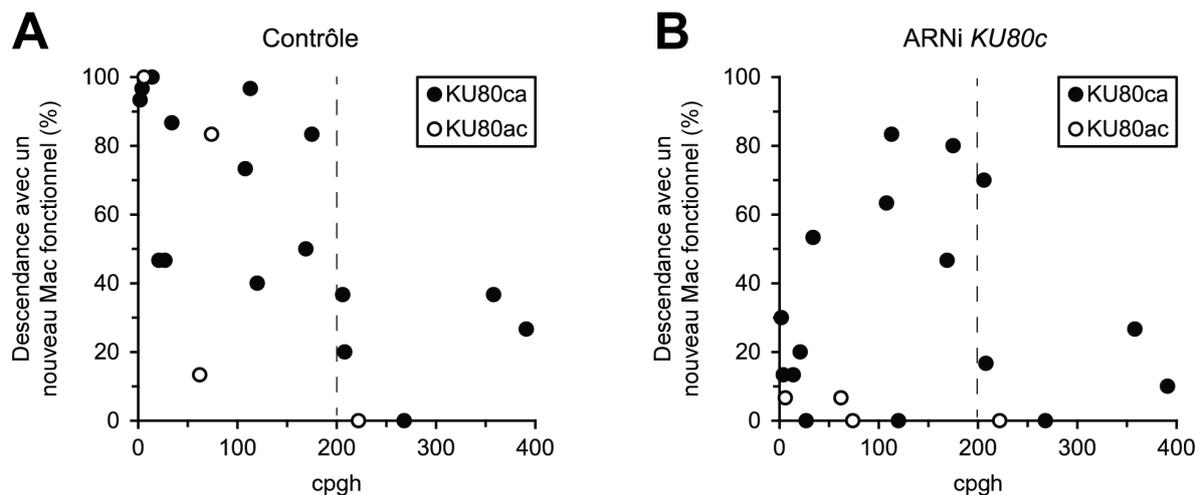


FIGURE V.10 – Transgènes FLAG-KU80ca et FLAG-KU80ac et protéines associées théoriques

Les transgènes sont sous le contrôle du promoteur endogène de *KU80c*. L'étiquette fusionnée est trois Flag successifs. La zone barrée représente les modifications de la séquence du transgène rendant l'ARNm résistant à l'ARNi, sans modifier la séquence de la protéine.

contrôle, la chimère Flag-Ku80ca semble présenter un seuil à 200 CPGH (représenté par la ligne pointillée). Avant ce seuil elle n'est pas ou faiblement toxique (entre 40 et 100% de descendance sexuelle fonctionnelle) et devient progressivement toxique au-delà (entre 0 et 40% de descendance sexuelle fonctionnelle). La chimère Flag-Ku80ac semble elle présenter un profil de toxicité similaire à celui obtenu pour Flag-Ku80a (Figure V.4A). Le nombre de clones injectés pour Flag-Ku80ac est cependant relativement faible (5 clones différents) et de nouveaux test seront nécessaires pour s'assurer de l'effet toxique de cette chimère.

FIGURE V.11 – Complémentations de la déplétion de *Ku80c* par *Flag-Ku80ca* ou *Flag-Ku80ac* en fonction du CPGH

Pourcentage de descendance sexuelle viable en fonction du niveau d'injection dans des cellules complémentées par *FLAG-KU80ca* ou *FLAG-KU80ac*.

V.3.1 Le domaine α/β de Ku80c est responsable de la spécialisation de la protéine pour les RPG

Le domaine α/β de Ku80c est essentiel pour les RPG En ARNi *KU80c*, Flag-Ku80ac ne permet aucune complémentation. Le domaine $\alpha - \beta$ de Ku80c (1-241) est donc nécessaire pour les RPG. De plus, Flag-Ku80ca permet, lui, une survie croissante de 0 à 200 CPGH, avant d'être affecté par le seuil de toxicité à 200 CPGH, déjà observé en condition contrôle. Ku80ca est donc capable de compléter au moins partiellement l'absence du Ku80c endogène.

Couplé à un corps de Ku80a, le domaine $\alpha - \beta$ de Ku80c (1-241) est suffisant pour produire une descendance sexuelle viable et donc pour activer les RPG.

De plus, la complémentation avec la chimère Ku80ca montre que le corps de Ku80a (242-739, tonneau β , piliers-pont et domaines C-terminaux) est fonctionnel pour le NHEJ. En effet, la complémentation implique forcément l'excision des IES et donc la réparation des cassures par la chimère Ku80ca. Ceci devra malgré tout être vérifié par PCR à minima. Enfin, la chimère Flag-Ku80ca présentant une certaine toxicité au-delà de 200 CPGH, il serait intéressant de rechercher un possible effet mineur sur les RPG par séquençage du génome réarrangé.

Le domaine α/β de Ku80c permet la localisation stable de Pgm dans les ébauches Le comportement de ces chimères et de leur effet sur Pgm a ensuite été observé en IF en pré-extraction (**Figure V.12**). Tout comme Flag-Ku80c, Flag-Ku80ca localise spécifiquement dans les ébauches et forme des foci. De la même manière, Flag-Ku80ac présente un comportement identique à Flag-Ku80a et est presque complètement absent des ébauches. La capacité de Ku80c à s'accrocher stablement aux ébauches est donc dépendante de son domaine α/β . De plus, il permet une plus grande stabilisation nucléaire de Pgm qu'avec le domaine α/β de Ku80a.

V.3.2 Le domaine central en tonneau β , piliers et pont est suffisant pour l'interaction avec Pgm

Ce domaine est donc au cœur de la spécialisation de Ku80c dans les RPG, soit en modifiant l'activité de la protéine soit en lui permettant d'interagir avec un facteur des

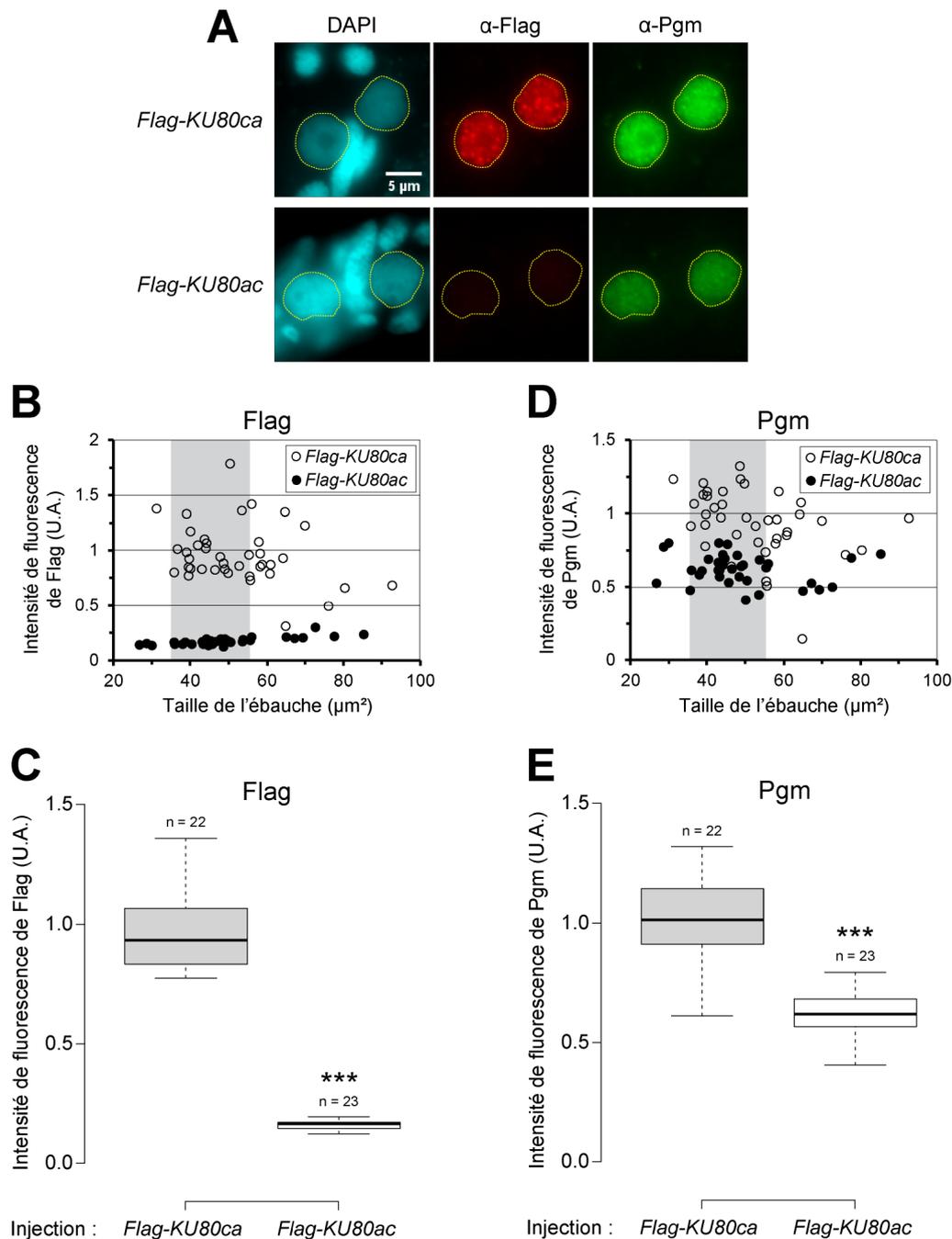


FIGURE V.12 – Localisation des constructions Flag chimériques en pré-extraction et effet sur la stabilité de Pgm

(A) IF de Pgm, Flag-Ku80ca et ac dans des cellules autogames à T5 en déplétion du Ku8oc endogène. Microscopie à épifluorescence. (B) Intensité de fluorescence dans l'ébauche de Flag-Ku80ca ou Flag-Ku80ac en unité arbitraire en fonction de la taille de l'ébauche mesurée en μ m². Un bruit de fond correspondant au signal de la lame en fluorescence verte a été soustrait à l'ensemble des données (zone sans cellule). La zone d'intérêt a été choisie entre 35 et 55 μ m² (zone grise). (C) textitBoxplot de l'intensité de fluorescence du Flag des ébauches comprises entre 35 et 55 μ m². Le n représente le nombre d'ébauches par conditions. (D) Intensité de fluorescence dans l'ébauche de Pgm en unité arbitraire en fonction de la taille de l'ébauche mesurée en μ m². La zone d'intérêt a été choisie entre 35 et 55 μ m² (zone grise). (E) Boxplot de l'intensité de fluorescence du Pgm des ébauches comprises entre 35 et 55 μ m². Le n représente le nombre d'ébauches par conditions. *** pour $p < 0.001$ dans un test statistique de Mann-Whitney-Wilcoxon montrant la différence significative par rapport au contrôle. CPGH clone FLAG-KU80ca : 175. CPGH clone FLAG-KU80ac : 222. Taux de survie de la descendance sexuelle : FLAG-KU80c 80% ; FLAG-KU80a ARNi- 0%.

RPG. Afin d'expliquer le couplage avec Pgm, nous avons cherché à savoir si ce domaine α/β était responsable de l'interaction Ku/Pgm décrite dans MARMIGNON ET AL. (2014).

Pour cela, Ku80c a été découpé en différents sous-domaines par Julien Bischerour (**Figure V.13A**) en vue de leur production en cellules d'insecte sous forme de fusions avec l'épitope HA.

Les capacités d'interaction de ces différents domaines de Ku80c ont été testées par co-immunoprécipitation avec MBP-Pgm, également produit en cellules d'insecte (**Figure V.13B**).

Sur l'ensemble des domaines testés, tous sont capables d'interagir avec Pgm à l'exception justement du domaine α/β . De manière surprenante, l'interaction Ku/Pgm est dépendante du domaine central tonneau β , piliers et pont. La spécialisation du domaine α/β ne réside donc pas dans sa capacité d'interaction directe à Pgm.

V.3.3 Colocalisation des foci Flag-Ku80ca et Pgm

L'étude de la chimère Flag-Ku80ca a révélé qu'elle aussi formait des foci, visibles après une pré-extraction. Concernant Flag-Ku80c, nous n'avons pas pu avoir la preuve directe de la colocalisation des foci avec ceux de Pgm (**Figure IV.14**) suite à divers problèmes techniques (stabilité de l'anticorps, température du laboratoire, ...). Ces problèmes techniques ont pu être résolus lors de l'utilisation de cellules transformées par *FLAG-KU80ca* et ont permis d'obtenir des motifs en foci de Flag-Ku80ca et de Pgm (**Figure V.14**). La majorité des foci Flag-Ku80ca colocalisent avec les foci Pgm. Ce résultat confirme l'intérêt de procéder à des études plus approfondies sur la composition de ces foci et de leur rôle dans les RPG.

V.3.4 Résultats principaux du chapitre

Paramecium tetraurelia ainsi que les autres paramécies du groupe *aurelia* ont systématiquement conservé au moins une version des protéines Ku80a et une version des Ku80c, issues de la WGD intermédiaire. La version Ku80c est la seule capable chez *P. tetraurelia* d'activer les RPG. En absence de Pgm (donc de CDB programmées), Flag-Ku80a est complètement délocalisée du noyau alors que Flag-Ku80c persiste, indiquant une capacité probablement déterminante pour les RPG de s'associer avec une structure nucléaire, en-

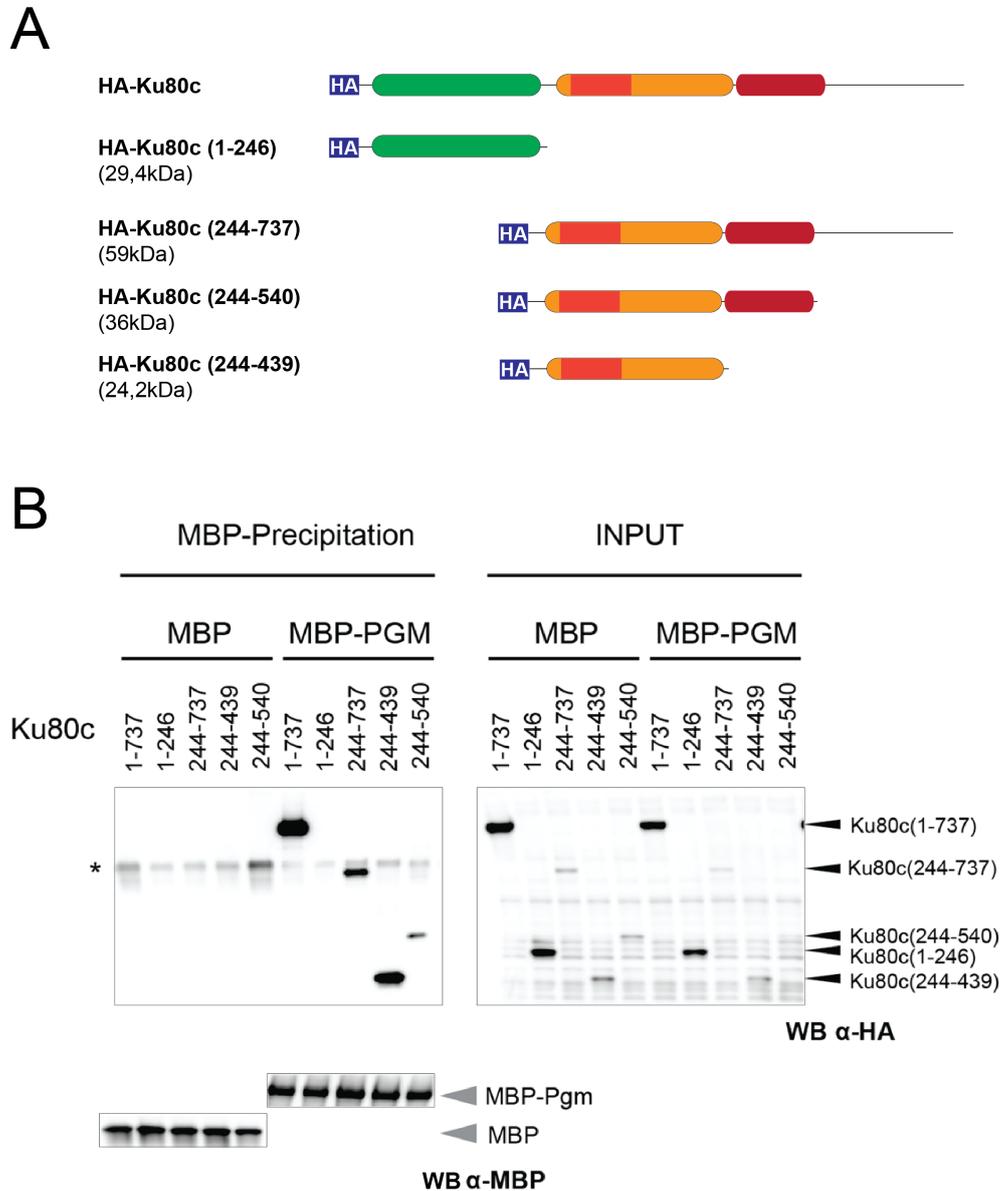


FIGURE V.13 – Co-immunoprécipitation de sous-domaines de Ku80c avec Pgm

(A) Découpages de Ku80c en différents sous-domaines. Les protéines sont fusionnées à une étiquette HA. (B) Co-immunoprécipitation des différents découpages de Ku80c avec MBP-Pgm à partir d'extraits de cellules d'insecte, révélés par WB en utilisant un anticorps anti-HA. L'extrait total pré-immunoprécipitation est présenté dans le panneau de droite (INPUT).

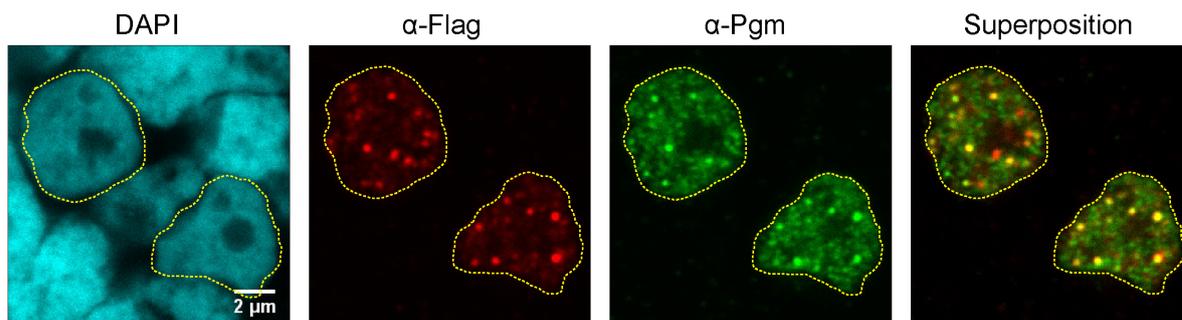


FIGURE V.14 – Colocalisation de Flag-Ku80c avec Pgm en IF avec pré-extraction

IF α -Flag et α -Pgm sur des cellules transformées avec *FLAG-KU80c* en autogamie à T5. Microscopie confocale.

core inconnue. Cette capacité et celle d'activer Pgm sont portées par le domaine α/β de Ku80c.

Troisième partie

Discussion

Chapitre VI

Caractérisation du couplage entre Ku et Pgm

VI.1 L'IMPORTANCE DES INTERRELATIONS DE LOCALISATION

VI.1.1 Réinterprétation des résultats de Marmignon et al. (2014)

La paramécie est le premier exemple de couplage extrême entre la coupure de l'ADN et la réparation des cassures qui en résultent. En absence de Ku, protéine essentielle du NHEJ, Pgm ne présente plus aucune activité catalytique *in vivo*. La caractérisation des effets d'une absence de Ku sur Pgm est donc une étape essentielle vers la compréhension des mécanismes impliqués dans le couplage. L'un de ces effets concerne la localisation nucléaire de Pgm.

La première description de l'effet d'une déplétion de Ku sur la localisation de Pgm a été réalisée dans (MARMIGNON ET AL. 2014) (**Figure VI.1**).

Ces résultats montrant une très forte fluorescence de Pgm-GFP à J3 ou J4 (temps tardifs de l'autogamie) en ARNi *KU80c*, les RPG se produisant à J2 en condition contrôle. L'interprétation avait à l'époque été de conclure que Pgm ne requiert pas la présence de Ku80c pour se localiser dans les ébauches, avec des données montrant au contraire une accumulation de signal.

Les connaissances acquises depuis permettent de réinterpréter ces résultats pour plusieurs raisons :

- Les jours 3 et 4 correspondent à des temps tardifs de l'autogamie. Lorsque les RPG

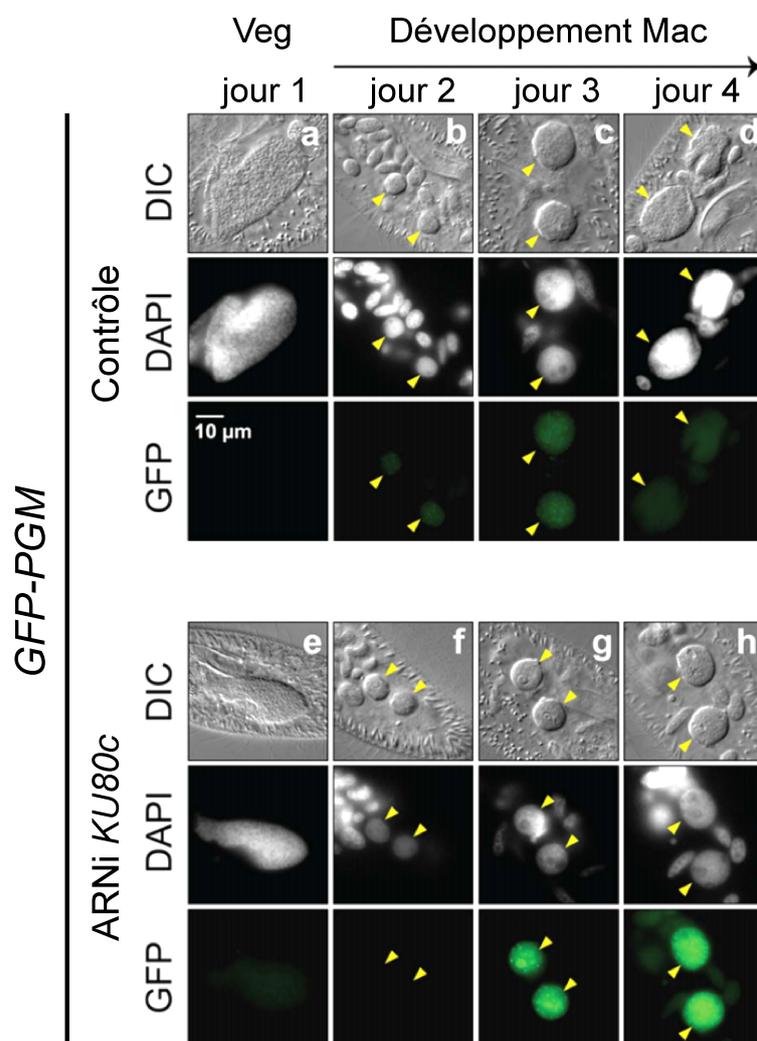


FIGURE VI.1 – Localisation de Pgm-GFP en absence de Ku d'après MARMIGNON ET AL. (2014)

Fluorescence GFP en protocole de pré-extraction de Pgm-GFP en ARNi contrôle ou *KU80c*. Les images à jours 2 correspondent approximativement à un T₅ de l'autogamie et donc au moment des RPG.

sont inhibés par une déplétion de Ku80c, plusieurs centaines de gènes normalement induits de façon transitoire pendant l'autogamie restent surexprimés aux temps tardifs (Evangelia Eleftheriou et Olivier Arnaiz, communication personnelle). Le transgène *PGM-GFP*, sous le contrôle du promoteur endogène de *Pgm*, est lui aussi surexprimé. Des données préliminaires de l'équipe indiquent qu'il s'agit d'une dérégulation transcriptionnelle (Mélanie Bazin, communication personnelle).

- Les premiers résultats présentés dans ce manuscrit ont mis l'accent sur l'importance d'observer la localisation des protéines au moment où les RPG ont lieu, afin de justement éviter la superposition de tels phénomènes de surexpression tardive.

- Lorsque l'on se concentre sur les images à jour 2 de la **figure VI.1**, on devine la présence de signal GFP dans la condition contrôle et une absence totale de signal dans l'ARNi *KU80c*. Cette diminution du signal serait donc cohérente avec les résultats présentés dans ma thèse qui indiquent une destabilisation de la localisation nucléaire de *Pgm* aux stades intermédiaires de l'autogamie, où se produisent les RPG. Les conclusions présentées dans MARMIGNON ET AL. (2014) sont donc à réexaminer.

Une déplétion de Ku80c a donc deux effets principaux sur *Pgm* : 1) destabilisation de *Pgm* dans les ébauches. 2) induction transcriptionnelle tardive de la transcription de *Pgm*.

VI.1.2 La localisation nucléaire stable de *Pgm* dépend fortement de Ku

L'observation poussée des ébauches à T5 dans une gamme de tailles précise nous a donc permis d'analyser de manière poussée les effets d'une déplétion de Ku80c sur la localisation de *Pgm* et inversement.

Il fut un temps envisagé que Ku permette le transport nucléaire de *Pgm* et que cette caractéristique soit la raison de la dépendance de *Pgm* pour Ku. Les résultats de la localisation de *Pgm* en ARNi *KU* sans pré-extraction viennent contredire directement cette hypothèse. Sans pré-extraction et en déplétion de Ku, *Pgm* est toujours majoritairement présent dans les ébauches. Il ne dépend donc pas de Ku pour son entrée dans l'ébauche. *Pgm* est cependant complètement perdu en réalisant une pré-extraction sur des cellules déplétées de Ku. En absence de Ku, *Pgm* est donc capable de rentrer dans l'ébauche, mais, inactif, son état d'accrochage nucléaire est modifié et ne résiste pas à la pré-extraction.

Des expériences de DUBOIS ET AL. (2017) utilisant un mutant catalytique de Pgm montrent qu'un Pgm incapable d'introduire des CDB est toujours capable de s'accrocher stablement à l'ébauche. Il y a donc bien un accrochage pré-coupe de Pgm à la chromatine, accrochage qui a été montré dans ce manuscrit comme étant Ku-dépendant.

La conséquence de cette perte d'accrochage est la rétention de l'ensemble des IES avec une inactivation complète de Pgm en absence de Ku.

L'activité catalytique de Pgm n'a pas encore pu être observée *in vitro* malgré les nombreuses tentatives de Julien Bischerour. Sur l'ensemble des conditions testées, l'ajout de Ku au mélange réactionnel n'a pas encore été réalisé. Dans l'hypothèse où la présence de Ku permettrait l'introduction de CDB *in vitro*, cette expérience permettrait une démonstration directe du rôle activateur de Ku.

VI.1.3 DNA-PKcs permet-il le recyclage de Pgm ?

Les données de localisation de Pgm en conditions de pré-extraction dans une ARNi DNA-PKcs montrent une augmentation du signal et donc l'augmentation de l'accrochage nucléaire de Pgm. Le phénotype moléculaire sur l'excision des IES dans une ARNi DNA-PKcs est ambigu et montre la présence de coupure non réparées ainsi que de la rétention d'IES. Ces phénotypes pourraient-ils être dus à un piégeage de Pgm sur les sites de coupure ? Afin d'assurer l'introduction des très nombreuses CDB, une protéine Pgm pourrait avoir à être recyclée après chaque cassure. Dans l'hypothèse où DNA-PKcs permettrait ce recyclage, une absence de DNA-PKcs pourrait donc provoquer la séquestration des Pgm sur des bornes d'IES déjà clivées et ainsi empêcher l'introduction des cassures sur une partie des IES.

Il est montré dans GRATIAS AND BÉTERMIER (2001) que toutes les IES ne sont pas excisées au même moment, certaines dans les premiers instants des RPG, d'autres plus tardivement. L'équipe, et particulièrement Vinciane Régnier, cherche à déterminer la cinétique d'excision de chaque IES, en utilisant une approche de tri d'ébauches par cytométrie en flux en fonction de leur ploïdie, et donc de l'avancement des RPG. Le contenu en ADN des différentes classes d'ébauches sera ensuite séquencé et permettra ainsi de déterminer pour chaque IES le moment de son excision.

Dans l'hypothèse où l'absence de DNA-PKcs pourrait empêcher le recyclage de Pgm,

les premières IES excisées en conditions standard devraient également être celles où les CDB sont détectées en ARNi *DNA-PKcs*. De la même manière, les IES "tardives" devraient être maintenues dans ces conditions. Un séquençage global du génome d'ébauche sera donc nécessaire pour avancer sur le rôle de *DNA-PKcs* dans les RPG chez la paramécie.

VI.1.4 Flag-Ku80c ne dépend majoritairement pas de Pgm pour se localiser stablement dans l'ébauche

En absence de Pgm, et donc de CDB, il était attendu que Ku ne soit plus accroché fortement aux ébauches et soit donc fortement délocalisé lors de l'utilisation du protocole de pré-extraction. Or, la déplétion de Pgm n'a perturbé que 38% du signal Flag-Ku80c par rapport à un contrôle où plusieurs milliers de CDB sont présentes. Ku80c est donc capable d'interagir étroitement avec un élément associé l'ébauche indépendamment de Pgm et des CDB. Ku interagissant avec Pgm, il pourrait être une plateforme permettant la stabilisation de Pgm sur la chromatine. La détermination de cet élément associé à l'ébauche sera donc une étape clé de la compréhension du mécanisme de couplage entre Ku et Pgm. On pourrait imaginer qu'il s'agisse d'une protéine, d'une séquence ADN ou bien encore d'ARN, guidant Ku sur la chromatine. De la même manière que pour BRITTON ET AL. (2013), un traitement RNase permettrait de vérifier si la localisation stable de Ku en déplétion de Pgm est dépendante des ARN. Dans l'hypothèse de l'interaction avec une protéine, il serait possible d'utiliser une technique de détection d'interaction *in vivo* actuellement développée au laboratoire. Cette technique est la Bio-ID et utilise une étiquette biotine-ligase biotinylant les protéines dans un environnement proche autour de la protéine de fusion. L'ensemble des protéines ayant été biotinylées sont ensuite détectées par spectrométrie de masse. Cette technique risquerait cependant de donner un grand nombre d'interactants et de nombreuses expériences seraient ensuite nécessaires pour déterminer le partenaire responsable de la stabilisation nucléaire de Ku.

Si cette existence de partenaire est prouvée, protéique, ADN ou ARN, il peut être proposé que ce partenaire puisse reconnaître les IES et qu'il guide ainsi Ku et Pgm sur le site de coupure et de réparation.

Les PgmL ont été montrés comme des partenaires essentiels du complexe d'excision (BISCHEROUR ET AL. 2018). L'étude du couplage devra donc également étudier plus en détail les relations entre les PgmL et Ku.

VI.1.5 Les foci Ku et Pgm

La nature des foci Ku et Pgm reste mystérieuse. Au nombre (très approximatif) de 100 par ébauche, il est peu probable qu'ils représentent des événements uniques, vu le nombre massif d'événements de réparation lors des RPG. Il s'agirait donc plutôt de lieux où sont réalisés un grand nombre de coupures et d'événements de réparation. Des tentatives de suivi de ces foci sur cellules vivantes ont été réalisées pour déterminer si ces structures étaient stables dans le temps ou transitoires. Pour cela, la construction GFP-Ku80c et une technique d'accrochage de cellules vivantes à une lame recouverte de mucus de moule, permettant la fixation des cellules, ont été utilisées. Malheureusement, cette technique de fixation n'a pour l'instant permis de garder vivantes que quelques minutes des paramécies autogames sous le microscope. La durée de vie des foci et leur dynamique dans les ébauches n'a donc pour l'instant pas pu être évaluée.

Chapitre VII

Spécialisation de Ku80c dans les RPG

VII.1 KU80A EST INCAPABLE DE REMPLACER KU80C

L'ensemble des paramécies du complexe des *aurelia* ont subi au moins trois WGD. La WGD intermédiaire a provoqué l'apparition de deux groupes de Ku80, Ku80a/b et Ku80c/d, dont chaque espèce a conservé au moins un membre de chaque. Cela semblait indiquer un rôle important et différent de ces deux groupes.

Pour tester si c'était effectivement le cas, j'ai cherché à savoir si Ku80a pouvait remplacer Ku80c lors des RPG. En absence du Ku80c endogène, une fusion Flag-Ku80a ne permet pas la survie de la descendance sexuelle et les IES testées indiquent une absence d'excision. De plus, malgré une localisation nucléaire de GFP-Ku80a en protocole sans pré-extraction, la totalité du signal est perdue en utilisant le protocole avec pré-extraction. En absence d'extrémités d'ADN cassées, Ku80a n'est pas stablement associé aux ébauches, contrairement à Ku80c, qui persiste lors d'une ARNi PGM. Ce résultat suggère que l'hypothétique partenaire de Ku80c lui permettant d'interagir fortement avec l'ébauche n'interagit pas avec Ku80a.

Ku80a n'est donc pas capable de remplacer Ku80c, indiquant donc clairement qu'il y a eu une spécialisation de cette version pour les RPG depuis la duplication de Ku80 lors de la WGD intermédiaire. Un séquençage du génome d'ébauches devra cependant être réalisé pour confirmer l'effet global de l'inefficacité de Ku80a dans les RPG.

Cette spécialisation est une aubaine pour l'étude des mécanismes associés aux RPG. J'ai proposé précédemment de rechercher les partenaires de Ku80c *in vivo* pour tenter

de trouver ses partenaires et donc les protéines permettant sa spécialisation. Une telle méthode produit généralement de grandes quantités de résultats et un long travail de criblage est ensuite nécessaire pour trouver LA protéine permettant la spécialisation.

Ku80a est une protéine très proche de Ku80c et pouvant se localiser dans les ébauches, mais non stablement. L'étude de ses partenaires permettrait ainsi, par soustraction à celles découvertes pour Ku80c, d'éliminer le bruit de fond de nombreuses protéines telles que celles du NHEJ. Le fameux partenaire théorique de Ku80c devrait donc ne pas être trouvé parmi les interactants de Ku80a.

VII.1.1 Le domaine α/β de Ku80c est au cœur de la spécialisation

L'utilisation des protéines chimériques avait pour but de rechercher un possible sous-domaine suffisant pour activer les RPG. Un tel sous-domaine activateur a pu être identifié. La chimère possédant le domaine α/β de Ku80c et le corps de Ku80a (Ku80ca) s'est montrée active dans les RPG et très proche phénotypiquement de Ku80c, la seule différence observée étant une baisse du niveau de complémentation au-delà de 200 CPGH. La chimère Ku80ac ne se localise pas stablement dans l'ébauche. Cette capacité est donc apparemment portée uniquement par le domaine α/β de Ku80c, qui est donc probablement le sous-domaine définissant la spécialisation de Ku80c dans les RPG. Le domaine α/β de Ku80c est donc très probablement responsable de l'interaction avec le partenaire hypothétique. Les études d'interactomes pourront donc se limiter à l'usage des domaines α/β de Ku80c pour la recherche du partenaire et de Ku80a pour l'élimination du bruit de fond.

VII.2 DÉTAIL DE L'INTERACTION ENTRE KU80C ET PGM

L'interaction entre Ku et Pgm est proposée comme essentielle pour les RPG. Il aurait été facile d'imaginer que le domaine α/β de Ku80c soit responsable de l'interaction avec Pgm. Les co-immunoprécipitations sur les différents découpages ont cependant montré le rôle essentiel du domaine tonneau β , piliers-pont dans l'interaction avec Pgm. elles ont également montré l'absence d'interaction entre le domaine α/β et Pgm.

Des données préliminaires du laboratoire semblent indiquer que la protéine Ku80a pleine taille interagi avec Pgm. Ces résultats privilégieraient donc une hypothèse où l'in-

teraction Ku/Pgm est essentielle et est assurée par le tonneau β , piliers-pont de Ku δ oc et probablement de Ku δ oaa également (dans un contexte Ku δ oca). Cette hypothèse, du rôle fonctionnel de l'interaction Ku/Pgm, doit malgré tout être démontrée.

VII.2.1 L'interaction Ku/Pgm est-elle essentielle lors des RPG ?

La question du rôle de l'interaction Ku/Pgm est au cœur de l'étude du couplage entre coupure et réparation. Une expérience de complémentation fonctionnelle utilisant une protéine Ku δ oc mutante ayant perdu sa capacité d'interaction avec Pgm pourrait tester l'importance de cette interaction. D'après le modèle décrit précédemment, Ku pourrait toujours se fixer stablement à la chromatine mais Pgm perdrait cette capacité. Ainsi, il ne serait pas capable d'introduire les CDB et les RPG seraient bloqués.

La première stratégie que j'ai déployé lors de mon arrivé au sein de l'équipe fut de développer un test de double-hybride inverse chez la levure. Un tel test aurait permis de réaliser une mutagenèse aléatoire de Ku δ oc et de ne sélectionner que les mutants de perte d'interaction avec Pgm. Les mutations clés identifiées, une version mutante de Ku δ oc n'interagissant plus avec Pgm aurait été testée lors des RPG. Malheureusement, les essais de développement du double-hybride ne m'ont pour l'instant pas permis de reproduire l'interaction Ku/Pgm chez la levure, empêchant donc la recherche mutants de perte d'interaction.

La deuxième stratégie repose sur la collaboration de l'équipe avec le groupe d'Orsolya Barbaras à l'EMBL, spécialisé dans la structure des transposases. Ils ont pour projet de former un cristal du complexe formé entre Ku et Pgm. Si une telle structure était obtenue, il serait possible de réaliser une mutagenèse dirigée abolissant l'interaction.

VII.3 MODÈLE DU COUPLAGE ENTRE COUPURE ET RÉPARATION DE L'ADN LORS DES RPG CHEZ *P. tetraurelia*

L'ensemble des données obtenues lors de mes travaux de thèse m'ont permis de proposer un modèle du couplage lors des RPG (**Figure VII.1**).

Une particule, protéique ou ARN, interagissant avec le α/β de Ku δ oc est fortement accrochée à la chromatine, potentiellement au niveau d'une IES. Ku rentre dans le noyau

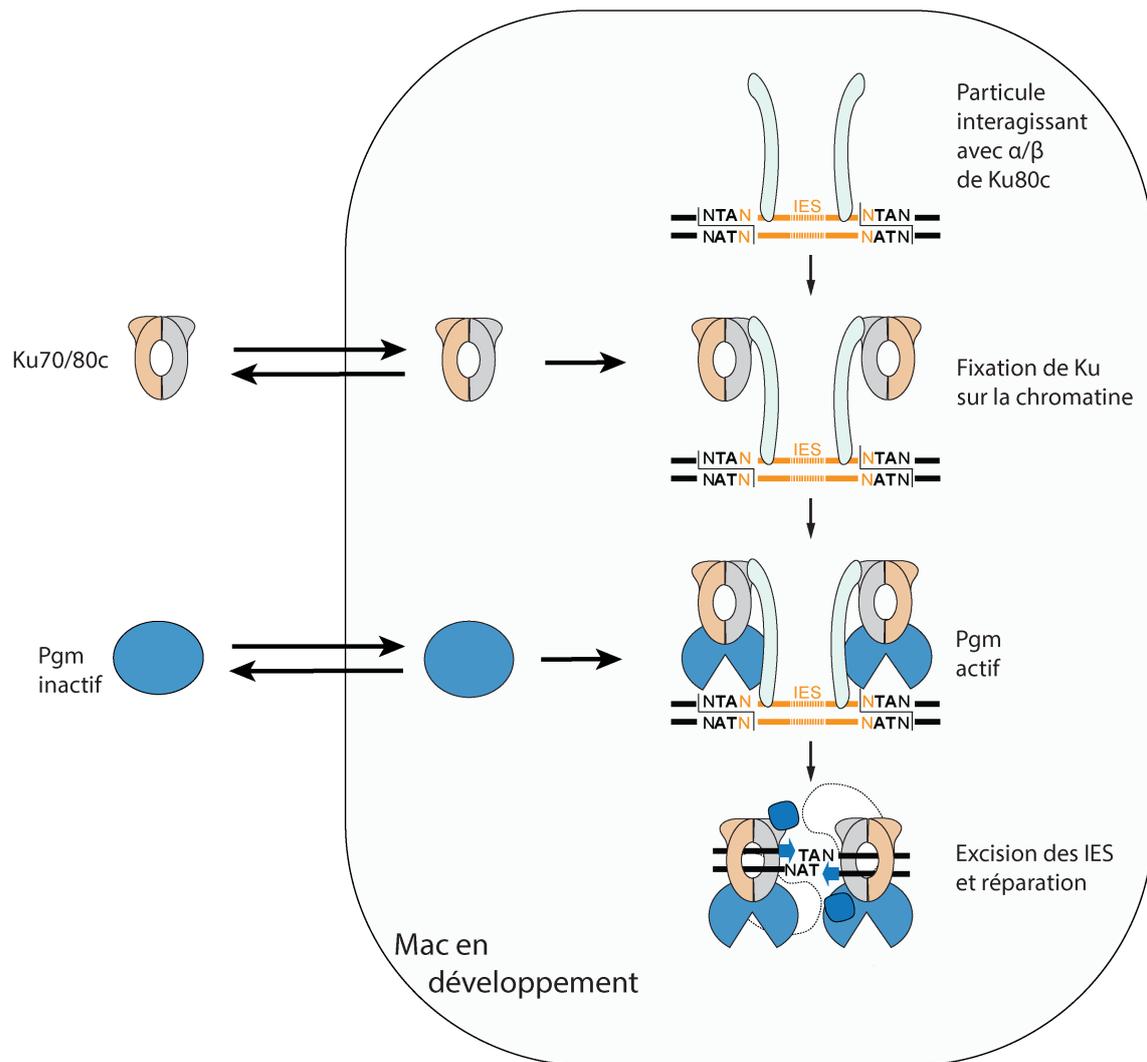


FIGURE VII.1 – Modèle du couplage entre coupure et réparation de l'ADN lors des RPG chez *P. tetraurelia*

et est stabilisé sur la chromatine par la particule. Pgm, inactif, rentre dans le noyau et est stabilisé par son interaction avec Ku au niveau du tonneau β , piliers-pont. Cette interaction provoque son activation. A proximité de l'IES, il introduit les cassures (avec l'aide des Pgm1, non représentés sur le schéma), Ku reconnaît les cassures et déclenche la réparation par le NHEJ classique.

Chapitre VIII

L'importance du couplage

Chez la paramécie, 47% des gènes sont interrompus par au moins une IES. La nécessité d'obtenir des phases ouvertes de lecture fonctionnelles a poussé la paramécie à développer un mécanisme de réarrangement extrêmement efficace. Pgm requiert la présence de Ku70/Ku80c pour introduire les CDB. Cette stricte dépendance permet d'assurer la présence du NHEJ avant même que la cassure ait eu lieu. Quelles seraient les conséquences d'une abolition de ce couplage extrême? Des expériences ont été menées en ce sens par Julien Bischerour. Un transgène ne contenant que le domaine α/β de Ku80c a été injecté dans des paramécies. Ce domaine étant responsable de la spécialisation de Ku80c, il était possible de prévoir que sa présence pourrait activer Pgm. Malheureusement, les quelques données disponibles pour le moment semblent indiquer que ce sous-domaine seul n'est pas capable d'être stablement localisé dans les ébauches. La capacité de localisation nucléaire de ce domaine est donc certainement conditionné par l'ensemble de l'hétérodimère Ku.

Une autre approche pour tester le découplage a été d'utiliser un mutant Ku70-6E. Ce mutant, décrit dans BRITTON ET AL. (2013) avec la protéine humaine porte six mutations dans ses domaines piliers/pont, modifiant la charge positive de l'anneau en charge négative, rendant donc Ku déficient dans sa capacité à interagir avec les extrémités d'ADN mais restant globalement stable structurellement. La transposition de ces mutations sur la version paramécie de Ku70 a permis d'obtenir un découplage entre coupure et réparation. Pgm introduit les cassures mais le NHEJ est soit inactif, en ARNi *KU70*, soit partiellement inactif, en condition de compétition entre le Ku70 endogène et le mutant Ku70-6E. Dans

le premier cas, les CDB sont accumulées et non réparées, semblables à celles d'un phénotype d'ARNi *LIG4*. Dans le second cas, des événements de réparation ont lieu mais pas suffisamment pour assurer le bon déroulement des RPG. Il sera intéressant d'étudier plus en détail les événements de réparation. Le découplage a-t-il par exemple favorisé les réparations illégitimes? Cette étude permettra de mettre en exergue le rôle du couplage entre coupure et réparation de l'ADN chez la paramécie.

Il existe aujourd'hui très peu d'exemples de tels couplages. Chez *Tetrahymena*, Ku n'est pas essentiel à l'excision des IES par Tpb2, une autre transposase domestiquée de type PiggyBac. Cependant, ses IES sont presque exclusivement inter-géniques et la pression de sélection pour un mécanisme de couplage extrême est donc moindre. Les très rares IES intra-géniques de *Tetrahymena* dépendent de Tpb1 et Tpb6 (CHENG ET AL. 2016), d'incroyables protéines de fusion entre une protéine Ku80 et une transposase PiggyBac (Figure VIII.1).

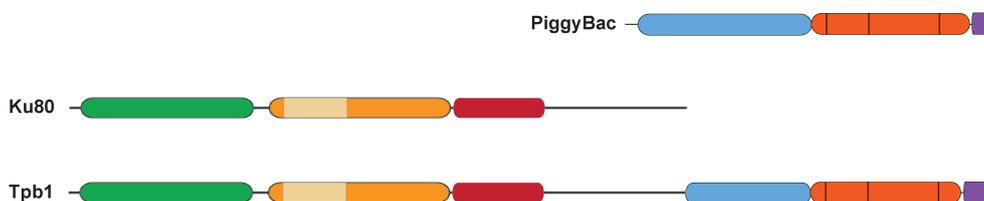


FIGURE VIII.1 – Domaines de Tpb1

Tetrahymena a donc adopté une stratégie différente de la paramécie pour répondre à la nécessité de couplage lors des RPG. Les événements de RPG imposent une forte pression aux mécanismes de réparation. Dans le cas de la recombinaison V(D)J, Ku n'est pas non plus essentiel à l'introduction des CDB par RAG₁/RAG₂ mais il semblerait qu'il existe malgré tout un couplage, cette fois-ci assuré par le mécanisme de coupure. RAG₂ a été montré comme important dans la stabilisation de la synapse entre les deux extrémités d'ADN (LESCALE ET AL. 2016) préparant un contexte favorable pour une réparation efficace par le NHEJ.

Les réarrangements programmés ne sont pas uniquement présents chez les ciliés ou

dans le contexte de synthèse des anticorps. Différents organismes procèdent à des éliminations programmées d'ADN (WANG AND DAVIS (2014) pour revue) mais les mécanismes moléculaires sont encore mal connus. L'étude de ces mécanismes pourrait permettre de découvrir de nouvelles stratégies de couplage ou bien d'observer de potentielles convergences évolutives, avec des stratégies se rapprochant de celles observées chez la paramécie, *Tetrahymena* ou lors de la recombinaison V(D)J.

Quatrième partie

Matériels et Méthodes

CULTURE PARAMÉCIE

La souche sauvage 51 de *P. tetraurelia* (GRATIAS AND BÉTERMIER 2003) ou sa version mutante 51 nd7-1 (DUBOIS ET AL. 2017) ont été cultivées dans un milieu standard fait d'infusion de poudre d'herbe de blé (*wheatgrass powder*) inoculé avec *Klebsiella pneumoniae* et supplémenté avec du β -sitostérol (0.8 $\mu\text{g}/\text{mL}$). L'autogamie a été provoquée comme décrit dans DUBOIS ET AL. (2017).

ARN INTERFÉRENCE

Les ARNi durant l'autogamie ont été réalisées en utilisant la procédure de *feeding* comme décrite dans DUBOIS ET AL. (2017). En résumé, les paramécies sont mises en culture avec des bactéries *E. coli* HT115 exprimant un ARNdb à partir d'un plasmide L4440 contenant un fragment de la séquence du gène visé par l'interférence. En ingérant les bactéries, les paramécies prennent en charge cet ARNdb et déclenchent l'ARNi.

Les séquences utilisées pour les différentes ARNi sont présentées en annexe. La condition contrôle pour l'ensemble des expériences d'ARNi présentées consiste en un *feeding* avec des bactéries possédant le vecteur L4440 vide.

La présence d'un nouveau Mac fonctionnel a été testée quatre jours après le début de la carence alimentaire, comme décrit dans DUBOIS ET AL. (2017).

MICRO-INJECTION D'ADN

Les transgènes utilisés dans les différentes expériences de complémentation ont été intégrés dans un plasmide pUC18. Chaque plasmide a été linéarisé avec BsaI et micro-injecté avec un plasmide de complémentation ND7 dans le Mac de paramécies végétatives 51 nd7-1, comme décrit dans DUBOIS ET AL. (2017).

CALCUL DU CPGH PAR QPCR

Les cellules injectées sont repiquées dans un milieu standard permettant leur croissance végétative. Une partie des cellules filles est utilisée pour l'ensemble des expériences de complémentation et l'autre partie est utilisée pour une extraction d'ADN total

afin d'évaluer les CPGH par qPCR. L'extraction d'ADN est réalisée sur environ 200 cellules végétatives en utilisant le kit d'extraction *NucleoSpin Tissue* (Macherey Nagel). La qPCR est réalisée sur un LightCycler 480 à l'aide du kit SYBR Green I de Roche. Pour chaque clone injecté testé et pour un clone non injecté, une qPCR est réalisée sur un locus génomique lambda afin de normaliser la quantité d'ADN entre chaque clone testé. Dans les travaux présentés ici, le locus utilisé a été PTET.51.1.G1060180 (disponible sur paramecium.i2bc.paris-saclay.fr) avec les amorces OMB063 et OMB064 (disponibles en annexes). Une autre qPCR est effectuée avec les mêmes clones mais cette fois ci autour du locus du gène injecté. Dans un clone non injecté, les copies endogènes du gène étudié seront les seules amplifiées alors que dans un clone injecté, les copies endogènes ainsi que les transgènes seront amplifiés. Le calcul du CPGH est ensuite effectué en utilisant les données obtenues dans les deux qPCR selon la formule présentée dans la **Figure VIII.2A**.

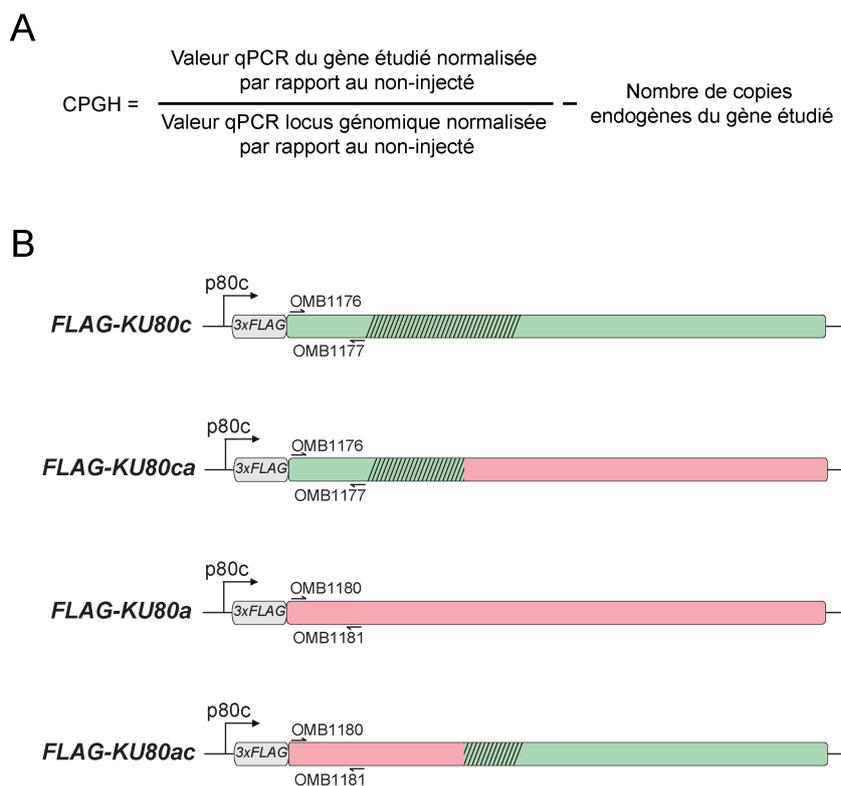


FIGURE VIII.2 – Formule de calcul du CPGH et emplacement des amorces pour l'amplification de *KU8oc* et *KU8oa*

Les amorces utilisées pour amplifier *KU8oc* ou *KU8oca* sont OMB1176 et OMB1177

(**Figure VIII.2B**) (séquence en annexe). OMB₁₁₈₀ et OMB₁₁₈₁ permettent l'amplification de *KU80a* et *KU80ac* (également en annexe).

IMMUNOFLUORESCENCE

Sans pré-extraction Les cellules ont été fixées dans un tampon PHEM (60 mM PIPES, 25 mM Hepes, 10 mM EGTA, 2 mM MgCl₂ pH 6.9) + 2% formaldéhyde pendant 10 minutes puis perméabilisées par du PHEM + 1% Triton pendant 15 minutes. Elles ont ensuite été lavées trois fois 10 minutes dans du TBST (10 mM Tris pH 7.4, 0.15 M NaCl, 0.1% Tween20) + 3% BSA.

L'incubation des anticorps a été réalisée dans le tampon TBST + BSA. Les anticorps ainsi que leur concentration sont décrits dans BISCHEROUR ET AL. (2018). Les noyaux ont été marqués au DAPI (Sigma) avec 0.5 µg/ml.

Avec pré-extraction Les cellules ont été perméabilisées dans un tampon PHEM + 1% Triton pendant 4 minutes puis fixées par un tampon PHEM + 2% formaldéhyde pendant 15 minutes. Les étapes suivantes de lavage et d'immunomarquage sont communes avec le protocole sans pré-extraction.

ANALYSE DES IMAGES DE MICROSCOPIE

Les différentes conditions d'IF ont été observées avec un microscope Olympus à épifluorescence. L'intensité de fluorescence a été mesurée sur la zone de diamètre maximal des Mac en développement en utilisant le logiciel ImageJ (<https://imagej.nih.gov/ij/index.html>). Afin de faciliter l'analyse, un outil d'automatisation a été développé par Romain Le Bars (MicMacEpiv21.ijm prochainement disponible).

Cette macro nécessite de placer l'ensemble des images d'une même condition dans un dossier. Les différents canaux doivent être enregistrés dans des fichiers .tif indépendants (**Figure VIII.3**), la macro ne prenant pas en charge les images "multicanal".

Une fois le dossier sélectionné, la macro superpose les canaux vert et rouge (si un des deux canaux n'est pas présent, la macro se contentera de celui à sa disposition). Elle génère ensuite un masque en fonction de l'intensité de la fluorescence, permettant de

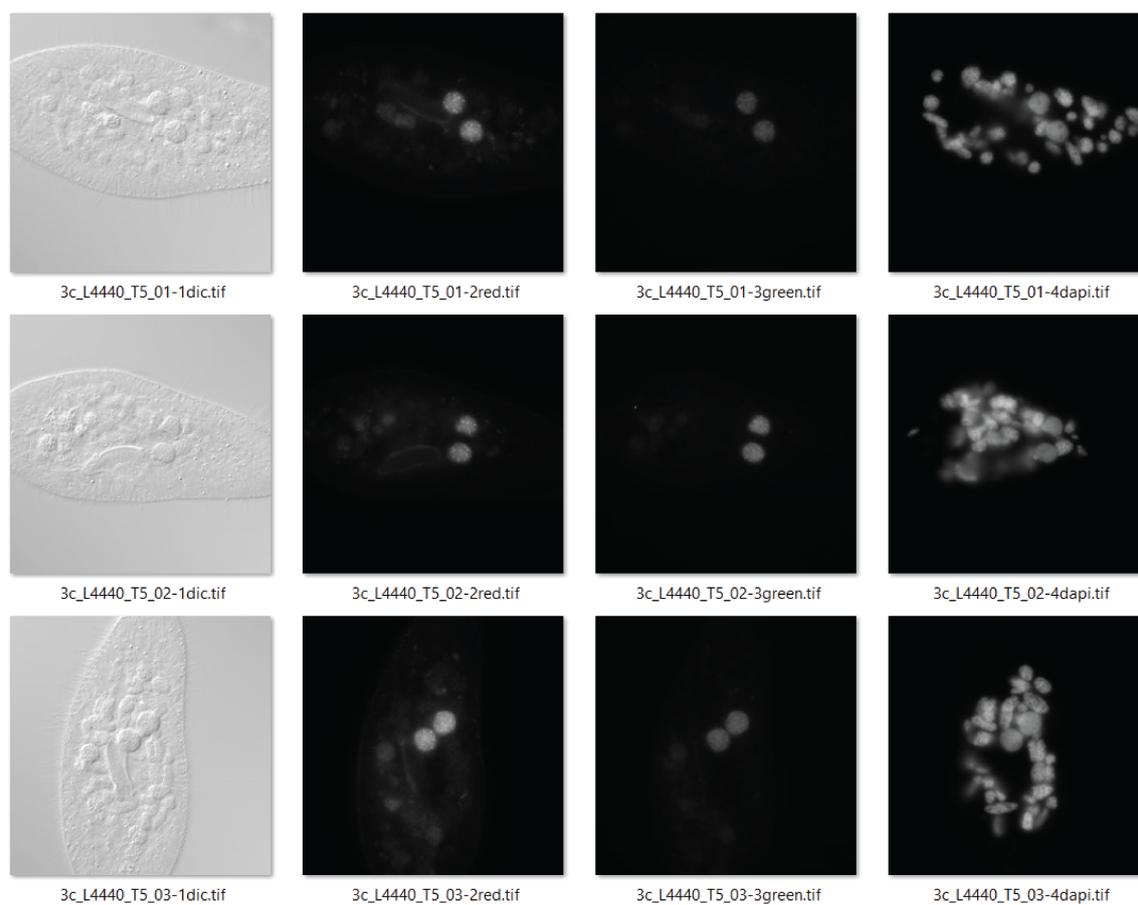


FIGURE VIII.3 – Exemple d'organisation de fichiers avant analyse par la macro *MicMacEpiov21*

créer des sélections ou ROI (*Region Of Interest*) autour des ébauches. Ces ROI sont ensuite placées sur l'image DAPI, afin que l'utilisateur puisse vérifier si les ROI générées automatiquement correspondent bien aux ébauches. Dans le cas contraire, l'utilisateur peut procéder à une détection semi-automatique ou complètement manuelle. Une fois que les ROI correspondent de manière certaine aux ébauches, la macro procède automatiquement à la mesure de la taille des ébauches (en surface) ainsi qu'à l'intensité moyenne de fluorescence dans les canaux vert et rouge (si présent). Une fois ces analyses effectuées, la macro réalise un nouveau masque à partir des canaux vert et rouge et sélectionne cette fois-ci la zone de l'image sans cellule, afin de réaliser la mesure de fluorescence de la lame et donc d'obtenir une valeur de bruit de fond qui pourra être soustraite aux données. Pour finir, la macro enregistre les données dans un tableau .csv et réalise une carte .tif indiquant les ROI créées pour l'analyse de l'image (**Figure VIII.4**).

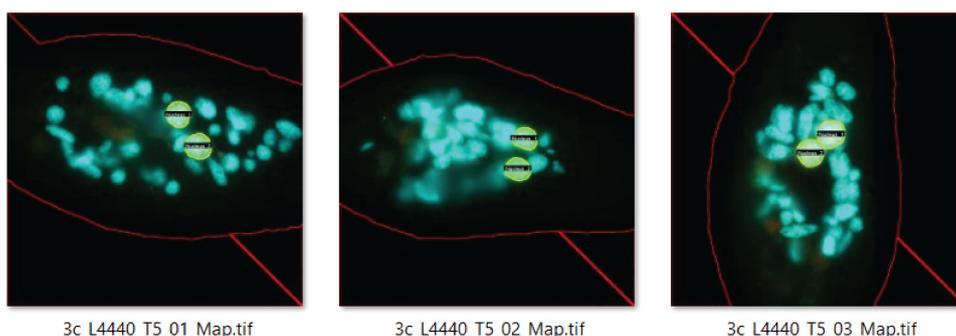


FIGURE VIII.4 – Exemple des cartes des ROI après analyse par la macro *MicMacEpi21*

Les ébauches sont entourées en jaune et numérotées, les lignes rouges délimitent les zones extérieures à la cellule qui sont utilisées pour la mesure du bruit de fond.

Les données sont ensuite représentées sur un graphique, avec l'intensité de fluorescence en fonction de la taille de l'ébauche (**Figure VIII.5**). Cette représentation permet d'observer une courbe en cloche asymétrique, avec une augmentation soudaine de la fluorescence autour de $40 \mu\text{m}^2$, un pic vers $50 \mu\text{m}^2$ puis une descente progressive jusqu'à environ $75 \mu\text{m}^2$. Cette courbe permet de définir la zone d'intérêt (zone grise) pour la comparaison entre les différentes conditions d'ARNi.

Pour faciliter la lecture et la comparaison des données, l'ensemble des valeurs comprises dans la zone d'intérêt est représenté sous forme de boîte à moustache, ou *boxplot*, à l'aide de l'outil [http : //shiny.chemgrid.org/boxplotr/](http://shiny.chemgrid.org/boxplotr/). La significativité de la différence

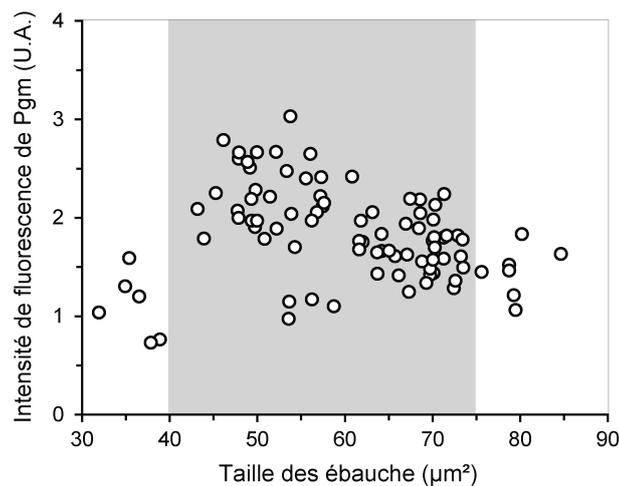


FIGURE VIII.5 – Exemple de graphique de la fluorescence Pgm en fonction de la taille de l'ébauche en autogamie contrôle après analyse par la macro *MicMacEpi021*

entre les moyennes est calculée par un test statistique de Wilcoxon-Mann-Whitney en utilisant l'outil <https://ccb-compute2.cs.uni-saarland.de/wtest/>

WESTERN BLOT

L'extraction protéique et l'analyse par WB ont été réalisées sur environ 100 000 cellules autogames, collectées par centrifugation et transférées dans de l'azote liquide. Les cellules concentrées et congelées ont été directement lysées après addition d'un volume équivalent de 10% SDS bouillant contenant 1x de *Protease Inhibitor Cocktail Set 1* (Merk Chemicals) et incubées 3 minutes à 100°C. Les SDS-PAGE et WB ont été réalisés comme décrit dans DUBOIS ET AL. (2017). Le signal a été visualisé avec le système Chemidoc Touch Imaging (Bio-Rad) et les analyses densitométriques ont été réalisées avec le logiciel Image Lab Software (Bio-Rad).

PCR AUTOUR DES IES

Les PCR autour des IES permettant de vérifier leur rétention ou non ont été réalisées comme décrit dans MARMIGNON ET AL. (2014).

Cinquième partie

Annexes

Pt-Ku80c	1	MSGKEATLILLDVGSAMYEQYQCGSKLSRLELAVDCIGLMIQQKIFNYKNHEVGLILF
Pt-Ku80a	1	MAGKEATLVLLDVGASMYEPYKQAQGGKIRLELAVDCIGMMIQQKIFNYKNHEVGLVLF
Pt-Ku80b	1	MAGKEATLVLLDVGASMYEPYKQAQGGKIRLELAVDCIGMMIQQKIFNYKNHEVGLVLF
----- α/β domain -----		
Pt-Ku80c	61	GTEEAEDGKTLFYIQDLSIPDLDFFRNLSLDPNHVVGQVGGDIFDALDKAVHALDDHAKT
Pt-Ku80a	61	GTEEAEDGNTFYIQTLSSPDLEFYRNLTLPNHDIKPKIIGGDIFFDALDKSVSTLDEYVKT
Pt-Ku80b	61	GTEEAEDGNTFYIQVMSPPDLEFYRNLTLPNHDPVKIKGGDIFDALDKAVSTLDEYVKA
Pt-Ku80c	121	KKMEKKIFILTAGCGQTDYSEKQITIKLIKMIKVDVKINFIALDFMNDYNGDMDDPBKPE
Pt-Ku80a	121	KKMDKKIFVLTAGFGQTYNEKKIAKLIKMIKVDVKINFIALDFMNEYDAELDDPSKPE
Pt-Ku80b	121	KKMEKKIFVLTAGFGQTDYNEKKIKLIKMIKVDVKINFIALDFMNEYDVELDDPBKPE
Pt-Ku80c	181	EFALNRRMLTASVQCQEQSINSRYVFLMVQELRNMRIFPANVAFELYSQFHTRSLQAR
Pt-Ku80a	181	NQETLNDRMLNAVYESQEQSINSRLVYYMVQELRSHMRIFPANVAFELYSQFHTRKMQAR
Pt-Ku80b	181	NQATLNDRMLNAVYQHREQSINSRLVYQMVQELRNHMRIFPANVAFELYSQFHTRKMQAR
Pt-Ku80c	241	ASFRGDFQINDEISVQVLIYKRCFEERLPLLRKHSTLGEFQTDITNKNHVRNDLIHYNPED
Pt-Ku80a	241	ASFRGDFQINDETSISVLYYKRCFEERLPLLRKHSTLGEFSSEPTRNVVRNDLIHYNPED
Pt-Ku80b	241	ASFRGDFQINDETSIQVLIYKRCFEERLPLSLKKSATGEYSSEPTRNIVRNDLIHYNPED
----- β barrel -----		
Pt-Ku80c	301	PNMTPTEKDNIIKGYQGRNLVPVDQIMEDKMKYQCPRFQFQLLGFVDRSHIPRYYYTSTIV
Pt-Ku80a	301	PNMTPIERENIKGYLYGRSLIPVDSIMEDKMKYQCRSFQLLGFVDRKSQIPRHYFMSV
Pt-Ku80b	301	PNMNPTEKDNIIKGYQGRNLIPVDSIMEKMKYQCRSFQLLGFVDRSIPRHYFFNV
Pt-Ku80c	361	DMVIAVENORQOKALALVIALIATKVAARFVGREKTAPKLIIMLLPHKSKNSQCFWMI
Pt-Ku80a	361	DMVVAIDCEKARKKSLSLI IAL IATKVAARFVGREKSSPKMVVLLPHKSKSYQCFWMI
Pt-Ku80b	361	DMVIAIDCEKARKSLSALI IAL IATKVAARFVGRQKSSPKMIIMLLPHKSKSYQCFWMI
Pt-Ku80c	421	SLPTSEDIRHFQFAALKRSTPEQQAVASAVLIDKMDLEALPNESGEPPEELKMKYIANPTR
Pt-Ku80a	421	ALPTSEDIRHFQFAALKRSTPEQQAVASLIDKMDLEALPNESGEPPEELKMKYIANPTR
Pt-Ku80b	421	SLPTSEDIRHFQFSTLRKSTPEQQSAVALIDKMNLEALPNESGEPPEELKMKYIANPTR
Pt-Ku80c	481	QYFQQVVMHKAITRSDVIPPISPLILEYLHPEKRVYDYAKAALQKVKAAAFKFKNEIKKQ
Pt-Ku80a	481	QYFQQVVMHKAITRTDVIPPISPLILEYLHPEQRVYNYAQDAIQRVKNAPFKFKVNEIKKP
Pt-Ku80b	481	QYFQQVVMHKAITRTDVIPPISPLILDYLHPERRVYNYAQDAIQRVKNAPFKFKVNDIKKP
Pt-Ku80c	541	QDKKVFWKQLFEDQST...EQIQQQVEDEVVEINQEEEEEMVNMFAKQKLGFNDDIIQEIIGS
Pt-Ku80a	541	QDKKVFWKQLFDEQTTQQQQQQQIEEEVVEINREEEEEEMVNMFAKQKLGFNDDIIQEIIGS
Pt-Ku80b	541	QDKKVFWKQLFDEQTA...QQQVPEQIEEEVVEINREEEEEEMVNMFAKQKLGFNDDIIQEIIGS
----- C-ter (DNA-PKcs binding) -----		
Pt-Ku80c	599	VDPTISDFKKMITEKRVDLVDIALQQIQKVI IQFVDQS LKGSFYPKALECLKEMRKACIIE
Pt-Ku80a	601	VDPISDFKKMITEKRVDLVDLSDALQQIQKVI IQGLVDQSVKGSFFPKALECLKEMRKACISE
Pt-Ku80b	600	VDPISDFKKMITEKRVDLVDLSDALQQIQKVINALVDQSVKGSFFPKALECLKEMRKACISE
Pt-Ku80c	659	DEAPVFNKYLHVLKEKYSQLVFVAQIVQQGITLISNIENQKSHVSVDEAQEFLNKEDLSH
Pt-Ku80a	661	DEAPVFNKFLVFLKDKYNQSLFVAQIVQQGITLISNIENQKSGVTAEAAQDFLNKEDNKH
Pt-Ku80b	660	DEAPVFNKFLVFLKDKYNQQLFVAQIVQQGITLISDIENHKSSVTAEAAQDFLNKEDNKH
Pt-Ku80c	719	KQLVLDQLQHEEEDLLAETID
Pt-Ku80a	721	QQMVDQLQHEEEDLLADIE
Pt-Ku80b	720	QQMVDQLQHEEEDLLADIE

FIGURE 6 – Alignement des protéines Ku80 de *P. tetraurelia* d'après MARMIGNON ET AL. (2014)

SÉQUENCES DES AMORCES

OMB063 AGACAAGTAGGGAATCCACTTCTAGTAATC

OMB064 TAATGTATTGATAAGGCTTGCTCTACAGCC

OMB1176 AGCTCCTGATGGCAAGACAT

OMB1177 AGTCTGTCCACAACCAGCTG

OMB1180 GGACGCTGAGGATGGGAATA

OMB1181 TGTCTAACCAAATCCTGCTGTG

SÉQUENCES UTILISÉES POUR LES ARNi

Pgm Publiée dans BAUDRY ET AL. (2009)

Ku70 Publiée dans MARMIGNON ET AL. (2014)

Ku80c Publiée dans MARMIGNON ET AL. (2014)

DNA-PKcs Construction de Sophie Malinsky

tctagattatggatgcatttaaagatactctatctacttaattctataaaatcaagactacaatgacat
 ctctcaagaataatgaaatcctttgccaccattgattttgtctctttgtacaataatctagcaaaactcctatcca
 tctgatgcatactattaaatatacactttgtgattatcaatcttaactgaatatctcaatgacaaataaagac
 tatcatcattgaactatttaatagcagtcacccagctaggcatttctctctattgtatgtctattgagttg
 ctattaccagctatctattactagaaattctgtttgtttgccataaaatctcaaaagataacattggatcatcaat
 ttatgcttattaactcatccatctcgatctttacactcaactgtactttacagattctcattctctctcaattt
 ccattgtggtttctgtaactactcaactgataatgaagctgtatagataatcttgacaatctctcctagaaata
 atcctactattactcttgacttcagtatacagtggttttctcaagtttgctcaactaaaaattaattaccat
 tctttattatgaaattatcccaataagatatctcttcttctgaataaatgagcataaaatagtttctcttag
 tctatgtttcatcaacacataaatgtatgtaataaaggcactgcttaataatctctactatttcagataattcac
 ttaagagagaatttcttttgatttatattcaaaatgtaatctctcaaatgaaattgctagctttgtacaactc
 acaatcaatttcttggtatttattatcagcttagaactcagattttaaataatgtaattgtttttatttattcagttg
 acatctctgtaggctattccaatagatcaaaaataagacctcattgtaattgaaacaacatgttctccaatt
 atcatccatgttacattcaactattacaaaagataaggtgtgaattctgaaactcaacaataagttggaatta
 ctcaaaagaggaatcaataatttattacccaaagctattctcacattatgagggtgctagtatctaattgctcgtctt
 ttattataacatgagccatttcaattattggtaagattcttggttaattataagaattcttaataaacaagtcaaa
 atttaaactctgtcttaacagtggaataaacttcaatatattaactatcccaaaagctt

Lig4 Publiée dans KAPUSTA ET AL. (2011)

XRCC4 Publiée dans KAPUSTA ET AL. (2011)

SÉQUENCES DES TRANSGÈNES UTILISÉS

Flag-KU8oc TTAAATAGATAAATAATACAGGCAATTTTTTAATATACAAATTTCAA
 TAATGAATTTGTAGATATATATAAAAAATAAGTGCAAAGATGAATTCTAGAGAC
 TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGAC
 GATGATAAGGGATCCATGTCTGGAAAAGAAGCTACTCTAATATTATTGGATGTT
 GGAGCCTCAATGTATGGTCAATATTAATAGGGAGGAAGTAAAAAGTTGAGCAGA
 TTGGAATTGGCAGTAGATTGTTTAGGATTGATGATTCAATAAAAGATTTTTAAT
 TATAAGAATCACGAAGTCGGATTGATTTTATTTGGAACAGAAGAAGCTCCTGAT
 GGCAAGACATTGTACATCCAAGATTTGTCAATCCCAGATTTAGATTTCTTTAGG
 AATATTAGTGATTTACCTAACCATGATGTTGGATAATAAGTGGGTGGAGATATA
 TTTGATGCATTGGACAAGGCAGTACATGCTTTAGATGATCATGCAAAGACAAAA
 AAAATGGAAAAGAAGATCTTTATATTAACAGCTGGTTGTGGACAGACTGATTAC
 AGTGAAGAAGTAGATAACCAAATTAATCAAGATGATTGAGAAAAGTGGATGTTAAG
 ATAAATTCATAGCTCTTGATTTTATGAATGATTACAATGGAGATATGGATGAT
 CCTGAAAACCTGAAGAATTCGAAGCACTTAATAATAGAATGCTTACAGCTTCA
 TATCAATGTTAAGAACAATCTATCAATTCTCGTTATGTTTTTTTGATGGTCCAA
 GAATTGAGGAACAACATGAGAATATTCCCAGCAAATGTTGCCTTTGAATTGTAT
 TCCCAATTCATACCAGATCTTTGTAAGCTAGAGCATCATTACAGGGGGGATTC
 CAAATTAATGATGAGATCTCTGTACAAGTTCTAATATATAAACGATGTTTTGAA
 GAGAGATTGCCTACTCTTAGAAAACATTCTACCCTTGGAGAATTTCAAACCTGAT
 ACCAATAAGAATCATGTGAGGAATGATCTCATCTATTACAATCCAGAAGATCCC
 AATATGACTCCAATAGAAAAGGACAACATAATAAGAGGATACTAATATGGTAGG
 AATCTGGTTCAGTGGATTAATCATGGAGGACAAAATGAAATACTAGTGTCTCT
 CGTTAATTTTAATTATTAGGCTTTGTAGATCGATCCCATATACCTAGATATTAT
 TATACCTCAACTGTTGACATGGTCATTGCTGTAGAAAATCAGAAGTAATAGAAA
 GCTCTGGCTGCTCTAGTCATTGCACTCATTGCAACTAGGAAAGTAGCTCTTGCC
 AGATTTGTAGGCAGAGAGAAAACAGCACCTAAATTGATTATGTTATTACCCCAT
 AAATCCAAGAACTCTTAGTGCTTCTGGATGATATCCTTGCCTACAACCTGAAGAC
 ATCAGGCATTTCAATTTGCTGCTTTGAAGAGATCTACTCCTCCATAATAGATG
 GCAGTCTCTGCTATGATTGACTGTATGGATCTAGAAAAGATGCCACGGAAGAT
 GGACAGTTTGAAGAACTTCTTAAAATGAAGTATGTTGCCAATCCTACGAGATAA
 TACTTCCAATAAGTGGTCATGCACAAAGCCATCACTAGATCTGATGTATTGCCT
 CCTATCTCCCCTCTCATCTTGGAGTATTTACATCCTGAAAAGAGGGTTTACGAT
 TATGCAAAGAGGCACTTCAAAAAGTCAAGGCTGCTTTCAAATTCAAAATCAAC
 GAAATCAAAAACAAGGGGATAAGAAAGTGTTTTGGAATAACTATTTGAAGAT
 CAATCTACAGAATAAATTTAATAGTAAGTGGAGGATGAAGTAGTTGAAATCAAT
 TAAGAAGAAGAAGAGATGGTTAATATGTTTGCCAAATAAAAATTAGGATTCAT
 GACGATATAGTTAAGGAAATCGGCACTGTTGATCCCCTTCAGATTCAGGAAG
 ATGATTACTGAAAAGAGAGTGCCTTAGTTGACACTGCCTTACAATAAATTCAA
 AAAGTCATTATTTAATTTGTGGATCAATCTCTCAAAGGCAGTTTCTATCCAAA
 GCACTTGAGTGTCTCAAAGAAATGAGGAAAGCCTGTATCACAGAAGATGAAGCC
 CCAGTATTCAATAAATACTTGCATGTTCTTAAAGAGAAGTACAGTCAACTTGTC
 TTCTGGGCTCAGATTGTCCAATAAGGAATCACTTTAATCAGCAACATTGAGAAT
 TAGAAATCTCATGTTTCTGTTGATGAAGCACAGAAGTAATTCATTATTATAAT
 ATTAGTTCTAAATAAAGAGGATATTAGTCATAAATAATTAGTTGATCAATTAT
 AACACGAAGAAGAGGATCTATTGGCTGAAATTGATTGATTATTATTAGTCTAT
 ATCTAACAATTTTATTCCAATTTTATATTAATAA

Flag-KU8oc résistant à l'ARNi

TTAAATAGATAAATAATACAGGCAATTTTTTAATATACAAATTTCAA
TAATGAATTTGTAGATATATATAAAAAATAAGTGCAAAGATGAATTCTAGAGAC
TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGAC
GATGATAAGGGATCCATGTCTGGAAAAGAAGCTACTCTAATATTATTGGATGTT
GGAGCCTCAATGTATGGTCAATATTAATAGGGAGGAAGTAAAAAGTTGAGCAGA
TTGGAATTGGCAGTAGATTGTTTAGGATTGATGATTCAATAAAAGATTTTTAAT
TATAAGAATCACGAAGTCGGATTGATTTTTATTTGGAACAGAAGAAGCTCCTGAT
GGCAAGACATTGTACATCCAAGATTTGTCAATCCCAGATTTAGATTTCTTTAGG
AATATTAGTGATTTACCTAACCATGATGTTGGATAATAAGTGGGTGGAGATATA
TTTGATGCATTGGACAAGGCAGTACATGCTTTAGATGATCATGCAAAGACAAAA
AAAATGGAAAAGAAGATCTTTATATTAACAGCTGGTTGTGGACAGACTGATTAC
AGTGAAAAGTAGATAACCAAATTAATCAAGATGATTGAGAAAAGTGGATGTTAAG
ATAAATTCATAGCTCTTGATTTTTATGAATGATTACAATGGAGATATGGATGAT
CCTGAAAAACCTGAAGAATTCGAAGCTCTTAACAATAGGATGCTAACAGCCTCA
TATCAGTGTTAGGAACAATCCATTAATTCACGTTACGTTTTCTTGATGGTTCAA
GAGTTGAGGAATAACATGAGGATATTTCCAGCTAATGTTGCTTTTGAGTTATAT
TCCCAGTTTCATACTAGATCACTATAGGCTAGAGCTTCATTTAGGGGGGACTTC
CAGATTAACGATGAAATCTCTGTGCAAGTGCTAATTTATAAGCGATGTTTCGAG
GAAAGATTGCCAACTCTAAGAAAGCATTCAACCCTTGGCGAATTCCAAACAGAT
ACTAATAAAAACCATGTGAGAAATGATCTTATCTACTATAATCCAGAGGATCCT
AATATGACACCAATCGAAAAGACAACATCATAAGGGGATATTAATATGGCAGG
AATCTCGTACCAGTTGATTAATAATGGAGGATAAAATGAAGTACTAGTGTCTCT
CGTTAATTTTAATTATTAGGCTTTGTAGATCGATCCCATATACCTAGATATTAT
TATACCTCAACTGTTGACATGGTCATTGCTGTAGAAAATCAGAAGTAATAGAAA
GCTCTGGCTGCTCTAGTCATTGCACTCATTGCAACTAGGAAAGTAGCTCTTGCC
AGATTTGTAGGCAGAGAGAAAACAGCACCTAAATTGATTATGTTATTACCCCAT
AAATCCAAGAACTCTTAGTGCTTCTGGATGATATCCTTGCCTACAACCTGAAGAC
ATCAGGCATTTCAATTTGCTGCTTTGAAGAGATCTACTCCTCCATAATAGATG
GCAGTCTCTGCTATGATTGACTGTATGGATCTAGAAAAGATGCCACCGGAAGAT
GGACAGTTTGAAGAACTTCTTAAAATGAAGTATGTTGCCAATCCTACGAGATAA
TACTTCCAATAAGTGGTCATGCACAAAGCCATCACTAGATCTGATGTATTGCCT
CCTATCTCCCCTCTCATCTTGGAGTATTTACATCCTGAAAAGAGGGTTTACGAT
TATGCAAAGAGGCACTTCAAAAAGTCAAGGCTGCTTTCAAATTCAAAATCAAC
GAAATCAAAAACAAGGGGATAAGAAAGTGTTTTGGAAATAACTATTTGAAGAT
CAATCTACAGAATAAATTTAATAGTAAGTGGAGGATGAAGTAGTTGAAATCAAT
TAAGAAGAAGAAGAGATGGTTAATATGTTTGCCAAATAAAAATTAGGATTCAT
GACGATATAGTTAAGGAAATCGGCACTGTTGATCCCCTTCAGATTTTCAGGAAG
ATGATTACTGAAAAGAGAGTGCAGCTTAGTTGACACTGCCTTACAATAAATTCAA
AAAGTCATTATTTAATTTGTGGATCAATCTCTCAAAGGCAGTTTCTATCCAAA
GCACTTGAGTGTCTCAAAGAAATGAGGAAAGCCTGTATCACAGAAGATGAAGCC
CCAGTATTCAATAAATACTTGCATGTTCTTAAAGAGAAGTACAGTCAACTTGTC
TTCTGGGCTCAGATTGTCCAATAAGGAATCACTTTAATCAGCAACATTGAGAAT
TAGAAATCTCATGTTTCTGTTGATGAAGCACAAGAAGTAATTCATTATTATAAT
ATTAGTTCCTAAATAAAGAGGATATTAGTCATAAATAATTAGTTGATCAATTAT
AACACGAAGAAGAGGATCTATTGGCTGAAATTGATTGATTATTATTAGTCTAT
ATCTAACAATTTTATTCCAATTTTTATATTAATAA

Flag-KU8oa TTAAATAGATAAATAATACAGGCAATTTTTTAATATACAAATTTCAA
 TAATGAATTTGTAGATATATATAAAAAATAAGTGCAAAGATGAATTTCTAGAGAC
 TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGAC
 GATGATAAGGGATCCATGGCAGGAAAAGAAGCAACTTTAGTTTTATTAGATGTA
 GGTGCCTCAATGTATGAACCATACAAGTAGGCTCAAGGAAAGAAAATCACTAGA
 TTAGAATTAGCTGTTGATTGCATAGGGATGATGATTTAATAAAAAGATCTTTAAT
 TACAAGAATCATGAAGTTGGATTAGTTTTGTTTGGAACAGAGGACGCTGAGGAT
 GGAATACTTTTTACATCTAAACATTGTCGTCTCCAGATCTAGAGTTTTATAGA
 AATCTGACAGAATTACCTAATCATGATATACCTAAAATAATTGGTGGAGATATT
 TTTGATGCTTTGGACAAATCAGTTAGCACTTTAGATGAGTATGTGAAGACCAAG
 AAGATGGATAAGAAGATATTTGTGCTCACAGCAGGATTTGGTTAGACAGAATAT
 AATGAGAAGAAGATTGCAAAGCTTATTAATAATGATTGAAAAAGTAGATGTGAAG
 ATCAATTTCATAGCATTGGATTTTCATGAATGAATATGATGCTGAATTGGATGAT
 CCTTCCAAACCTGAGAATCAAGAACTCTAAATGATCGAATGCTTAATGCTGTT
 TACGAAAGTTAAGAGCAATCAATCAATTCACGTTTAGTTTTATTACATGGTTCAA
 GAACTAAGGAGTCATATGAGAATATTCCCAGCTAATATCGCATTGGAATTGTAT
 TCACAATTTACACCAAACAATGCAAGCAAGAGCCTCATTAGAGGAGATTTT
 CAAATAAATGATGAGACATCCATCTCTGTATTGGTTTACAAAAGATGCACAGAG
 GAGAAATTACCAAGTTTAAAAAGCATTCTGCTACTGGAGAATTCAGCAGTGAA
 CCCACTAGAAATGTAGTTAGAAATGACACCATTCAATCAATCCAGAAGACCCA
 AACATGACTCCAATTGAAAGAGAGAATATCATTAAGGGTTACTTATATGGCAGA
 AGCCTAATTCCTGTGGATAGTATAATGGAAGATAAAATGAAGTATTAATGTGTT
 AGGTCATTTAATTATTGGGATTTGTTGACAAGTCATAAATTCCTAGACATTAT
 TTTATGTCTAGTGTGGACATGGTTGTAGCTATTGATTGTGAAAAAGCAAAGAAA
 TCTTTATCTTCTTTGATCATAGCCTTAATTGCCACAAAGAAAATTGCTATTGCA
 AGATTTGTTGGCAGAGAAAAGAGTTCTCCAAAATGGTGGTGTATTGCCTCAT
 AAATCAAAGAGCTATTAGTGTTTTTGGATGATTGCCTTACCCACTTCTGAAGAC
 ATCAGGCATTTCCAATTTGCTGCTTTAAGAAAGTCCACTCCTCATCAATAAATA
 GCTGTAGCCTCCTTGATTGATAAAATGGATCTAGAAGCACTTCCAAACGAATCT
 GGAGAACCAGAAGAATTACTTAAAATGAAATACATTGCAAATCCAACCTAGACAA
 TATTTCTAATAAGTAGTCATGCATAAAGCTATCACAAGAACTGATGTCATTCT
 CCAATCTCTCTTTGATTCTAGAATATTTGCATCCCGAATAACGAGTATACAAT
 TATGCTTAGGATGCTATATAGAGAGTGAAGAATGCTTTTAAATTCAAAGTTAAT
 GAAATCAAGAAACCCTAAGATAAAAAAGTATTTGGAAACAATTATTCGATGAA
 CAACTACATAATAACAACAAGCTTAACAATAAATAGAGGAAGAAGTGGTTGAA
 ATCAATAGAGAGGAGGAAGAGATGGTCAATATGTTTGCCAAATAAAAATTGGGA
 TTCAATGATGATATCATTTAAGAAATTGGTTCTGTGGATCCAATCTCAGATTT
 AAGAAAATGATCACTGAAAAGAGAGTCGATTTGGTTGATTCTGCCCTTCAACAA
 ATTCAGAAGGTTATAATTGGATTAGTCGATTAATCAGTAAAAGGCAGTTTCTTC
 CCCAAGGCCTTGGAATGCTTAAAAGAAATGAGAAGAGCTTGTATTTCAGAAGAT
 GAAGCTCCAGTTTTCAACAAGTTCCTCTTTGTAAGGATAAGTACAATCAA
 TCTATATTTTGGGCCCAAATTGTACAATAAGGCATCACTTTGATAAGCAACATA
 GAGAATTAATAATCAGGTGTCACTGCTGAGGAAGCATAAGATGTATATATATTT
 AATTAATTATTAGTTCTGAACAAAGAAGATAACAACATTAGTAAATGGTTGA
 TCAATTATAACATGAGGAAGAAGATTTATTAGCTGATATTGAATGATTTATTAT
 TAGTCTATATCTAACAATTTTATTCCAATTTTATATTAATAA

GFP-KU8oc Publié dans MARMIGNON ET AL. (2014)

GFP-KU80a TTAAATAGATAAATAATACAGGCAATTTTTTAATATACAAATTTCAA
TAATGAATTTGTAGATATATATAAAAAATAAGTGCAAAGATGAATTTCTAGAGGA
GAAGAACTTTTCACTGGTGTGTTCCAATTTCTTGTGAACTTGATGGTGATGTT
AATGGACATAAATTTTCTGTCTCTGGTGAGGGTGAAGGTGATGCAACTTATGGA
AAATTAACCTTAAAATTTATTTGCACTACTGGAAAATTACCTGTTCCATGGCCA
ACACTTGTCACTACTTTAACTTATGGAGTCCAATGTTTTTCAAGGTACCCTGAC
CACATGAAACAACATGACTTTTTTCAAATCTGCCATGCCAGAAGGATATGTCCAA
GAAAGAACTATATTCTTCAAAGATGATGGAACTACAAGACAAGAGCTGAAGTC
AAATTTGAAGGAGATACCCTTGTCAATAGAATTGAGCTTAAAGGAATTGATTTT
AAAGAAGATGGAAACATTTTAGGCCATAAATTGGAATACAACACTATAACTCACAT
AATGTATACATCATGGCAGACAAACAAAAAATGGAATCAAAGTCAACTTCAAA
ATTAGACACAACATTGAAGATGGATCAGTTCAATTAGCAGACCATTATCAACAA
AATACTCCTATTGGAGATGGACCAGTACTTTTACCAGACAACCATTACTTATCA
ACACAATCTACCTTATCAAAAGATCCAAATGAAAAGAGAGATCACATGGTGTTA
TTAGAGTTTGTAACTGCTGCTGGAATTACACATGGCATGGATGAATTATACAAA
GCTAGCTCAGGAGGAGGATCAGGAGGATCAGGGGGAGGATCCATGGCAGGAAAA
GAAGCAACTTTAGTTTTATTAGATGTAGGTGCCTCAATGTATGAACCATACAAG
TAGGCTCAAGGAAAGAAAATCACTAGATTAGAATTAGCTGTTGATTGCATAGGG
ATGATGATTTAATAAAAGATCTTTAATTACAAGAATCATGAAGTTGGATTAGTT
TTGTTTGGAACAGAGGACGCTGAGGATGGGAATACTTTTTACATCTAAACATTG
TCGCTCCAGATCTAGAGTTTTATAGAAATCTGACAGAATTACCTAATCATGAT
ATACCTAAAATAATTGGTGGAGATATTTTTGATGCTTTGGACAAATCAGTTAGC
ACTTTAGATGAGTATGTGAAGACCAAGAAGATGGATAAGAAGATATTTGTGCTC
ACAGCAGGATTTGGTTAGACAGAATATAATGAGAAGAAGATTGCAAAGCTTATT
AAAATGATTGAAAAAGTAGATGTGAAGATCAATTCATAGCATTGGATTTTCATG
AATGAATATGATGCTGAATTGGATGATCCTTCCAAACCTGAGAATCAAGAACT
CTAAATGATCGAATGCTTAATGCTGTTTACGAAAGTTAAGAGCAATCAATCAAT
TCACGTTTAGTTTATTACATGGTTCAAGAACTAAGGAGTCATATGAGAATATTC
CCAGCTAATATCGCATTGGAATTGTATTCACAATTCACACCAAACAAATGCAA
GCAAGAGCCTCATTTAGAGGAGATTTTCAAATAAATGATGAGACATCCATCTCT
GTATTGGTTTACAAAAGATGCACAGAGGAGAAATTACCAAGTTTAAAAAAGCAT
TCTGCTACTGGAGAATTCAGCAGTGAACCCACTAGAAATGTAGTTAGAAATGAC
ACCATTCATTACAATCCAGAAGACCCAAACATGACTCCAATTGAAAGAGAGAAT
ATCATTAAAGGGTTACTTATATGGCAGAAGCCTAATTCCTGTGGATAGTATAATG
GAAGATAAAATGAAGTATTAATGTGTTAGGTCATTTTAATTATTGGGATTTGTT
GACAAGTCATAAATTCCTAGACATTATTTTATGTCTAGTGTGGACATGGTTGTA
GCTATTGATTGTGAAAAAGCAAAGAAATCTTATCTTCTTTGATCATAGCCTTA
ATTGCCACAAAGAAAATTGCTATTGCAAGATTTGTTGGCAGAGAAAAGAGTTCT
CCAAAATGGTGGTGTATTGCCTCATAAATCAAAGAGCTATTAGTGTTTTTGG
ATGATTGCCTTACCCACTTCTGAAGACATCAGGCATTTCCAATTTGCTGCTTTA
AGAAAGTCCACTCCTCATCAATAAATAGCTGTAGCCTCCTTGATTGATAAAATG
GATCTAGAAGCACTTCCAAACGAATCTGGAGAACCAGAAGAATTACTTAAAATG
AAATACATTGCAAATCCAACTAGACAATATTTCTAATAAGTAGTCATGCATAAA
GCTATCACAAGAACTGATGTCATTCCTCCAATCTCTCCTTTGATTCTAGAATAT
TTGCATCCCGAATAACGAGTATACAATTATGCTTAGGATGCTATATAGAGAGTG
AAGAATGCTTTTAAATTCAAAGTTAATGAAATCAAGAAACCCTAAGATAAAAAA
GTATTTTGGAAACAATTATTCGATGAACAACTACATAATAACAACAAGCTTAA
CAATAAATAGAGGAAGAAGTGGTTGAAATCAATAGAGAGGAGGAAGAGATGGTC
AATATGTTTGCCAAATAAAAAATTGGGATTCATGATGATATCATTTAAGAAATT
GGTTCTGTGGATCCAATCTCAGATTTCAAGAAAATGATCACTGAAAAGAGAGTC

GATTTGGTTGATTCTGCCCTTCAACAAATTCAGAAGGTTATAATTGGATTAGTC
GATTAATCAGTAAAAGGCAGTTTCTTCCCAAGGCCTTGGAAATGCTTAAAAGAA
ATGAGAAGAGCTTGTATTTTCAAGAAGATGAAGCTCCAGTTTTCAACAAGTTCCTC
TTTGTACTTAAGGATAAGTACAATCAATCTATATTTTGGGCCCAAATTGTACAA
TAAGGCATCACTTTGATAAGCAACATAGAGAATTAATAATCAGGTGTCAGTGTGCT
GAGGAAGCATAAGATGTATATATATTTAATTAATTATTAGTTCCTGAACAAAGA
AGATAACAAACATTAGTAAATGGTTGATCAATTATAACATGAGGAAGAAGATTT
ATTAGCTGATATTGAATGATTTATTATTAGTCTATATCTAACAATTTTATTCCA
ATTTTTATATTAATAA

Flag-KU8oca TTAAATAGATAAATAATACAGGCAATTTTTTAATATACAAATTTCAA
TAATGAATTTGTAGATATATATAAAAAATAAGTGCAAAGATGAATTCTAGAGAC
TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGAC
GATGATAAGGGATCCATGTCTGGAAAAGAAGCTACTCTAATATTATTGGATGTT
GGAGCCTCAATGTATGGTCAATATTAATAGGGAGGAAGTAAAAAGTTGAGCAGA
TTGGAATTGGCAGTAGATTGTTTAGGATTGATGATTCAATAAAAGATTTTTAAT
TATAAGAATCACGAAGTCGGATTGATTTTATTTGGAACAGAAGAAGCTCCTGAT
GGCAAGACATTGTACATCCAAGATTTGTCAATCCCAGATTTAGATTTCTTTAGG
AATATTAGTGATTTACCTAACCATGATGTTGGATAATAAGTGGGTGGAGATATA
TTTGATGCATTGGACAAGGCAGTACATGCTTTAGATGATCATGCAAAGACAAAA
AAAATGGAAAAGAAGATCTTTATATTAACAGCTGGTTGTGGACAGACTGATTAC
AGTGAAAAGTAGATAACCAAATTAATCAAGATGATTGAGAAAAGTGGATGTTAAG
ATAAATTCATAGCTCTTGATTTTATGAATGATTACAATGGAGATATGGATGAT
CCTGAAAAACCTGAAGAATTCGAAGCTCTTAACAATAGGATGCTAACAGCCTCA
TATCAGTGTTAGGAACAATCCATTAATTCACGTTACGTTTTCTTGATGGTTCAA
GAGTTGAGGAATAACATGAGGATATTTCCAGCTAATGTTGCTTTTGAGTTATAT
TCCCAGTTTCATACTAGATCACTATAAGGCTAGAGCCTCATTAGAGGAGATTT
CAAATAAATGATGAGACATCCATCTCTGTATTGGTTTACAAAAGATGCACAGAG
GAGAAATTACCAAGTTTAAAAAAGCATTCTGCTACTGGAGAATTCAGCAGTGAA
CCCCTAGAAATGTAGTTAGAAATGACACCATTCAATCAATCCAGAAGACCCA
AACATGACTCCAATTGAAAGAGAGAATATCATTAAAGGGTTACTTATATGGCAGA
AGCCTAATTCCTGTGGATAGTATAATGGAAGATAAAATGAAGTATTAATGTGTT
AGGTCATTTAATTATTGGGATTTGTTGACAAGTCATAAATTCCTAGACATTAT
TTTATGTCTAGTGTGGACATGGTTGTAGCTATTGATTGTGAAAAAGCAAAGAAA
TCTTTATCTTCTTTGATCATAGCCTTAATTGCCACAAAGAAAATTGCTATTGCA
AGATTTGTTGGCAGAGAAAAGAGTTCTCCAAAATGGTGGTGTATTGCCTCAT
AAATCAAAGAGCTATTAGTGTTTTTGGATGATTGCCTTACCCACTTCTGAAGAC
ATCAGGCATTTCCAATTTGCTGCTTTAAGAAAGTCCACTCCTCATCAATAAATA
GCTGTAGCCTCCTTGATTGATAAAATGGATCTAGAAGCACTTCCAAACGAATCT
GGAGAACCAGAAGAATTACTTAAAATGAATACATTGCAAATCCAACCTAGACAA
TATTTCTAATAAGTAGTCATGCATAAAGCTATCACAAGAACTGATGTCATTCT
CCAATCTCTCTTTGATTCTAGAATATTTGCATCCCGAATAACGAGTATACAAT
TATGCTTAGGATGCTATATAGAGAGTGAAGAATGCTTTTAAATTCAAAGTTAAT
GAAATCAAGAAACCCTAAGATAAAAAAGTATTTTGGAAACAATTATTCGATGAA
CAAACCTACATAATAACAACAAGCTTAACAATAAATAGAGGAAGAAGTGGTTGAA
ATCAATAGAGAGGAGGAAGAGATGGTCAATATGTTTGCCAAATAAAAATTGGGA
TTCAATGATGATATCATTAAAGAAATTGGTTCTGTGGATCCAATCTCAGATTT
AAGAAAATGATCACTGAAAAGAGAGTCGATTTGGTTGATTCTGCCCTTCAACAA
ATTCAGAAGGTTATAATTGGATTAGTCGATTAATCAGTAAAAGGCAGTTTCTTC
CCCAAGGCCTTGGAAATGCTTAAAAGAAATGAGAAGAGCTTGTATTTCAAGAAGAT

GAAGCTCCAGTTTTCAACAAGTTCCTCTTTGTACTTAAGGATAAGTACAATCAA
TCTATATTTTGGGCCCAAATTGTACAATAAGGCATCACTTTGATAAGCAACATA
GAGAATTA AAAATCAGGTGTC ACTGCTGAGGAAGCATAAGATGTATATATATTT
AATTAATTATTAGTTCCTGAACAAAGAAGATAACAAACATTAGTAAATGGTTGA
TCAATTATAACATGAGGAAGAAGATTTATTAGCTGATATTGAATGATTTATTAT
TAGTCTATATCTAACAAATTTTATTCCAATTTTATATTAATAA

Flag-KU8oac TTAAATAGATAAATAATACAGGCAATTTTTTAATATACAAATTTCAA
TAATGAATTTGTAGATATATATAAAAAATAAGTGCAAAGATGAATTCTAGAGAC
TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGAC
GATGATAAGGGATCCATGGCAGGAAAAGAAGCAACTTTAGTTTTATTAGATGTA
GGTGCCTCAATGTATGAACCATACAAGTAGGCTCAAGGAAAGAAAATCACTAGA
TTAGAATTAGCTGTTGATTGCATAGGGATGATGATTTAATAAAAGATCTTTAAT
TACAAGAATCATGAAGTTGGATTAGTTTTGTTTGGAACAGAGGACGCTGAGGAT
GGGAATACTTTTTACATCTAAACATTGTCGCTCCAGATCTAGAGTTTTATAGA
AATCTGACAGAATTACCTAATCATGATATACCTAAAATAATTGGTGGAGATATT
TTTGATGCTTTGGACAAATCAGTTAGCACTTTAGATGAGTATGTGAAGACCAAG
AAGATGGATAAGAAGATATTTGTGCTCACAGCAGGATTTGGTTAGACAGAATAT
AATGAGAAGAAGATTGCAAAGCTTATTAATAATGATTGAAAAAGTAGATGTGAAG
ATCAATTTTCATAGCATTGGATTTTCATGAATGAATATGATGCTGAATTGGATGAT
CCTTCAAACCTGAGAATCAAGAACTCTAAATGATCGAATGCTTAATGCTGTT
TACGAAAGTTAAGAGCAATCAATCAATTCACGTTTAGTTTTATTACATGGTTCAA
GAACTAAGGAGTCATATGAGAATATCCCAGCTAATATCGCATTGGAATTGTAT
TCACAATTTACACCAAACAATGCAAGCAAGAGCCTCATTAGGGGGGACTTC
CAGATTAACGATGAAATCTCTGTGCAAGTGCTAATTTATAAGCGATGTTTCGAG
GAAAGATTGCCAACTCTAAGAAAGCATTCAACCCTTGGCGAATTCCAAACAGAT
ACTAATAAAAACCATGTGAGAAATGATCTTATCTACTATAATCCAGAGGATCCT
AATATGACACCAATCGAAAAAGACAACATCATAAGGGGATATTAATATGGCAGG
AATCTCGTACCAGTTGATTAATAATGGAGGATAAAATGAAGTACTAGTGTCTCT
CGTTAATTTTAATTATTAGGCTTTGTAGATCGATCCCATATACCTAGATATTAT
TATACCTCAACTGTTGACATGGTCATTGCTGTAGAAAATCAGAAGTAATAGAAA
GCTCTGGCTGCTCTAGTCATTGCACTCATTGCAACTAGGAAAGTAGCTCTTGCC
AGATTTGTAGGCAGAGAGAAAACAGCACCTAAATTGATTATGTTATTACCCCAT
AAATCCAAGAACTCTTAGTGCTTCTGGATGATATCCTTGCCTACAACCTGAAGAC
ATCAGGCATTTCAATTTGCTGCTTTGAAGAGATCTACTCCTCCATAATAGATG
GCAGTCTCTGCTATGATTGACTGTATGGATCTAGAAAAGATGCCACCGGAAGAT
GGACAGTTTGAAGA ACTTCTTAAAATGAAGTATGTTGCCAATCCTACGAGATAA
TACTTCCAATAAGTGGTCATGCACAAAGCCATCACTAGATCTGATGTATTGCCT
CCTATCTCCCCTCTCATCTTGGAGTATTTACATCCTGAAAAGAGGGTTTACGAT
TATGCAAAGAGGCACTTCAAAAAGTCAAGGCTGCTTTCAAATTCAAAATCAAC
GAAATCAAAAAACAAGGGGATAAGAAAGTGTTTTGGAAATAACTATTTGAAGAT
CAATCTACAGAATAAATTTAATAGTAAGTGGAGGATGAAGTAGTTGAAATCAAT
TAAGAAGAAGAAGAGATGGTTAATATGTTTGCCAAATAAAAATTAGGATTCAT
GACGATATAGTTAAGGAAATCGGCACTGTTGATCCCCTTCAGATTTTCAGGAAG
ATGATTACTGAAAAGAGAGTGC ACTTAGTTGACACTGCCTTACAATAAATTCAA
AAAGTCATTATTTAATTTGTGGATCAATCTCTCAAAGGCAGTTTCTATCCAAA
GCACTTGAGTGTCTCAAAGAAATGAGGAAAGCCTGTATCACAGAAGATGAAGCC
CCAGTATTCAATAAATACTTGCATGTTCTTAAAGAGAAGTACAGTCAACTTGTC
TTCTGGGCTCAGATTGTCCAATAAGGAATCACTTTAATCAGCAACATTGAGAAT
TAGAAATCTCATGTTTCTGTTGATGAAGCACAGAAGTAATTCATTATTATAAT

ATTAGTTCCTAAATAAAGAGGATATTAGTCATAAATAATTAGTTGATCAATTAT
AACACGAAGAAGAGGATCTATTGGCTGAAATTGATTGATTATTATTAGTCTAT
ATCTAACAATTTTATTCCAATTTTATATTAATAA

Sixième partie

Article en préparation

Functional diversification of *Paramecium* Ku80 paralogs allows coupling between cleavage and repair during programmed DNA elimination

Arthur Abello ^{1,3}, Vinciane Régnier ^{1,2,3}, Olivier Arnaiz ¹, Romain Le Bars ¹, Mireille Bétermier ^{1*}, Julien Bischerour ^{1*}

¹ Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France

² Univ Paris Diderot, Sorbonne Paris Cité, Paris, France

³ These authors participated equally to this work

Keywords: Double-strand break repair / NHEJ / Whole genome duplication / transposase domestication / piggyBac / Programmed genome rearrangements / ciliates

* To whom correspondence should be addressed. Tel: +33169823164. E-mail: mireille.betermier@i2bc.paris-saclay.fr; julien.bischerour@i2bc.paris-saclay.fr

ABSTRACT

INTRODUCTION

Gene duplication and diversification have long been considered a driving force for the evolution of organisms throughout the tree of life (reviewed in (Taylor and Raes 2004)). Different gene duplication mechanisms have been described. Segmental duplication is frequent in many organisms and affects a region of a chromosome harboring one or more genes or part of a gene (domain duplication). Much more dramatic whole genome duplications (WGD) are rare events that lead to duplication of all the genes in the genome simultaneously and therefore have great potential for the emergence of novel functions (Ohno 1970) or the subdivision of ancestral functions between the two duplicated gene copies (sub-functionalization, see (Force et al. 1999; Lynch and Force 2000)). The *Paramecium aurelia* group of ciliate species constitutes a remarkable system to study the evolutionary fate of duplicated genes. In these unicellular eukaryotes, three successive WGDs have occurred during evolution (Aury et al. 2006) and numerous gene duplicates from the intermediate (WGD2) and most recent (WGD1) WGDs have been retained in extant genomes (~25% and ~50%, respectively) (Aury et al. 2006; McGrath et al. 2014). Very few examples of functional diversification of WGD duplicates, however, have been reported in these species (Shi et al 2017). Instead, extensive analyses of *Paramecium* genomes and gene expression levels have indicated that the vast majority of post-WGD gene pairs (also called ohnologs) have likely preserved their original function and tend to return to their initial single-copy state overtime (Aury et al. 2006), the main determinants of retention in two copies in extant genomes being expression level and dosage constraints (Gout et al. 2010). Within a pair of ohnologs, progressive dosage unbalance between duplicates is thought to allow pseudogenization and eventual loss of the copy with the lower expression level (Gout and Lynch 2015).

Like other ciliates, *Paramecium* species all exhibit a characteristic nuclear dimorphism, with one polyploid somatic macronucleus (MAC, 800n), essential for gene expression but fragmented and eventually destroyed during sexual processes (conjugation between partners of compatible mating types or autogamy, a self-fertilization process), and two identical germline micronuclei (MIC, 2n) that undergo meiosis and transmit the parental genetic information to the next generation (reviewed in (Bétermier and Duharcourt 2014)). In each offspring cell, a new MAC develops from a copy of the MIC through a process involving large-scale programmed genome rearrangements (PGR). In *P. tetraurelia*, PGR eliminate 25 to 30% of germline DNA, including numerous repeated sequences, such as transposable

elements (TEs) or minisatellites, and ~45,000 short and noncoding Internal Eliminated Sequences (IESs), each found in the germline genome as a single-copy element - although most IESs have evolved from ancestral TEs (Arnaiz et al. 2012; Guerin et al. 2017). IESs, flanked by one conserved TA dinucleotide at each end, are precisely excised from the somatic genome, leaving one TA at their excision site. Because IESs interrupt 47% of all genes in the germline, their efficient and precise excision is mandatory to ensure that the somatic genome is correctly assembled, thus allowing progeny survival. IES excision is mediated through the concerted action of six distinct domesticated PiggyBac transposases: one catalytic subunit (PiggyMac or Pgm) and five architectural partners (PgmL1 to PgmL5) that act together to cleave both DNA strands around each flanking TA (Baudry et al. 2009; Dubois et al. 2017; Bischerour et al. 2018). Once IESs are removed, tens of thousands of DNA double strand breaks (DSBs) introduced in somatic chromosomes are massively repaired by the Ligase IV- and Xrcc4-dependent classical non-homologous end joining pathway (C-NHEJ), which directly ligates broken DNA ends without the need of sequence homology and allows the precise closure of IES excision sites ((Kapusta et al. 2011), reviewed in (Dubois et al. 2012; Bétermier et al. 2014)). Linear excised IES molecules are either directly circularized, also in a C-NHEJ-dependent manner, or form concatemers before circle formation (Bétermier et al. 2000; Gratias and Bétermier 2001; Allen et al. 2017).

The C-NHEJ pathway has been conserved in all kingdoms of life and strictly depends upon the Ku protein, a key DSB repair factor that binds broken DNA ends, protects them against extensive resection and recruits downstream C-NHEJ proteins (reviewed in (Grundy et al. 2014)). Most bacteria and archaea harbor a single *KU* gene encoding a protein homodimer, while the related *KU70* and *KU80* genes encode the two subunits of the eukaryotic Ku heterodimer. *KU70* and *KU80* were duplicated from a single ancestral *KU* gene and their products are not interchangeable within the heterodimer, providing a nice example of functional diversification of duplicated genes (Aravind and Koonin 2001). Interestingly, some bacteria harbor multiple *KU* genes, which suggests that gene duplication and diversification may have occurred also during bacterial evolution (Dupuy et al. 2018). Eukaryotic *Ku70* and *Ku80* are each organized in three domains: a globular N-terminal α/β domain – absent from bacterial Ku proteins - exposed to the outer surface of the dimer, a core dimerization domain containing a β -barrel and the DNA binding ring, and a more divergent C-terminal domain (Walker et al. 2001). As a consequence of the WGDs, *P. tetraurelia* harbors two *KU70* and three *KU80* genes, which has conferred to this species additional potential for the

diversification of *KU* genes (Marmignon et al. 2014). As expected for WGD1 ohnologs, *KU70a* and *KU70b* encode 98% identical proteins with redundant functions, but *KU70a* is consistently expressed at higher levels than *KU70b*. *KU80a* and *KU80b* are WGD1 ohnologs with constitutively low expression levels throughout the life cycle, while *KU80c*, a more distant WGD2 duplicate, is strongly and specifically induced during MAC development. Remarkably, *KU80c* expression is absolutely required for the licensing of Pgm-mediated DNA cleavage at IES ends, indicating that DNA cleavage and DSB repair are tightly coupled during PGR in *P. tetraurelia* (Marmignon et al. 2014).

In the present study, we have examined whether the essential need for *KU80c* during MAC development is simply due to its differential pattern of expression relative to *KU80a* and *KU80b*, or whether the Ku80c protein itself, which shares only 74 and 73% identity with Ku80a and Ku80b, respectively, has specialized for IES excision. We first show that Pgm total intracellular levels remain unchanged and that Pgm still enters the developing new MAC upon RNAi-mediated *KU80c* knockdown (KD). However, Pgm is no longer stably retained in the developing MAC, in association with genome-wide IES retention. We also report that overexpression of Ku80a under the control of *KU80c* transcription signals neither restores the development of a functional MAC nor IES excision in a *KU80c* KD, confirming that, in addition to being specifically produced during MAC development, the Ku80c protein has acquired a specific function during PGR. Finally, using chimeric Ku80a/Ku80c or Ku80c/Ku80a fusions, we provide evidence that the N-terminal α/β domain of Ku80c is the major determinant that confers its functional specificity to the protein. We discuss how nuclear dimorphism and the pressure for efficient and precise IES excision in *Paramecium* may have contributed to the sub-functionalization of WGD paralogs involved in C-NHEJ-mediated DSB repair.

RESULTS

Pgm nuclear stability is reduced in Ku-depleted cells.

RNAi-mediated KD of the development-specific copy of the *KU80* gene (*KU80c*) or its partner *KU70* was shown to trigger death of sexual progeny in *P. tetraurelia*, IES retention and the inhibition of DSB introduction at IES ends (Marmignon et al. 2014). We previously reported that a Pgm-GFP fusion protein still localizes in the new developing MACs of Ku-depleted cells. However, these first observations were made at late time-points during autogamy (24 to 36 hrs after the T0 time-point, defined as the stage where 50% of the cell population harbor a fragmented old MAC), well after IES excision has been completed in control conditions. In order to get a better understanding of the molecular mechanisms underlying the Ku-dependent activation of Pgm activity, we re-examined the endogenous Pgm signal at earlier time-points during MAC development in *Ku70/80*-depleted cells, using specific anti-Pgm antibodies (Dubois et al. 2017).

Early autogamous cells (5 to 10 hours after the T0 time-point) were fixed either with or without a Triton pre-extraction step (Fig. 1A and 1B, Fig. 1-figure supplement 1). As shown previously, the pre-extraction step allows the removal of labile nuclear proteins, whereas fixation prior to Triton treatment allows unbiased observation of nuclear proteins, although with higher background and reduced sensitivity (Bischerour et al. 2018). Cells were submitted to control (L4440), *PGM*, *KU70* or *KU80c* KDs and the efficiency of KD was confirmed based on the absence of viable progeny with a functional new MAC (Table S1). Total Pgm cellular amounts were confirmed to be unchanged in *KU70* and *KU80* KDs on western blots (Fig. 1C). In *Ku80c*- and *Ku70*-depleted cells, a moderate decrease of the Pgm signal was observed in the absence of pre-extraction (25% and 44%, respectively), whereas it completely vanished from control Pgm-depleted cells (Fig. 1A and 1B), suggesting that *Ku70/80c* is not strictly required for Pgm entry in the new developing MAC. When the same cells were submitted to Triton pre-extraction, the Pgm signal went down to the background level detected in a *PGM* KD, suggesting that Pgm is loosely associated with the new developing MAC in the absence of *Ku70/80c*. This phenotype is reminiscent of the depletion of any of the PgmL proteins, other essential Pgm partners that were characterized recently (Bischerour et al. 2018). *Ku70/80c*, like PgmLs, contribute to the retention of Pgm in the new developing MAC, most likely by its stabilization inside the excision complex.

To confirm our previously published observations using the GFP-Pgm fusion (Marmignon et al. 2014), we performed an immunostaining experiment of endogenous Pgm in wild-type cells at a later autogamy stage (30 hrs after the T0 time-point), including a Triton pre-extraction step (Fig. 1A right panel). While the signal of endogenous Pgm was not detected anymore in the control, as expected, it clearly reappeared in the *KU80c* KD and, to a lesser but significant extent, in the *PGM* KD (6.1 and 2.9-fold above the control, respectively). Aberrant up-regulation of *PGM* and *KU* genes at late autogamy stages was reported in *PGM* or *KU80c* KDs, both conditions under which DNA rearrangements are impeded (Marmignon et al. 2014). Overexpression of *PGM* and *KU* at late time-points, together with the depletion of RNAi-inducing feeding media upon prolonged starvation, may account for the observed accumulation of Pgm and Ku80c in the new MAC during late autogamy. In all following experiments, therefore, we focused on the early time-window, during which Pgm transiently accumulates in the new MAC and DNA elimination takes place in control cells.

***DNA-PKcs* KD increases Pgm abundance in the new MAC**

During NHEJ-mediated DSB repair, Ku70/80 associates with the DNA-PKcs catalytic subunit to form the DNAPK complex that recruits and activates downstream repair factors involved in the processing of ends prior to ligation (reviewed in (Davis et al. 2014)). *P. tetraurelia* encodes an ortholog of the mammalian DNA-PKcs, which is overexpressed during autogamy (Arnaiz et al. 2017). We tested whether the entire DNA-PK complex is required to stabilize the localization of Pgm in the new developing MAC by knocking down the *DNA-PKcs* gene. In agreement with previous data showing that DNA-PKcs is required for MAC development (S Malinsky, personal communication cited in (Bétermier and Duharcourt 2014)), DNA-PKcs-depleted cells failed to produce viable post-autogamous progeny (Table S1). In contrast to Ku-depleted cells, DNA-PKcs depleted cells even showed a 50% increase of the nuclear Pgm signal with respect to a control (Fig. 1-figure supplement 2). This is consistent with previous observations that Pgm-dependent DSBs are indeed introduced at IES ends in DNA-PKcs-depleted cells, although perhaps not as efficiently as in a control and not repaired as efficiently for a subset of IESs. Our data suggest that the Ku70/80 heterodimer alone is required to stabilize Pgm in the developing new MAC.

Ku80c is required for the excision of all IESs genome-wide

The retention of a small set of IES in a *KU80c* KD was demonstrated using PCR approaches (Marmignon et al. 2014), suggesting that destabilization of Pgm nuclear localization is correlated with IES retention in the new developing MAC. To survey IES retention at the genome-wide scale, we sequenced total genomic DNA from late autogamous cells, in which expression of the *KU80c* gene was knocked down. As a control, the same karyonidal population (i.e. cells issued from the same MAC differentiation event) was subjected to a control KD against the *ND7* gene, which is not involved in DNA rearrangements (Skouri and Cohen 1997; Baudry et al. 2009). IES retention scores were calculated following the mapping of Illumina sequencing reads on IES sequences and their excision junctions and analyzed using previously developed bioinformatics tools (Denby Wilkes et al. 2016). No IES was found to be significantly retained in the *ND7* KD control, whereas almost the entire set of IES was retained in the two *KU80c* KD replicates, indicating that all IESs depend on Ku80c for their elimination (Fig. 2B and Fig. 2-figure supplement 1). We also examined IES excision polymorphism in the three samples, since different types of excision errors are introduced at each sexual cycle in the new developing MAC in wild-type cells (Duret et al. 2008) and the frequency of a specific type of error was shown to increase in cells depleted of PgmL1 or PgmL3 (Bischerour et al. 2018). The nature and frequency of IES excision errors in *KU80c* KDs were similar to those observed in the control *ND7* KD, supporting our conclusion that the Pgm-associated excision complex is totally inactive in Ku80c-depleted cells (Fig. 2-figure supplement 1).

Pgm is dispensable for the stable localization of Ku80c in the new developing MAC

Pgm is not retained in the developing MAC in the absence of Ku70/80c. Reciprocally, we investigated whether the nuclear localization of Ku80c and its stability was impacted by the depletion of Pgm. In the absence of specific antibodies against *P. tetraurelia* Ku proteins, we constructed a transgene expressing the Ku80c protein fused to a 3xFLAG tag at its amino-terminal extremity (FLAG-Ku80c) under the control of the *KU80c* regulatory sequences. The FLAG signal was monitored by immunostaining of Ku70- or Pgm-depleted cells using the Triton pre-extraction protocol (Figure 3). In the control, FLAG-Ku80c was observed exclusively in the new developing MAC and forms foci, as previously reported for a GFP-Ku80c fusion (Marmignon et al. 2014). In the *KU70* KD, the signal was strongly reduced, supporting the idea that Ku80c works exclusively with its Ku70 partner inside the nucleus. In the *PGM* KD, in which DSB introduction and IES excision are abolished (Baudry et al. 2009;

Arnaiz et al. 2012), the FLAG-Ku80c signal still accounted for 70% of the control, suggesting that the main fraction of Ku still localizes stably in the nucleus without Pgm, hence that the nuclear localization of Ku does not depend upon the introduction of programmed DSBs. A contribution of DSBs to intra-nuclear Ku80c stability, however, was inferred from the analysis of Xrcc4-depleted cells, in which DSBs are introduced normally at IES ends but are not repaired and accumulate in the developing new MAC (Kapusta et al. 2011). Under these conditions, indeed, the FLAG-Ku80c signal strongly increased, suggesting that the presence of unrepaired DSBs can contribute to accumulation of FLAG-Ku80c in the developing MAC.

Developmental expression of Ku80a does not complement a *KU80c* KD

P. tetraurelia harbors two types of *KU80* paralogs originating from WGD2, which exhibit different expression patterns during the sexual cycle (Marmignon et al. 2014): *KU80c* is specifically induced during DNA rearrangements, whereas *KU80a&b* are constitutively expressed at low level. In addition, WGD2 *KU80* ohnologs have diverged in their amino acid sequences, with 74% identity between Ku80a and Ku80c (Fig. 4A). Interestingly, all species from the *Paramecium aurelia* group whose MAC genome has been sequenced have conserved at least one paralog of each type, whereas more distant species (e.g. *P. caudatum*, *P. multimicronucleatum* and *P. polycaryum*) that did not undergo the last two WGDs encode a single version of Ku80 (Fig. 4-figure supplement 1)). The conservation of Ku80c across all *P. aurelia* species suggests that the *KU80c* gene has undergone sub-functionalization not only through the regulation of its transcription, but also through the function of its encoded protein itself. To test this hypothesis, we examined whether Ku80a can substitute for a depletion of Ku80c during sexual processes. The *KU80a* open reading frame was cloned under the control of the *KU80c* regulatory sequences and fused at its 5' end to sequences encoding the 3xFLAG peptide (Fig. 4A and Materials and Methods). In parallel, the nucleic acid sequence of *KU80c* was modified in order to become insensitive to RNAi (Fig. 4A). The *FLAG-KU80a* and *FLAG-KU80c* transgenes expressing each fusion were microinjected into the MAC of vegetative cells and the resulting transformants were grown and starved to induce autogamy. Cells were collected 5 to 10 hrs following the T0 time-point and the expression of FLAG-Ku80 was checked on western blots using anti-FLAG antibodies (Fig. 4B, grey bars; Fig. 4-figure supplement 2). Six transformants with different expression levels of each fusion were selected for further analysis. In control non-injected cells (Ni) or in transformants with too low expression levels (c1 and c2), little or no viable post-autogamous progeny was recovered

(Figure 4B, black bars), while higher expression levels allowed functional complementation and the recovery of progeny harboring a functional new MAC (c3 to c6). These data confirm that the FLAG-Ku80c fusion protein is functional. In contrast, at similar expression levels, FLAG-Ku80a failed to complement a *KU80c* KD (see a5 vs c3, a6 vs c4 or c5). The same experiment was repeated several times and functional complementation was never observed (Table S1). Moreover, expression of the *FLAG-KU80a* transgene exhibited a dominant negative phenotype in control KD conditions (Table S1), raising the possibility that Flag-Ku80a titrates out an essential partner of Ku80c, such as Ku70 for instance.

To gain more insight into the step at which Ku80a is defective during MAC development, we tested whether IES excision takes place normally in FLAG-Ku80a-expressing cells depleted of their endogenous Ku80c. We extracted the total genomic DNA of late autogamous cells at ~30 hrs following the T0 time-point and performed a molecular analysis of the retention of one particular IES, IES 51A1835 (Preer et al. 1992). This experiment revealed the presence of the non-excised (IES+) form both in Ku80-depleted cells expressing FLAG-Ku80a and in a non-injected control, while IES+ molecules were not detected in Ku80-depleted cells expressing FLAG-Ku80c (Fig. 4C), strongly suggesting that Ku80a is unable to substitute for Ku80c during IES excision. To extend this observation genome-wide, we performed Illumina sequencing of DNA samples extracted from complemented cells expressing equivalent levels of FLAG-Ku80a and FLAG-Ku80c (a4 and c4, respectively, see Fig. 4B) and calculated IES retention scores for both conditions. We found that almost no IES (7 out of 45,000) was retained in the fully complemented c4 clone, while a large majority (~44,000 out of 45,000) were significantly retained in a4 (Fig. 2B). Taken together, these experiments confirm that FLAG-Ku80c can fully restore IES excision in a *KU80c* KD, while FLAG-Ku80a is not functional.

Ku80a does not allow stable Pgm nuclear localization

In view of the previously reported interaction between Pgm and Ku80c (Marmignon et al. 2014), we speculated that Ku80a might be unable to interact with Pgm. To test this hypothesis, a recombinant HA-Ku80a fusion was co-expressed with MBP-Pgm in insect cells and their interaction was tested by MBP pulldown (Fig. 4E and Fig. 4-figure supplement 3). We found that both HA-Ku80a and HA-Ku80c co-precipitate with MBP-Pgm, ruling out that

a differential interaction with Pgm may simply explain the functional difference between Ku80a and Ku80c.

We further analyzed the localization and stability of Pgm through immunostaining using the Triton pre-extraction protocol, in Ku80c-depleted cells complemented with FLAG-Ku80a (a3) or FLAG-Ku80c (c3) (Fig. 4D). FLAG-Ku80c-expressing cells showed a wild type localization pattern for both FLAG-Ku80c and Pgm, with an intense signal for each protein and few large foci in the new developing MAC. In contrast, we could not detect a signal for Pgm or Ku80 in FLAG-Ku80a-expressing cells depleted of their endogenous Ku80c (Fig. 4D), suggesting that Ku80a itself does not localize in the new MAC (at least not in a stable manner) and in turn fails to stabilize Pgm nuclear localization. To discriminate between the defective nuclear import of Ku80a, or lack of its retention in the nuclear compartment, we repeated the experiment omitting the Triton pre-extraction step. The control FLAG-Ku80c fusion could not be detected under these conditions, whereas it was successfully observed with a pre-extraction step (Figure 4-figure supplement 4). In contrast, Pgm was readily detected in both conditions, suggesting that the anti-FLAG antibodies fail to bind the FLAG tag located at the amino terminus of Ku80. To circumvent this problem, we used GFP-Ku80c and GFP-Ku80a fusions to monitor the localization of GFP fluorescence. We selected transformants with similar expression levels of GFP-Ku80a and GFP-Ku80c (a5 and c13, respectively; see Figure 4-figure supplement 5) and repeated the experiment with the two alternative protocols. GFP-Ku80a and GFP-Ku80c fluorescence signals were both detected in the developing new MACs of cells that were prepared without the pre-extraction step (Fig. 4-figure supplement 4). In contrast, the GFP-Ku80a signal completely vanished following Triton pre-extraction, while it was still detected for GFP-Ku80c, further demonstrating that Ku80a is efficiently imported into the new developing MAC but only loosely associates with this nuclear compartment.

The α/β domain of Ku80c plays an essential role during DNA rearrangements.

The crystal structure of the Ku70/80 has revealed a domain organization comprising an amino-terminal α/β domain, a central β -barrel domain, and a C-terminal helical arm (Fig. 5A) (Walker et al. 2001). Using the entire set of Ku80 sequences we could find in the group of *P. aurelia* species, we generated a consensus for Ku80a and Ku80c proteins (Fig. 5-figure supplement 1). Differences between the Ku80a and Ku80c consensus sequences distributed

evenly along both proteins, without highlighting a particular domain that might account for their functional divergence. To better define which domain is potentially implicated in the sub-functionalization of Ku80c, we tested the activity of Ku80ca and Ku80ac chimeric proteins, in which the Ku80a and Ku80c α/β domains were swapped (Fig. 5A). The Ku80ca chimera carries the α/β domain of Ku80c (residues 1- 240) fused to the C-terminal part of Ku80a (residues 241-737). The Ku80ac chimera is formed by the α/β domain of Ku80a (residues 1- 240) and the C-terminal part of Ku80c (residues 241-737). Both chimeric constructs were expressed as N-terminal FLAG fusions in *P. tetraurelia*, under the control of the *KU80c* regulatory sequences and their expression was verified on western blots using anti-FLAG antibodies (Fig. 4-figure supplement 2). We then compared sexual progeny survival in a complementation assay, after cells expressing each chimeric protein were subjected to a *KU80c* KD during autogamy (Table S1). In this assay, the Ku80ac chimera failed to complement a *KU80c* KD at every level of injection that we tested (Fig. 5B). In contrast, expression of Ku80ca allowed efficient complementation of a *KU80c* KD. Genomic DNA of late autogamous cells was purified and the retention of IES 1835A was analyzed by PCR (Fig. 5C). As expected, Ku80c-depleted cells complemented with 3xFLAG-Ku80ac showed accumulation of IES+ PCR products, suggesting that Ku80ac failed to activate the excision machinery, as reported above for Ku80a (Fig. 4C). On the other hand, cells complemented with FLAG-Ku80ca showed no accumulation of IES+ forms, indicating that elimination of IES 51A1835 proceeded normally in the presence of Ku80ca (Fig. 5C). Altogether, these results point to an essential role of the α/β domain of Ku80c in its functional specialization during genome rearrangements, while the C-terminal domains of Ku80a and Ku80c can replace each other with no conspicuous effect on progeny survival.

The C-terminal part of Ku80c also contributes to its functional specialization

Although the FLAG-Ku80ca protein can complement Ku80c depleted cells, we noticed that the offspring of FLAG-Ku80c and FLAG-Ku80ca complemented cells have slightly different phenotypes. Indeed, cells from the progeny of FLAG-Ku80ca complemented cells do not grow as fast as those originating from FLAG-Ku80c complemented cells and are highly heterogeneous in shape. Nevertheless, they succeeded in resuming vegetative growth and were initially considered as post-autogamous cells with a functional new MAC. To more closely examine whether the Ku80ca chimera is fully active, we sequenced the entire genome

of late autogamous cells originating from two independent transformants expressing different levels of FLAG-Ku80ca (ca5 and ca8, see Fig. 5B and 6A). As a control, we also sequenced the genome of post-autogamous cells that were only partially complemented by FLAG-Ku80c (c2, see Fig. 4B and 6A). While excision of a vast majority of IESs was restored in the three conditions, we found that 633, 476 and 195 IES (hereafter designated as “sensitive” IESs), were significantly retained in the c2, ca5 and ca8 transformants, respectively (Fig. 5D). As expected, the number of retained IESs and the expression level of the Ku80 proteins correlate negatively. We observed 80% overlap between the two sets of Ku80ca-sensitive IES (Fig. 6B), in favor of a specific subset of IESs being retained in the presence of Ku80ca. On the other hand, 23% of Ku80ca-sensitive IES (108 out of 468) were also retained in cells that were only partially complemented with FLAG-Ku80c (c2). To distinguish whether IES retention is attributable specifically to the absence of the C-terminal part of Ku80c or simply to the presence of limiting amounts of Ku80 in the different complementation experiments, we further analyzed the characteristics of these three groups of IESs. We first compared the cumulative distribution of IES lengths for the c2, ca5, and ca8 sets of retained IES (Fig. 6C). Whereas the c2 set of retained IES was found to be slightly biased toward longer IESs, both ca5 and ca8 sets of retained IESs were significantly shorter, with 65 and 80 % of retained IESs being smaller than 50bp (40% for the whole set of *P. tetraurelia* IESs and only 25% for c2). Apart from their short size, we were not able to find any specific motif associated to IESs whose excision cannot be restored by Ku80ca.

In parallel, we noticed that Ku80ca-sensitive IESs were often found in IES-rich regions of the genome. To quantify this observation, we scored for each IES the number of IESs found in its vicinity (within a 1-kb distance on both sides). The cumulative distributions of the neighboring IES counts for all IESs (black dots) or for IESs retained in the c2, ca5 and ca8 samples (red dots) are presented in Fig. 6D. No bias is observed for the c2 sample with respect to the entire set of IESs in the genome, whereas the distributions of the ca5 and ca8 sample are biased toward a larger number of neighboring IESs, confirming that Ku80ca-sensitive IES tend to be located in IES-dense genomic regions. Because Ku80ca-sensitive IES tend to be short with respect to the whole IES population (Fig. 6C), we wondered whether their presence in IES-dense regions could be linked to their small size. To address this question, we generated a random sample of IESs with a size distribution identical to those retained in the ca5 sample (Fig. 6C, Random IES sample) and analyzed the cumulative distribution of the IES count in their vicinity (Fig. 6D, Random IES sample). We observed a slight bias towards IESs located in IES-rich regions, but far less pronounced compared to the

ca5 and ca8 samples. Our data, therefore, confirm that the size distribution of Ku80ca-sensitive IESs doesn't account for their peculiar distribution in IES-rich genomic regions.

DISCUSSION

MATERIALS AND METHODS

Injection of FLAG-KU80 or GFP-KU80 transgenes

The level of transgene injection (copy par haploid genome or CPHG) was determined by QPCR in cells maintained at the vegetative state.

Immunofluorescence analysis

Immunostaining using polyclonal guinea pig α -Pgm 2659-GP or monoclonal α -FLAG M2 (Sigma-Aldrich) was performed as described previously (Bischerour et al. 2018). Unless specified, the protocol includes a triton pre-extraction step prior to fixation of cells. Observations were made using a Zeiss Axioplan 2 epifluorescence microscope with a 63x oil objective or a BX63 epifluorescence microscope with a 60x oil objective.

Quantification of the signal was performed using the imageJ software (<https://imageJ.nih.gov>). An imageJ macro was developed allowing for automatic detection of new MACs using a threshold based on green (and red, when present) fluorescence and allowing for manual adjustments if necessary. Fluorescence intensity analysis, boxplot representation and statistical analysis were performed as described previously (Bischerour et al. 2018). In statistical analysis, ** stands for $p < 0,01$ and *** for $p < 0,0001$.

Localisation of GFP fusion

Autogamous cells were fixed for 10 min in PHEM (60mM PIPES, 25mM Hepes, 10mM EGTA, 2mM $MgCl_2$ pH 6.9) +1.2% paraformaldéhyde, washed once in TBST (10 mM Tris pH 7.4, 0.15 M NaCl, 0.1% Tween20) + 3% BSA, and stained with 0.5 μ g/ml DAPI (SIGMA). Alternatively, cells were extracted with ice-cold PHEM + 0.7% Triton during 4 min, and fixed for 10 min in in PHEM +1.2% paraformaldéhyde + 0.3% Triton before the TBST wash and DAPI staining.

FUNDING

This work was supported by the Centre National de la Recherche Scientifique (CNRS) [intramural funding to M.B.], the French National Agency for Research (ANR) [grant number ANR-14-CE10-0005-01 to M.B.], the ARC Foundation for Cancer Research [grant number PJA20151203521 to J.B.] and the I2BC [grant « SUBOKU » to J.B.]

ACKNOWLEDGEMENTS

This work has benefited from the facilities and expertise of the high throughput sequencing core facility of I2BC (Centre de Recherche de Gif - <http://www.i2bc-saclay.fr>). We are grateful to Sophie Malinsky (Institute of Biology of the Ecole Normale Supérieure, Paris) for the generous gift of the *DNA-PKcs* RNAi-inducing plasmid. We thank Cindy Mathon and Pascaline Tirand for excellent technical assistance, Loïc Escoriza, Nathalie Mathy and Coralie Zangarelli for their help in the handling of *Paramecium* cultures, Maya Houmel for performing preliminary experiments during her Bachelor internship, Marc Guérineau, Linda Sperling, Anne-Marie Tassin and all members of the Duharcourt (Institut Jacques Monod, Paris), Meyer (IBENS, Paris) and Tassin (I2BC, Gif-sur-Yvette) laboratories for stimulating discussions.

REFERENCES

- Allen SE, Hug I, Pabian S, Rzeszutek I, Hoehener C, Nowacki M. 2017. Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell* **168**: 990-999 e997.
- Aravind L, Koonin EV. 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* **11**: 1365-1374.
- Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury JM, Denby Wilkes C, Garnier O, Labadie K, Lauderdale BE, Le Mouel A et al. 2012. The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet* **8**: e1002984.
- Arnaiz O, Van Dijk E, Bétermier M, Lhuillier-Akakpo M, de Vanssay A, Duharcourt S, Sallet E, Gouzy J, Sperling L. 2017. Improved methods and resources for *Paramecium* genomics: transcription units, gene annotation and gene expression. *BMC Genomics* **18**: 483.
- Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Segurens B, Daubin V, Anthouard V, Aiach N et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**: 171-178.
- Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M. 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev* **23**: 2478 - 2483.
- Bétermier M, Bertrand P, Lopez BS. 2014. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* **10**: e1004086.
- Bétermier M, Duharcourt S. 2014. Programmed Rearrangement in Ciliates: *Paramecium*. *Microbiol Spectr* **2**.
- Bétermier M, Duharcourt S, Seitz H, Meyer E. 2000. Timing of developmentally programmed excision and circularization of *Paramecium* internal eliminated sequences. *Mol Cell Biol* **20**: 1553-1561.
- Bischerour J, Bhullar S, Denby Wilkes C, Regnier V, Mathy N, Dubois E, Singh A, Swart E, Arnaiz O, Sperling L et al. 2018. Six domesticated PiggyBac transposases together carry out programmed DNA elimination in *Paramecium*. *Elife* **7**.
- Davis AJ, Chen BP, Chen DJ. 2014. DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. *DNA Repair (Amst)* **17**: 21-29.

- Denby Wilkes C, Arnaiz O, Sperling L. 2016. ParTIES: a toolbox for *Paramecium* interspersed DNA elimination studies. *Bioinformatics* **32**: 599-601.
- Dubois E, Bischerour J, Marmignon A, Mathy N, Regnier V, Betermier M. 2012. Transposon Invasion of the *Paramecium* Germline Genome Countered by a Domesticated PiggyBac Transposase and the NHEJ Pathway. *Int J Evol Biol* **2012**: 436196.
- Dubois E, Mathy N, Regnier V, Bischerour J, Baudry C, Trouslard R, Betermier M. 2017. Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic Acids Res* **45**: 3204-3216.
- Dupuy P, Sauviac L, Bruand C. 2018. Stress-inducible NHEJ in bacteria: function in DNA repair and acquisition of heterologous DNA. *Nucleic Acids Res*.
- Duret L, Cohen J, Jubin C, Dessen P, Gout JF, Mousset S, Aury JM, Jaillon O, Noel B, Arnaiz O et al. 2008. Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res* **18**: 585-596.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.
- Gout JF, Kahn D, Duret L, Paramecium Post-Genomics C. 2010. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet* **6**: e1000944.
- Gout JF, Lynch M. 2015. Maintenance and Loss of Duplicated Genes by Dosage Subfunctionalization. *Mol Biol Evol* **32**: 2141-2148.
- Gratias A, Bétermier M. 2001. Developmentally programmed excision of internal DNA sequences in *Paramecium aurelia*. *Biochimie* **83**: 1009-1022.
- Grundy GJ, Moulding HA, Caldecott KW, Rulten SL. 2014. One ring to bring them all--the role of Ku in mammalian non-homologous end joining. *DNA Repair (Amst)* **17**: 30-38.
- Guerin F, Arnaiz O, Boggetto N, Denby Wilkes C, Meyer E, Sperling L, Duharcourt S. 2017. Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genomics* **18**: 327.
- Kapusta A, Matsuda A, Marmignon A, Ku M, Silve A, Meyer E, Forney JD, Malinsky S, Betermier M. 2011. Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet* **7**: e1002049.
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459-473.

- Marmignon A, Bischerour J, Silve A, Fojcik C, Dubois E, Arnaiz O, Kapusta A, Malinsky S, Betermier M. 2014. Ku-mediated coupling of DNA cleavage and repair during programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *PLoS Genet* **10**: e1004552.
- McGrath CL, Gout JF, Doak TG, Yanagi A, Lynch M. 2014. Insights into three whole-genome duplications gleaned from the *Paramecium caudatum* genome sequence. *Genetics* **197**: 1417-1428.
- Ohno S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, Berlin, Heidelberg.
- Preer LB, Hamilton G, Preer JR. 1992. Micronuclear DNA from *Paramecium tetraurelia*: serotype 51 A gene has internally eliminated sequences. *J Protozool* **39**: 678-682.
- Skouri F, Cohen J. 1997. Genetic approach to regulated exocytosis using functional complementation in *Paramecium*: identification of the ND7 gene required for membrane fusion. *Mol Biol Cell* **8**: 1063-1071.
- Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* **38**: 615-643.
- Walker JR, Corpina RA, Goldberg J. 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**: 607-614.

FIGURE LEGENDS

Figure 1. Expression and localization of Pgm in *Ku80C* and *Ku70* RNAi.

(A) Immunostaining of Pgm in early or late autogamous cells subjected to control (L4440), *KU80c*, *PGM* or *KU70* RNAi. At early time point, immunostaining was performed with or without triton pre-extraction. Scale bar is 5 μm .

(B) Plots of Pgm fluorescence intensity in developing MACs subjected to the different RNAi conditions and collection time points shown in (A). At early time point, quantification was performed on developing MACs ranging in size between 55-110 μm^2 (without pre-extraction) or 40-75 μm^2 (with pre-extraction). At late time point, all developing MACs $>65 \mu\text{m}^2$ were analyzed. 27 to 76 developing MACs were quantified for each RNAi condition.

(C) Western blot analysis of Pgm expression in early autogamous cells subjected to control (L4440), *PGM*, *KU70* or *KU80c* RNAi.

Figure 2. Analysis of IES retention scores through whole-genome sequencing

(A) Distribution of IES retention scores in total genomic DNA extracted from late autogamous cells subjected to control (*ND7*) or *KU80c* RNAi. Significantly retained IESs relative to a set of reference MAC genomes are highlighted in red.

(B) Distribution of IES retention scores in *Ku80c*-depleted cells expressing FLAG-*Ku80c* (c4) or FLAG-*Ku80a* (a6). See Fig. 4 for a more detailed description of the two clones.

Figure 3. Expression and localization of FLAG-*Ku80c* in *KU70*, *PGM* and *XRCC4* RNAi.

(A) Immunostaining of FLAG-*Ku80c* in early autogamous cells subjected to control (L4440), *KU70*, *PGM*, or *XRCC4* RNAi. Scale bar is 5 μm .

(B) Plot of FLAG-*Ku80c* fluorescence intensity in 25-60 μm^2 developing MACs. 12 to 20 MACs were analyzed. Quantification could not be performed for developing MACs subjected to *XRCC4* RNAi, since the same acquisition time was used for imaging all the RNAi conditions, and led to saturated FLAG-*Ku80c* signal in the case of *XRCC4* RNAi.

Figure 4. *Ku80a* does not complement *Ku80c* depleted cells during DNA rearrangements

(A) Schematic representation of the DNA tools used for complementation assay. Modified nucleic acid sequence of *KU80c* that confers resistance to RNAi is represented by hatched lines. FLAG-*KU80c* and FLAG-*KU80a* transgenes are under the control of the endogenous *KU80c* regulatory sequences.

(B) Recovery of post-autogamous cells in FLAG-KU80c and FLAG-KU80a transformants. % of survival of the progeny (in black) and relative abundance of FLAG signal (in grey) quantified by western blot 5 hours after the beginning of autogamy and normalized by the Pgm signal.

(C) Molecular analysis of IES 51A1835 retention by PCR in non-injected cells (NI), FLAG-KU80c and FLAG-KU80a transformants, submitted to control (L4440) or *KU80c* RNAi. IES- and IES+ are showed and heteroduplexes of IES+ and IES- strands formed during the PCR reaction are signaled by *.

(D) Co-immunostaining with α -FLAG and α -Pgm antibodies in early autogamous FLAG-KU80c or FLAG-KU80a expressing cells and subjected to *KU80C* RNAi. Scale bar is 5 μ m.

(E) Pull down of HA-Ku80a and HA-Ku80c fusions with MBP-Pgm using recombinant proteins expressed in insect cells. In each panel, the HA-tagged protein that was co-expressed with MBP or MBP-Pgm is indicated on the left and the band revealed on western blot using anti-HA antibodies is indicated on the right. MBP and MBP-Pgm expression was checked by western blot using anti-MBP antibodies and is presented in figure 3 supplemental X.

Figure 5. The Ku80c α/β domain fulfills an essential function during DNA rearrangements.

(A) Schematic representation of the DNA tools used for complementation assay. The Ku80ca chimera is constituted of the residues 1 to 240 of Ku80c and residues 241 to 737 of Ku80a and reciprocally for the Ku80ac chimera. FLAG-KU80ca and FLAG-KU80ac transgenes are under the control of the endogenous KU80c regulatory sequences.

(B) Recovery of post-autogamous cells in FLAG-KU80ca and FLAG-KU80ac transformants. % of survival of the progeny (in black) and relative abundance of FLAG-Ku80 signal (in grey) quantified by western blot 5 hours after the beginning of autogamy and normalized by the Pgm signal.

(C) Molecular analysis of IES 51A1835 IES retention by PCR in non-injected cells, FLAG-KU80ca and FLAG-KU80ac transformants, submitted to control (L4440) or *KU80c* RNAi. IES- and IES+ are showed and heteroduplexes of IES+ and IES- are signaled by *.

(D) Distribution of genome-wide IES retention scores in Ku80c-depleted cells complemented by FLAG-Ku80c (c2) or FLAG-Ku80ca (ca5 and ca8). See Fig. 4 for a more detailed description of the c2 clone.

Figure 6. Analysis of IES retention in Ku80ca complemented cells reveals subtle functional properties of the c-terminal domain of Ku80c

(A) Recovery of post-autogamous cells in FLAG-KU80c (c2&c4) and FLAG-KU80ca (ca5&ca8) transformants (data from Figure 3B). % of survival of the progeny (in black) and relative abundance of FLAG-Ku80 signal (in grey) quantified by western blot 5hours after the beginning of autogamy and normalized by the Pgm signal.

(B) Venn diagram representing the overlap between the set of retained IES in c2, ca5 and ca8 transformants.

(C) Cumulative distribution of IES length for the entire set of IESs (in black) or subset of retained IESs in c2, ca5 and ca8 transformants (in red). The cumulative distribution of a random control set of IESs with the same number and the same distribution of size is shown on the right panel and was used for further analyses (see D).

(D) Cumulative distribution of IES density in a 1kb window vicinity around IESs for the entire set of IES (in black) or the subset of retained IES in c2, ca5 and ca8 transformants and in random control set of IES described in C (in red). For each IES we counted the number of IES that are present at less than 1 kb of distance.

FIGURE SUPPLEMENTS

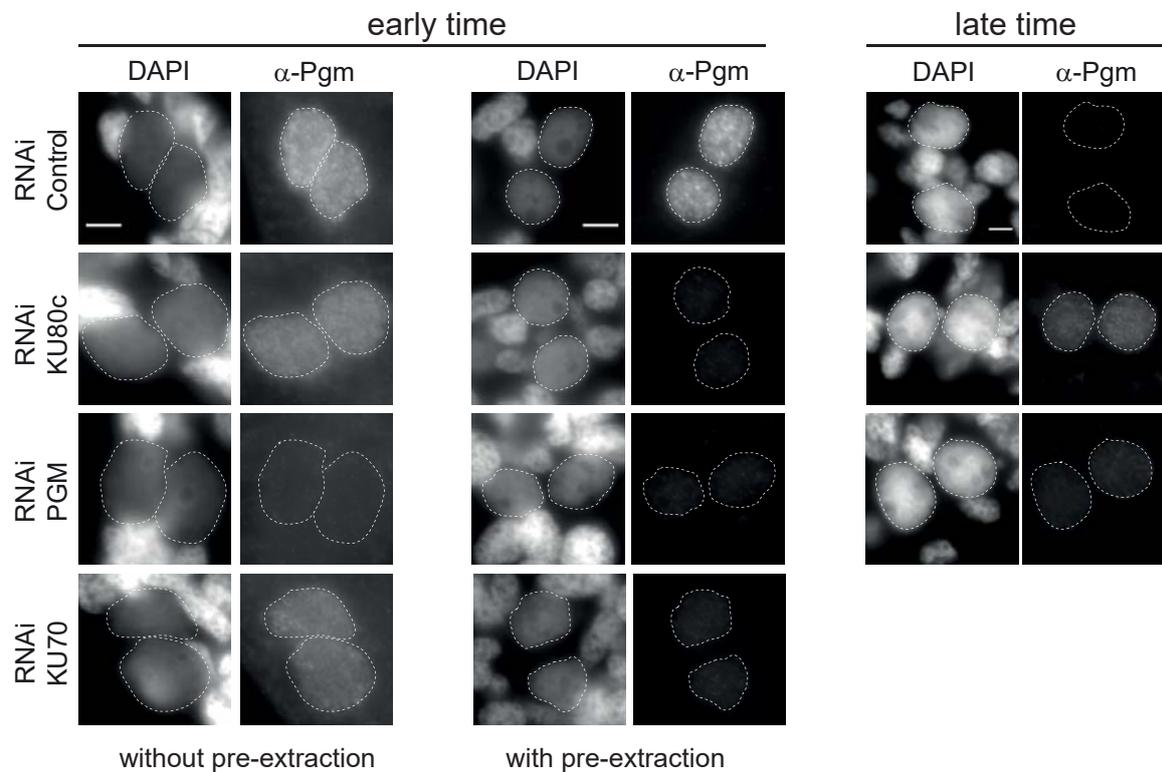
Fig. 4-figure supplement 4. Nuclear localization of GFP-Ku80 or FLAG-Ku80 fusions with or without Triton pre-extraction

(A) Early autogamous cells expressing the FLAG-Ku80c fusion imaged after paraformaldehyde fixation with or without a Triton pre-extraction step. Scale bar is 5 μm

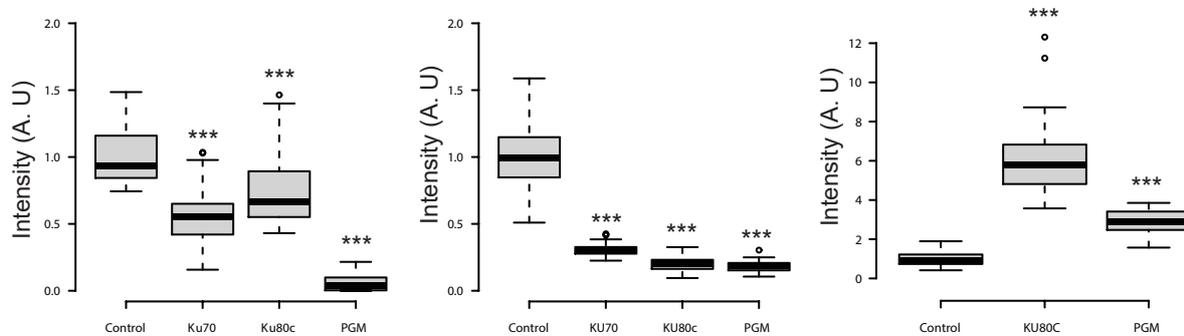
(B) Early autogamous cells expressing GFP-Ku80c or GFP-Ku80a fusion proteins imaged after paraformaldehyde fixation with or without a triton pre-extraction step. Scale bar is 5 μm .

(C) Plots of GFP fluorescence intensity for developing MACs ranging between 55-110 μm^2 (without pre-extraction) or 40-65 μm^2 in size (with pre-extraction). 28 to 56 developing MACs were quantified for each condition.

A



B



C

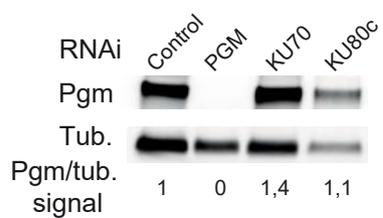


Figure 1-figure supplement 1

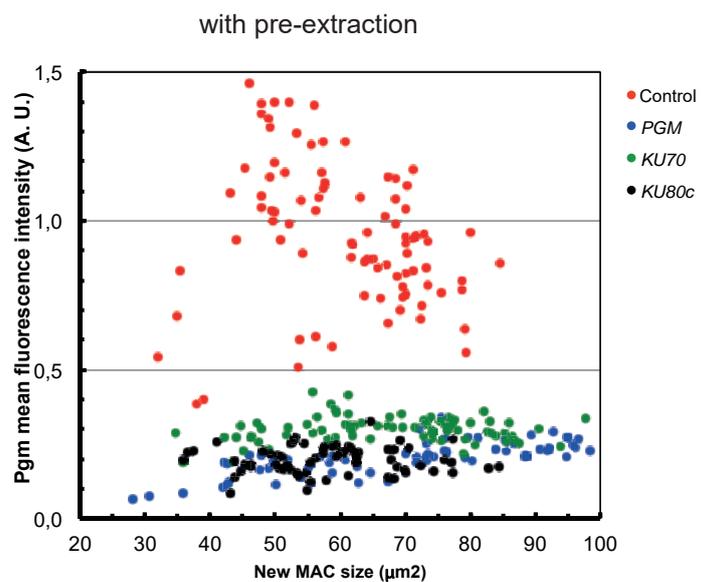
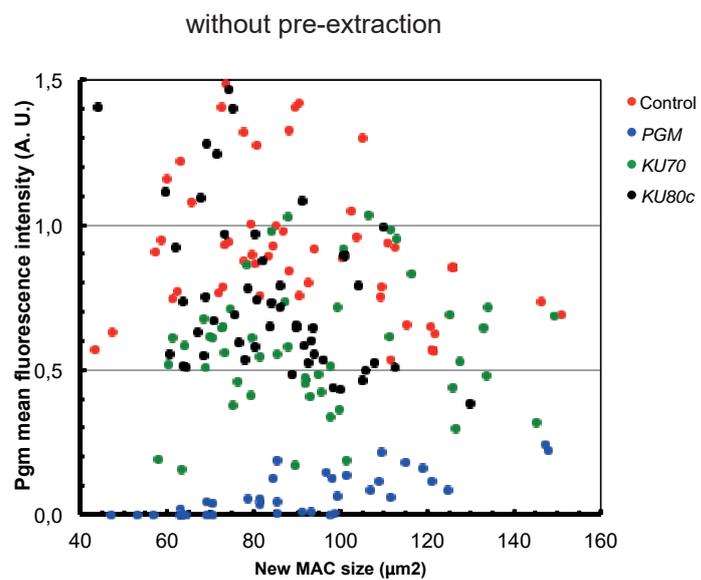
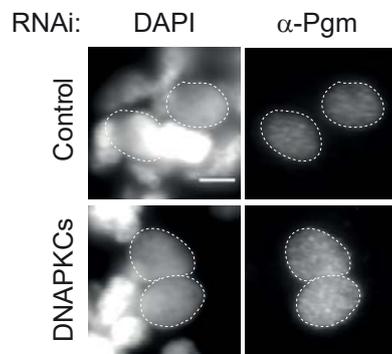


Figure 1-figure supplement 2

A



B

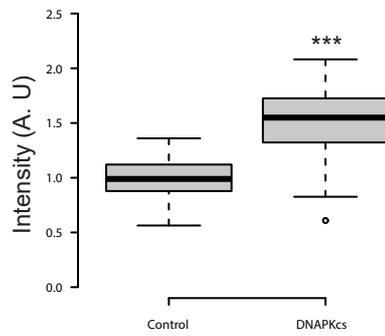


Figure 2

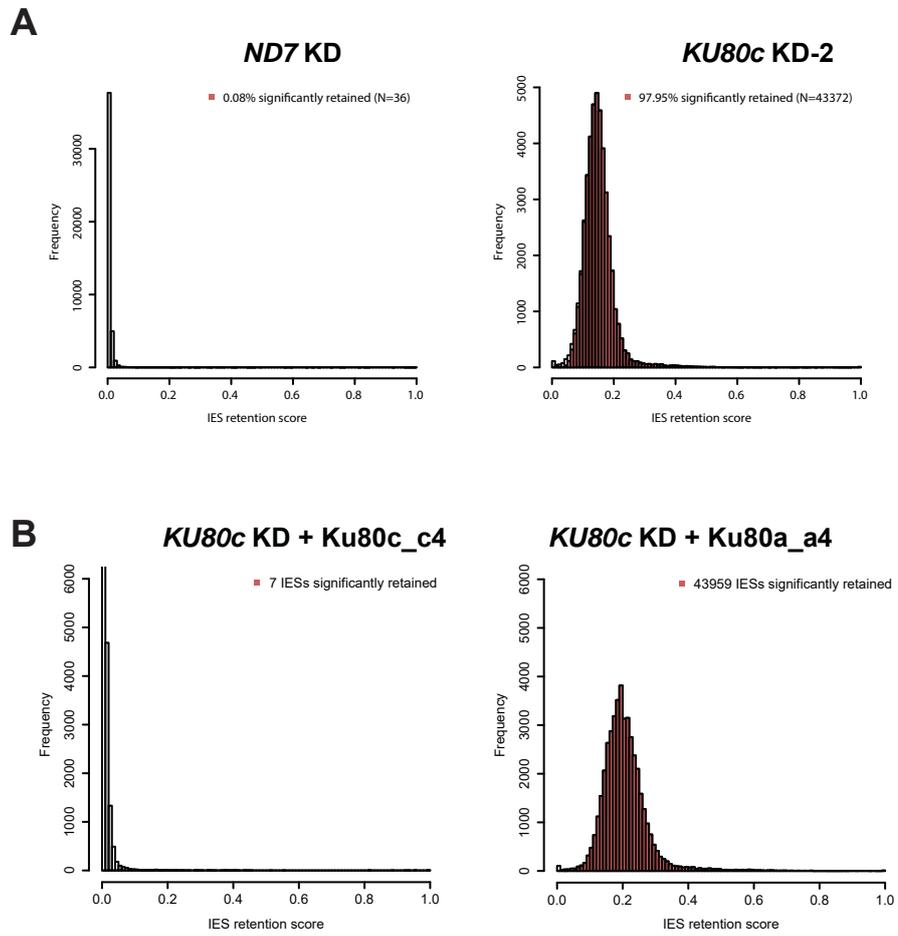
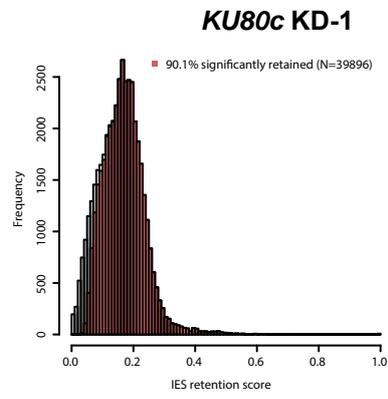
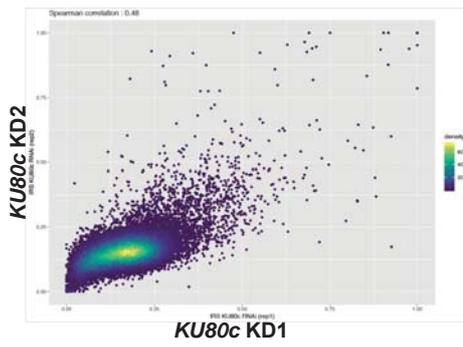


Figure 2-figure supplement 1

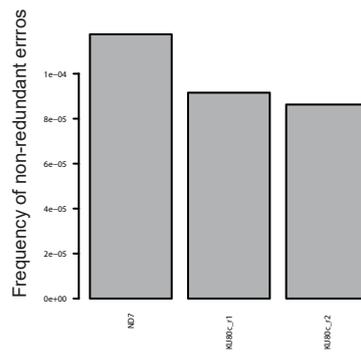
A



B



C



D

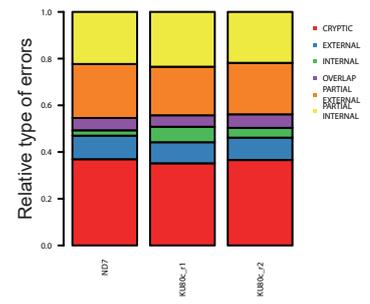
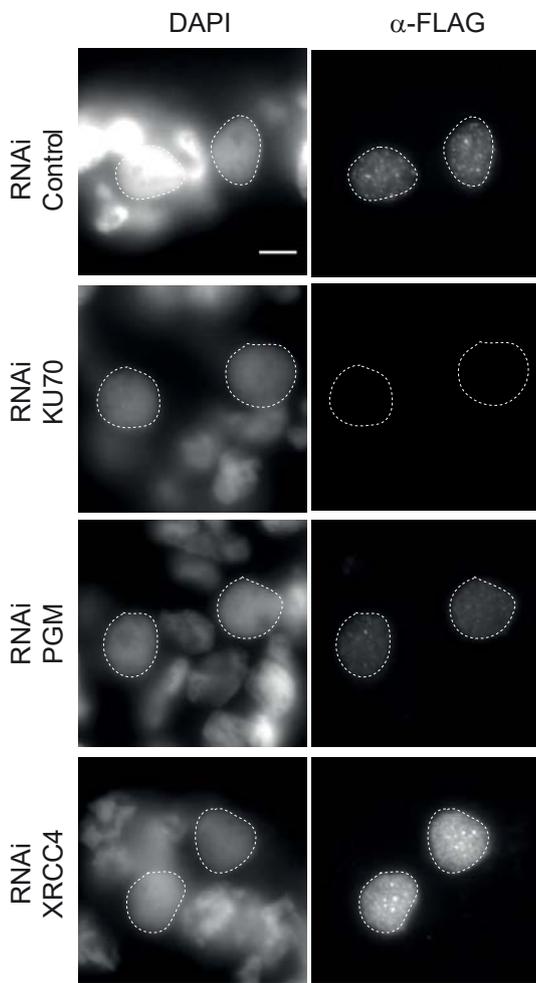


Figure 3

A



B

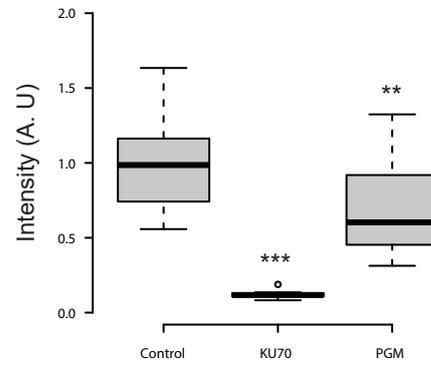


Figure 4

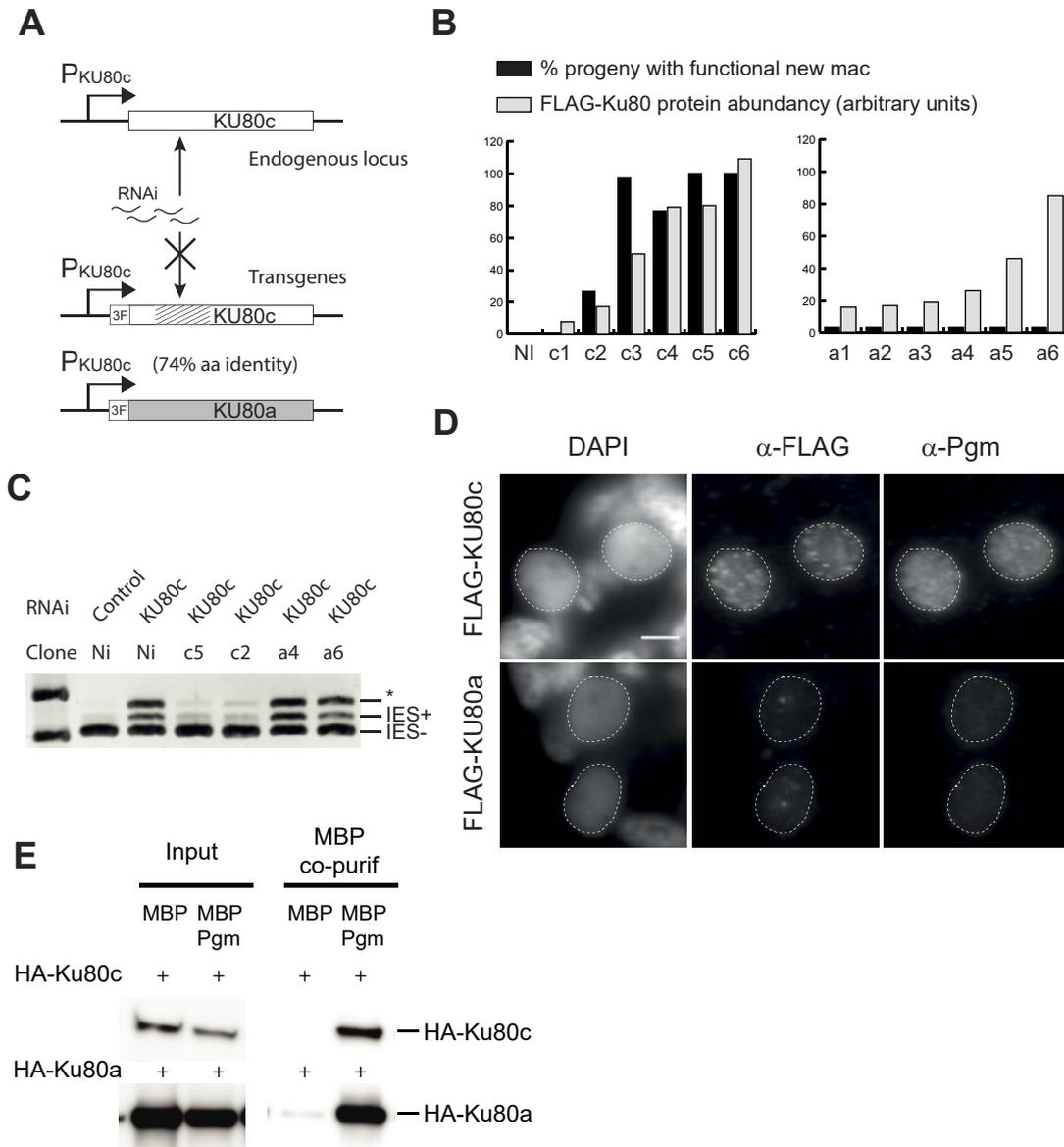


Figure 4-figure supplement 1

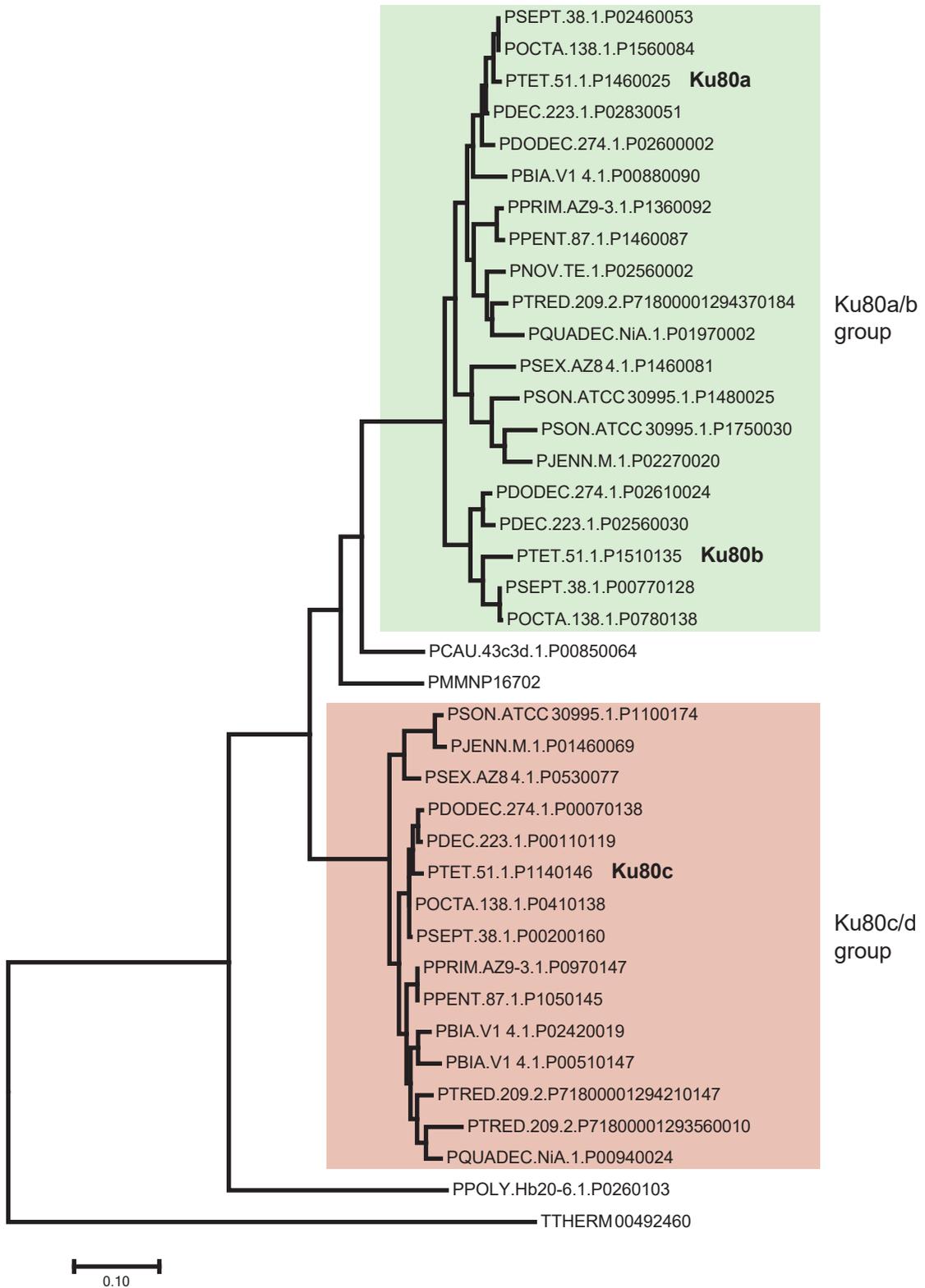


Figure 4-figure supplement 2

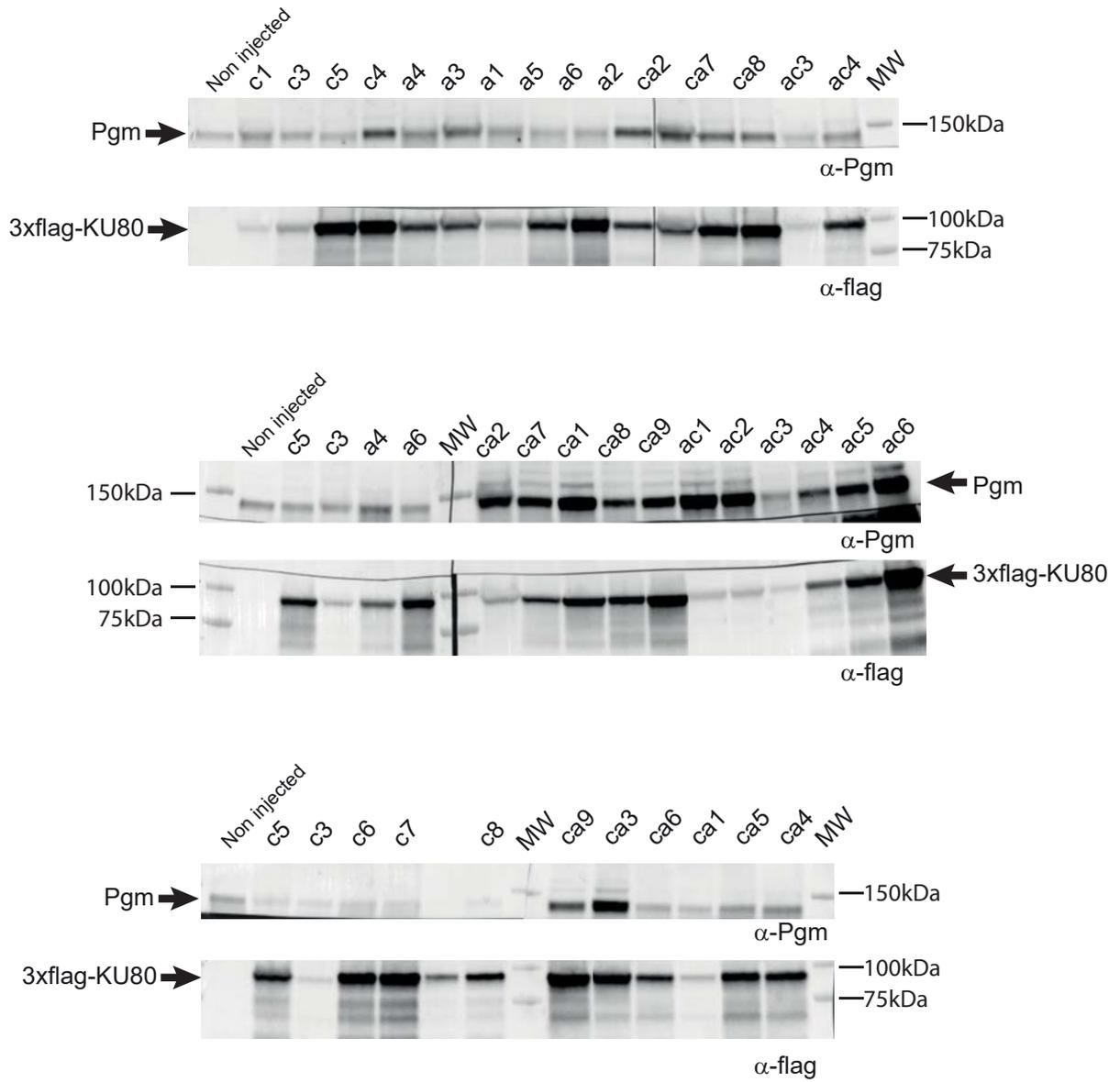


Figure 4-figure supplement 3

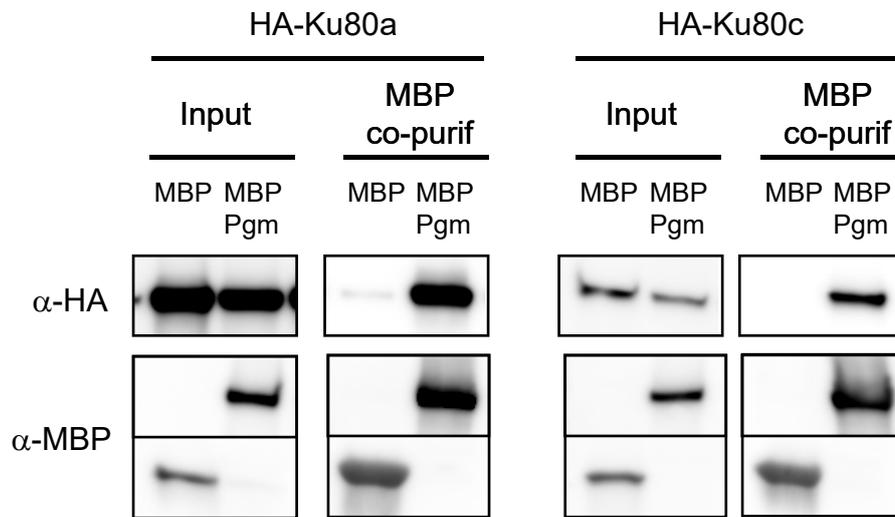
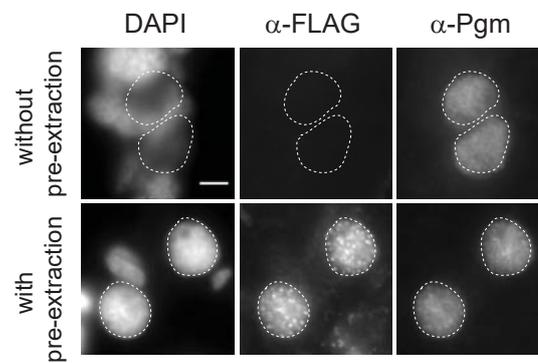
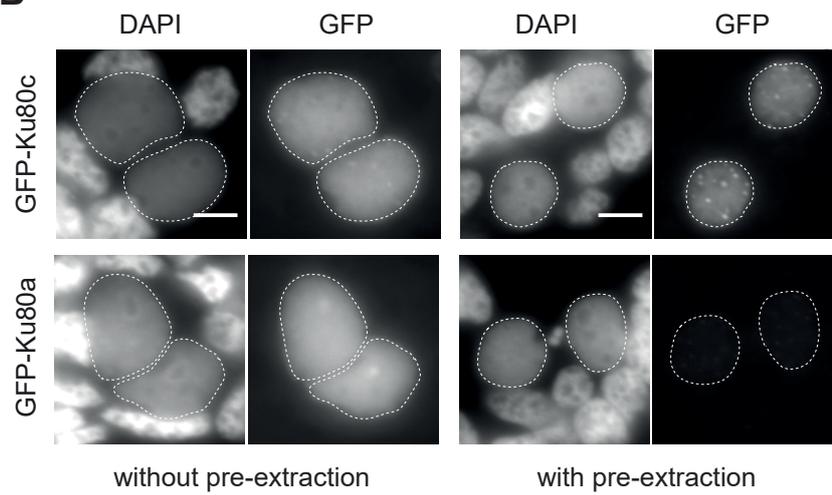


Figure 4-figure supplement 4

A



B



C

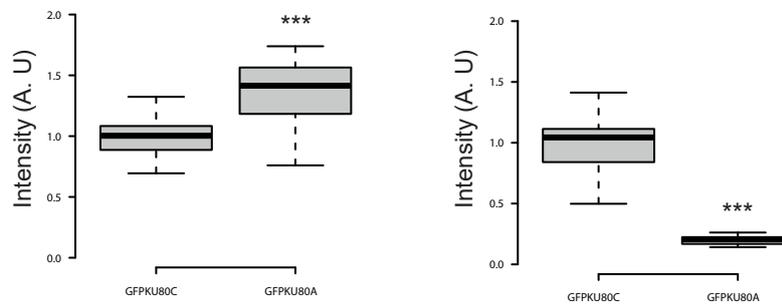
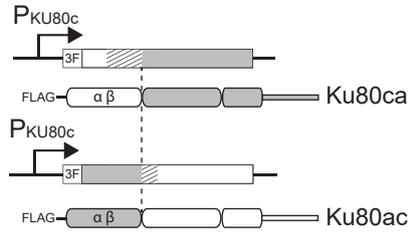
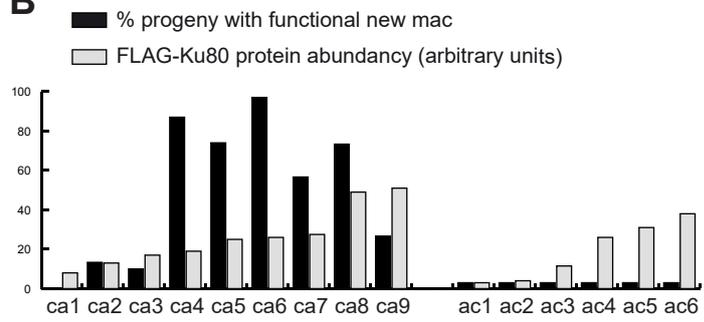


Figure 5

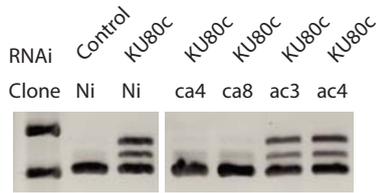
A



B



C



D

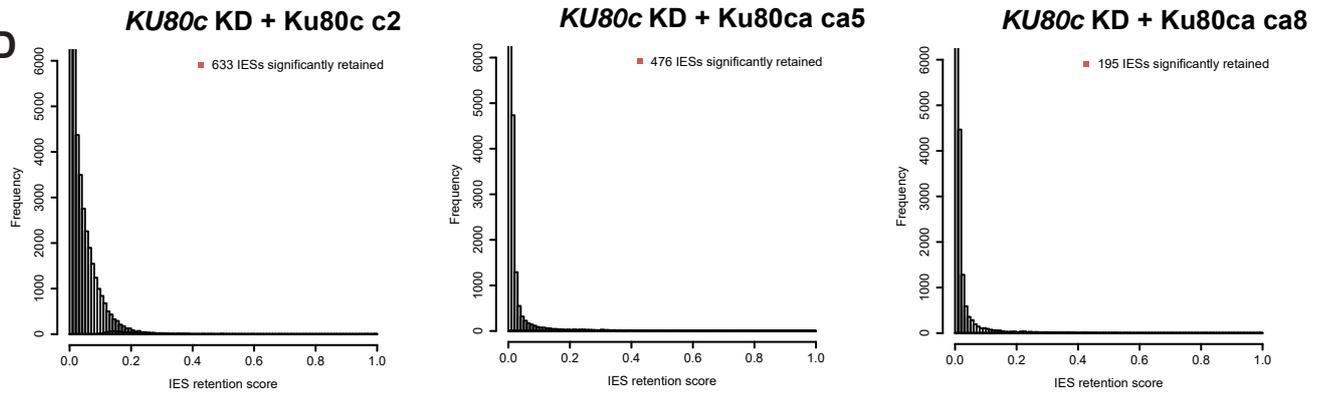


Figure 5-figure supplement 1

Reference sequence (1): PJENN.M.1.P01460069
 Identities normalised by aligned length.
 Colored by: property

	cov	pid	1	80
1 PJENN.M.1.P01460069	100.0%	100.0%	MSGKEASLILLDVGASMYGEYQMGG-GKKLSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	MSGKEASLILLDVGASMYGEYQMGG-GKKLSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	MSGKEATLILLDVGASMYSEYQGGG-SKKLTRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	MSGREATLILLDVGASMYQYQGGG-NRKRMSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	MSGKEATLILLDVGASMYQYQGGG-SRKRMSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	MSGKEATLILLDVGASMYQYQGGG-NRKRMSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
7 PTET.51.3.P0520142	99.9%	91.7%	MSGKEATLILLDVGASMYQYQGGG-SKKLSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
8 POCTA.138.1.P0410138	99.9%	92.0%	MSGKEATLILLDVGASMYQYQGGG-SKKLSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	MSGKEATLILLDVGASMYQYQGGG-SKKLSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
10 PDEC.223.1.P00110119	99.9%	91.9%	MSGKEATLILLDVGASMYQYQGGG-SKKLSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
11 PDODEC.274.1.P00070138	99.9%	91.7%	MSGKEATLILLDVGASMYQYQGGG-SKKLSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	MSGKEATLILLDVGASMYQYQGGG-NRKRMSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
13 PPENT.87.1.P1050145	99.9%	91.5%	MSGREATLILLDVGASMYQYQGGG-NRKRMSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	MSGREATLILLDVGASMYQYQGGG-NRKRMSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
15 PPRIMP30731	99.9%	91.5%	MSGREATLILLDVGASMYQYQGGG-NRKRMSRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
16 PDEC.223.1.P02560030	100.0%	74.1%	MAGKEATLVLLDVGASMYEPYKQAO-GKNITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
17 PDODEC.274.1.P02610024	100.0%	74.1%	MAGKEATLVLLDVGASMYEPYKQAO-GKNITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
18 PTET.51.3.P0760120	100.0%	73.0%	MAGKEATLVLLDVGASMYEPYKQAO-GKNITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
19 POCTA.138.1.P0780138	100.0%	74.1%	MAGKEATLVLLDVGASMYEPYKQAO-GKNITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	MAGKEATLVLLDVGASMYEPYKQAO-GKNITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
23 PPRIMP29158	99.9%	75.1%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
24 PPENT.87.1.P1460087	99.9%	74.8%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	MAGKEATLVLLDVGASMYQYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
27 PTET.51.3.P1590130	100.0%	74.0%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
28 POCTA.138.1.P1560084	100.0%	74.3%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
30 PDODEC.274.1.P02600002	100.0%	74.0%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
31 PDEC.223.1.P02830051	100.0%	74.3%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
32 PNOV.TE.1.P02560002	100.0%	74.5%	MAGKEATLVLLDVGASMYEPYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	MAGKEATLVLLDVGASMYQYKQAO-GKKITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	MAGKEATLVLLDVGASMYEPYKQAO-GKEITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	MAGKEATLVLLDVGSSMYEKYQGGG-GKEITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
36 PJENN.M.1.P02270020	99.9%	72.6%	MAGKEATLVLLDVGASMYQYKQAO-GKQITRLELAVDCLGLMIQQRIFNYKNHEVGLLILFGTEEPDGKTLVIQDLSIP	
consensus/100%			Muu+EAoLLILLDVGASMYt.Y..st.s+phpRLLELAVDCLGHMIQQRIFNYKNHEVGLLILFGTE-A.DGpThyIQ.hs.P	
consensus/90%			MuG+EAATLILLDVGASMYt.YpQut.skploRLELAVDCLGHMIQQRIFNYKNHEVGLLILFGTE-A.DGpThyIQ.hs.P	
consensus/80%			MuGKEATLILLDVGASMYt.YpQut.ukcloRLELAVDCLGHMIQQRIFNYKNHEVGLLILFGTE-A.DGpThyIQsLS.P	
consensus/70%			MuGKEATLILLDVGASMYp.YpQut.ukkloRLELAVDCLGHMIQQRIFNYKNHEVGLLILFGTEEA.DGpThyIQsLS.P	

	cov	pid	81	160
1 PJENN.M.1.P01460069	100.0%	100.0%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHVLDDHTKTKRMEKKIYLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHVLDDHTKTKRMEKKIYLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
7 PTET.51.3.P0520142	99.9%	91.7%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
8 POCTA.138.1.P0410138	99.9%	92.0%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
10 PDEC.223.1.P00110119	99.9%	91.9%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
11 PDODEC.274.1.P00070138	99.9%	91.7%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
13 PPENT.87.1.P1050145	99.9%	91.5%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
15 PPRIMP30731	99.9%	91.5%	DLDFFRNVSELPNHDVGLVGGDIFDALDKAVHALDDHAKTKRMEKKIFLLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
16 PDEC.223.1.P02560030	100.0%	74.1%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
17 PDODEC.274.1.P02610024	100.0%	74.1%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
18 PTET.51.3.P0760120	100.0%	73.0%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
19 POCTA.138.1.P0780138	100.0%	74.1%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
23 PPRIMP29158	99.9%	75.1%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
24 PPENT.87.1.P1460087	99.9%	74.8%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	NLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
27 PTET.51.3.P1590130	100.0%	74.0%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
28 POCTA.138.1.P1560084	100.0%	74.3%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
30 PDODEC.274.1.P02600002	100.0%	74.0%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
31 PDEC.223.1.P02830051	100.0%	74.3%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
32 PNOV.TE.1.P02560002	100.0%	74.5%	NLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	NLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDEHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDDHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDDHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
36 PJENN.M.1.P02270020	99.9%	72.6%	DLEFYRNLTSLPNDVQVGGDIFDALDKAVSLLDDHVKTKRMEKKIFVLTAGCGQTDYSEKQITKLIKMIKVDVKIN	
consensus/100%			sh-Farnls-L.pdhhs...GGDIFDALDKVSLDtasKs+h-KKlallTuGhGpTtYSEK.IsKLIKMIKVDVKIN	
consensus/90%			DL-Farnlo-LssHDhsp...GGDIFDALDKVSLD-asKTKKk-KKIFLLTAGHGQ-TYSEKpIsKLIKMIKVDVKIN	
consensus/80%			DL-Farnlo-LsNHdLsp.LGGDIFDALDKVSLD-HsKTKKk-KKIFLLTAGHGQ-TYSEKpIsKLIKMIKVDVKIN	
consensus/70%			DL-Farnlo-LPNHDLsp.LGGDIFDALDKVSLD-HsKTKKk-KKIFLLTAGHGQ-TYSEKpIsKLIKMIKVDVKIN	

	cov	pid	161	240
1 PJENN.M.1.P01460069	100.0%	100.0%	FIALDFMNDYNGMDDPDKPEEYEAALNRMLTASVYCCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA	
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	FIALDFMNDYNGMDDPDKPEEYEAALNRMLTASVYCCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA	

Figure 5-figure supplement 1

3	PSEX.AZ8_4.1.P0530077	99.9%	93.2%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
4	PQUADEC.Nia.1.P00940024	99.9%	91.2%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
5	PBIA.V1_4.1.P00510147	99.9%	91.1%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
6	PTRED.209.2.P71800001294210147	99.9%	90.8%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
7	PTET.51.3.P0520142	99.9%	91.7%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
8	POCTA.138.1.P0410138	99.9%	92.0%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
9	POCT.K8.1.P71800002769600152	99.9%	92.0%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
10	PDEC.223.1.P00110119	99.9%	91.9%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
11	PDODEC.274.1.P00070138	99.9%	91.7%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
12	PBIA.V1_4.1.P02420019	99.9%	90.8%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
13	PPENT.87.1.P1050145	99.9%	91.5%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
14	PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
15	PPRIMP30731	99.9%	91.5%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
16	PDEC.223.1.P02560030	100.0%	74.1%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
17	PDODEC.274.1.P02610024	100.0%	74.1%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
18	PTET.51.3.P0760120	100.0%	73.0%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
19	POCTA.138.1.P0780138	100.0%	74.1%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
20	POCT.K8.1.P71800002769920120	100.0%	73.8%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
21	PSEX.AZ8_4.1.P1460081	99.9%	74.5%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
22	PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
23	PPRIMP29158	99.9%	75.1%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
24	PPENT.87.1.P1460087	99.9%	74.8%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
25	PBIA.V1_4.1.P00880090	100.0%	75.1%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
26	PQUADEC.Nia.1.P01970002	100.0%	74.5%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
27	PTET.51.3.P1590130	100.0%	74.0%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
28	POCTA.138.1.P1560084	100.0%	74.3%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
29	POCT.K8.1.P71800002770710083	100.0%	74.2%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
30	PDODEC.274.1.P02600002	100.0%	74.0%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
31	PDEC.223.1.P02830051	100.0%	74.3%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
32	PNOV.TE.1.P02560002	100.0%	74.5%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
33	PTRED.209.2.P71800001294370184	100.0%	74.7%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
34	PSON.ATCC_30995.1.P1480025	99.9%	72.9%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
35	PSON.ATCC_30995.1.P1750030	99.9%	72.4%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
36	PJENN.M.1.P02270020	99.9%	72.6%	FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
consensus/100%				FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
consensus/90%				FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
consensus/80%				FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA
consensus/70%				FIALDFMNDYNGMDDPDKPEEFALNNRMLTASYQCQEQSINSRYVFLMVQELRNNMRIFPANVAFELYSQFHTRSLQA

1	PJENN.M.1.P01460069	100.0%	100.0%	cov	pid 241	:	:	:	3	:	320
2	PSON.ATCC_30995.1.P1100174	100.0%	98.2%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
3	PSEX.AZ8_4.1.P0530077	99.9%	93.2%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
4	PQUADEC.Nia.1.P00940024	99.9%	91.2%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
5	PBIA.V1_4.1.P00510147	99.9%	91.1%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
6	PTRED.209.2.P71800001294210147	99.9%	90.8%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
7	PTET.51.3.P0520142	99.9%	91.7%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
8	POCTA.138.1.P0410138	99.9%	92.0%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
9	POCT.K8.1.P71800002769600152	99.9%	92.0%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
10	PDEC.223.1.P00110119	99.9%	91.9%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
11	PDODEC.274.1.P00070138	99.9%	91.7%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
12	PBIA.V1_4.1.P02420019	99.9%	90.8%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
13	PPENT.87.1.P1050145	99.9%	91.5%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
14	PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
15	PPRIMP30731	99.9%	91.5%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
16	PDEC.223.1.P02560030	100.0%	74.1%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
17	PDODEC.274.1.P02610024	100.0%	74.1%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
18	PTET.51.3.P0760120	100.0%	73.0%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
19	POCTA.138.1.P0780138	100.0%	74.1%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
20	POCT.K8.1.P71800002769920120	100.0%	73.8%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
21	PSEX.AZ8_4.1.P1460081	99.9%	74.5%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
22	PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
23	PPRIMP29158	99.9%	75.1%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
24	PPENT.87.1.P1460087	99.9%	74.8%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
25	PBIA.V1_4.1.P00880090	100.0%	75.1%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
26	PQUADEC.Nia.1.P01970002	100.0%	74.5%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
27	PTET.51.3.P1590130	100.0%	74.0%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
28	POCTA.138.1.P1560084	100.0%	74.3%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
29	POCT.K8.1.P71800002770710083	100.0%	74.2%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
30	PDODEC.274.1.P02600002	100.0%	74.0%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
31	PDEC.223.1.P02830051	100.0%	74.3%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
32	PNOV.TE.1.P02560002	100.0%	74.5%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
33	PTRED.209.2.P71800001294370184	100.0%	74.7%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
34	PSON.ATCC_30995.1.P1480025	99.9%	72.9%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
35	PSON.ATCC_30995.1.P1750030	99.9%	72.4%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
36	PJENN.M.1.P02270020	99.9%	72.6%	RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
consensus/100%				RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
consensus/90%				RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
consensus/80%				RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							
consensus/70%				RASFRGDFQINDEISIQVLIYKRCFEERLPLKXKHSILGQFQIDINKNHVRNDLIYYPEDPMTPIEKENIIRGQYGR							

1	PJENN.M.1.P01460069	100.0%	100.0%	cov	pid 321	:	:	:	4	400
2	PSON.ATCC_30995.1.P1100174	100.0%	98.2%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
3	PSEX.AZ8_4.1.P0530077	99.9%	93.2%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
4	PQUADEC.Nia.1.P00940024	99.9%	91.2%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
5	PBIA.V1_4.1.P00510147	99.9%	91.1%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
6	PTRED.209.2.P71800001294210147	99.9%	90.8%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
7	PTET.51.3.P0520142	99.9%	91.7%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
8	POCTA.138.1.P0410138	99.9%	92.0%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
9	POCT.K8.1.P71800002769600152	99.9%	92.0%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
10	PDEC.223.1.P00110119	99.9%	91.9%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
11	PDODEC.274.1.P00070138	99.9%	91.7%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						
12	PBIA.V1_4.1.P02420019	99.9%	90.8%	NLVPVQIMEDMKYQCPQFQLLGFVDRSHIPRYYTSTVDMVIAVENQKQKALAALVIALIARVAVIARFVGREKT						

Figure 5-figure supplement 1

Accession	Cov	Pid	Sequence
13 PPENT.87.1.P1050145	99.9%	91.5%	NLVPVDDIMEDRMKYQCPRQFQLLGFVDRSHIPRYYYISTVDMVIAVENQKQKQKALAALVIALIATRKVAIAIRFVGREKT
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	NLVPVDDIMEDRMKYQCPRQFQLLGFVDRSHIPRYYYISTVDMVIAVENQKQKQKALAALVIALIATRKVAIAIRFVGREKT
15 PPRIMP30731	99.9%	91.5%	NLVPVDDIMEDRMKYQCPRQFQLLGFVDRSHIPRYYYISTVDMVIAVENQKQKQKALAALVIALIATRKVAIAIRFVGREKT
16 PDEC.223.1.P02560030	100.0%	74.1%	NLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
17 PPODEC.274.1.P02610024	100.0%	74.1%	NLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
18 PTET.51.3.P0760120	100.0%	73.0%	NLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGRQKS
19 POCTA.138.1.P0780138	100.0%	74.1%	NLIPVDSLMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	NLIPVDSLMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	NLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	NLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREK
23 PPRIMP29158	99.9%	75.1%	NLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREK
24 PPENT.87.1.P1460087	99.9%	74.8%	NLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREK
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	NLIPVGLMEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREK
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
27 PTET.51.3.P1590130	100.0%	74.0%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
28 POCTA.138.1.P1560084	100.0%	74.3%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
30 PPODEC.274.1.P02600002	100.0%	74.0%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
31 PDEC.223.1.P02830051	100.0%	74.3%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
32 PNOV.TE.1.P02560002	100.0%	74.5%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREK
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKA
36 PJENN.M.1.P02270020	99.9%	72.6%	SLIPVDSIMEEDRMKYQCSRSPQLLGFVDRSQIPRHYFMFNVDMVIAIDCEKARKSLALI IALIA TKKVAIAIRFVGREKS
consensus/100%			SLIPVDTIM--MKYQC.R.FQLLGFV-+.PIPR.Yah.sVdMlul-spptpKsLuulIuLIAT+KlslARFVGRp+s
consensus/90%			SLIPVDPIM--MKYQCsrpQLLGFV+opIPRyah.sVdMlul-spKtpKuLuulIuLIAT+KVALARFVGREKS
consensus/80%			SLIPVDPIMEDRMKYQCSRSPQLLGFV+SpIPRyahSsVdMlul-sPKtpKuLuulIuLIAT+KVAIAIRFVGREKO
consensus/70%			SLIPVDPIMEDRMKYQCSRSPQLLGFV+SpIPRyahSsVdMlul-sPKtpKuLuulIuLIAT+KVAIAIRFVGREKO
cov pid 401			
1 PJENN.M.1.P01460069	100.0%	100.0%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
7 PTET.51.3.P0520142	99.9%	91.7%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
8 POCTA.138.1.P0410138	99.9%	92.0%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
10 PDEC.223.1.P00110119	99.9%	91.9%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
11 PPODEC.274.1.P00070138	99.9%	91.7%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
13 PPENT.87.1.P1050145	99.9%	91.5%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
15 PPRIMP30731	99.9%	91.5%	APKLIIMLLPHKSKNSQCFWMI SLPTTEDIRHFQF AALKRSTPNQOMAVASAMIDSMLEKMPTEDEGQFELLRMKYVANPT
16 PDEC.223.1.P02560030	100.0%	74.1%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
17 PPODEC.274.1.P02610024	100.0%	74.1%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
18 PTET.51.3.P0760120	100.0%	73.0%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
19 POCTA.138.1.P0780138	100.0%	74.1%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
23 PPRIMP29158	99.9%	75.1%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
24 PPENT.87.1.P1460087	99.9%	74.8%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
27 PTET.51.3.P1590130	100.0%	74.0%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
28 POCTA.138.1.P1560084	100.0%	74.3%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
30 PPODEC.274.1.P02600002	100.0%	74.0%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
31 PDEC.223.1.P02830051	100.0%	74.3%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
32 PNOV.TE.1.P02560002	100.0%	74.5%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
36 PJENN.M.1.P02270020	99.9%	72.6%	SPKLMVLLPHKSKNSQCFWMI SLPTSEDIRHFQF AALKRSTPNQOAVASALIDKMDLELPNQSGBEELLRMKYIANPT
consensus/100%			uPKhhllLPHKSKS.c.PFWMI SLPTtoEDIRHFQFuAL++SsP.QQhAVuuhIDpMLePhPspGp.EELLRMKYIANPT
consensus/90%			uPKhhllLPHKSKS.c.PFWMI SLPTtoEDIRHFQFuAL++SsP.QQhAVuuhIDpMLePhPspGp.EELLRMKYIANPT
consensus/80%			uPKhhllLPHKSKS.c.PFWMI SLPTtoEDIRHFQFuAL++SsP.QQhAVuuhIDpMLePhPspGp.EELLRMKYIANPT
consensus/70%			uPKhhllLPHKSKS.c.PFWMI SLPTtoEDIRHFQFuAL++SsP.QQhAVuuhIDpMLePhPspGp.EELLRMKYIANPT
cov pid 481			
1 PJENN.M.1.P01460069	100.0%	100.0%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
7 PTET.51.3.P0520142	99.9%	91.7%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
8 POCTA.138.1.P0410138	99.9%	92.0%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
10 PDEC.223.1.P00110119	99.9%	91.9%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
11 PPODEC.274.1.P00070138	99.9%	91.7%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
13 PPENT.87.1.P1050145	99.9%	91.5%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
15 PPRIMP30731	99.9%	91.5%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
16 PDEC.223.1.P02560030	100.0%	74.1%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
17 PPODEC.274.1.P02610024	100.0%	74.1%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
18 PTET.51.3.P0760120	100.0%	73.0%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
19 POCTA.138.1.P0780138	100.0%	74.1%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	RQYFQQVIMHKAITRSDVLPPI SPLILEYLHPEIRVYNYAKEALQKIAAFAKFKVNEIKKQDKKVFWKQLFEEQST--TQ

Figure 5-figure supplement 1

Accession	Cov	Pid	Sequence
23 PPRIMP29158	99.9%	75.1%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
24 PPENT.87.1.P1460087	99.9%	74.8%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTTQQQ
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
27 PTET.51.3.P1590130	100.0%	74.0%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTTQQQ
28 POCTA.138.1.P1560084	100.0%	74.3%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTTQQQ
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTTQQQ
30 PPODEC.274.1.P02600002	100.0%	74.0%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTTQQQ
31 PDEC.223.1.P02830051	100.0%	74.3%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTTQQQ
32 PNOV.TE.1.P02560002	100.0%	74.5%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
36 PJENN.M.1.P02270020	99.9%	72.6%	RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
consensus/100%			RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
consensus/90%			RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
consensus/80%			RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT--Q
consensus/70%			RQYFQQVVMHKAITRTDVIPIISPLILEYLHPEQRVYNYAQDAIQRVKNAFKFKVNEIKKQDDKQVFWKQLFDEQTT..p
cov pid 561			6
1 PJENN.M.1.P01460069	100.0%	100.0%	QI---QPDDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	QI---QPDDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	QI---QPDDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	QV---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	QV---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	QI---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
7 PTET.51.3.P0520142	99.9%	91.7%	QI---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
8 POCTA.138.1.P0410138	99.9%	92.0%	QI---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	QI---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
10 PDEC.223.1.P00110119	99.9%	91.9%	QI---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
11 PPODEC.274.1.P00070138	99.9%	91.7%	QI---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	QV---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
13 PPENT.87.1.P1050145	99.9%	91.5%	QV---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	QV---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
15 PPRIMP30731	99.9%	91.5%	QV---QQQVEDEVVEINQEEEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
16 PDEC.223.1.P02560030	100.0%	74.1%	QV---PEQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
17 PPODEC.274.1.P02610024	100.0%	74.1%	QV---PEQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
18 PTET.51.3.P0760120	100.0%	73.0%	QV---PEQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
19 POCTA.138.1.P0780138	100.0%	74.1%	QI---PQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	QV---PQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
23 PPRIMP29158	99.9%	75.1%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
24 PPENT.87.1.P1460087	99.9%	74.8%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	QI---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
27 PTET.51.3.P1590130	100.0%	74.0%	QA---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
28 POCTA.138.1.P1560084	100.0%	74.3%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
30 PPODEC.274.1.P02600002	100.0%	74.0%	QI---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
31 PDEC.223.1.P02830051	100.0%	74.3%	QI---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
32 PNOV.TE.1.P02560002	100.0%	74.5%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
36 PJENN.M.1.P02270020	99.9%	72.6%	QV---QQQIEEEVVEINREEEEMVSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDLVDVTDALQQIQKVIISFVDSQL
consensus/100%			Qh... .p-ElVInpEEEEEMSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDL.VDQALQQIQKVIISFVDSQL
consensus/90%			Qh... .p1l-ElVInpEEEEEMSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDL.VDQALQQIQKVIISFVDSQL
consensus/80%			Ql... .Q00lE-EVVEInpEEEEEMSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDL.VDQALQQIQKVIISFVDSQL
consensus/70%			Ql... .Q00lE-EVVEInpEEEEEMSMFAKQKLGFNDDIIEKEIGTVDPISDFKMMITEKRIDL.VDQALQQIQKVIISFVDSQL
cov pid 641			7
1 PJENN.M.1.P01460069	100.0%	100.0%	KGSFYFPALECLKEMRRACISEDETVPFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	KGSFYFPALECLKEMRRACISEDETVPFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
7 PTET.51.3.P0520142	99.9%	91.7%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
8 POCTA.138.1.P0410138	99.9%	92.0%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
10 PDEC.223.1.P00110119	99.9%	91.9%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
11 PPODEC.274.1.P00070138	99.9%	91.7%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
13 PPENT.87.1.P1050145	99.9%	91.5%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
15 PPRIMP30731	99.9%	91.5%	KGSFYFPALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
16 PDEC.223.1.P02560030	100.0%	74.1%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
17 PPODEC.274.1.P02610024	100.0%	74.1%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
18 PTET.51.3.P0760120	100.0%	73.0%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
19 POCTA.138.1.P0780138	100.0%	74.1%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
23 PPRIMP29158	99.9%	75.1%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
24 PPENT.87.1.P1460087	99.9%	74.8%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
27 PTET.51.3.P1590130	100.0%	74.0%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
28 POCTA.138.1.P1560084	100.0%	74.3%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
30 PPODEC.274.1.P02600002	100.0%	74.0%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
31 PDEC.223.1.P02830051	100.0%	74.3%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI
32 PNOV.TE.1.P02560002	100.0%	74.5%	KGSFFPFALECLKEMRRACISEDEAPVFNKYLHLKDKYSSQLVFWAQIVQGGIILISNENQKSNVSADEAQEFLNKEDI

Figure 5-figure supplement 1

Accession	Cov	Pid	721	800
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	KG S FFP N AM C LKEM R KACISEDEAPVFNKFLFVLKDKYNQSMF W AQIVQGGI L IS K NIENQK S TVS A EEAQDFLNKEDN	
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	KG S FFP N AL C LKEM R KACISEDEVPVFNKFLFVLKDKYNQ Q V F WAQIVQGGI L IS N NIENQK S SVT A DEAQDFLNKEDN	
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	KG S FFP N AL C LKEM R KACISEDEAPIFNKFLFVLKDKYNQ Q IF F WAQIVQGGI L IS N NIENQK S SVT A DEAQDFLNKEDN	
36 PJENN.M.1.P02270020	99.9%	72.6%	KG S FFP N AL C LKEM R KACISEDEAPVFNKFLFVLKDKYNQ Q IF F WAQIVQGGI L IS N NIENQK S SVT A EEAQDFLNKEDN	
consensus/100%			KG S aaP R Ah C LKE H R + ACIOEDESs L FNKaL.hLK-KasQ.hFWAp I VQGGI L ISp L EN.kstlss--AQ-FLNKED.	
consensus/90%			KG S FaP N AL C LKEM R + ACIOEDES S PVFNKaLa L LK-KYSQ.hFWAQIVQGGI L IS S NIENQK S Sp L os-EAQ-FLNKED.	
consensus/80%			KG S FaP N AL C LKEM R KACIOEDES S PVFNKaLa V LK-KYSQ.lFWAQIVQGGI L IS N NIENQK S S V oa-EAQ-FLNKED.	
consensus/70%			KG S FaP N AL C LKEM R KACIOE D EAPVFNKaLa V LK-KYSQ.lFWAQIVQGGI L IS N NIENQK S S V oa-EAQ-FLNKED.	
	cov	pid	721	800
1 PJENN.M.1.P01460069	100.0%	100.0%	SHKQLVDQLQHEEEDLLA E ID-----	
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	SHKQLVDQLQHEEEDLLA E ID-----	
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	NHRQLVDQLQHEEEDLLA E ID-----	
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	SHKQLVDQLQHEEEDLLA E ID-----	
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	SHKQLVDQLQHEEEDLLA E ID-----	
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	SHKQLVDQLQHEEEDLLA E ID-----	
7 PTET.51.3.P0520142	99.9%	91.7%	SHKQLVDQLQHEEEDLLA E ID-----	
8 POCTA.138.1.P0410138	99.9%	92.0%	SHKQLVDQLQHEEEDLLA E ID-----	
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	SHKQLVDQLQHEEEDLLA E ID-----	
10 PDEC.223.1.P00110119	99.9%	91.9%	SHKQLVDQLQHEEEDLLA E ID-----	
11 PDODEC.274.1.P00070138	99.9%	91.7%	SHKQLVDQLQHEEEDLLA E ID-----	
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	SHKQLVDQLQHEEEDLLA E ID-----	
13 PPENT.87.1.P1050145	99.9%	91.5%	SHKQLVDQLQHEEEDLLA E ID-----	
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	SHKQLVDQLQHEEEDLLA E ID-----	
15 PPRIMP30731	99.9%	91.5%	SHKQLVDQLQHEEEDLLA E ID-----	
16 PDEC.223.1.P02560030	100.0%	74.1%	KHQQMV D QLQNEEEDLLA D IE-----	
17 PDODEC.274.1.P02610024	100.0%	74.1%	KHQQMV D QLQNEEEDLLA D IE-----	
18 PTET.51.3.P0760120	100.0%	73.0%	KHQQMV D QLQNEEEDLLA D IE-----	
19 POCTA.138.1.P0780138	100.0%	74.1%	KHQQMV D QLQNEEEDLLA D IE-----	
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	KHQQMV D QLQNEEEDLLA D IE-----	
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	KHQQMV D QLQVEEEDLLA D IE-----	
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	KHQQMV D QLQHEEEDLLA D IE-----	
23 PPRIMP29158	99.9%	75.1%	KHQQMV D QLQHEEEDLLA D IE-----	
24 PPENT.87.1.P1460087	99.9%	74.8%	KHQQMV D QLQHEEEDLLA D IE-----	
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	KHQQMV D QLQHEEEDLLA D IE-----	
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	KHQQMV D QLQHEDEDLLA D IE-----	
27 PTET.51.3.P1590130	100.0%	74.0%	KHQQMV D QLQHEEEDLLA D IE-----	
28 POCTA.138.1.P1560084	100.0%	74.3%	KHQQMV D QLQHEEEDLLA D IE-----	
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	KHQQMV D QLQHEDEDLLA D IE-----	
30 PDODEC.274.1.P02600002	100.0%	74.0%	KHQQMV D QLQHEEEDLLA D IE-----	
31 PDEC.223.1.P02830051	100.0%	74.3%	KHQQMV D QLQHEEEDLLA D IE-----	
32 PNOV.TE.1.P02560002	100.0%	74.5%	KHQQMV D QLQHEEEDLLA D IE-----	
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	KHQQMV D QLQHEEEDLLA D IE-----	
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	KHQQMV D QLQNEEEDLLA D IE-----	
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	KHQQMV D TLQHEEEDLLA D IE-----	
36 PJENN.M.1.P02270020	99.9%	72.6%	KHQQMV D TLQHEEEDLLA D IE-----	
consensus/100%			pHpphV D pLQ.E-EDLLA-I-----	
consensus/90%			pHpqhV D QLQpEEEDLLA-I-----	
consensus/80%			pHpqhV D QLQaEEEDLLA-I-----	
consensus/70%			pHpqhV D QLQHEEEDLLA-I-----	
	cov	pid	801	813
1 PJENN.M.1.P01460069	100.0%	100.0%	-----	
2 PSON.ATCC_30995.1.P1100174	100.0%	98.2%	-----	
3 PSEX.AZ8_4.1.P0530077	99.9%	93.2%	-----	
4 PQUADEC.NiA.1.P00940024	99.9%	91.2%	-----	
5 PBIA.V1_4.1.P00510147	99.9%	91.1%	-----	
6 PTRED.209.2.P71800001294210147	99.9%	90.8%	-----	
7 PTET.51.3.P0520142	99.9%	91.7%	-----	
8 POCTA.138.1.P0410138	99.9%	92.0%	-----	
9 POCT.K8.1.P71800002769600152	99.9%	92.0%	-----	
10 PDEC.223.1.P00110119	99.9%	91.9%	-----	
11 PDODEC.274.1.P00070138	99.9%	91.7%	-----	
12 PBIA.V1_4.1.P02420019	99.9%	90.8%	-----	
13 PPENT.87.1.P1050145	99.9%	91.5%	-----	
14 PPRIM.AZ9-3.1.P0970147	99.9%	91.5%	-----	
15 PPRIMP30731	99.9%	91.5%	-----	
16 PDEC.223.1.P02560030	100.0%	74.1%	-----	
17 PDODEC.274.1.P02610024	100.0%	74.1%	-----	
18 PTET.51.3.P0760120	100.0%	73.0%	-----	
19 POCTA.138.1.P0780138	100.0%	74.1%	-----	
20 POCT.K8.1.P71800002769920120	100.0%	73.8%	-----	
21 PSEX.AZ8_4.1.P1460081	99.9%	74.5%	-----	
22 PPRIM.AZ9-3.1.P1360092	99.9%	75.1%	-----	
23 PPRIMP29158	99.9%	75.1%	-----	
24 PPENT.87.1.P1460087	99.9%	74.8%	-----	
25 PBIA.V1_4.1.P00880090	100.0%	75.1%	-----	
26 PQUADEC.NiA.1.P01970002	100.0%	74.5%	-----	
27 PTET.51.3.P1590130	100.0%	74.0%	-----	
28 POCTA.138.1.P1560084	100.0%	74.3%	-----	
29 POCT.K8.1.P71800002770710083	100.0%	74.2%	-----	
30 PDODEC.274.1.P02600002	100.0%	74.0%	-----	
31 PDEC.223.1.P02830051	100.0%	74.3%	-----	
32 PNOV.TE.1.P02560002	100.0%	74.5%	-----	
33 PTRED.209.2.P71800001294370184	100.0%	74.7%	-----	
34 PSON.ATCC_30995.1.P1480025	99.9%	72.9%	-----	
35 PSON.ATCC_30995.1.P1750030	99.9%	72.4%	-----	
36 PJENN.M.1.P02270020	99.9%	72.6%	-----	
consensus/100%			-----	
consensus/90%			-----	
consensus/80%			-----	
consensus/70%			-----	

Figure 6

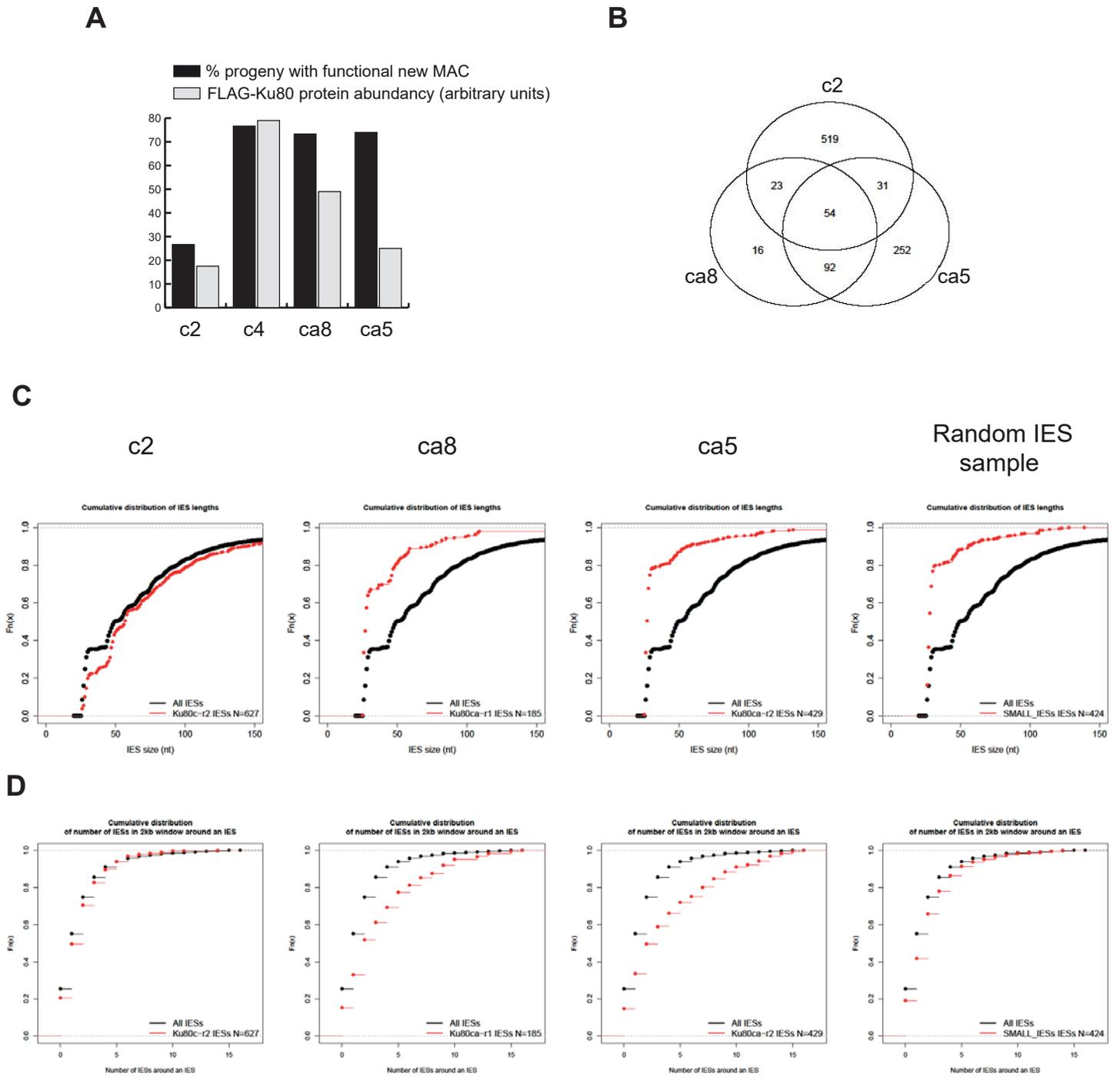


Table S1

date expérience	injection	niveau Inj	Feeding	canyionide	Survie	WB04-10-18	WB09-10-18	WB28-11-18	Moyenne WB	Nom clone
juil-18	Ni	0	DNAPKcs	5c1	0			0	0	Ni5c1
sept-18	A	71	KU80	6c3	0	16		16	16	InjA3
sept-18	A	62	KU80	5a1	0	19		19	19	InjA10
sept-18	A	44	KU80	5a1	0	24	28	26	26	InjA9
sept-18	A	72	KU80	6c3	0	46		46	46	InjA5
sept-18	A	100	KU80	6c3	0	82	88	85	85	InjA7
sept-18	AC	16	KU80	6c3	0		3	3	3	InjAC3
sept-18	AC	37	KU80	6c4	0		4	4	4	InjAC8
sept-18	AC	56	KU80	6c4	0	15	8	11.5	11.5	InjAC6
sept-18	AC	132	KU80	5a1	0	30	22	26	26	InjAC4
sept-18	AC	213	KU80	6c3	0	31	31	31	31	InjAC1
sept-18	AC	428	KU80	5a1	0	38	38	38	38	InjAC5
sept-18	C	13	KU80	5a1	0	7		7	7	InjC7
nov-18	C	20	KU80	1b1	0			8	8	C2-Ku80
sept-18	C	37	KU80	6c3	26.6666667	18	17	17.5	17.5	InjC4
sept-18	C	120	KU80	6c3	76.6666667	80	79	78	79	InjC3
nov-18	C	800	KU80	1b1	100			80	80	C5-Ku80
nov-18	C	930	KU80	1b1	100			109	109	C6-Ku80
nov-18	C	1415	KU80	2c1	87			294	294	C14-Ku80
nov-18	CA	43	KU80	1b1	0			8	8	CA2-Ku80
sept-18	CA	13	KU80	6c3	13.3333333	16	10	13	13	InjCA3
sept-18	CA	43	KU80	5a1	10		17	17	17	InjCA10
nov-18	CA	135	KU80	2c1	87			19	19	CA7-Ku80
nov-18	CA	175	KU80	1b1	74			25	25	CA6-Ku80
nov-18	CA	284	KU80	1b1	97			26	26	CA1-Ku80
sept-18	CA	38	KU80	5a1	56.6666667	35	20	27.5	27.5	InjCA9
sept-18	CA	77	KU80	5a1	73.3333333	55	43	49	49	InjCA11
sept-18	CA	316	KU80	5a1	26.6666667		51	51	51	InjCA8
sept-18	Ni	0	KU80	5a1	0			0	0	Ni5a1
sept-18	Ni	0	KU80	6c3	0			0	0	Ni6c3
sept-18	Ni	0	KU80	6c4	0			0	0	Ni6c4
juil-18	Ni	0	KU80	5c1	0			0	0	Ni5c1
nov-18	C	20	L4400	1b1	97			8	8	C2-L4440
nov-18	C	800	L4400	1b1	100			80	80	C5-L4440
nov-18	C	930	L4400	1b1	100			109	109	C6-KL4440
nov-18	C	1415	L4400	2c1	33			294	294	C14-L4440
nov-18	CA	43	L4400	1b1	100			8	8	CA2-L4440
nov-18	CA	135	L4400	2c1	100			19	19	CA7-L4440
nov-18	CA	175	L4400	1b1	93			25	25	CA6-L4440
nov-18	CA	284	L4400	1b1	90			26	26	CA1-L4440

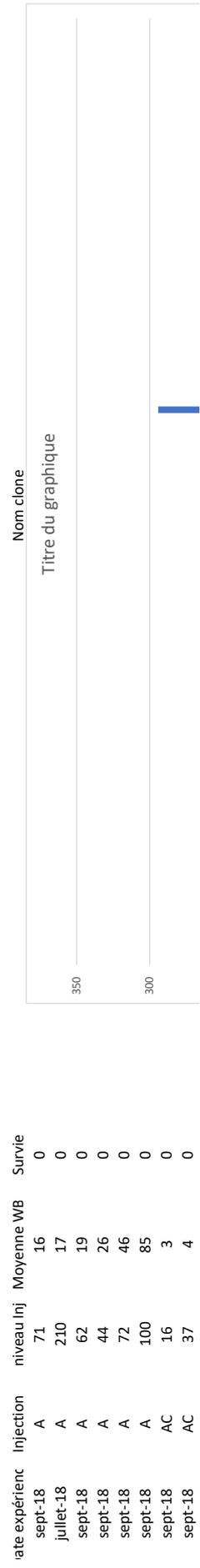


Table S1

sept-18	CA7	38	27.5	56.6666667
sept-18	CA8	77	49	73.3333333
sept-18	CA9	316	51	26.6666667
sept-18	AC1	16	3	0
sept-18	AC2	37	4	0
sept-18	AC3	56	11.5	0
sept-18	AC4	132	26	0
sept-18	AC5	213	31	0
sept-18	AC6	428	38	0

Bibliographie

- Ahnesorg, P., Smith, P., and Jackson, S. P. *XLFI interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining*. *Cell*, 124(2) :301–313 (2006). doi :10.1016/j.cell.2005.12.031. (Cité page 21.)
- Allen, S. E., Hug, I., Pabian, S., Rzeszutek, I., Hoehener, C., and Nowacki, M. *Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs*. *Cell*, 168(6) :990–999.e7 (2017). doi :10.1016/j.cell.2017.02.020. (Cité pages 43 et 59.)
- Andres, S. N., Appel, C. D., Westmoreland, J. W., Williams, J. S., Nguyen, Y., Robertson, P. D., Resnick, M. A., and Williams, R. S. *Tetrameric Ctp1 coordinates DNA binding and DNA bridging in DNA double-strand-break repair*. *Nature Structural & Molecular Biology*, 22(2) :158–166 (2015). doi :10.1038/nsmb.2945. (Cité page 8.)
- Aravind, L. and Koonin, E. V. *SAP - a putative DNA-binding motif involved in chromosomal organization*. *Trends in Biochemical Sciences*, 25(3) :112–114 (2000). (Cité page 18.)
- Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Denby Wilkes, C., Garnier, O., Labadie, K., Lauderdale, B. E., Le Mouël, A., Marmignon, A., Nowacki, M., Poulain, J., Prajer, M., Wincker, P., Meyer, E., Duharcourt, S., Duret, L., Bétermier, M., and Sperling, L. *The Paramecium germline genome provides a niche for intragenic parasitic DNA : evolutionary dynamics of internal eliminated sequences*. *PLoS genetics*, 8(10) :e1002984 (2012). doi :10.1371/journal.pgen.1002984. (Cité pages 41, 42 et 45.)
- Arnaiz, O. and Sperling, L. *ParameciumDB in 2011 : new tools and new data for functional and comparative genomics of the model ciliate Paramecium tetraurelia*. *Nucleic Acids Research*, 39(Database issue) :D632–636 (2011). doi :10.1093/nar/gkq918. (Cité page 94.)
- Arnoult, N., Correia, A., Ma, J., Merlo, A., Garcia-Gomez, S., Maric, M., Tognetti, M., Benner, C. W., Boulton, S. J., Saghatelian, A., and Karlseder, J. *Regulation of DNA Repair pathway choice in S/G2 by the NHEJ inhibitor CYREN*. *Nature*, 549(7673) :548–552 (2017). doi :10.1038/nature24023. (Cité page 30.)
- Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Câmara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.-M., Kissmehl, R., Klotz, C., Koll, F., Le Mouël, A., Lepère, G., Malinsky, S., Nowacki, M., Nowak, J. K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Bétermier, M., Weissenbach, J., Scarpelli, C., Schächter, V., Sperling, L., Meyer, E., Cohen, J., and Wincker, P. *Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia*. *Nature*, 444(7116) :171–178 (2006). doi :10.1038/nature05230. (Cité pages 40, 41, 43 et 93.)

- Barsoum, E., Martinez, P., and Aström, S. U. *Alpha3, a transposable element that promotes host sexual reproduction*. *Genes & Development*, 24(1) :33–44 (2010). doi :10.1101/gad.557310. (Cité page 28.)
- Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., and Bétermier, M. *PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia*. *Genes & Development*, 23(21) :2478–2483 (2009). doi :10.1101/gad.547309. (Cité pages 43, 45, 60 et 140.)
- Betermier, M. and Duharcourt, S. *Programmed Rearrangement in Ciliates : Paramecium*. *Microbiology Spectrum*, 2(6) (2014). doi :10.1128/microbiolspec.MDNA3-0035-2014. (Cité pages 58 et 60.)
- Bhat, K. P. and Cortez, D. *RPA and RAD51 : fork reversal, fork protection, and genome stability*. *Nature Structural & Molecular Biology*, 25(6) :446 (2018). doi :10.1038/s41594-018-0075-z. (Cité page 8.)
- Bhattacharyya, S., Soniat, M. M., Walker, D., Jang, S., Finkelstein, I. J., and Harshey, R. M. *Phage Mu Gam protein promotes NHEJ in concert with Escherichia coli ligase*. *Proceedings of the National Academy of Sciences of the United States of America*, 115(50) :E11614–E11622 (2018). doi :10.1073/pnas.1816606115. (Cité page 26.)
- Bischerour, J., Bhullar, S., Denby Wilkes, C., Régnier, V., Mathy, N., Dubois, E., Singh, A., Swart, E., Arnaiz, O., Sperling, L., Nowacki, M., and Bétermier, M. *Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium*. *eLife*, 7 (2018). doi :10.7554/eLife.37927. (Cité pages 46, 47, 48, 49, 51, 74, 118 et 133.)
- Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M., and Meyer, E. *Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling*. *Nucleic Acids Research*, 39(10) :4249–4264 (2011). doi :10.1093/nar/gkq1283. (Cité page 56.)
- Bourgain, F. M. and Katinka, M. D. *Telomeres inhibit end to end fusion and enhance maintenance of linear DNA molecules injected into the Paramecium primaurelia macronucleus*. *Nucleic Acids Research*, 19(7) :1541–1547 (1991). (Cité page 83.)
- Britton, S., Coates, J., and Jackson, S. P. *A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair*. *The Journal of Cell Biology*, 202(3) :579–595 (2013). doi :10.1083/jcb.201303073. (Cité pages 23, 117 et 125.)
- Brown, J. S., Lukashchuk, N., Sczaniecka-Clift, M., Britton, S., le Sage, C., Calsou, P., Beli, P., Galanty, Y., and Jackson, S. P. *Neddylaton Promotes Ubiquitylation and Release of Ku from DNA-Damage Sites*. *Cell Reports*, 11(5) :704–714 (2015). doi :10.1016/j.celrep.2015.03.058. (Cité page 16.)
- Bétermier, M., Duharcourt, S., Seitz, H., and Meyer, E. *Timing of developmentally programmed excision and circularization of Paramecium internal eliminated sequences*. *Molecular and Cellular Biology*, 20(5) :1553–1561 (2000). (Cité pages 43 et 48.)
- Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondanèche, M.-C., Sanal, O., Plebani, A., Stéphan, J.-L., Hufnagel, M., le Deist, F., Fischer, A., Durandy, A., de Villartay, J.-P., and Revy, P. *Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly*. *Cell*, 124(2) :287–299 (2006). doi :10.1016/j.cell.2005.12.030. (Cité page 21.)

- Buonomo, S. B. C., Clyne, R. K., Fuchs, J., Loidl, J., Uhlmann, F., and Nasmyth, K. *Disjunction of Homologous Chromosomes in Meiosis I Depends on Proteolytic Cleavage of the Meiotic Cohesin Rec8 by Separin*. *Cell*, 103(3) :387–398 (2000). doi :10.1016/S0092-8674(00)00131-8. (Cité page 12.)
- Caron, F. and Meyer, E. *Does Paramecium primaurelia use a different genetic code in its macronucleus?* *Nature*, 314(6007) :185–188 (1985). (Cité page 41.)
- Chalker, D. L., Meyer, E., and Mochizuki, K. *Epigenetics of Ciliates*. Cold Spring Harbor Perspectives in Biology, 5(12) (2013). doi :10.1101/cshperspect.a017764. (Cité page 63.)
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N., and Lieber, M. R. *Non-homologous DNA end joining and alternative pathways to double-strand break repair*. *Nature Reviews. Molecular Cell Biology*, 18(8) :495–506 (2017). doi :10.1038/nrm.2017.48. (Cité pages 22 et 23.)
- Chanut, P., Britton, S., Coates, J., Jackson, S. P., and Calsou, P. *Coordinated nuclease activities counteract Ku at single-ended DNA double-strand breaks*. *Nature Communications*, 7 (2016). doi :10.1038/ncomms12889. (Cité page 33.)
- Cheng, C.-Y., Young, J. M., Lin, C.-Y. G., Chao, J.-L., Malik, H. S., and Yao, M.-C. *The piggyBac transposon-derived genes TPB1 and TPB6 mediate essential transposon-like excision during the developmental rearrangement of key genes in Tetrahymena thermophila*. *Genes & Development*, 30(24) :2724–2736 (2016). doi :10.1101/gad.290460.116. (Cité pages 61 et 126.)
- Chou, C. H., Wang, J., Knuth, M. W., and Reeves, W. H. *Role of a major autoepitope in forming the DNA binding site of the p70 (Ku) antigen*. *The Journal of Experimental Medicine*, 175(6) :1677–1684 (1992). (Cité page 18.)
- Craxton, A., Somers, J., Munnur, D., Jukes-Jones, R., Cain, K., and Malewicz, M. *XLS (c9orf142) is a new component of mammalian DNA double-stranded break repair*. *Cell Death and Differentiation*, 22(6) :890–897 (2015). doi :10.1038/cdd.2015.22. (Cité page 21.)
- Crickard, J. B. and Greene, E. C. *The biochemistry of early meiotic recombination intermediates*. *Cell Cycle (Georgetown, Tex.)* (2018). doi :10.1080/15384101.2018.1553355. (Cité page 10.)
- de Massy, B. *Initiation of Meiotic Recombination : How and Where ? Conservation and Specificities Among Eukaryotes*. *Annual Review of Genetics*, 47(1) :563–599 (2013). doi :10.1146/annurev-genet-110711-155423. (Cité page 9.)
- Dubois, E., Mathy, N., Régnier, V., Bischerour, J., Baudry, C., Trouslard, R., and Bétermier, M. *Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements*. *Nucleic Acids Research*, 45(6) :3204–3216 (2017). doi :10.1093/nar/gkw1359. (Cité pages 45, 71, 72, 87, 116, 131 et 136.)
- Dupuy, P., Sauviac, L., and Bruand, C. *Stress-inducible NHEJ in bacteria : function in DNA repair and acquisition of heterologous DNA*. *Nucleic Acids Research* (2018). doi :10.1093/nar/gky1212. (Cité page 26.)
- Elick, T. A., Bauser, C. A., and Fraser, M. J. *Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase*. *Genetica*, 98(1) :33–41 (1996). (Cité page 43.)

- Furrer, D. I., Swart, E. C., Kraft, M. F., Sandoval, P. Y., and Nowacki, M. *Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA*. *Cell Reports*, 20(2) :505–520 (2017). doi :10.1016/j.celrep.2017.06.050. (Cité page 59.)
- Garnier, O., Serrano, V., Duharcourt, S., and Meyer, E. *RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia**. *Molecular and Cellular Biology*, 24(17) :7370–7379 (2004). doi :10.1128/MCB.24.17.7370-7379.2004. (Cité page 97.)
- Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.-B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A. E., Hendrickson, E. A., Giovannangeli, C., Jasin, M., and Brunet, E. *Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining*. *Molecular Cell*, 55(6) :829–842 (2014). doi :10.1016/j.molcel.2014.08.002. (Cité page 24.)
- Gilley, D. and Blackburn, E. H. *Lack of telomere shortening during senescence in *Paramecium**. *Proceedings of the National Academy of Sciences of the United States of America*, 91(5) :1955–1958 (1994). (Cité page 36.)
- Gilley, D., Preer, J. R., Aufderheide, K. J., and Polisky, B. *Autonomous replication and addition of telomere-like sequences to DNA microinjected into *Paramecium tetraurelia* macronuclei*. *Molecular and Cellular Biology*, 8(11) :4765–4772 (1988). (Cité page 83.)
- Gout, J.-F., Johri, P., Arnaiz, O., Doak, T. G., Bhullar, S., Couloux, A., Guerin, F., Malinsky, S., Sperling, L., Labadie, K., Meyer, E., Duharcourt, S., and Lynch, M. *Universal trends of post-duplication evolution revealed by the genomes of 13 *Paramecium* species sharing an ancestral whole-genome duplication*. *bioRxiv*, page 573576 (2019). doi :10.1101/573576. (Cité page 94.)
- Gratias, A. and Bétermier, M. *Developmentally programmed excision of internal DNA sequences in *Paramecium aurelia**. *Biochimie*, 83(11) :1009–1022 (2001). (Cité pages 73 et 116.)
- Gratias, A. and Bétermier, M. *Processing of double-strand breaks is involved in the precise excision of *paramecium* internal eliminated sequences*. *Molecular and Cellular Biology*, 23(20) :7152–7162 (2003). (Cité pages 43 et 131.)
- Guirouilh-Barbat, J., Lambert, S., Bertrand, P., and Lopez, B. S. *Is homologous recombination really an error-free process?* *Frontiers in Genetics*, 5 :175 (2014). doi :10.3389/fgene.2014.00175. (Cité pages 12 et 13.)
- Guérin, F., Arnaiz, O., Boggetto, N., Denby Wilkes, C., Meyer, E., Sperling, L., and Duharcourt, S. *Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements*. *BMC genomics*, 18(1) :327 (2017). doi :10.1186/s12864-017-3713-7. (Cité pages 41, 45 et 60.)
- Haley, B., Paunesku, T., Protić, M., and Woloschak, G. E. *Response of heterogeneous ribonuclear proteins (hnRNP) to ionizing radiation and their involvement in DNA damage repair*. *International journal of radiation biology*, 85(8) :643–655 (2009). doi :10.1080/09553000903009548. (Cité page 23.)
- Hammel, M., Rey, M., Yu, Y., Mani, R. S., Classen, S., Liu, M., Pique, M. E., Fang, S., Mahaney, B. L., Weinfeld, M., Schriemer, D. C., Lees-Miller, S. P., and Tainer, J. A.

- XRCC4 Protein Interactions with XRCC4-like Factor (XLF) Create an Extended Grooved Scaffold for DNA Ligation and Double Strand Break Repair.* *The Journal of Biological Chemistry*, 286(37) :32638–32650 (2011). doi :10.1074/jbc.M111.272641. (Cité page 21.)
- Hickman, A. B. and Dyda, F. *Mechanisms of DNA Transposition.* *Microbiology Spectrum*, 3(2) :MDNA3-0034-2014 (2015). doi :10.1128/microbiolspec.MDNA3-0034-2014. (Cité page 27.)
- Hillers, K. J. *Crossover interference.* *Current biology : CB*, 14(24) :R1036–1037 (2004). doi :10.1016/j.cub.2004.11.038. (Cité page 12.)
- Hodgson, A., Terentyev, Y., Johnson, R. A., Bishop-Bailey, A., Angevin, T., Croucher, A., and Goldman, A. S. H. *Mre11 and Exo1 contribute to the initiation and processivity of resection at meiotic double-strand breaks made independently of Spo11.* *DNA Repair*, 10(2) :138–148 (2011). doi :10.1016/j.dnarep.2010.11.008. (Cité page 8.)
- Hoehener, C., Hug, I., and Nowacki, M. *Dicer-like Enzymes with Sequence Cleavage Preferences.* *Cell*, 173(1) :234–247.e7 (2018). doi :10.1016/j.cell.2018.02.029. (Cité page 59.)
- Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H. A., Zhang, Y., Yu, W., Pontarotti, P., Escrava, H., Le Petillon, Y., Liu, X., Chen, S., Schatz, D. G., and Xu, A. *Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination.* *Cell*, 166(1) :102–114 (2016). doi :10.1016/j.cell.2016.05.032. (Cité page 28.)
- Ishida, N., Nakagawa, T., Iemura, S.-I., Yasui, A., Shima, H., Katoh, Y., Nagasawa, Y., Natsume, T., Igarashi, K., and Nakayama, K. *Ubiquitylation of Ku80 by RNF126 Promotes Completion of Nonhomologous End Joining-Mediated DNA Repair.* *Molecular and Cellular Biology*, 37(4) (2017). doi :10.1128/MCB.00347-16. (Cité page 18.)
- Jangam, D., Feschotte, C., and Betrán, E. *Transposable Element Domestication As an Adaptation to Evolutionary Conflicts.* *Trends in genetics : TIG*, 33(11) :817–831 (2017). doi :10.1016/j.tig.2017.07.011. (Cité page 28.)
- Jette, N. and Lees-Miller, S. P. *The DNA-dependent protein kinase : a multifunctional protein kinase with roles in DNA double strand break repair and mitosis.* *Progress in biophysics and molecular biology*, 117(0) :194–205 (2015). doi :10.1016/j.pbiomolbio.2014.12.003. (Cité page 18.)
- Jones, D. T., Taylor, W. R., and Thornton, J. M. *The rapid generation of mutation data matrices from protein sequences.* *Computer applications in the biosciences : CABIOS*, 8(3) :275–282 (1992). (Cité page 94.)
- Kakarougkas, A. and Jeggo, P. A. *DNA DSB repair pathway choice : an orchestrated handover mechanism.* *The British Journal of Radiology*, 87(1035) (2014). doi :10.1259/bjr.20130685. (Cité page 32.)
- Kaminski, A. M., Tumbale, P. P., Schellenberg, M. J., Williams, R. S., Williams, J. G., Kunkel, T. A., Pedersen, L. C., and Bebenek, K. *Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis.* *Nature Communications*, 9 (2018). doi :10.1038/s41467-018-05024-8. (Cité page 20.)
- Kapitonov, V. V. and Jurka, J. *RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons.* *PLoS Biology*, 3(6) (2005). doi :10.1371/journal.pbio.0030181. (Cité page 28.)

- Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J. D., Malinsky, S., and Bétermier, M. *Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining*. PLoS genetics, 7(4) :e1002049 (2011). doi :10.1371/journal.pgen.1002049. (Cité pages 50, 80, 85 et 140.)
- Katinka, M. D. and Bourgain, F. M. *Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of Paramecium primaurelia*. The EMBO journal, 11(2) :725–732 (1992). (Cité page 83.)
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. *The Pyre2 web portal for protein modeling, prediction and analysis*. Nature Protocols, 10(6) :845–858 (2015). doi :10.1038/nprot.2015.053. (Cité page 104.)
- Kim, H.-S., Chen, Q., Kim, S.-K., Nickoloff, J. A., Hromas, R., Georgiadis, M. M., and Lee, S.-H. *The DDN catalytic motif is required for Metnase functions in non-homologous end joining (NHEJ) repair and replication restart*. The Journal of Biological Chemistry, 289(15) :10930–10938 (2014). doi :10.1074/jbc.M113.533216. (Cité page 28.)
- Koike, M., Ikuta, T., Miyasaka, T., and Shiomi, T. *Ku80 can translocate to the nucleus independent of the translocation of Ku70 using its own nuclear localization signal*. Oncogene, 18(52) :7495–7505 (1999a). doi :10.1038/sj.onc.1203247. (Cité page 18.)
- Koike, M., Ikuta, T., Miyasaka, T., and Shiomi, T. *The nuclear localization signal of the human Ku70 is a variant bipartite type recognized by the two components of nuclear pore-targeting complex*. Experimental Cell Research, 250(2) :401–413 (1999b). doi :10.1006/excr.1999.4507. (Cité page 18.)
- Kowalczykowski, S. C. *An Overview of the Molecular Mechanisms of Recombinational DNA Repair*. Cold Spring Harbor Perspectives in Biology, 7(11) (2015). doi :10.1101/cshperspect.a016410. (Cité page 5.)
- Lepère, G., Bétermier, M., Meyer, E., and Duharcourt, S. *Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia*. Genes & Development, 22(11) :1501–1512 (2008). doi :10.1101/gad.473008. (Cité page 56.)
- Lescale, C., Abramowski, V., Bedora-Faure, M., Murigneux, V., Vera, G., Roth, D. B., Revy, P., de Villartay, J.-P., and Deriano, L. *RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair*. Nature Communications, 7 (2016). doi :10.1038/ncomms10529. (Cité pages 28 et 126.)
- Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., and Duharcourt, S. *Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements*. PLoS genetics, 10(9) :e1004665 (2014). doi :10.1371/journal.pgen.1004665. (Cité page 59.)
- Li, X. and Heyer, W.-D. *Homologous recombination in DNA repair and DNA damage tolerance*. Cell research, 18(1) :99–113 (2008). doi :10.1038/cr.2008.1. (Cité page 12.)
- Lin, I.-T., Chao, J.-L., and Yao, M.-C. *An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in Tetrahymena thermophila*. Molecular Biology of the Cell, 23(11) :2213–2225 (2012). doi :10.1091/mbc.E11-11-0952. (Cité page 61.)
- Lin, Z., Kong, H., Nei, M., and Ma, H. *Origins and evolution of the recA/RAD51 gene family : Evidence for ancient gene duplication and endosymbiotic gene transfer*. Proceedings

- of the National Academy of Sciences of the United States of America, 103(27) :10328–10333 (2006). doi :10.1073/pnas.0604232103. (Cité page 10.)
- Little, A. J., Matthews, A., Oettinger, M., Roth, D. B., and Schatz, D. G. *Chapter 2 - The Mechanism of V(D)J Recombination*. In F. W. Alt, T. Honjo, A. Radbruch, and M. Reth, editors, *Molecular Biology of B Cells (Second Edition)*, pages 13–34. Academic Press, London (2015). doi :10.1016/B978-0-12-397933-9.00002-3. (Cité pages 26 et 27.)
- Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y., and Chalmers, R. *The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase*. *Molecular and Cellular Biology*, 27(3) :1125–1132 (2007). doi :10.1128/MCB.01899-06. (Cité page 28.)
- Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. *Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination*. *Cell*, 108(6) :781–794 (2002). (Cité page 22.)
- Mahaney, B. L., Hammel, M., Meek, K., Tainer, J. A., and Lees-Miller, S. P. *XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair*. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 91(1) :31–41 (2013). doi :10.1139/bcb-2012-0058. (Cité page 21.)
- Makharashvili, N., Tubbs, A. T., Yang, S.-H., Wang, H., Barton, O., Zhou, Y., Deshpande, R. A., Lee, J.-H., Lobrich, M., Sleckman, B. P., Wu, X., and Paull, T. T. *Catalytic and Noncatalytic Roles of the CtIP Endonuclease in Double-Strand Break End Resection*. *Molecular Cell*, 54(6) :1022–1033 (2014). doi :10.1016/j.molcel.2014.04.011. (Cité page 8.)
- Maliszewska-Olejniczak, K., Gruchota, J., Gromadka, R., Denby Wilkes, C., Arnaiz, O., Mathy, N., Duharcourt, S., Bétermier, M., and Nowak, J. K. *TFIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements*. *PLoS genetics*, 11(7) :e1005383 (2015). doi :10.1371/journal.pgen.1005383. (Cité page 59.)
- Malu, S., Malshetty, V., Francis, D., and Cortes, P. *Role of non-homologous end joining in V(D)J recombination*. *Immunologic Research*, 54(1) :233–246 (2012). doi :10.1007/s12026-012-8329-z. (Cité page 29.)
- Marmignon, A., Bischerour, J., Silve, A., Fojcik, C., Dubois, E., Arnaiz, O., Kapusta, A., Malinsky, S., and Bétermier, M. *Ku-mediated coupling of DNA cleavage and repair during programmed genome rearrangements in the ciliate Paramecium tetraurelia*. *PLoS genetics*, 10(8) :e1004552 (2014). doi :10.1371/journal.pgen.1004552. (Cité pages vi, ix, x, 51, 52, 53, 54, 55, 77, 82, 83, 85, 87, 100, 108, 113, 114, 115, 136, 139, 140 et 143.)
- Martini, E., Roche, D. M. J., Marheineke, K., Verreault, A., and Almouzni, G. *Recruitment of Phosphorylated Chromatin Assembly Factor 1 to Chromatin after UV Irradiation of Human Cells*. *The Journal of Cell Biology*, 143(3) :563–575 (1998). doi :10.1083/jcb.143.3.563. (Cité page 74.)
- McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., and Oettinger, M. A. *Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps*. *Cell*, 83(3) :387–395 (1995). (Cité page 28.)
- McGrath, C. L., Gout, J.-F., Doak, T. G., Yanagi, A., and Lynch, M. *Insights into Three Whole-Genome Duplications Gleaned from the Paramecium caudatum Genome Sequence*. *Genetics*, 197(4) :1417–1428 (2014). doi :10.1534/genetics.114.163287. (Cité page 93.)

- McKee, B. D., Yan, R., and Tsai, J.-H. *Meiosis in male Drosophila*. *Spermatogenesis*, 2(3) :167–184 (2012). doi :10.4161/spmg.21800. (Cité page 11.)
- Mimori, T., Akizuki, M., Yamagata, H., Inada, S., Yoshida, S., and Homma, M. *Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap*. *The Journal of Clinical Investigation*, 68(3) :611–620 (1981). (Cité page 16.)
- Mitra, R., Fain-Thornton, J., and Craig, N. L. *piggyBac can bypass DNA synthesis during cut and paste transposition*. *The EMBO journal*, 27(7) :1097–1109 (2008). doi :10.1038/emboj.2008.41. (Cité page 45.)
- Moon, A. F., Pryor, J. M., Ramsden, D. A., Kunkel, T. A., Bebenek, K., and Pedersen, L. C. *Structural accommodation of ribonucleotide incorporation by the DNA repair enzyme polymerase Mu*. *Nucleic Acids Research*, 45(15) :9138–9148 (2017). doi :10.1093/nar/gkx527. (Cité page 23.)
- Morellet, N., Li, X., Wieninger, S. A., Taylor, J. L., Bischerour, J., Moriau, S., Lescop, E., Bardiaux, B., Mathy, N., Assrir, N., Bétermier, M., Nilges, M., Hickman, A. B., Dyda, F., Craig, N. L., and Guittet, E. *Sequence-specific DNA binding activity of the cross-brace zinc finger motif of the piggyBac transposase*. *Nucleic Acids Research*, 46(5) :2660–2677 (2018). doi :10.1093/nar/gky044. (Cité page 45.)
- Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., and de Villartay, J. P. *Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency*. *Cell*, 105(2) :177–186 (2001). (Cité page 22.)
- Mundia, M. M., Desai, V., Magwood, A. C., and Baker, M. D. *Nascent DNA Synthesis During Homologous Recombination Is Synergistically Promoted by the Rad51 Recombinase and DNA Homology*. *Genetics*, 197(1) :107–119 (2014). doi :10.1534/genetics.114.161455. (Cité page 8.)
- Neale, M. J., Pan, J., and Keeney, S. *Endonucleolytic processing of covalent protein-linked double-strand breaks*. *Nature*, 436(7053) :1053–1057 (2005). doi :10.1038/nature03872. (Cité page 10.)
- Nowacki, M., Higgins, B. P., Maquilan, G. M., Swart, E. C., Doak, T. G., and Landweber, L. F. *A functional role for transposases in a large eukaryotic genome*. *Science (New York, N.Y.)*, 324(5929) :935–938 (2009). doi :10.1126/science.1170023. (Cité page 62.)
- Nowacki, M., Zagorski-Ostoja, W., and Meyer, E. *Nowa1p and Nowa2p : novel putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium tetraurelia*. *Current biology : CB*, 15(18) :1616–1628 (2005). doi :10.1016/j.cub.2005.07.033. (Cité page 56.)
- Nowak, J. K., Gromadka, R., Juszczuk, M., Jerka-Dziadosz, M., Maliszewska, K., Mucchielli, M.-H., Gout, J.-F., Arnaiz, O., Agier, N., Tang, T., Aggerbeck, L. P., Cohen, J., Delacroix, H., Sperling, L., Herbert, C. J., Zagulski, M., and Bétermier, M. *Functional study of genes essential for autogamy and nuclear reorganization in Paramecium*. *Eukaryotic Cell*, 10(3) :363–372 (2011). doi :10.1128/EC.00258-10. (Cité page 56.)
- Ochi, T., Blackford, A. N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V. M., Robinson, C. V., Blundell, T. L., and Jackson, S. P. *PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair*. *Science*

- (New York, N.Y.), 347(6218) :185–188 (2015). doi :10.1126/science.1261971. (Cité page 21.)
- Pannunzio, N. R., Watanabe, G., and Lieber, M. R. *Nonhomologous DNA end-joining for repair of DNA double-strand breaks*. *Journal of Biological Chemistry*, 293(27) :10512–10523 (2018). doi :10.1074/jbc.TM117.000374. (Cité page 22.)
- Pellestor, F., Gatinois, V., Puechberty, J., Geneviève, D., and Lefort, G. *Le chromothripsis - Une forme insoupçonnée de complexification extrême des remaniements chromosomiques*. *médecine/sciences*, 30(3) :266–273 (2014). doi :10.1051/medsci/20143003014. (Cité pages 24 et 25.)
- Pitcher, R. S., Brissett, N. C., and Doherty, A. J. *Nonhomologous end-joining in bacteria : a microbial perspective*. *Annual Review of Microbiology*, 61 :259–282 (2007). doi :10.1146/annurev.micro.61.080706.093354. (Cité pages 24 et 25.)
- Preer, J. R., Preer, L. B., Rudman, B. M., and Barnett, A. J. *Deviation from the universal code shown by the gene for surface protein 51a in Paramecium*. *Nature*, 314(6007) :188–190 (1985). (Cité page 41.)
- Pryor, J. M., Conlin, M. P., Carvajal-Garcia, J., Luedeman, M. E., Luthman, A. J., Small, G. W., and Ramsden, D. A. *Ribonucleotide incorporation enables repair of chromosome breaks by nonhomologous end joining*. *Science*, 361(6407) :1126–1129 (2018). doi :10.1126/science.aat2477. (Cité page 23.)
- Rajaei, N., Chiruvella, K. K., Lin, F., and Aström, S. U. *Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast*. *Proceedings of the National Academy of Sciences of the United States of America*, 111(43) :15491–15496 (2014). doi :10.1073/pnas.1406027111. (Cité page 28.)
- Raval, P., Kriatchko, A. N., Kumar, S., and Swanson, P. C. *Evidence for Ku70/Ku80 association with full-length RAG1*. *Nucleic Acids Research*, 36(6) :2060–2072 (2008). doi :10.1093/nar/gkn049. (Cité page 28.)
- Ropars, V., Drevet, P., Legrand, P., Baconnais, S., Amram, J., Faure, G., Márquez, J. A., Piétrement, O., Guerois, R., Callebaut, I., Le Cam, E., Revy, P., de Villartay, J.-P., and Charbonnier, J.-B. *Structural characterization of filaments formed by human Xrcc4–Cernunnos/XLF complex involved in nonhomologous DNA end-joining*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(31) :12663–12668 (2011). doi :10.1073/pnas.1100758108. (Cité page 21.)
- Sandoval, P. Y., Swart, E. C., Arambasic, M., and Nowacki, M. *Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting*. *Developmental Cell*, 28(2) :174–188 (2014). doi :10.1016/j.devcel.2013.12.010. (Cité page 59.)
- Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S. P. *Human CtIP promotes DNA end resection*. *Nature*, 450(7169) :509–514 (2007). doi :10.1038/nature06337. (Cité page 8.)
- Seol, J.-H., Shim, E. Y., and Lee, S. E. *Microhomology-mediated end joining : Good, bad and ugly*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 809 :81–87 (2018). doi :10.1016/j.mrfmmm.2017.07.002. (Cité page 31.)

- Shibata, A., Moiani, D., Arvai, A. S., Perry, J. J. P., Harding, S. M., Genois, M.-M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., Ismail, A., Ismalaj, E., Petricci, E., Matthew, J. N., Bristow, R. G., Masson, J.-Y., Wyman, C., Jeggo, P., and Tainer, J. A. *DNA Double Strand Break Repair Pathway Choice Is Directed by Distinct MRE11 Nuclease Activities*. *Molecular cell*, 53(1) :7–18 (2014). doi :10.1016/j.molcel.2013.11.003. (Cité page 6.)
- Sibanda, B. L., Chirgadze, D. Y., Ascher, D. B., and Blundell, T. L. *DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair*. *Science (New York, N.Y.)*, 355(6324) :520–524 (2017). doi :10.1126/science.aak9654. (Cité page 19.)
- Swart, E. C., Denby Wilkes, C., Sandoval, P. Y., Hoehener, C., Singh, A., Furrer, D. I., Arambasic, M., Ignarski, M., and Nowacki, M. *Identification and analysis of functional associations among natural eukaryotic genome editing components*. *F1000Research*, 6 :1374 (2017). doi :10.12688/f1000research.12121.1. (Cité page 59.)
- Symington, L. S. *Mechanism and Regulation of DNA End Resection in Eukaryotes*. *Critical reviews in biochemistry and molecular biology*, 51(3) :195–212 (2016). doi :10.3109/10409238.2016.1172552. (Cité pages 6, 7 et 10.)
- Volff, J.-N. *Turning junk into gold : domestication of transposable elements and the creation of new genes in eukaryotes*. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, 28(9) :913–922 (2006). doi :10.1002/bies.20452. (Cité page 28.)
- Walker, J. R., Corpina, R. A., and Goldberg, J. *Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair*. *Nature*, 412(6847) :607–614 (2001). doi :10.1038/35088000. (Cité pages 16 et 17.)
- Wang, H., Li, Y., Truong, L. N., Shi, L. Z., Hwang, P. Y.-H., He, J., Do, J., Cho, M. J., Li, H., Negrete, A., Shiloach, J., Berns, M. W., Shen, B., Chen, L., and Wu, X. *CtIP Maintains Stability at Common Fragile Sites and Inverted Repeats by End Resection-Independent Endonuclease Activity*. *Molecular Cell*, 54(6) :1012–1021 (2014). doi :10.1016/j.molcel.2014.04.012. (Cité page 8.)
- Wang, J. and Davis, R. E. *Programmed DNA elimination in multicellular organisms*. *Current Opinion in Genetics & Development*, 27 :26–34 (2014). doi :10.1016/j.gde.2014.03.012. (Cité page 127.)
- Wang, X., Liu, H., Shi, L., Yu, X., Gu, Y., and Sun, X. *LINP1 facilitates DNA damage repair through non-homologous end joining (NHEJ) pathway and subsequently decreases the sensitivity of cervical cancer cells to ionizing radiation*. *Cell Cycle*, 17(4) :439–447 (2018). doi :10.1080/15384101.2018.1442625. (Cité page 23.)
- Watson, J. D. and Crick, F. H. *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. *Nature*, 171(4356) :737–738 (1953). (Cité page 3.)
- Wu, P.-Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S. N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., and Junop, M. S. *Structural and Functional Interaction between the Human DNA Repair Proteins DNA Ligase IV and XRCC4*. *Molecular and Cellular Biology*, 29(11) :3163–3172 (2009). doi :10.1128/MCB.01895-08. (Cité page 20.)
- Wu, Q., Ochi, T., Matak-Vinkovic, D., Robinson, C. V., Chirgadze, D. Y., and Blundell, T. L. *Non-homologous end-joining partners in a helical dance : structural studies of XLF-XRCC4 interactions*. *Biochemical Society transactions*, 39(5) :1387–92, suppl 2 p following 1392 (2011). doi :10.1042/BST0391387. (Cité page 21.)

- Xing, M., Yang, M., Huo, W., Feng, F., Wei, L., Jiang, W., Ning, S., Yan, Z., Li, W., Wang, Q., Hou, M., Dong, C., Guo, R., Gao, G., Ji, J., Zha, S., Lan, L., Liang, H., and Xu, D. *Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway*. *Nature Communications*, 6 :6233 (2015). doi :10.1038/ncomms7233. (Cité page 21.)
- Yamtich, J. and Sweasy, J. B. *DNA Polymerase Family X : Function, Structure, and Cellular Roles*. *Biochimica et biophysica acta*, 1804(5) :1136–1150 (2010). doi :10.1016/j.bbapap.2009.07.008. (Cité page 22.)
- Yao, M.-C., Chao, J.-L., and Cheng, C.-Y. *Programmed Genome Rearrangements in Tetrahymena*. *Microbiology Spectrum*, 2(6) (2014). doi :10.1128/microbiolspec.MDNA3-0012-2014. (Cité page 61.)
- Yerlici, V. T. and Landweber, L. F. *Programmed Genome Rearrangements in the Ciliate Oxytricha*. *Microbiology Spectrum*, 2(6) (2014). doi :10.1128/microbiolspec.MDNA3-0025-2014. (Cité page 62.)
- Zhang, Y., He, Q., Hu, Z., Feng, Y., Fan, L., Tang, Z., Yuan, J., Shan, W., Li, C., Hu, X., Tanyi, J. L., Fan, Y., Huang, Q., Montone, K., Dang, C. V., and Zhang, L. *Long noncoding RNA LINP1 regulates double strand DNA break repair in triple negative breast cancer*. *Nature structural & molecular biology*, 23(6) :522–530 (2016). doi :10.1038/nsmb.3211. (Cité page 23.)

Titre : Spécialisation de Ku80c dans le couplage entre coupure et réparation de l'ADN lors des réarrangements programmés du génome chez *Paramecium tetraurelia*

Mots clés : Couplage, Cassure double-brin, Réparation, NHEJ, Ciliés, Paramécie

Résumé : Au cours de son cycle sexuel, le cilié *Paramecium tetraurelia* procède à de massifs réarrangements programmés de son génome (RPG). Ils consistent, entre autres choses, en l'excision de 45 000 séquences précisément délimitées, appelées IES (Internal Eliminated Sequences). La transposase domestiquée Piggymac (Pgm) introduit les cassures double-brin (CDB) à l'extrémité des IES. La réparation très précise de ces dommages est réalisée par la voie de réparation des extrémités non-homologues (NHEJ). Un des acteurs de cette voie est l'hétérodimère Ku70/Ku80. Suite à des duplications globales du génome, la paramécie possède trois gènes *KU80*. Un seul de ces gènes est induit lors des RPG (*KU80c*) et une expérience d'ARN interférence (ARNi) contre *KU80c* montre une complète inhibition de l'introduction des CDB. De plus, des expériences de Co-IP en système hétérologue montrent que Ku70/Ku80c interagit avec Pgm. Ces résultats prouvent le rôle essentiel de Ku dans l'introduction des CDB lors des RPG et

soulèvent la question du mécanisme impliqué.

Au cours de ma thèse j'ai caractérisé le couplage entre Ku et Pgm en analysant des expériences d'immunofluorescence avec ou sans pré-extraction, permettant de déterminer les interdépendances de ces protéines pour leur localisation et pour leur stabilité nucléaire. Ces approches ont permis de démontrer que Pgm requiert la présence de Ku pour être stablement localisé dans les noyaux lors des RPG. *Ku80c* partage 74% de sa séquence protéique avec *Ku80a*. Des expériences de complémentations fonctionnelles surexprimant *Ku80a* lors des RPG ont montré que *Ku80a* n'est pas capable ni de se localiser stablement dans les noyaux ni de participer à la stabilisation nucléaire de Pgm. De plus, les RPG sont inhibés. Ces résultats montrent que *Ku80c* s'est spécialisé dans le couplage avec Pgm pour l'introduction des CDB lors des RPG. L'utilisation de protéines chimériques a permis de déterminer que la spécialisation de *Ku80c* est portée par son domaine N-terminal α/β .

Title : Specialization of Ku80c in the coupling between DNA break and repair during programmed genome rearrangements in *Paramecium tetraurelia*

Keywords : Coupling, Double strand break, Repair, NHEJ, Ciliates, Paramecia

Abstract : During its sexual cycle, the ciliate *Paramecium tetraurelia* undergoes massive Programmed Genome Rearrangements (PGR). They consist, among others, in excision of 45,000 precisely delimited sequences, called IES (Internal Eliminated Sequences). A domesticated transposase, PiggyMac (Pgm), introduces double-strand DNA breaks (DSB) at IES ends. The Non Homologous End Joining pathway (NHEJ) handles highly precise repair of DSB. One of the actors of this pathway is the heterodimer Ku70/Ku80. In *P. tetraurelia*, the *KU80* gene is present in three paralogous copies. Only *KU80c* is specifically expressed during PGR and RNA interferences against *KU80c* showed a complete inhibition of DNA cleavage. Furthermore, a Co-IP experiment in a heterologous system showed that both Ku70/Ku80c interact with Pgm. These results provide evidence that Ku is an essential partner of Pgm for DSB introduction; raising the question of the activating mechanism involved.

During my PhD, I characterized the coupling between Ku and Pgm by analyzing immunofluorescence experiments, with or without pre-extraction, allowing the determination of inter-dependencies between those proteins for their nuclear localization and stability. Those methods demonstrated that Pgm requires the presence of Ku for a stable nuclear localization during the PGR.

Ku80c shares 74% of the protein sequence with *Ku80a*. Functional complementation assays overexpressing *Ku80a* during the PGR showed that *Ku80a* is not capable to stably localize in nuclei nor to participate in Pgm nuclear stability. Furthermore, PGR are inhibited. Those results show that *Ku80c* has specialized for the DSB introduction during PGR. The use of chimeric proteins allowed to determine that *Ku80c* specialization was carried out by its N terminal domain.

