



HAL
open science

Dynamique de la réplication de l'ADN et complexe pré-réplicatif chez *Leishmania* sp.: apport du système CRISPR/Cas9

Lauriane Sollelis

► **To cite this version:**

Lauriane Sollelis. Dynamique de la réplication de l'ADN et complexe pré-réplicatif chez *Leishmania* sp.: apport du système CRISPR/Cas9. Médecine humaine et pathologie. Université Montpellier, 2016. Français. NNT : 2016MONTT062 . tel-01559354

HAL Id: tel-01559354

<https://theses.hal.science/tel-01559354>

Submitted on 10 Jul 2017

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

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

THÈSE

Pour obtenir le grade de
Docteur

Délivré par

L'Université Montpellier

Préparée au sein de l'école doctorale CBS2

Et de l'unité de recherche

UMR MIVEGEC (UM - CNRS 5290 - IRD 224)

Spécialité : **Parasitologie**



**Dynamique de la Réplication de l'ADN et
Complexe Pré-Réplicatif Chez *Leishmania sp.*
Apport du Système CRISPR/Cas9.**

Présentée par **Lauriane SOLLELIS**

Soutenue le 20 Décembre 2016 devant le jury composé de

Dr Yvon Sterkers, MCU-PH, Université Montpellier

Directeur de thèse

Pr. Artur Scherf, DRCE, CNRS, Institut Pasteur

Co-Directeur Parafrap

Dr Catherine Merrick, Senior Lecturer, Keele U.

Rapporteur

Dr Gerald Spâth, DR, Institut Pasteur

Rapporteur

Dr. Etienne Schwob, DR, CNRS

Examineur

Pr Patrick Bastien, PU-PH, Université Montpellier

Examineur



Remerciements,

Je tiens tout d'abord à remercier le Dr Catherine Merrick et le Dr Gerald Spâth pour avoir accepté d'être les rapporteurs de ce travail de thèse ainsi que les autres membres du jury le Pr. Artur Scherf, le Dr. Etienne Schwob et le Pr. Patrick Bastien.

Je remercie chaleureusement mon directeur de thèse le Dr Yvon Sterkers. Merci Yvon de m'avoir toujours soutenue depuis le début de cette aventure, d'avoir toujours su m'écouter et laisser une grande liberté dans mon travail. Je te serai toujours reconnaissante pour ça et surtout d'avoir tout fait pour que je puisse participer au BOP !

Je remercie aussi tous les membres du laboratoire avec lesquels j'ai travaillé tout au long de ces trois années, notamment Slavica, Maria Rosa, Michèle, Gilles, Diane, Elodie, Patrick L et toutes les personnes du côté hôpital ! Yves merci pour tes constructions improbables et c'était agréable d'avoir partagé le bureau avec toi!

Nada et Lucien c'était vraiment un plaisir de travailler avec vous, j'espère ne pas vous avoir traumatisés avec mon CRISPR !

Rafa, toujours là, toujours prêt à rendre service, merci pour tous ces moments de bonne humeur passés avec toi au labo ou pas !!

Laurence, Christophe, c'était toujours agréable de pouvoir discuter avec vous.

Jota Je ne sais pas par où commencer, il s'est passé tellement de choses depuis le 1^{er} Avril 2014.... Je te dois beaucoup, j'ai vraiment beaucoup appris à tes côtés entre le CRISPR et le régime vegan (avec ou sans souffrance). Merci pour tout ce que tu m'as transmis, ton soutien et ton enthousiasme. Nos discussions vont me manquer. *Nb : Prends bien soin de nos amis les vers pendant mon absence !*

Artur, je vous remercie de m'avoir accueilli dans votre laboratoire, c'était toujours très enrichissant pour moi. Merci également pour votre soutien pendant ces trois années. Merci à toute l'équipe de BIHP notamment Aurélie, Anne, Jess, Shruthi, Cameron, Kiki, Doro, Julien ... Vous m'avez toujours très bien accueillie dans la joie et la bonne humeur et j'ai beaucoup appris à vos côtés.

Merci également à Etienne et son équipe de m'avoir accueillie dans leur laboratoire et de m'avoir initiée au monde des levures.

Je voudrais également remercier tous mes collègues Parafrapers, Ana, Gigi, Shuai, Kannan, Fabien, Liwen, Amandine, Marion, Valeria et Claire pour tous ces bons moments passés à chacune de nos retrouvailles ! Merci à Patrick et Slavica pour l'organisation de nos workshops.

Je tiens également à remercier mon ancien labo DIMNP, toujours présent, à chaque fois que j'avais besoin d'un conseil, d'un réactif, ou tout simplement de passer un bon

moment la porte était toujours grande ouverte ! Merci à tous, Laurence, Maryse, Kai,
Marjo, Rachel, Michel, Luc, Catherine, Seb, Caty, Sharon, Anne

Maude Lele, merci pour tes conseils de dernière minute désolée pour le bizutage

Merci à mes amis et à famille pour leur soutien, leur aide et surtout leur compréhension.

Maryse de m'avoir accueillie et fait découvrir Paris pendant mes passages éclairs sur
Paris.

Tables des matières

1	Préambule	9
2	Introduction générale.....	10
2.1	Le parasite: <i>Leishmania</i> spp :	11
2.2	Description microscopique.....	12
2.3	Cycle de vie de <i>Leishmania</i>	14
2.4	Vecteurs et réservoirs	17
2.4.1	Vecteurs.....	17
2.4.2	Réservoirs	18
2.5	Epidémiologie & manifestations cliniques	18
2.5.1	Manifestations cliniques	18
2.5.2	Zones d'endémie	21
2.6	Moyens de lutte et prophylaxie	22
2.7	Particularités génomiques.....	23
2.7.1	Organisation du génome nucléaire	23
2.7.2	Spécificités d'espèces	23
2.7.3	Organisation des gènes et transcription	26
2.7.1	Base J	28
2.7.2	Génome mitochondrial	30
2.7.3	Plasticité du génome	31
2.8	De l'origine à la réplication.....	38
2.8.1	Bactéries	40
2.8.2	Eucaryotes modèles	41
2.8.3	Archaea.....	44
2.8.4	Trypanosomatidés.....	44
2.9	CRISPR/Cas9 : outil de génie génétique	48
2.9.1	Découverte	48
2.9.2	Fonctionnement	49
2.9.3	De la bactérie à l'outil génétique.	52
3	Problématique et objectifs de mon travail de thèse.....	54
4	Implémentation du système CRISPR-Cas9 chez <i>Leishmania</i> spp	55
4.1	Introduction.....	56
4.2	Résultats	57
4.2.1	Knock Out (article CRISPR-cas9, Sollelis et al. Cell Microbiol 2015).....	57

4.2.2	« Tagging » au locus endogène par CRISPR/Cas9.....	71
4.2.3	Systèmes inductibles pour CRISPR/Cas9	75
4.3	Conclusion	83
5	Etude de la réplication de l'ADN chez <i>Leishmania</i>	84
5.1	Dynamique de réplication chez <i>Leishmania</i> (Article 2: Stanojcic et al. Sci Rep 2016)	85
5.1.1	Introduction.....	85
5.1.2	Résultats	85
5.1.3	Conclusion	99
5.2	Identification et caractérisation des protéines Orc1/Cdc6 et Orc1b impliquées dans la réplication de l'ADN.	101
5.2.1	Introduction.....	101
5.2.2	Résultats	101
5.2.3	Conclusion	117
5.3	Identification des origines de réplication chez <i>Leishmania</i>	118
5.3.1	Introduction.....	118
5.3.2	ChIP-seq.....	119
5.3.3	Séquençage du brin naissant.....	121
5.4	Discussion générale sur la réplication du parasite <i>Leishmania</i>	126
5.4.1	Analyse critique des données de MFA-seq chez <i>Leishmania</i>	128
5.4.2	Analyse critique des données de séquençage du brin naissant.....	130
5.4.3	Synthèse des données publiées	132
6	Conclusion & Perspectives	134
7	Annexes	136
7.1	Annexe I :CRISPR-Cas9 for <i>Leishmania major</i>	137
7.2	Annexe III : Nascent Strands preparation.....	142
8	Bibliographie.....	145

Tables des Figures

Figure 2-1 Vestiges de la civilisation précolombienne.....	10
Figure 2-2 Classification simplifiée du genre <i>Leishmania</i>	11
Figure 2-3. <i>L. mexicana</i> : schéma d'un cycle de division cellulaire.....	13
Figure 2-4 Le cycle biologique de <i>Leishmania</i> . (Kaye & Scott, 2011)	14
Figure 2-5 Forme promastigote et amastigote de <i>Leishmania</i>	15
Figure 2-6 Cycle biologique du parasite <i>Leishmania</i> chez un vecteur compétent.....	17
Figure 2-7 Femelle phlébotome prenant son repas sanguin.	18
Figure 2-8 Espèces de <i>Leishmania</i> provoquent un large éventail de maladies.	19
Figure 2-9 Mappemonde représentant les zones endémiques de leishmaniose.....	22
Figure 2-10 Diagramme de Venn représentant le nombre de gènes spécifiques d'espèces chez trois espèces de <i>Leishmania</i> . (Smith et al 2007).	25
Figure 2-11 Transcription polycistronique chez <i>Leishmania</i>	28
Figure 2-12 Biosynthèse de la base J par la modification spécifique de la base thymine sur l'ADN.	29
Figure 2-13 Positions des bases J chez <i>L. major</i>	30
Figure 2-14 Modèle d'amplification génique chez <i>Leishmania</i>	32
Figure 2-15 Représentation de l'aneuploïdie dans quatre espèces de <i>Leishmania</i> déterminée par FISH.....	34
Figure 2-16 Représentation des évènements de répartition asymétrique des chromosomes chez <i>Leishmania</i>	36
Figure 2-17 Aneuploïdie dans des isolats cliniques de <i>L. donovani</i>	37
Figure 2-18 Le cycle cellulaire.....	39
Figure 2-19 Représentation schématique de l'initiation de la réplication chez les bactéries.	40
Figure 2-20 Dynamique de régulation de la réplication de l'ADN chez les Métazoaires. (Aladjem, 2007)	42
Figure 2-21 Fourche de réplication eucaryote.	43
Figure 2-22 Protéines putative du complexe ORC chez <i>T. brucei</i>	46
Figure 2-23 Position des H4K10Ac et des sites de liaison à l'ADN de Orc1/Cdc6 sur le chromosome 7 de <i>T. brucei</i>	47
Figure 2-24 Classification des systèmes CRISPR.	50
Figure 2-25. Crispr Type II chez <i>S. pyogenes</i>	51
Figure 4-1 CRISPR/Cas9 in <i>Leishmania</i>	57
Figure 4-2 Measure of relative expression of the sgRNA and U6 promotor by qRT-PCR.	70
Figure 4-3. Strategy for tagging genes at the locus in <i>Leishmania</i> with CRISPR/Cas9 editing tool. ..	72
Figure 4-4 Localization of Flabarin.	74
Figure 4-5. Induction of Cas9 expression by a T7TR system in <i>Leishmania infantum</i>	76
Figure 4-6 Plasmide pTubInd_Cas9.....	77
Figure 4-7 Expression of Cas9 under T7 promotor.	78
Figure 4-8 Measure of Cas9 mARN by qRT-PCR in Linf_T7TR_Cas9.	78
Figure 4-9. Possible Cre recombination events between two loxP sites.	80
Figure 4-10 Cas9_LoxP cassette for a CRISPR/Cas9 inducible by DiCre.	81
Figure 4-11 Reconstitution of a functional Cas9 by an inducible DiCre.	82
Figure 5-1 Localization of LmOrc1/Cdc6 in <i>L. major</i>	104
Figure 5-2 Localization of LmOrc1b in <i>L. major</i>	105
Figure 5-3 Tagging marker free of Linf_Orc1b_3HA_Cter with CRISPR/Cas9.....	107
Figure 5-4 RNA interference of TbORC1b in procyclic forms of <i>T. brucei</i>	108
Figure 5-5. Double Knockdown of TbOrc1b and TbOrc1/Cdc6 by RNA interference.	109

Figure 5-6. Complementation of TbOrc1/Cdc6 RNA interference by LmOrc1/Cdc6.	109
Figure 5-7. Complementation of TbOrc1b RNA interference by LmOrc1b.....	110
Figure 5-8 . Clustal omega alignments of <i>L major</i> , <i>T. brucei</i> and <i>T. cruzi</i> Orc1/Cdc6 proteins sequences.....	111
Figure 5-9 Clustal omega alignment of <i>L major</i> , <i>T. brucei</i> and <i>T. cruzi</i> Orc1b proteins sequences..	112
Figure 5-10 Growth defect after an induced KO of Orc1/Cdc6 and Orc1b in <i>L. infantum</i>	113
Figure 5-11 Genome edition of ORC1b was verified by PCR.	115
Figure 5-12 Effect of the depletion of LinfOrc1/Cdc6 and LinfOrc1b in <i>L. infantum</i> promastigote.	116
Figure 5-13 Functional study of LinfOrc1/Cdc6 and LinfOrc1b inducible KO using the CRISPR/Cas9 system.	117
Figure 5-14 Results of ChIP-Seq experiment.	120
Figure 5-15 Purification procedure of small nascent DNA strands.....	122
Figure 5-16 Small Nascent Strand isolation after ultracentrifugation.	123
Figure 5-17. Two types of library for WGS.	125
Figure 5-18 Preparation of stranded library with Accel-NGS 1S Plus DNA Library Kit from Swift Biosciences.	126
Figure 5-19 Higher ORI density at early replicating regions in <i>L. major</i> chromosomes.....	132

Table des tableaux

Tableau 2-1 Manifestations cliniques en fonction de l'espèce de <i>Leishmania</i> .	21
Tableau 5-1 Replication factors	102
Tableau 5-2: Replication factors localization using different tag.	103
Tableau 5-3 Résumé des principaux résultats de peignage moléculaire.	127

1 Préambule

Après avoir travaillé cinq années dans l'équipe du Dr Henri Vial à DIMNP sur le métabolisme des phospholipides chez *Plasmodium falciparum*, j'ai obtenu un financement du labex Parafrap pour effectuer ma thèse sous la direction du Dr Yvon Sterkers. Ces trois dernières années de travail m'ont permis de découvrir un autre parasite assez intrigant, *Leishmania*. Cette seconde expérience en parasitologie a confirmé ma volonté de poursuivre en post-doctorat dans ce domaine. En janvier, je pars donc rejoindre l'équipe du Dr Mattias Marti à Glasgow pour effectuer mon post-doctorat.

Ce manuscrit fait la synthèse de ces trois années de travail et s'articule autour de deux articles publiés qui sont accompagnés de données plus préliminaires s'inscrivant dans la continuité de ces publications.

J'aimerais également attirer votre attention sur le fait que mon jury est composé d'une anglophone, que je remercie particulièrement et pour qui j'ai mis les légendes des figures des parties 4 et 5 en anglais.

Je vous souhaite une bonne lecture.

2 Introduction générale

Dans les années 1900 deux médecins William Boog Leishman et Charles Donovan, en mission en Inde, ont indépendamment observé sur des biopsies de rate de patients la présence de «parasites ovales» colorés au Giemsa dans des macrophages. C'est ainsi que pour la première fois le parasite fut mis en évidence. Il a été nommé *Leishmania donovani* en l'honneur des deux médecins (Bailey & Bishop, 1959). Ce n'est que quelques années plus tard, en 1926, que Parrot et Donatien ont clairement établi le rôle des phlébotomes dans la transmission de la leishmaniose chez l'homme et le chien. Des travaux préliminaires incriminant le phlébotome comme vecteur avaient été initiés en 1786 par l'italien Scopoli et les frères Sargent de 1901 à 1904. Des poteries ornées de décors représentant des personnages souffrant de lésions cutanées ou de mutilations faciales typiques de leishmaniose cutanéomuqueuse ont été retrouvées sur des vestiges de la civilisation précolombienne (Figure 2-1) : ces vestiges témoignent d'une présence ancienne de la leishmaniose sur le continent sud-américain.

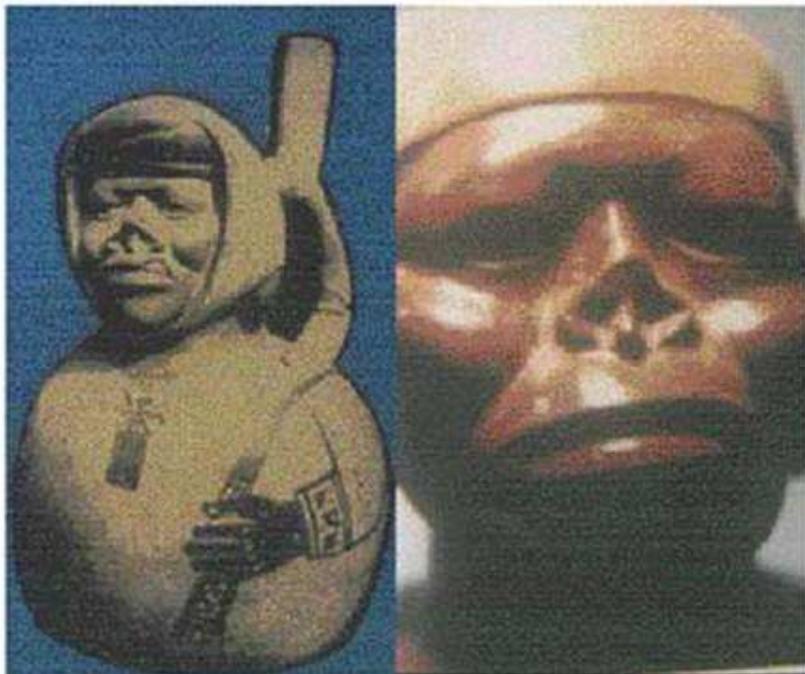


Figure 2-1 Vestiges de la civilisation précolombienne.
(Source the feria journal of medicine)

2.1 Le parasite: *Leishmania* spp :

Leishmania est un parasite protozoaire flagellé appartenant à l'ordre des Kinetoplastida et à la famille des *Trypanosomatidae*, comme *Trypanosoma cruzi* responsable de la maladie de Chagas et *Trypanosoma brucei* agent de la maladie du sommeil.

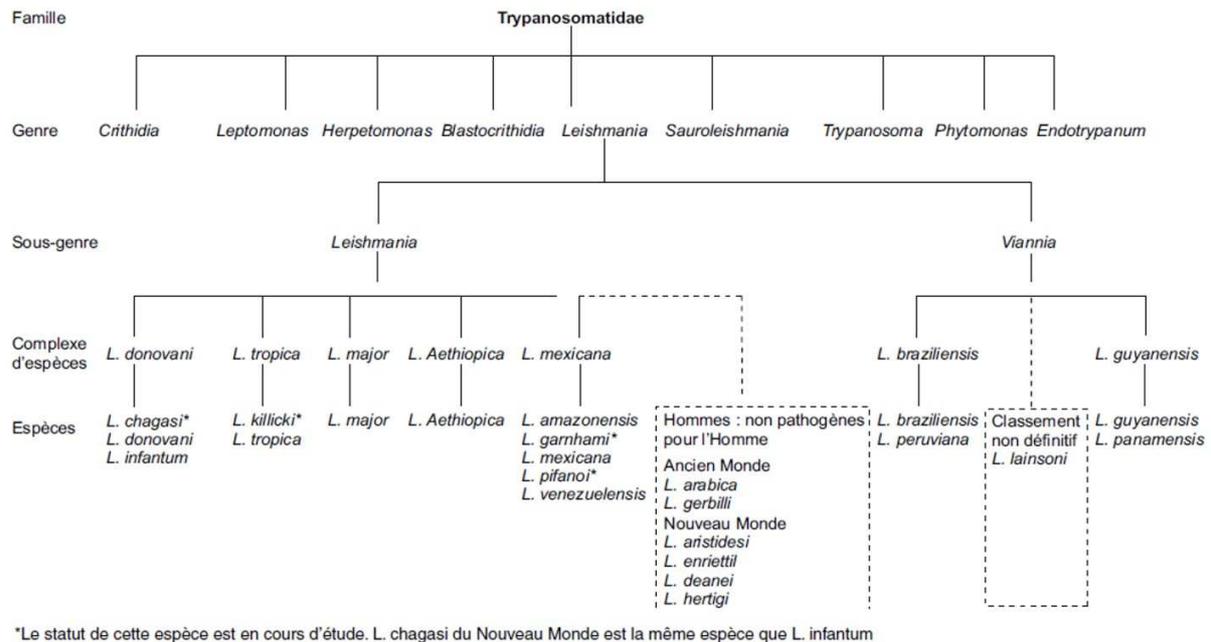


Figure 2-2 Classification simplifiée du genre *Leishmania*.
(D'après la Thèse de Fouad El Baidouri 2012)

La taxonomie utilisée pour classer les espèces de *Leishmania* mise en place par le Pr Rioux est toujours utilisée aujourd'hui (Rioux *et al.*, 1990). Elle est basée sur la caractérisation biochimique des enzymes (zymodème) des souches de *Leishmania*. Tout ce travail d'identification des souches a été réalisé au laboratoire de Parasitologie-Mycologie de Montpellier. Dans le genre *Leishmania* (Figure 2-2) on peut différencier deux sous genres, *Leishmania* et *Viannia*, selon que le parasite se développe respectivement dans la partie centrale ou postérieure de l'intestin du vecteur. On parle aussi d'espèces du Nouveau Monde pour celles retrouvées en Amérique (*Viannia*) et d'espèces de l'Ancien Monde pour celles présentes en Europe, Afrique et Asie (*Leishmania*). Seule une espèce, *L. chagasi* appartenant au Nouveau Monde, s'est avérée être la même espèce que *L. infantum* établie dans l'Ancien Monde (Maurício IL, 2000). La littérature rapporte la présence de souches hybrides entre deux espèces, par exemple

d'hybrides entre *L. major* and *L. infantum* isolés de patients au Portugal (Ravel *et al.*, 2006) (Chargui *et al.*, 2009). Avec la progression du séquençage à haut débit et quasi systématique des génomes, la caractérisation génétique des souches a pris une autre dimension, ce qui pourrait conduire à l'actualisation de la classification.

2.2 Description microscopique

Leishmania est un parasite dimorphique, présent au cours du cycle sous deux formes : les promastigotes, chez le phlébotome, et les amastigotes, chez l'hôte mammifère. L'organisation des organelles chez les kinétoplastidés est polarisée. Ces parasites possèdent des organelles uniques occupant une position subcellulaire bien définie, incluant le noyau (N) contenant l'ADN¹ génomique, un appareil de Golgi, le corps basal du flagelle et un flagelle (F) qui émerge au pôle antérieur du corps cellulaire, une poche flagellaire et enfin une mitochondrie unique qui s'étend tout au long de la cellule. L'ADN mitochondrial unique et condensé forme un réseau appelé le kinétoplaste (K) situé au pôle antérieur. Au cours de la division cellulaire, ces organelles vont se répliquer et se diviser de manière indépendante mais coordonnée.

En culture *in vitro*, les *Leishmania* promastigotes sont des cellules qui se divisent de façon asynchrone et ont une croissance exponentielle.

Au cours du cycle cellulaire, on observe différentes formes que l'on qualifie et quantifie selon la présence d'un ou deux noyaux, kinétoplastes, et flagelles, en proportions de 1K1N1F, 1K1N2F, 1K2N2F et 2K2N2F.

¹ acide désoxyribonucléique

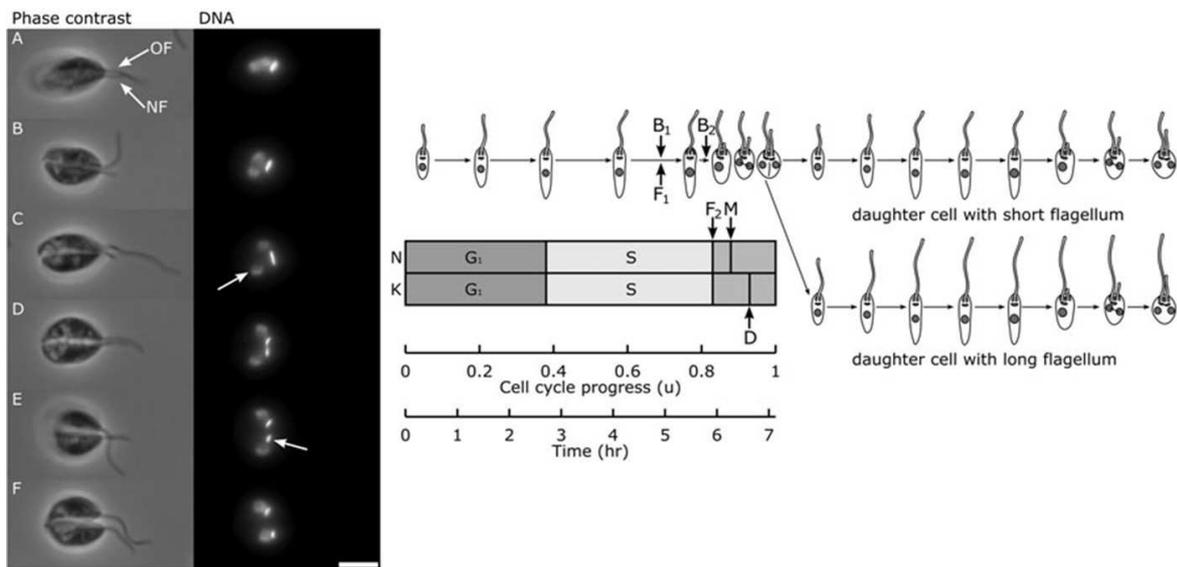


Figure 2-3. *L. mexicana* : schéma d'un cycle de division cellulaire.

Les images en contraste de phase et DAPI A à F montre la chronologie de division d'une leishmanie. La reproduction schématisée montre le cycle de division coordonné mais indépendant du flagelle (F), noyau (N) et kinétoplaste (K). (Wheeler et al., 2011)

La première description détaillée du cycle cellulaire a été réalisée chez *L. mexicana* (Figure 2-3). La mise en place du nouveau (NF) commence tôt dans le cycle en fin de phase G1 (A), cependant on ne le voit émerger de la poche flagellaire que plus tardivement (5,2h). Les réplifications du noyau et du kinétoplaste sont coordonnées mais indépendantes au cours de la phase S (Figure 2-3B), particulièrement longue chez *Leishmania* (2,9h) comparée à *T. brucei* (1,51h) (Woodward & Gull, 1990). Le noyau se divise en premier (Figure 2-3C) suivi du kinétoplaste (Figure 2-3E), enfin l'étape de cytotodiérèse permet l'individualisation des deux cellules filles (Figure 2-3F). Le cycle de division de *L. donovani* est identique à *L. mexicana* (Minocha et al., 2011, Minocha N, 2011).

En culture, en fonction de la densité cellulaire, la taille et la morphologie de la cellule vont évoluer, reproduisant schématiquement les étapes observées chez le vecteur. Lorsque la densité cellulaire augmente de 3.10^6 C/mL à 1.30^7 C/mL, le corps cellulaire s'allonge et s'affine tout comme le flagelle qui s'allonge. Le temps de génération varie de 7 (*L. mexicana*) à 10 (*L. major*) heures en fonction des espèces.

2.3 Cycle de vie de *Leishmania*

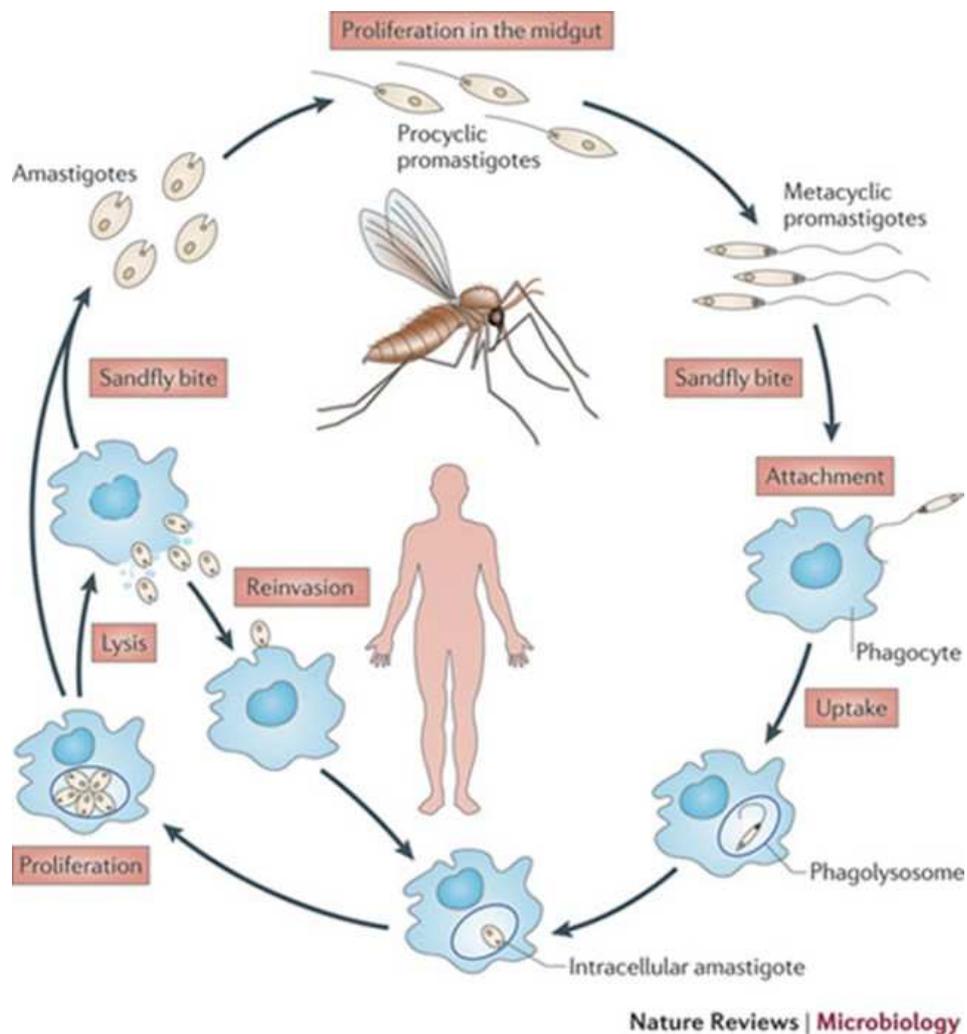


Figure 2-4 Le cycle biologique de *Leishmania*. (Kaye & Scott, 2011)

Leishmania est un parasite dixène, c'est-à-dire ayant besoin de deux hôtes pour réaliser son cycle biologique : un insecte et un hôte mammifère (Figure 2-4). Le parasite *Leishmania* est transmis à l'hôte mammifère par une piqure de **phlébotome**. Lors de son repas sanguin cet insecte hématophage va inoculer dans la peau de l'hôte la forme **promastigote métacyclique** du parasite (Figure 2-5A). Ces promastigotes vont être phagocytés par les macrophages recrutés au site de la piqûre. Une fois le parasite établi dans le macrophage, il va se différencier, s'arrondir et perdre sa mobilité (flagelle très court) : il s'agit du stade **amastigote** (Figure 2-5B). Cette forme non mobile va se multiplier dans un phagolysosome jusqu'à rupture du macrophage. Les amastigotes libérés vont de nouveau pénétrer par endocytose dans un macrophage sain et effectuer

plusieurs cycles de division. Le temps de génération est assez long pour cette forme. En fonction de l'espèce de *Leishmania* responsable de l'infection et d'autres paramètres encore mal connus, les cellules immunitaires infectées vont soit rester au site de l'infection (LC) soit pouvoir migrer dans les organes périphériques comme la rate (LV), le foie et la moelle osseuse et y être séquestrées. Les manifestations cliniques seront différentes en fonction du lieu de résidence final des parasites.

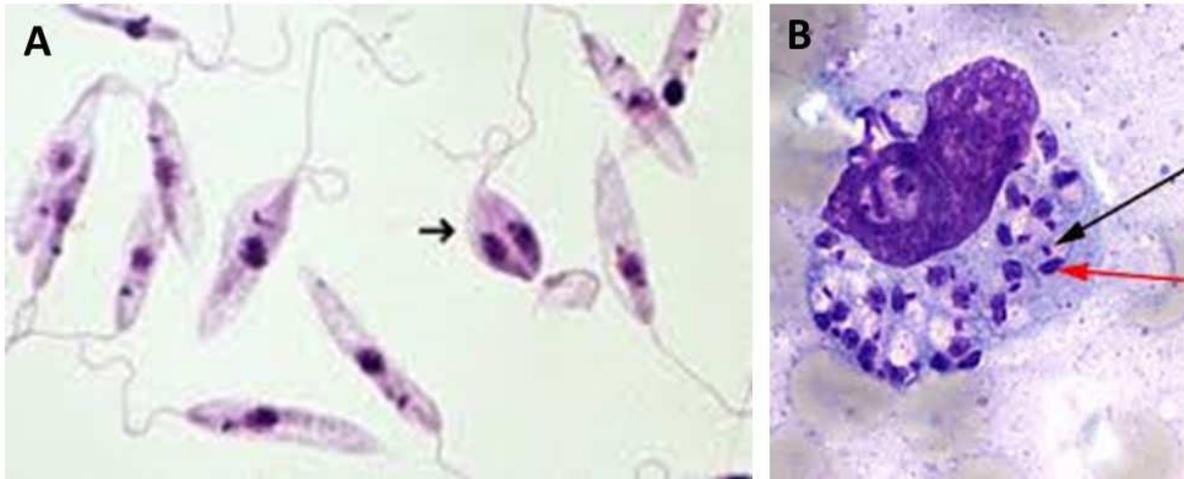


Figure 2-5 Forme promastigote et amastigote de *Leishmania*.

A-Promastigotes en culture, au centre une cellule en division (flèche), noter les deux flagelles dont un plus court et les deux noyaux visibles (mitose fermée); B Amastigotes intracellulaires, noyau et kinétoplaste visibles (flèches rouge et noire respectivement). (source collection du laboratoire)

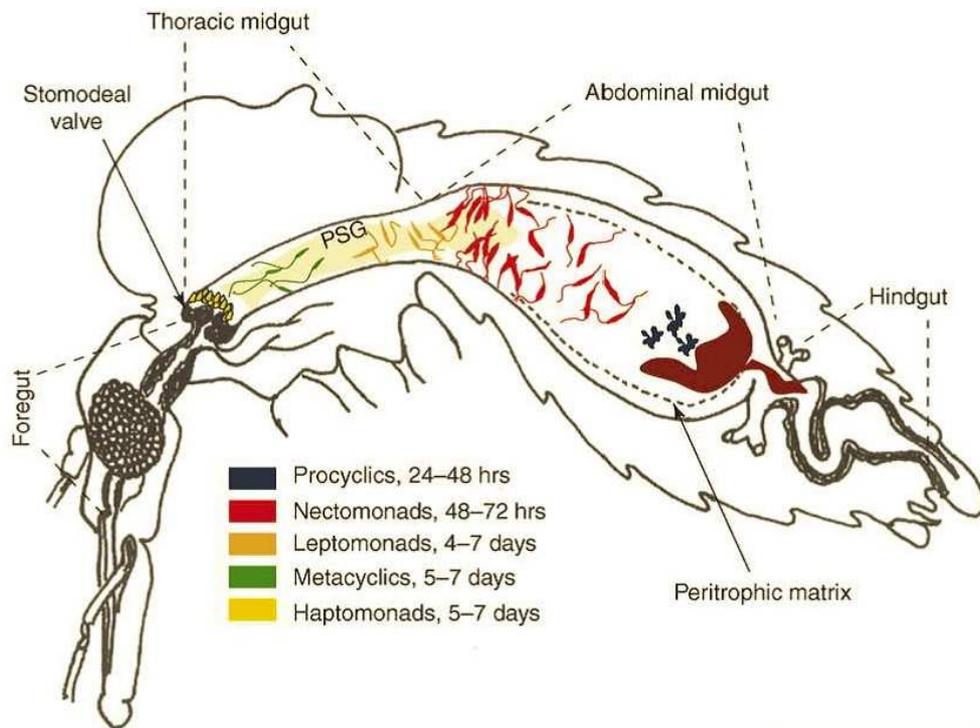
La physiopathologie de la maladie est encore mal connue; cependant, une récente étude a montré que la présence de parasites « apoptotic-like » serait indispensable pour que les parasites vivants puissent infecter et se développer dans les macrophages (Crauwels P 2015).

Lors du prochain repas sanguin, l'insecte vecteur va ingérer des macrophages contenant des amastigotes qui vont poursuivre leur cycle dans l'intestin du phlébotome (Figure 2-6). Le parasite va se multiplier et se différencier en **promastigote procyclique** : il s'allonge et retrouve un flagelle d'une longueur suffisante pour être mobile et se différencie en plusieurs stades promastigotes mal connus jusqu'à atteindre sa forme mature **métacyclique**. Dans le vecteur, les parasites se retrouvent piégés dans la membrane péritrophique, sorte de poche contenant le repas sanguin. Cette poche a pour fonction de protéger l'intestin du phlébotome contre les enzymes digestives qu'il sécrète

lors de la digestion et qui peuvent altérer sa paroi intestinale. Le parasite va devoir s'adapter rapidement à ce nouvel environnement hostile (Figure 2-6). Cette adaptation va se traduire de différentes façons. Tout d'abord il va changer de morphologie au cours de son trajet dans l'intestin et il va réguler à son avantage les enzymes digestives et diminuer les contractions de l'intestin qui permettent la circulation du repas sanguin. En retardant la digestion de l'hôte, le parasite augmente ses chances de persister dans l'intestin pour s'y différencier et s'y multiplier. La dégradation de la membrane péritrophique par les chitinases de l'hôte et du parasite permet la libération des parasites dans la lumière de l'intestin. Ces parasites pourront migrer dans la partie antérieure. La présence du parasite dans l'intestin entraîne une diminution de l'absorbance des nutriments et provoque ainsi un changement du comportement alimentaire du phlébotome.

Les molécules de surface du parasite les plus importantes sont des phosphoglycane et des glycolipides, en particulier protéophosphoglycane (PPG) et lipophosphoglycane (LPG). La production de PPG permet de bloquer l'absorption des nutriments au niveau de l'intestin ce qui va conduire le phlébotome à se nourrir plus souvent ; ceci concourt à assurer une transmission plus efficace du parasite. Le LPG, quant à lui, permet aux parasites d'adhérer aux parois intestinales pour ne pas être éliminés durant la défécation. La structure polymorphe du LPG est considérée comme le facteur principal de compatibilité entre espèces de *Leishmania* et de phlébotome.

Lors de la métacyclogénèse (mal connue) les résidus sucrés du LPG vont être masqués, le parasite pourra se détacher de la paroi de l'intestin et être transmis à un hôte mammifère. C'est à ce moment que le parasite migre jusqu'à la valve œsophagienne et au pharynx, où il ne se divise plus et adopte une morphologie petite, fine et allongée. Ces formes sont les formes infectieuses qui seront transmises à un nouvel hôte au cours d'un futur repas sanguin. La différenciation entre le stade amastigote, puis promastigote procyclique (différents stades) jusqu'à la forme promastigote métacyclique "non répliquative" infectieuse, va durer de 6 à 9 jours selon les espèces (Kamhawi, 2006).



TRENDS in Parasitology

Figure 2-6 Cycle biologique du parasite *Leishmania* chez un vecteur compétent. Illustrant les différentes formes morphologiques des promastigotes au cours de leur séjour dans l'intestin du phlébotome. (Tirée de Kamhawi S, 2006)

Au cours de son cycle de vie *Leishmania* est confronté à différents environnements et doit constamment s'adapter aux changements de conditions physiologiques, notamment de pH, température, concentration en oxygène et réponse immunitaire de l'hôte...

2.4 Vecteurs et réservoirs

2.4.1 Vecteurs

Il existe approximativement 900 espèces de phlébotomes (Figure 2-7), dont 70 sont impliquées dans la transmission de la leishmaniose. Les phlébotomes sont retrouvés principalement dans des régions chaudes du globe incluant: l'Europe du sud, l'Asie, l'Afrique, l'Australie et le sud et le centre de l'Amérique. Les espèces de *Leishmania* issues de l'ancien monde sont transmises par des phlébotomes du genre *Phlebotomus* alors que le genre *Lutzomyia* est le vecteur des espèces du nouveau monde. Par exemple, *P. perniciosus* et *P. ariasi*, sont responsables de la transmission de la leishmaniose canine dans le bassin méditerranéen et l'Inde; et *Lu. longipalpis* est le principal vecteur

incriminé dans la transmission de *L. chagasi* (syn. *L. infantum*) en Amérique du Sud (Ready, 2013).



Figure 2-7 Femelle phlébotome prenant son repas sanguin.

On peut différencier des espèces de phlébotome permissives et des espèces restrictives au développement de certaines espèces de *Leishmania* ; ceci est en lien direct avec la compatibilité entre les lectines du vecteur et les LPG du parasite. Par exemple, *P. papatasi* et *P. sergenti* sont dites restrictives puisqu'elles permettent seulement le développement et la transmission de *L. major* et de *L. tropica*, respectivement. Contrairement à *Phlebotomus halepensis*, *Phlebotomus argentipes* et *Lutzomyia longipalpis* qui soutiennent en laboratoire le développement de plusieurs espèces de *Leishmania*, incluant *L. major* et *L. tropica* (Kamhawi et al., 2000, Sacks & Kamhawi, 2001).

2.4.2 Réservoirs

Différentes espèces de rongeurs sont les principaux réservoirs pour la leishmaniose cutanée à *L. major*; et de nombreux mammifères sont réservoirs des leishmanioses cutanées du Nouveau Monde. Le vrai réservoir de la leishmaniose cutanéomuqueuse à *L. braziliensis* n'est pas connu. Concernant la leishmaniose viscérale, il n'y a pas de réservoir animal pour *L. donovani* (kala-azar) mais le chien (et dans une moindre mesure les canidés sauvages) constituent le réservoir de *L. infantum*.

2.5 Epidémiologie & manifestations cliniques

2.5.1 Manifestations cliniques

Source OMS (<http://www.who.int/mediacentre/factsheets/fs375/fr/>)

Les leishmanioses sont endémiques dans 98 pays répartis sur les 5 continents et différentes espèces sont responsables de différentes formes de la maladie. L'OMS estime que 12 millions de personnes sont touchées par la leishmaniose et un vingtième de la population mondiale serait à risque. Plus de 20 espèces sont connues pour infecter l'homme. Les manifestations cliniques sont variées, on retrouve la leishmaniose viscérale (LV), la leishmaniose dermique post-kala-azar (PKDL), la leishmaniose cutanée (LC) et la leishmaniose cutanéomuqueuse (LCM). Les lésions cutanées sont principalement dues à la réponse immunitaire de l'hôte envers le parasite et non au parasite lui-même. La forme clinique de la maladie va dépendre de l'espèce de leishmanie mais aussi du type de réponse immunitaire que va développer l'hôte.



Figure 2-8 Espèces de *Leishmania* provoquent un large éventail de maladies.

A lésion de la peau provoquée par la leishmaniose cutanée, B leishmaniose muco-cutanée, C leishmaniose viscérale.

(Source Philippe Desjeux, O.M.S)

La **leishmaniose viscérale** (LV) également connue sous le nom de « kala-azar » en Inde est la forme la plus sévère, et est mortelle en l'absence de traitement. Selon les estimations, 200 000 à 400 000 nouveaux cas de LV surviennent chaque année, entraînant environ 30.000 décès. La LV se caractérise par des poussées de fièvre, une perte de poids, une hépatosplénomégalie (augmentation du volume du foie et de la rate) et une anémie. La LV est fortement endémique dans le sous-continent indien et en Afrique de l'Est où l'espèce responsable est *L. donovani* alors que dans le Bassin Méditerranéen les cas de LV sont dus à *L. infantum* (Figure 2-8C).

La **leishmaniose dermique post-kala-azar** (LDPKA ou PKDL) est une complication tardive de la leishmaniose viscérale à *L. donovani*. Elle se manifeste par une éruption cutanée localisée sur le visage, la partie supérieure du bras, le tronc et d'autres parties du corps. Cette forme est principalement rencontrée en Afrique de l'Est et dans le sous-continent indien, où elle peut être respectivement développée par 50%, au plus, et 5-10% des patients atteints du kala-azar respectivement dans ces deux régions. La LDPKA survient généralement 6 mois, un an ou plusieurs années après la guérison apparente du kala-azar, mais elle peut également se déclarer plus tôt.

La **leishmaniose cutanée** (LC) est la forme la plus fréquente. Environ 95% des cas de LC surviennent dans les Amériques, dans le bassin méditerranéen, au Moyen-Orient et en Asie centrale. On estime qu'il y a 0,7 million à 1,3 million de nouveaux cas chaque année dans le monde. La LC se présente sous forme de lésions cutanées, principalement des ulcères, causant des cicatrices définitives sur les parties exposées du corps au niveau de la pique de l'insecte. Ces lésions représentent un handicap sévère pour les populations touchées notamment en termes d'intégration sociale. On distingue la leishmaniose cutanée localisée qui peut guérir spontanément en deux à six mois de la LC diffuse qui ne guérit pas spontanément et qui est difficile à traiter (Figure 2-8A).

La **leishmaniose muco-cutanée** essentiellement due à l'espèce *L. braziliensis*, détruit partiellement ou totalement les muqueuses du nez, de la bouche et de la gorge. Près de 90% des cas de leishmaniose muco-cutanée surviennent en Amérique du Sud et Centrale (Figure 2-8B).

La **leishmaniose canine** (LCan) est une zoonose importante dans le bassin Méditerranéen due à *L. infantum*. Dans cette région, les chiens atteints de leishmaniose sont un réservoir pour la LV humaine. La maladie déclarée est mortelle en l'absence de traitement. La majorité des chiens porteurs sont cependant asymptomatiques.

Tableau 2-1 Manifestations cliniques en fonction de l'espèce de *Leishmania*.
Adapté de (Kaye & Scott, 2011)

Manifestations cliniques	Espèces
<i>Ancien Monde, sous-genre Leishmania</i>	
Leishmaniose viscérale	<i>L. donovani</i> et <i>L. infantum</i>
Leishmaniose cutanée	<i>L. major</i> , <i>L. tropica</i> , <i>L. aethiopica</i> , <i>L. infantum</i>
Leishmaniose cutanée diffuse	<i>L. aethiopica</i>
<i>Nouveau Monde, sous-genre Leishmania</i>	
Leishmaniose viscérale	<i>L. infantum</i>
Leishmaniose cutanée	<i>L. mexicana</i> , <i>L. pifanoi</i> , <i>L. amazonensis</i> , <i>L. infantum</i>
Leishmaniose cutanée diffuse	<i>L. mexicana</i> , <i>L. amazonensis</i>
<i>Nouveau Monde, sous-genre Viannia</i>	
Leishmaniose cutanée	<i>L. braziliensis</i> , <i>L. guyanensis</i> , <i>L. panamensis</i> , <i>L. peruviana</i>
Leishmaniose cutanéomuqueuse	<i>L. braziliensis</i> et <i>L. panamensis</i>

2.5.2 Zones d'endémie

La leishmaniose est présente dans 5 des 6 continents et 98 pays sont touchés par cette maladie selon l'OMS (Figure 2-9). Toutefois, l'endémicité est hétérogène. Ainsi, 90% des cas de LV sont retrouvés dans 5 foyers : au Bangladesh, au Brésil, en Ethiopie, en Inde et au Soudan. La LC est plus largement répartie sur le globe, en effet 1/3 des cas sont retrouvés en Amérique du sud, dans le Bassin Méditerranéen et en Asie. L'Afghanistan, l'Algérie, le Brésil, la Colombie, le Costa Rica, l'Ethiopie, l'Iran, le Pérou, le Soudan et la Syrie sont les 10 pays où le plus grand nombre de cas sont recensés.

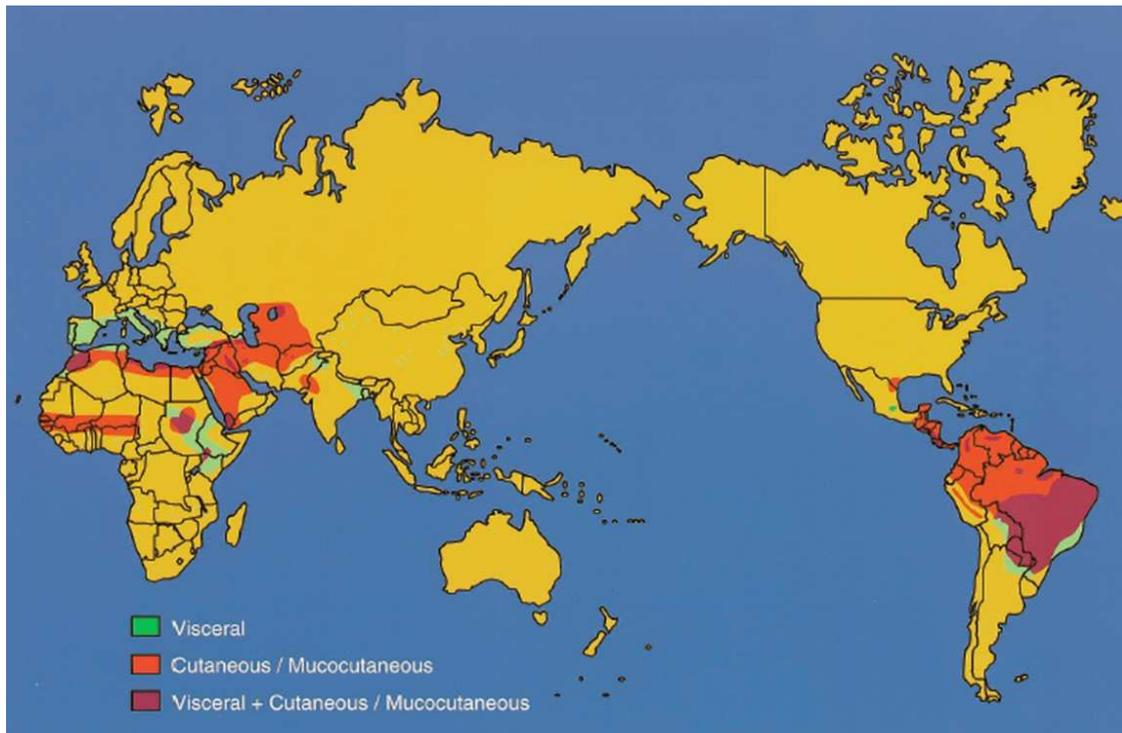


Figure 2-9 Mappemonde représentant les zones endémiques de leishmaniose.

En vert, orange et violet sont représentées respectivement les régions du monde atteintes par la leishmaniose viscérale, cutanée/cutanéomuqueuse et les trois réunies. (Handman, 2001)

2.6 Moyens de lutte et prophylaxie

La lutte anti-vectorielle et la protection contre les vecteurs sont actuellement le seul moyen de lutte mis en œuvre contre les leishmanioses. En Europe, la surveillance et la prise en charge des chiens par leurs propriétaires peut aussi réduire les réservoirs parasites et ainsi limiter la transmission à l'Homme. Différents produits comme la deltaméthrine dirigés contre les vecteurs de la leishmaniose sont utilisés pour protéger les chiens, mais l'instabilité de certaines molécules nécessite un traitement régulier des animaux. Concernant le traitement de la Lcan, le développement de vaccin (de Mendonca *et al.*, 2016) et l'utilisation de molécules telles que les dérivés antimoniaux, la miltéfosine, l'amphotéricine B ou encore l'allopurinol ont montré une certaine efficacité; mais les traitements sont longs, doivent être répétés tous les ans, voire continuellement, et ont des effets secondaires nocifs pour les animaux, pour revue voir (Reguera RM 2016). De plus, l'apparition de résistances chez le parasite n'est pas en faveur de l'éradication de la maladie ; il est donc essentiel de trouver de nouveaux principes actifs et d'identifier de nouvelles cibles thérapeutiques.

Chez l'homme, l'absence de vaccins efficaces, l'arsenal thérapeutique très limité et le développement de résistances médicamenteuses sont des inconvénients majeurs pour la gestion des cas cliniques et l'éradication de la maladie (den Boer *et al.*, 2011).

2.7 Particularités génomiques

2.7.1 Organisation du génome nucléaire

Les espèces de *Leishmania* de l'Ancien Monde possèdent 36 paires de chromosomes, dont la taille est comprise entre 0.28 à 2.8 Mb (Wincker *et al.*, 1996) alors que les espèces du Nouveau Monde possèdent 34 ou 35 paires de chromosomes. Les espèces *L. mexicana* et *L. amazonensis* ont subi deux évènements de fusion entre les chromosomes 8+29 d'une part et 20+36 d'autre part; pour le groupe *L. braziliensis* ce sont les chromosomes 20 et 34 qui ont fusionné (Britto *et al.*, 1998). Ces évènements de fusion ont été mis en évidence par Southern blot grâce à des sondes spécifiques de chaque chromosome après séparation par PFGE².

Nous verrons plus tard que le terme « paire de chromosomes homologues » n'est pas forcément le plus adapté à ce parasite.

Chez *Leishmania* l'organisation des régions subtélomériques est très conservée. Contrairement à *T. cruzi* et *T. brucei* où ces régions sont très variables. En effet, chez *T. brucei* ces régions contiennent une famille multigénique codant des glycoprotéines variantes de surface (VSG³). Ces protéines forment un manteau sur la membrane externe du parasite. Ces VSG permettent au parasite d'échapper au système immunitaire de l'hôte par le fameux mécanisme de variation antigénique (Horn, 2014).

2.7.2 Spécificités d'espèces

Le génome de *Leishmania spp* est riche en GC, environ 60% - jusqu'à 62.5% dans les régions codantes et 57.3% dans les régions intergéniques de *L. major*. La taille du génome de *L. major* est de 33Mb. En 2005, il a été le premier génome séquencé parmi les espèces de *Leishmania* (Ivens *et al.*, 2005). Les techniques de séquençage génomique (Fiebig *et al.*, 2015, Alonso *et al.*, 2016) et de RNA-seq (Alonso *et al.*, 2016) ont

² De l'anglais Pulse Field gel electrophoresis

³ De l'anglais Variant surface gene

contribués à améliorer l'annotation des génomes, pour exemple le génome de *L. major* Friedlin compte 9378 gènes (TriTrypDB 29 Released).

Chez ce parasite, on peut répartir les gènes en trois grandes catégories. La première englobe 85% des gènes, ils codent des protéines essentielles au métabolisme et sont communs à tous les kinétoplastidés. Il est possible de retrouver des homologues chez les Eucaryotes supérieurs. Dans une seconde catégorie, on classe les gènes retrouvés uniquement chez les kinétoplastidés, ils codent des protéines du cytosquelette, des enzymes métaboliques et des composants de la machinerie d'édition de l'ARN⁴, ces gènes ne possèdent pas d'homologue chez les Eucaryotes supérieurs. Enfin, la troisième catégorie regroupe les gènes spécifiques au genre *Leishmania*, ils représentent $\approx 12\%$ du génome. Ces gènes spécifiques restent encore à caractériser pour la plupart ; les protéines qu'ils codent ne possèdent pas de domaine conservé ou identifiable, ce qui suggère qu'une partie de ces gènes peuvent avoir un rôle important dans la pathogénicité.

Plus tard, il a été choisi de séquencer le génome de *L. infantum* et *L. braziliensis* dans l'optique d'identifier des gènes spécifiques d'espèce qui pourraient être responsables des manifestations cliniques de la maladie propre à chaque espèce (Peacock *et al.*, 2007). La comparaison des trois génomes de *L. major*, *L. infantum* et *L. braziliensis* a permis d'observer certaines des disparités et similitudes entre ces espèces (Figure 2-10). Malgré la fusion de certains chromosomes chez *L. braziliensis*, la synténie entre les gènes est parfaitement conservée entre les trois espèces, seuls ≈ 200 gènes seraient distribués différemment. *L. braziliensis* est cependant l'espèce la plus divergente d'un point de vue génétique et biologique. Cette dernière possède 47 gènes qui sont absents des deux autres espèces. En comparaison, *L. infantum* possède 27 gènes spécifiques d'espèce et seulement cinq pour *L. major*.

⁴ acide ribonucléique

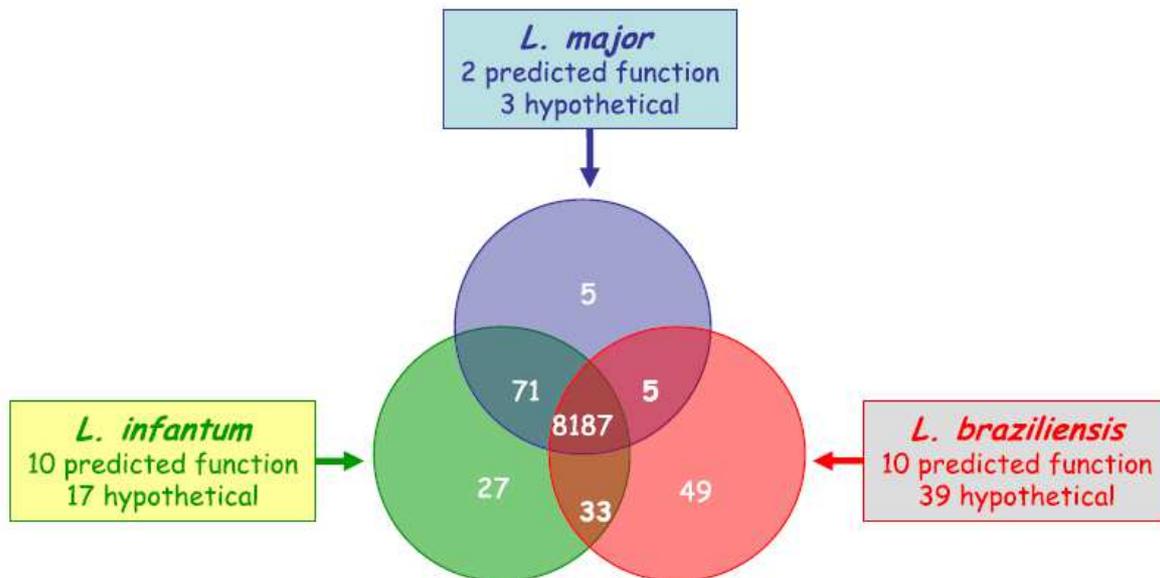


Figure 2-10 Diagramme de Venn représentant le nombre de gènes spécifiques d'espèces chez trois espèces de *Leishmania*. (Smith et al 2007).

Certains de ces gènes spécifiques d'espèces ont fait l'objet d'études fonctionnelles, comme le gène A2, décrit chez *L. infantum* et *L. donovani*, qui code pour une protéine spécifique des amastigotes et qui confère un changement de virulence lorsqu'il est introduit chez *L. major* (Zhang et al., 2003).

Une des différences majeure entre les génomes de ces espèces séquencées est le nombre de copies de gènes au sein de certaines familles géniques. Ces familles peuvent se trouver sur le même chromosome en répétition en tandem ou être disperser sur plusieurs chromosomes.

Leishmania à la différence de *T. brucei* est dépourvue de VSG en position sub-télomérique et ne possède pas de mécanisme de variation antigénique. En revanche cette espèce a développé une large famille de phosphoglycanase- β -galactosyle-transférases (SCGs). Ces enzymes sont impliquées dans la modification des chaînes de phosphoglycane composant des LPG⁵, protéine majeure de surface des promastigotes. D'autres protéines importantes sont les GP63: cette famille multigénique est localisée dans la zone codante du chromosome et codent des zinc-métalloprotéases. Ces dernières sont essentielles pour la virulence et l'invasion de la cellule hôte. Ce locus est limité à quatre et six copies pour *L. major* et *L. infantum*, respectivement, mais s'étend à huit fois plus de copies chez *L. braziliensis* sur le chromosome 10. Il a été suggéré que GP63

⁵ Lipo Proteo Glycane

pouvait être impliquée dans la préférence d'hôte lors de l'infection (Mauricio *et al.*, 2006).

D'autres gènes comme PSA-2, GP46 ou les amastines diffèrent en nombre de copies en fonction des espèces et sont associés à des facteurs de virulence (pour revue voir (Smith *et al.*, 2007)). Il a également été observé que certaines espèces de *Leishmania* sont porteuses de virus ARN (LRV⁶). Ces virus infectent principalement les espèces du Nouveau Monde (*Viannia*). Les patients infectés par ces espèces ont un haut risque que leur LC évolue en LCM. Les LRV semblent être responsables de l'exacerbation de la maladie (Zangger *et al.*, 2013) et seraient impliqués dans la persistance et la survie de *L. guyanensis* dans les macrophages (Eren *et al.*, 2016).

En 2012, toujours dans l'optique de trouver des gènes spécifiques d'espèces, Raymond *et coll.* ont séquencé le génome de *L. tarentolae*, parasite de lézard vivant sous forme de promastigote dans l'intestin ou circulant dans le sang de son hôte. *L. tarentolae* n'est pas pathogène pour l'Homme, la communauté s'attendait donc à identifier des gènes responsables de la virulence en comparant le génome de *L. tarentolae* avec celui des trois pathogènes humains. Cependant ce ne fut pas le cas, les génomes se sont, encore une fois, révélés hautement similaires (Raymond *et al.*, 2012). Toutefois les gènes absents chez le pathogène de lézard sont ceux exprimés spécifiquement dans les formes amastigotes chez les espèces spécifiques de l'homme, ce qui n'est pas étonnant puisque *L. tarentolae* ne serait pas capable de se différencier en amastigote chez le lézard.

Ainsi, la comparaison des génomes des différentes espèces de *Leishmania*, montre qu'il y a très peu de gènes spécifiques d'espèces ce qui a été une déception. Aujourd'hui, nous ne pouvons toujours pas expliquer clairement les différentes manifestations cliniques de la maladie, pourtant causées par des espèces bien distinctes.

2.7.3 Organisation des gènes et transcription

Chez les trypanosomatidés, les gènes ne contiennent pas d'introns. Par ailleurs, ils sont organisés de façon unique dans le monde animal, en longues unités transcriptionnelles directionnelles (DGCs⁷) séparées par des régions riches en A/T nommées *strand switch regions* (SSR). Ces SSR peuvent être convergentes ou divergentes selon que la région se

⁶ Leishmania RNA Virus

⁷ De l'anglais directional gene cluster

trouve entre deux DGCs qui finissent ou deux DGCs qui commencent (Martinez-Calvillo *et al.*, 2003). Certains chromosomes (Chr), comme les Chr 1/2/4/18/19/26, ne contiennent qu'une seule région switch qui est divergente. Au niveau de ces SSR peuvent se trouver certains gènes comme celui codant le mini exon (Chr.2) ou les ARN de transfert répartis principalement sur les chromosomes 9, 31, 36. Les plus grands chromosomes peuvent contenir jusqu'à 16 SSRs associés à une alternance de DGCs divergentes et convergentes en prenant en compte les SSR créées par les ARNt. Au total le génome de *L. major* contient 150 SSR. Les DGCs ne contiennent pas de groupe de gènes associés à une même fonction comme c'est le cas chez les procaryotes avec les opérons. On retrouve principalement des gènes sans relation fonctionnelle, et certains gènes en plusieurs copies, ce qui peut être un moyen pour réguler le niveau d'expression de ces gènes.

Le parasite *Leishmania*, comme les autres membres de la famille des trypanosomatidés possède un mécanisme unique de transcription des gènes. La plupart des gènes sont transcrits de façon polycistronique à partir d'un "promoteur" bidirectionnel (ou site d'initiation de la transcription) situé au niveau des SSR d'où divergent deux DGCs (Martinez-Calvillo *et al.*, 2003). Ces polycistrons sont transcrits par la RNA Polymérase II et simultanément maturés en ARN messagers (ARNm). Ce processus de maturation se compose de deux étapes couplées : (i) le trans-épissage qui correspond à l'ajout d'une séquence « cap » de 39 nucléotides (miniexon ou spliced leader, SL) en 5' et (ii) une polyadénylation classique en 3' du messenger (LeBowitz *et al.*, 1993) (Figure 2-11). Les sites d'initiation de la transcription (TSS⁸) des PTU⁹ sont matérialisés par des marques épigénétiques d'acétylation de l'Histone 3 (Thomas *et al.*, 2009) et les bases J marquent les site de terminaison (TTS¹⁰) des PTU (van Luenen *et al.*, 2012).

De façon surprenante, quasiment aucune régulation de la transcription n'a pu être mise en évidence chez les trypanosomatidés, laissant une grande place aux autres mécanismes de régulation de l'expression génique. En particulier aucun promoteur n'a pu être caractérisé.

⁸ De l'anglais Transcription start site

⁹ De l'anglais PolyTranscript Unit

¹⁰ De l'anglais Transcription termination site

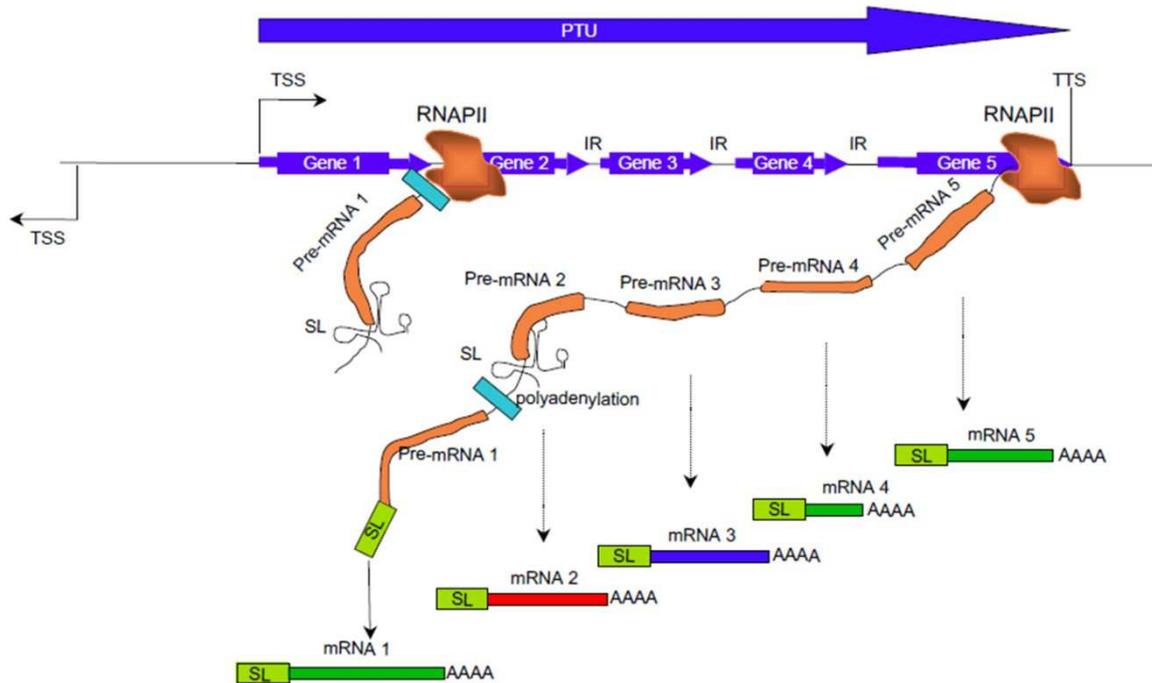


Figure 2-11 Transcription polycistronique chez *Leishmania*.

La transcription des PTU est initiée au TSS par le RNA polymérase II et se termine au niveau des TTS. Elle va produire un pré messager polycistronique. Chaque messager sera individualisé (maturé) par le mécanisme de trans-épissage qui consiste en l'ajout d'un spliced leader (SL) en 5' et d'une queue poly A en 3'. (Lombrana *et al.*, 2016)

2.7.1 Base J

Les trypanosomatidés présentent une modification particulière au niveau de l'ADN : la base J. Cette modification a en premier lieu été identifiée chez *T. brucei*. Ce sont les équipes d'Etienne Pays et de Piet Borst qui ont indépendamment suspecté une modification de l'ADN associée à l'inactivation des sites d'expression des VSGs (Bernards *et al.*, 1984); (Pays *et al.*, 1984). Cependant, cette modification n'était ni une méthylation, ni détectable par séquençage. Ce n'est que bien plus tard, entre 1991 et 1997 que la nature et la structure de cette modification a pu être élucidée (Gommers-Ampt *et al.*, 1991, Gommers-Ampt *et al.*, 1993).

La base J est également retrouvée chez les autres kinétoplastidés, notamment *Leishmania* et *T. cruzi* (dans les deux de cycles de vie) ainsi que chez les *Diplonema* et *Euglena* (Dooijes *et al.*, 2000, Borst, 2016). La base J résulte de la modification d'une thymidine en deux étapes (Figure 2-12): (i) une hydrolase (TH) va oxyder le groupement exocyclique-méthyl d'une thymidine formant le composé intermédiaire

HOMedU, (ii) le HOMedU est ensuite converti en β -D-glucosyl-5-(hydroxyméthyl) uracil (base J) par addition d'un glucose catalysé par une glucosyl-transférase.

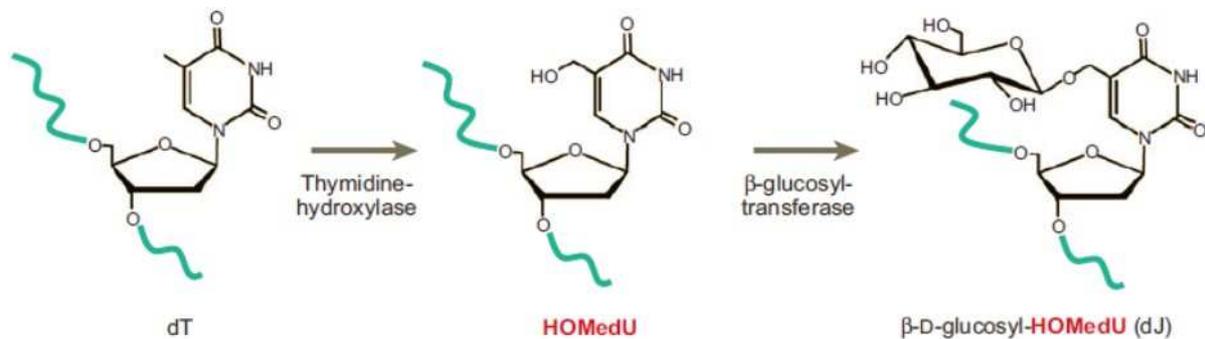


Figure 2-12 Biosynthèse de la base J par la modification spécifique de la base thymine sur l'ADN.

D'après (Borst & Sabatini, 2008)

Chez *Leishmania* cette base J représente environ 1% du génome nucléaire et 98% sont retrouvées au niveau des télomères. Le pourcentage restant de base modifiée est interne aux chromosomes et est appelée ij. Leurs localisations ont pu être déterminées par ChIP-seq à l'aide d'un anticorps anti-base J. L'analyse du ChIP-seq a mis en évidence trois localisations principales : (i) les pics se situant à la convergence de deux unités de transcription (PTUs¹¹) où se termine la transcription (28/29 des PTUs convergentes)(Figure 2-13A); (ii) dans les zones d'initiation de la transcription mais seules 16 des 58 régions divergentes contiennent des ij et il s'agit des SSR les plus grandes (van Luenen *et al.*, 2012), à cet endroit la présence de base J est toujours encadrée par deux pics d'histones H3 acétylées (H3ac)(Figure 2-13B); et enfin (iii) des pics de ij sont retrouvés dans des régions adjacentes à la présence d'H3ac (Figure 2-13C).

¹¹ De l'anglais polycistronic transcription units

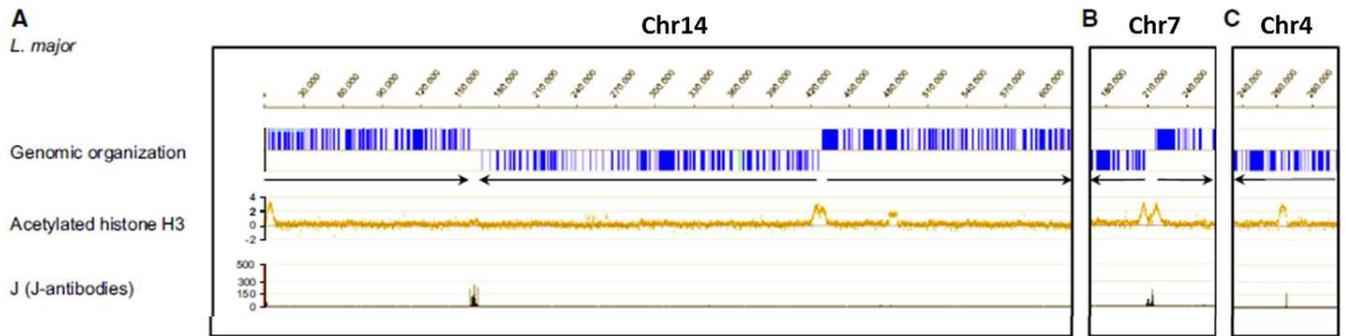


Figure 2-13 Positions des bases J chez *L. major*.

En bleu sont représentés les PTU, les flèches noires indiquent le sens de transcription. Le graphique jaune représente les positions des Histones H3 acétylées et le noir correspond à la position des bases J. (van Luenen et al., 2012)

Deux protéines JBP¹²¹ et JBP2 sont en lien avec la présence des bases J dans le génome. Chez des *Leishmania* KO pour JBP2 on observe une perte de base J qui conduit à un non arrêt de la terminaison de la transcription, autrement dit la RNA PolIII ne se décroche pas et continue de transcrire en arrivant au niveau des SSRs. Le rôle de ces deux protéines est encore mal connu et surtout on ne sait pas s'il existe des partenaires qui permettent de réguler leur activité et de reconnaître leurs cibles.

Chez *Leishmania* la base J serait un signal de terminaison de la transcription pour la RNA polymérase II et aurait un rôle dans la régulation de l'expression des gènes situés en fin de PTU (Reynolds *et al.*, 2016). Du fait de sa localisation aux télomères elles pourraient être impliquées dans leur maintenance. Chez *T. brucei* la modification a lieu sur le deuxième T du motif répété GGGTTA au niveau des télomères (Genest *et al.*, 2015, Cliffe LJ 2016).

2.7.2 Génome mitochondrial

Le génome mitochondrial ou kinétoplaste (ADNk) de *Leishmania* est composé de deux molécules circulaires concaténées (Simpson, 1987) : plusieurs milliers de mini-cercles et une cinquantaine de maxi-cercles. L'ADNk code principalement les ARN ribosomiques (rRNA) et des enzymes de la chaîne respiratoire. Les messagers issus subissent une maturation particulière appelée « editing » qui consiste en l'insertion/ délétion d'uridine

¹² J binding protéine

à des sites précis. Les ARN guide (gRNA) nécessaires au processus d'édition des ARNm sont codés par les mini-cercles. Pour revue voir (Estevez & Simpson, 1999).

La réplication de cet ADNk est possible grâce à l'intervention de la topoisomérase II qui va relaxer les cercles concaténés (Marini *et al.*, 1980). Les mini-cercles ainsi libres se répliqueront selon une structure en θ (Perez-Morga & Englund, 1993).

2.7.3 Plasticité du génome

2.7.3.1 Amplification génique

La littérature rapporte plusieurs indices de la plasticité génomique chez *Leishmania*, notamment des différences de taille entre chromosomes homologues, elles-mêmes reflétant des amplifications intra-chromosomiques dans des régions subtélomériques (Blaineau *et al.*, 1991, Myler *et al.*, 1994) ou dans le corps du chromosome (Iovannisci & Beverley, 1989).

Entre autre, le laboratoire de Marc Ouellette au Québec a particulièrement étudié les évènements de remaniements génomiques impliqués dans la résistance aux drogues chez *Leishmania* (Leprohon *et al.*, 2009). Ils ont observé des amplifications extra-chromosomiques circulaires (Figure 2-14 A), des amplifications linéaires (Figure 2-14 B), de l'aneuploïdie (amplification de chromosome entier) et dans certains cas des délétions de gènes. Ces remaniements sont bénéfiques pour le parasite, ils permettent de contrôler le nombre de copies d'un gène ou d'un groupe de gènes. Cette plasticité du génome pourrait jouer un rôle important lors de l'acquisition de résistance ou dans la survie du parasite soumis à un stress environnemental. Ainsi une population qui aura sélectionné le bon pool de gènes sera avantagée pour survivre dans ces conditions. Il semble que ces mécanismes soient un moyen de palier à l'absence de régulation transcriptionnelle et ainsi un moyen de réguler l'expression des gènes.

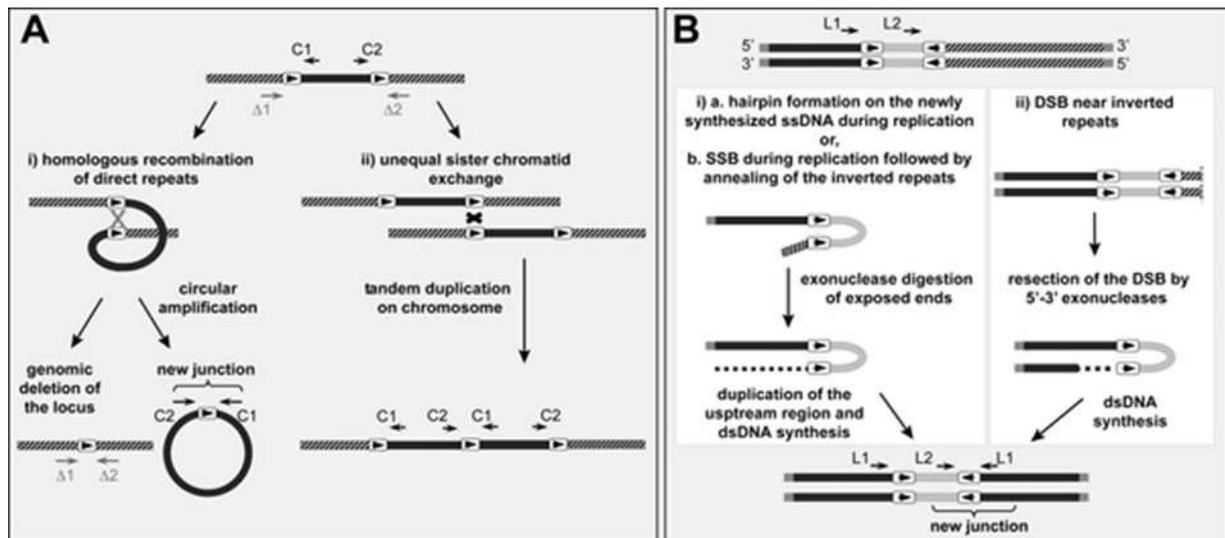


Figure 2-14 Modèle d'amplification génique chez *Leishmania*.

A amplifications extra-chromosomiques circulaires ; B amplifications linéaires (Ubeda et al., 2014). C1, C2, L1 et L2 sont les amorces permettant la mise en évidence le type d'amplification génique observé.

Les amplifications **extra-chromosomiques circulaires** sont rendues possibles par la présence de séquences répétées inversées situées de part et d'autre du gène amplifié par le parasite. En général la taille de ces séquences répétées sont comprises entre 0,2 et 0,25kb et possèdent un minimum de 85% d'identité. Elles sont généralement séparées de 1 à 100kb. L'origine de ces séquences répétées correspond à des rétrotransposons dégénérés qui servent à réguler les gènes au niveau post transcriptionnel (SIDER1) et traductionnelle (SIDER2), (Muller *et al.*, 2010).

Les **amplifications linéaires** sont quant à elles rendues possible par la présence de répétitions en tandems, codantes ou le plus souvent non-codantes, et elles se situent essentiellement dans les régions sub-téломériques, autorisant ainsi la stabilité des mini-chromosomes supplémentaires formés (Iovanisci & Beverley 1989, Blaineau et al. 1991; Dubessay et al. NAR 2001).

La présence de répétitions courtes, dispersées dans tout le génome et cartographiées à l'aide de la bio-informatique, a en fait permis d'expliquer nombre d'amplifications circulaires et linéaires (Ubeda *et al.*, 2014). Cette étude a montré qu'il existe au total 1886 couples de ces séquences répétées, directes ou inversées, chez *L. infantum*, chacune pouvant potentiellement donner lieu à des amplifications, et 60% étant communes avec *L. major* contre seulement 16% avec *L. braziliensis* (Ubeda *et al.*, 2014). La recombinaison RAD51 joue un rôle important dans la formation d'amplicons

circulaires et MRE11 est impliquée dans la formation des amplicons linéaires (Laffitte *et al.*, 2014). En parallèle de l'amplification génique par le même principe il a été observé des événements de délétion de gènes. Selon Ubeda et collaborateurs, L'amplification du chromosome entier ou **aneuploïdie** serait retrouvée lorsqu'il n'y a pas de séquence répétée (SR) à proximité du ou des gènes nécessaire(s) au parasite.

2.7.3.2 Aneuploïdie mosaïque

Le terme d'aneuploïde décrit le fait de posséder un chromosome en plus ou en moins par rapport à sa ploïdie de référence. Cet état de ploïdie anormale chez les eucaryotes supérieurs est considéré comme un phénomène délétère. Il est associé à des maladies génétiques (trisomie 21), la mort du fœtus et est observé dans différents cancers.

Depuis 25 ans, le laboratoire où j'effectue ma thèse s'intéresse aux chromosomes de *Leishmania*. Grâce à la technique du PFGE¹³ ils ont commencé à faire la lumière sur le caryotype de *Leishmania*. Ils se sont néanmoins aperçus que, les gels n'étaient pas si simples à interpréter. Ils ont observé sur les gels de PFGE des intensités différentes entre les bandes (Scholler *et al.*, 1986, Blaineau *et al.*, 1991), du polymorphisme dans la taille des chromosomes homologues entre espèces mais aussi au sien d'une même souche (Chr 1/2/27 chez *L. major*) (Pagès *et al.*, 1989). Le nombre de chromosomes est resté polémique jusqu'en 1996 où les travaux de Wincker *et al.* ont démontré la présence de groupes de liaisons correspondant aux 36 chromosomes dans plusieurs espèces de *Leishmania* (Wincker *et al.*, 1996). Mais l'équipe a de longue date suspecté une habileté du parasite à moduler son génome au point d'avoir une ploïdie variable (Bastien *et al.*, 1990, Bastien *et al.*, 1992).

De son côté Stephen Beverley et ses collaborateurs ont essayé d'obtenir le knock-out du gène de la DHFR-TS¹⁴ sur une souche virulente de *L. major*. Ils ont observé qu'en faisant un double remplacement allélique avec deux marqueurs de sélection, le locus sauvage était toujours présent. Ils ont mis en évidence la présence de trois et même quatre copies du chromosome portant la DHFR-TS (Cruz *et al.*, 1993), suggérant que le parasite soit capable d'amplifier le chromosome entier pour contrer la perte d'un gène qui semble essentiel. Dans la bibliographie, plusieurs cas de variation de ploïdie ont été reportés. On retrouve les chromosomes 1 (Martinez-Calvillo *et al.*, 2005) ; 7, 22, 28, 32

¹³ Pulse Field Gel Electrophoresis

¹⁴ Dihydrofolate reductase thymidylate synthase

chez *L. major* ou encore le chromosome 6 chez *L. infantum* impliqué dans la résistance au MTX¹⁵ (Ubeda *et al.*, 2008). Le chromosome 1 a longtemps été suspecté d'être trisomique (Ravel *et al.*, 1998) (Sunkin *et al.*, 2000). Il aura fallu attendre 2011 et l'optimisation de la méthode du FISH¹⁶ par (Sterkers *et al.*, 2011) pour clairement démontrer le caractère aneuploïde de sept autres chromosomes chez *Leishmania*. Ce travail a ensuite été étendu à plusieurs espèces de *Leishmania* cultivée *in vitro* au stade promastigote sans pression de sélection (Figure 2-15)(Lachaud *et al.*, 2013).

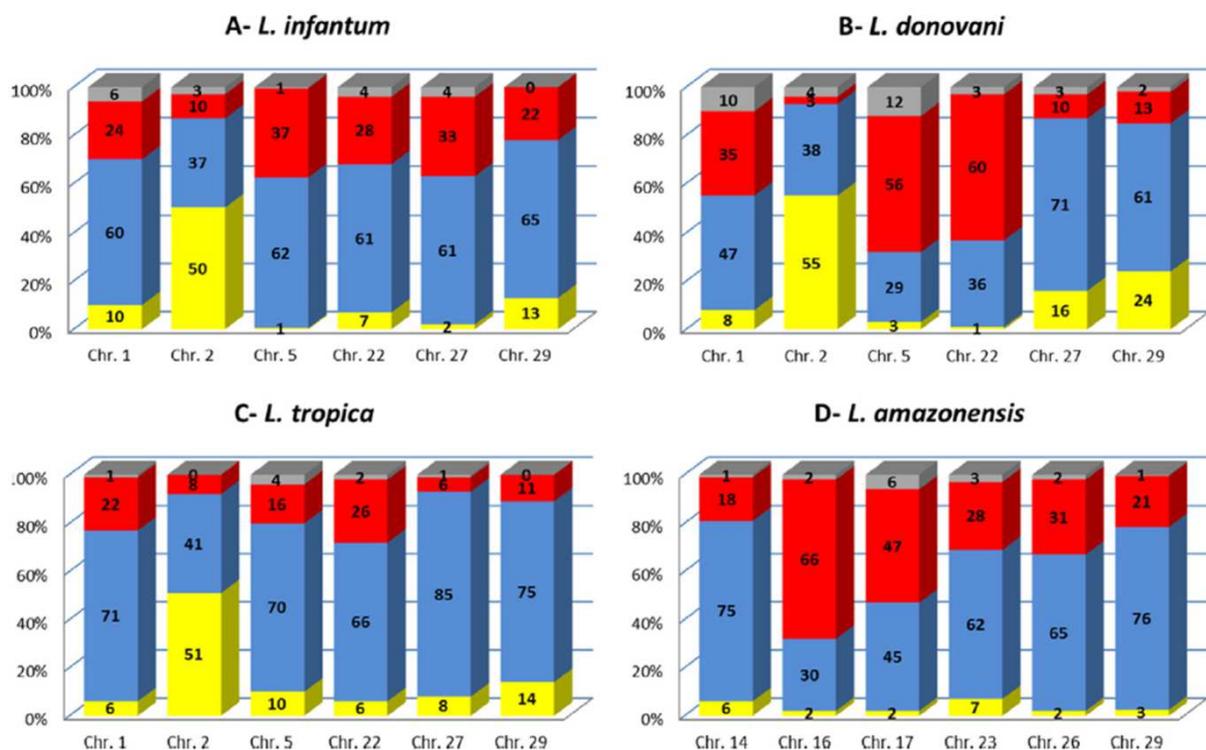


Figure 2-15 Représentation de l'aneuploïdie dans quatre espèces de *Leishmania* déterminée par FISH.

A- *L. infantum* (MHOM/FR/95/LEM3049), B- *L. donovani* (MHOM/IN/2003/LEM4537), C- *L. tropica* (MHOM/SY/92/LEM2514) and D- *L. amazonensis* (MHOM/GF/95/LEM3013). Les histogrammes représentent les pourcentages de chromosome mono- (jaune), di- (bleu), tri- (rouge) et tétrasomique (gris) retrouvés pour chaque espèces testées (Lachaud *et al.*, 2013).

Le FISH a permis de montrer non seulement l'aneuploïdie mais aussi d'introduire la notion de mosaïsme. En effet déterminant pour chaque cellule unique le nombre de

¹⁵ Methotrexate

¹⁶ Fluorescence in situ hybridization

chromosome homologue, Sterkers et ses collègues ont trouvé que chaque chromosome était présent avec au moins deux états de ploïdie différents parmi le -mono, le -di et la trisomie. Le pourcentage de somie au sein de la population est différent entre chromosomes mais aussi entre chromosomes homologues au sein des différentes souches et espèces de *Leishmania* (Sterkers et al., 2011, Lachaud et al., 2013).

Depuis plusieurs années, la stabilité des chromosomes surnuméraires pendant la mitose chez *Leishmania* cherche à être comprise, ceci est d'autant plus difficile que la position des centromères et les origines de réplication restaient inconnues jusqu'à aujourd'hui.

Pour comprendre le mécanisme qui conduit au maintien de l'aneuploïdie mosaïque chez *Leishmania* le laboratoire s'est intéressé à suivre les cellules pendant la division nucléaire.

Chez *L. major* MHOM/IL/81/Friedlin, le chromosome 2 est majoritairement monosomique et le chromosome 5 est majoritairement trisomique, au cours de la division cellulaire Sterkers et al. ont observé en FISH avec des sondes spécifiques des chromosomes 2 et 5 deux types de répartitions des chromosomes : des répartitions symétriques ou « classiques » : 1+1, 2+2 et 2+2, 3+3 pour les chr 2 et 5 respectivement. Mais surtout ils observent un pourcentage élevé de répartitions asymétriques [1+2] et [2+3] pour les chr 2 et 5 respectivement. La Figure 2-16 présente les données obtenues pour le chromosome 5. Dans tous les cas de divisions asymétriques, le nombre total de chromosome homologue est toujours un nombre impair : 3 ou 5. Ces divisions asymétriques à nombre impairs de chromosomes sont déterminantes pour le maintien de l'aneuploïdie mosaïque et permettent de soulever l'hypothèse d'une régulation non conventionnelle de la réplication (Sterkers et al., 2011).

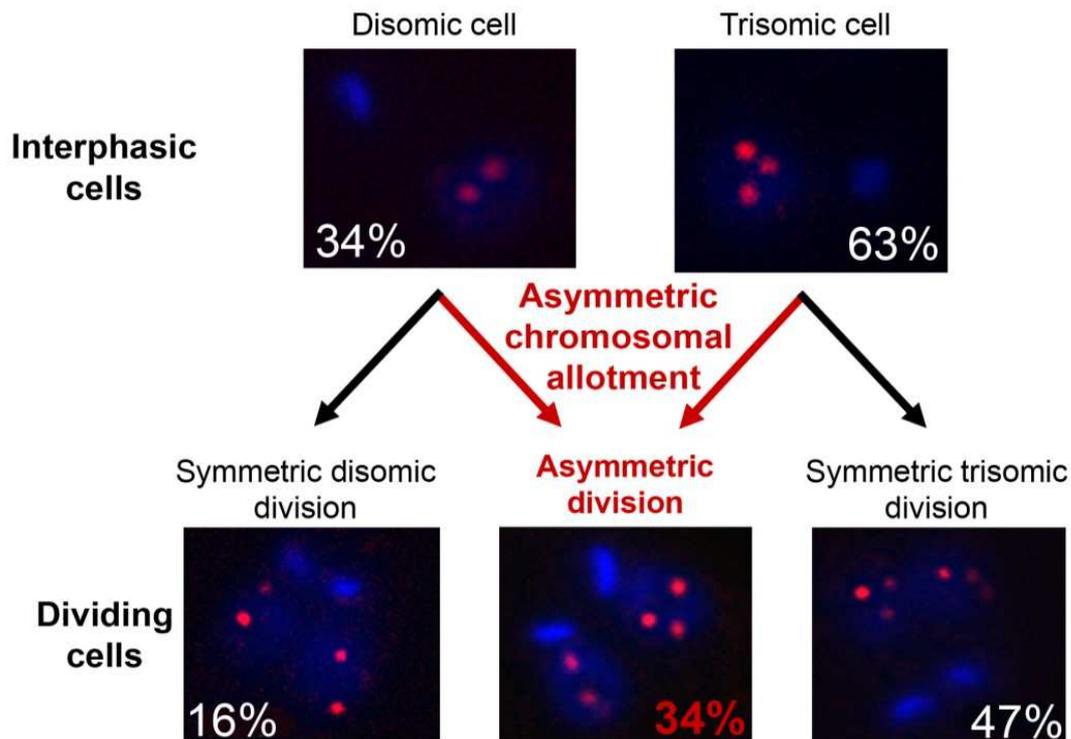


Figure 2-16 Représentation des évènements de répartition asymétrique des chromosomes chez *Leishmania*.

FISH avec une sonde spécifique du chromosome 5 sur des cellules en interphase. Le noyau et le kinetoplaste sont marqués au DAPI (bleu), les chromosomes homologues apparaissent sous forme de points (rouge). En interphase les cellules sont principalement trisomiques à 63% contre 34% de disomiques. En division trois figures de division coexistent : des divisions symétriques 2+2 à 16% et 3+3 dans 47% de cas et des divisions asymétriques 3+2 à hauteur de 34%. (Sterkers et al., 2011)

Le WGS¹⁷ permet de mesurer la profondeur de « reads » de chaque chromosome. Cette profondeur de « reads » correspond à la quantité de fois où chaque donnée de séquençage est présente à une position donnée sur le génome ; à ne pas confondre avec couverture de « reads » qui correspond à la proportion du génome couvert par les données de séquençage. Le WGS a permis de confirmer le caractère aneuploïde et d'étendre ce constat à tous les chromosomes (Rogers *et al.*, 2011). Le séquençage de 16 souches de *L. donovani* issus d'isolats cliniques a également montré une grande variation du nombre de copies de chromosomes entre les différentes souches comme le montre la **Error! Reference source not found.** (Downing *et al.*, 2011).

¹⁷ De l'anglais Whole Genome sequencing

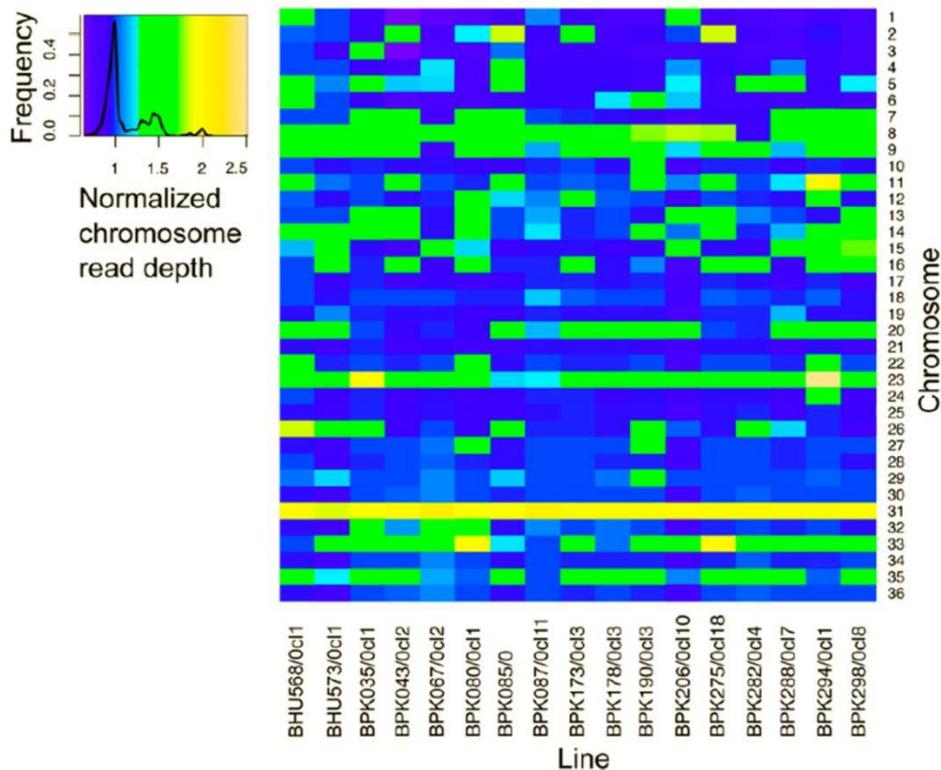


Figure 2-17 Aneuploïdie dans des isolats cliniques de *L. donovani*.

Les couleurs représente la fréquence de distribution des chromosome normalisé par rapport à la profondeur de reads de 1 (bleu) à 2 (jaune) et à 2,5 orange (Downing *et al.*, 2011).

Le WGS, méthode globale, permet d'obtenir une estimation de la ploïdie moyennée à l'échelle d'une population mais ne peut pas mettre en évidence le mosaïcisme. La technique du FISH, quant à elle, permet d'observer individuellement les chromosomes à l'échelle de la cellule unique. Malgré le côté fastidieux de la technique, le FISH reste la technique la plus adaptée pour observer l'état de ploïdie à l'échelle de la cellule unique.

L'amplification génique et l'aneuploïdie mosaïque constitutive témoignent d'une grande plasticité du génome de ce parasite. Ces évènements d'amplification sont des moyens de créer de la diversité génomique, car pour l'instant il n'y a pas de preuve évidente de l'existence d'un cycle sexué chez *Leishmania*, seuls de rares travaux ayant mis en évidence des évènements d'échange génétique entre parasites (Calvo-Alvarez *et al.*, 2014, Romano *et al.*, 2014). La capacité du parasite à amplifier certaines régions de son génome introduit la notion de dosage génique. Cette diversité que crée le parasite en faisant varier le nombre de copies de ses gènes peut être un moyen de réguler

l'expression génique en contrebalançant l'absence de régulation transcriptionnelle. On peut supposer que ces parasites se constituent un réservoir de gènes qui leur permettrait de s'adapter au plus vite à des changements d'environnement. Cette diversité pourrait constituer un facteur de virulence et avoir un rôle dans la pathogénicité.

Leishmania n'est pas la seule espèce de protozoaire où l'on retrouve de l'aneuploïdie, *T. cruzi* (Reis-Cunha *et al.*, 2015), *Giardia* (Tumova *et al.*, 2016) et l'aneuploïdie est fortement suspectée chez *Entamoeba histolytica* (Lohia *et al.*, 2007). Aujourd'hui on ne connaît pas les mécanismes moléculaires qui font et coordonnent cette aneuploïdie. Cependant cela peut expliquer les difficultés que nous rencontrons en génétique réverse sur ce parasite *Leishmania*. Aujourd'hui l'émergence du séquençage de cellule unique (SCS¹⁸) utilisé chez les eucaryotes supérieurs (Yong Wang, 2016) pourrait être une alternative au FISH et aider à apporter des réponses, cependant cette technique reste onéreuse.

2.8 De l'origine à la réplication

Plantes, bactéries, mammifères ou encore parasites, tous ces organismes doivent dupliquer leur information génétique pour la transmettre à la génération suivante. Cette transmission ne peut se faire sans un processus essentiel : la réplication de l'ADN. Ce processus est soumis à un control strict pour que le génome soit, entièrement répliqué, avec fidélité avant que la cellule puisse se diviser.

Les cellules sont programmées pour dupliquer leur ADN à un moment précis du cycle cellulaire : la phase S. Le cycle de division cellulaire (Figure 2-18) se décompose en quatre phases : (i) **phase G1** où la cellule se prépare à entrer en réplication en assemblant le complexe pré-réplicatif. (ii) La **phase S** où la cellule réplique son ADN. (iii) La **phase G2** qui est point de contrôle permettant de vérifier que la cellule est prête à se diviser (iv) pour donner deux nouvelles cellules filles au cours de la mitose **phase M**. Les cellules peuvent également rentrer dans un état de quiescence dit **stade G0** au cours de laquelle les cellules ne se divisent pas. Tout ce processus est étroitement orchestré par une multitude de protéines notamment les cyclines et les CDK¹⁹ qui représentent des

¹⁸ De l'anglais single cell sequencing

¹⁹ Cyclin Dependent Kinases

points de contrôle à chaque transition de phase. Malgré cette régulation précise, dans certains cas la réplication ne pourra pas aboutir à conserver l'intégrité du génome, la cellule recevra le signal pour entrer en apoptose et sera détruite par l'organisme. Dans d'autre cas les cellules proliféreront de façon anormale causant des cancers.

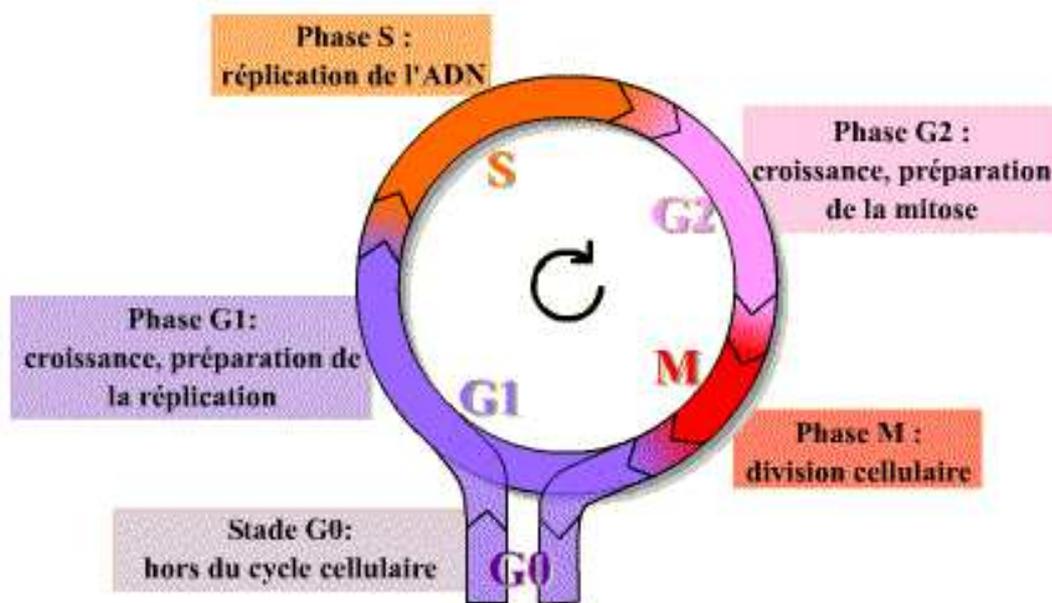


Figure 2-18 Le cycle cellulaire.

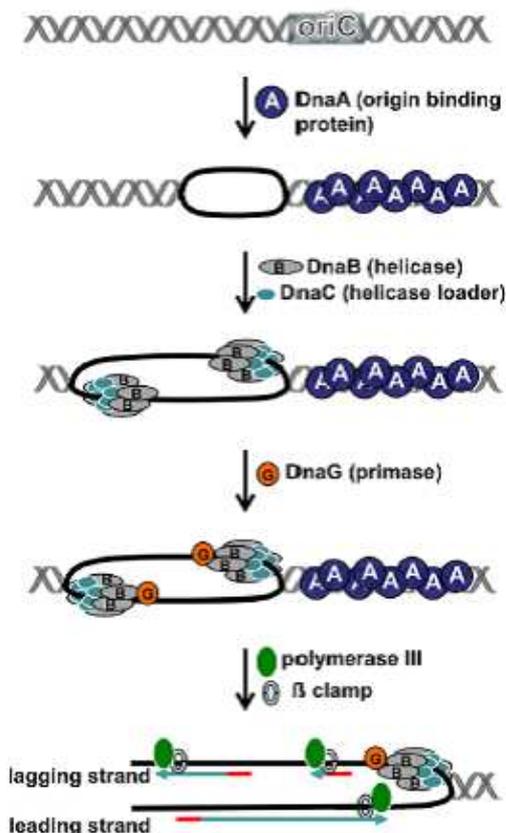
<http://www.snv.jussieu.fr/bmedia/cyclecellMB/index.htm>

Le processus de réplication de l'ADN est initié à des positions particulières sur le génome nommées **origine de réplication** (Ori). Plusieurs protéines formant le complexe de reconnaissance des origines permettent de définir ces origines. La réplication de l'ADN n'est pas initiée strictement de la même façon chez les bactéries, les archées ou les eucaryotes. Elle est malgré tout basée sur les mêmes principes fondamentaux. Dans cette sous-partie je m'intéresserai aux étapes préliminaires de la réplication de l'ADN : Comment la réplication est-elle initiée? Comment est-elle régulée ? Est-elle différente entre eucaryotes et procaryotes? Le processus d'activation des origines est commun aux bactéries, Archaea et eucaryotes, il utilise des **protéines de liaison aux origines** appartenant à la famille AAA+ qui généralement fonctionnent en multimère (Erzberger & Berger, 2006). Ces protéines peuvent être également impliquées dans d'autre processus cellulaires que la réplication tel que la ségrégation (Bell & Kaguni, 2013).

2.8.1 Bactéries

Les bactéries possèdent un génome de petite taille circulaire. Ces organismes répliquent leurs ADN à partir d'une seule origine de réplication (

Figure 2-19) appelée OriC (Jacob & Brenner, 1963). OriC représente une région d'environ 250 pb qui contient plusieurs répétitions d'une séquence consensus de 9 paires de bases, les DnaA box, et des régions riches en A/T. Le nombre de répétitions varie d'une espèce bactérienne à l'autre (Fuller *et al.*, 1984). Les monomères de protéines DnaA (protéine de liaison aux origines) vont venir se fixer sur ces DnaA box provoquant l'ouverture de la double hélice et la formation « d'une bulle ». Les DnaA sont des ATPases de la famille AAA+ capables de lier et d'hydrolyser de l'ATP. Elles vont permettre au complexe hélicase de se positionner sur l'ADN simple brin (Kaguni, 2011).



Le complexe hélicase est formé de six protéines DnaB qui sont recrutées et régulées par DnaC. L'hélicase bactérienne se positionne sur le brin retard et, progresse de 5' vers 3' (O'Donnell *et al.*, 2013). Il faut noter qu'il a été montré que des mutations dans la protéine DnaA pouvaient entraîner une hyper activation des origines de réplication (Simmons & Kaguni, 2003). Cette unique origine permet aux bactéries d'achever la réplication de leur unique chromosome circulaire (4.4 Mb) en 30 min pour *E.coli*.

Figure 2-19 Représentation schématique de l'initiation de la réplication chez les bactéries.

(Tan & Tomkins 2015 ARJ)

2.8.2 Eucaryotes modèles

Les premières protéines impliquées dans la reconnaissance des origines de réplication ont été identifiées dans un extrait cellulaire de *S. cerevisiae*. Ces protéines forment un complexe hétéro-hexamérique nommé Origin Recognition Complex (ORC), constitué de six sous-unités Orc1 à Orc6 (Bell & Stillman, 1992). Rapidement les orthologues des ORC sont identifiés chez divers organismes comme *Drosophila melanogaster* (Gossen *et al.*, 1995), *Arabidopsis thaliana* (Diaz-Trivino *et al.*, 2005) et *Homo sapiens* (Dhar & Dutta, 2000). Un autre facteur de réplication très conservé parmi les Eucaryotes modèles est Cdc6, cette protéine est indispensable au recrutement du complexe hélicase (Mcm2-7) au niveau des origines de réplication. En plus de Cdc6 des protéines ORC1-like ont été identifiées, Sir3 (formation d'hétérochromatine) chez *S. cerevisiae* et des paralogues d'ORC1 nommée ORC1a et ORC1b chez *A. thaliana*.

A l'exception de Orc6, Orc1-5 et Cdc6 possèdent une région conservée AAA+ incluant des domaines de liaison à l'ATP Walker A et Walker B. On retrouve des domaines Sensor1 et Sensor2 impliqués dans la régulation de l'activité ATPase. Le complexe ORC se stabilise par des interactions protéine-protéine engageant leurs domaines BAH (bromo-adjacent homology) ce qui permet au complexe de former un anneau autour de l'ADN.

Chez les eucaryotes supérieurs et *S. pombe*, il n'a pas été identifié de séquence consensus qui définirait précisément la position des origines de réplication sur le génome. Ces origines dépendraient davantage de l'organisation de la chromatine que d'une séquence d'ADN spécifique. Cependant ces origines seraient souvent associées à des zones de

transcription active (Masai *et al.*, 2010). A contrario, *S. cerevisiae* compte plusieurs origines de réplication qui sont définies par une séquence consensus de 11 pb appelées Autonomously Replicating Sequences ou ARS²⁰ (Struhl *et al.*, 1979) (Huberman *et al.*, 1987)

Chez les eucaryotes plusieurs origines sont organisées en cluster et leur activation au cours de la phase S dépend du stade de développement de la cellule (Mechali *et al.*, 2013). Le positionnement du complexe ORC (Stillman, 2005) définit les origines de réplication et initie le processus de réplication décrit dans la Figure 2-20.

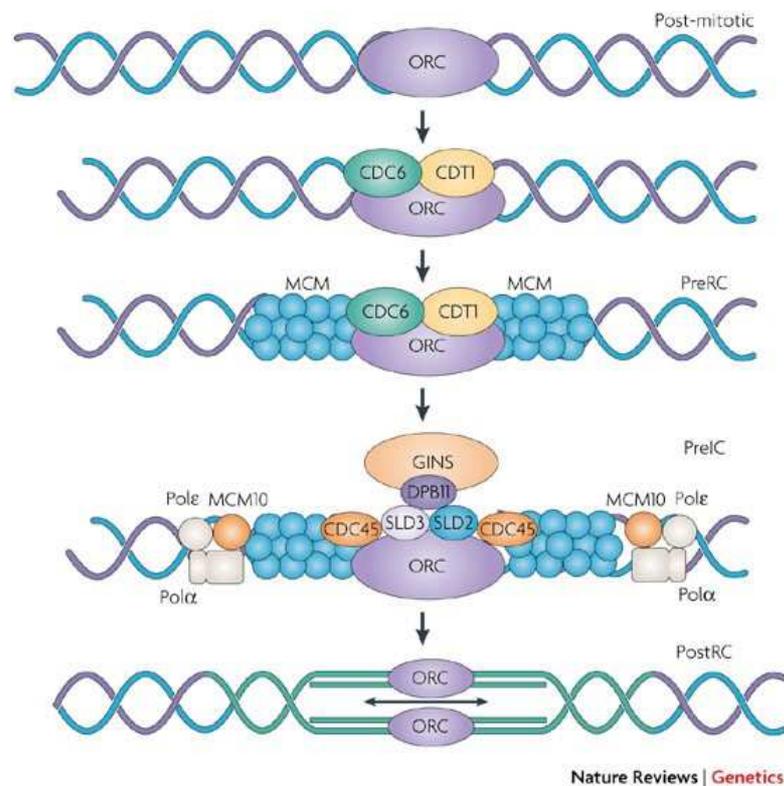


Figure 2-20 Dynamique de régulation de la réplication de l'ADN chez les Métazoaires. (Aladjem, 2007)

Le complexe ORC va servir de plateforme pour l'assemblage du complexe pré-répliatif. Chez la levure ces protéines ORC vont être associées aux origines tout au long du cycle cellulaire alors que chez les eucaryotes supérieurs, la liaison aux origines de la plus grande sous-unité du complexe ORC, Orc1, dépend du cycle cellulaire. Durant la phase de transition M-G1 le complexe ORC va recruter Cdc6 et Cdt1 qui vont faciliter le

²⁰ De l'anglais Autonomous Replication Sequence

recrutement du complexe hélicase formée de 6 protéines MCM (Licensing). Le complexe qui en résulte est nommé complexe pré-réplication (Pré-RC). Le Pré-RC va être activé pour créer le complexe pré-initiation (pré-IC) par le recrutement de facteurs additionnels Cdc45, SLD2-3, DPB11, le complexe GINS et MCM10. L'initiation de la réplication durant la phase S requiert des phosphorylations. L'hélicase est un complexe dynamique qui a la capacité de progresser le long de l'ADN en ouvrant la double hélice grâce à l'hydrolyse de l'ATP (Gai *et al.*, 2010). Le déplacement de part et d'autre de l'origine est orienté de 3' vers 5' et est régulé par d'autres acteurs. Ce complexe MCM a initialement été identifié chez la levure. Des mutations touchant les gènes MCM empêchent la cellule de maintenir un plasmide contenant une origine de réplication et un centromère (Maine *et al.*, 1984)

Chez les eucaryotes, le complexe du réplisome coordonne la progression de la fourche de réplication (Figure 2-20). La réplication du brin précoce et du brin tardif est accomplie par la polymérase ϵ et δ respectivement. Les ADN polymérases requièrent une amorce pour commencer leur synthèse ; un complexe primase va donc synthétiser une petite amorce ARN de quelques nucléotides au niveau des origines de réplication sur le brin précoce et également sur le brin tardif (au niveau de chaque extrémité 5' pour la synthèse des fragments d'Okazaki). Plusieurs facteurs vont réguler et coordonner la progression des polymérases avec le complexe Cdc45-MCM-GINS qui permet l'ouverture des brins d'ADN matrice. Les protéines PCNA et RPAs sont impliquées dans la stabilisation de la fourche de réplication (Leman & Noguchi, 2013).

Au cours de sa progression la fourche de réplication fait face à différents obstacles le long de son parcours : compaction de la chromatine autour des histones, cassure double brin de l'ADN, transcription des gènes... Tous ces événements peuvent bloquer la progression du réplisome qui devra être ré-initié.

Contrairement aux eucaryotes supérieurs, les bactéries peuvent re-répliquer leur ADN avant même qu'elles aient fini le premier cycle de réplication, (Dame *et al.*, 2011). Les eucaryotes possèdent de multiples origines, et la ré-initiation de certaines origines pourrait conduire à la création de chromosomes surnuméraires. Ces anomalies poseraient problème lors de la ségrégation des chromosomes (O'Donnell *et al.*, 2013).

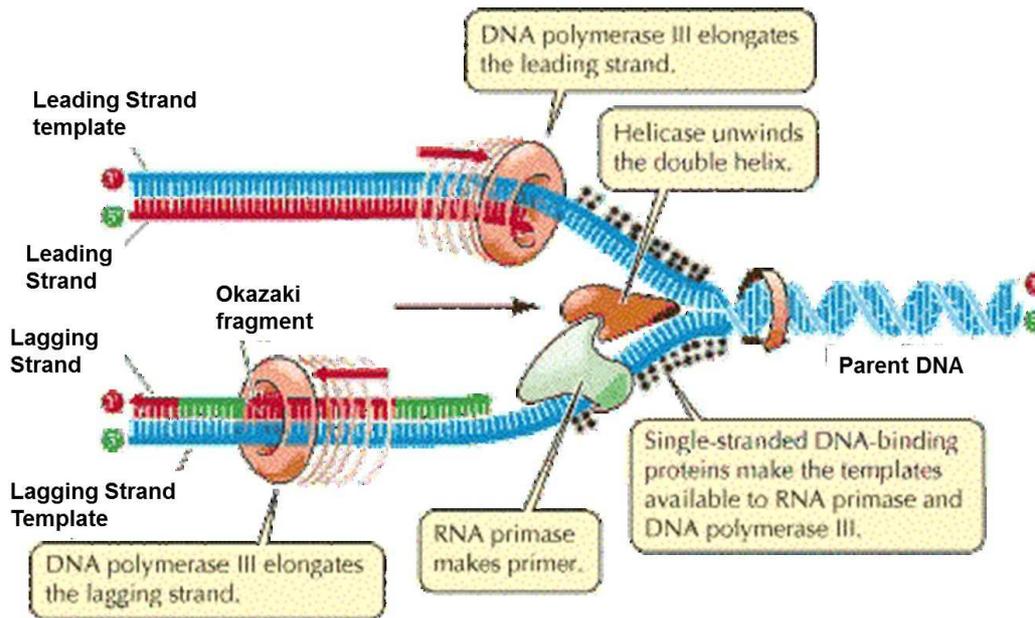


Figure 2-21 Fourche de réplication eucaryote.
<http://www.whatisdna.net/wiki/dna-polymerase/>
2.8.3 Archaea

Les *Archaea* possèdent un mécanisme de réplication de l'ADN à la frontière entre les bactéries et les eucaryotes. Une seule protéine, Orc1/Cdc6 est impliquée dans la reconnaissance des origines de réplication. C'est une protéine qui possède une homologie avec CDC6 et ORC1, elle appartient également à la famille AAA+ et est capable d'autophosphorylation (Grabowski & Kelman, 2001). Toutes les espèces d'*Archaea* ne possèdent pas le même nombre de copie du gène Orc1/Cdc6 : un seule pour l'espèce *Pyrococcus* et jusqu'à 9 chez *Halobacterium*. La protéine Orc1/Cdc6 reconnaît un motif en répétition inverse "origin recognition box" ORBs (Robinson *et al.*, 2004) (Lundgren *et al.*, 2004). Jusqu'à 3 ORBs ont été identifiés chez *S. solfataricus* mais une seule origine pour *P. Abyssii* adjacente au gène Orc1/Cdc6 comme chez les bactéries ou le gène DnaA est situé à côté de l'OriC.

Ici le complexe hélicase est formé d'hexamère de MCM ayant une activité hélicase 3'-5' comme chez les bactéries (Chong *et al.*, 2000). Aucun homologue de Cdt1 n'a jusqu'à ce jour été identifié chez les *Archaea*.

2.8.4 Trypanosomatidés

Lorsque j'ai commencé ma thèse en Décembre 2013 très peu de données étaient disponible dans la littérature concernant la réplication de l'ADN chez les trypanosomatidés, même si en 1987 North & Wyler suggéraient une différence entre la réplication des mammifères et celles des *Leishmania*. En effet ils ont montré que différentes molécules inhibitrices des polymérases étaient capables d'inhiber la synthèse d'ADN chez les cellules de mammifère mais n'étaient pas capable de l'inhiber chez des promastigotes de *Leishmania major* et *L. donovani* (North & Wyler, 1987).

2.8.4.1 Complexe pré-réplicatif

Après le séquençage des génomes de *L. major*, *T. brucei* et *T. cruzi* les gènes LmjF.28.0030 Tb11.02.5110b et TcCLB.508239.10 respectivement ont été annotés comme étant des orthologues de ORC1 (El-Sayed *et al.*, 2005).

Chez *L. major* (Kumar *et al.*, 2008) ont vérifié par Reverse Transcription-PCR que le gène LmjF.28.0030 était bien exprimé au stade promastigote. Ils ont également montré que la protéine possédait plusieurs motifs caractéristiques des ORC : AAA+ ATPase (résidus 53–205), Walker A (résidus 61-68 GMPGTGKT) et Walker B (résidus 147-152 VIVVDE) et l'ont nommé ORC1. Grâce à une expression transitoire ils ont montré une localisation diffuse dans le noyau de cette protéine.

Une seconde étude de la même équipe chez *L. donovani* a montré que ORC1 possédait un stretch d'acides aminés basiques KRSRR correspondant à un NLS putatif, cette séquence serait nécessaire pour l'adressage de Orc1 au noyau (Kumar & Saha, 2012).

En 2009 une équipe brésilienne s'est intéressée à caractériser les homologues de ORC1 chez *T. brucei* et *T. cruzi*. A l'aide d'un logiciel de prédiction de domaines fonctionnels,, les auteurs ont retrouvé deux domaines Orc1 et Cdc6 ; ils ont donc renommé la protéine Orc1/cdc6 en pensant que cette protéine pourrait être assimilée à une protéine bifonctionnelle Orc1/Cdc6 comme chez les *Archaea*. (Godoy *et al.*, 2009). Ce travail a montré que cette protéine avait une activité ATPase *in vitro*. Les auteurs ont réalisé une étude fonctionnelle d'Orc1/Cdc6 grâce à des complémentations chez la levure et une RNA interférence chez *T. brucei*. Les expériences de complémentation chez la levure d'un mutant ORC1 et CDC6 par la protéine du parasite Orc1/Cdc6 ont montré que seule la fonction Cdc6 a pu être restaurée chez les mutants de levures. Le RNAi de TbORC1/CDC6 a montré une altération de la croissance et l'apparition de cellules

anucléées (zoïdes). D'après ces résultats, les auteurs suggèrent qu'ORC1/CDC6 serait impliquée dans la réplication. La localisation de la protéine avec un anticorps spécifique a montré qu'Orc1/Cdc6 était nucléaire tout au long du cycle cellulaire et semble rester associée à l'ADN tout au long du cycle cellulaire.

Malgré la présence d'une unique protéine Orc1/Cdc6 dans le génome des trypanosomes, Dang et Li ont pu identifier dans les bases de données la présence d'une seconde protéine ORC putative qu'ils ont appelé **Orc1b**. Ils ont également mis en évidence la présence des six sous-unités du complexe hélicase **MCM** (Mcm2, Tb11.02.5730; Mcm3, Tb927.2.3930; Mcm4, Tb11.01.4070; Mcm5, Tb11.02.3270; Mcm6, Tb11.01.3510; Mcm7, Tb11.01.7810), ainsi que toutes les protéines du complexe **GINS** (Sld5, Tb927.3.4810; Psf1, Tb09.160.3540; Psf2, Tb11.01.2230; Psf3, Tb927.4.4680), et **Cdc45** (Tb11.02.0310) (Dang & Li, 2011).

Ils ont démontré que l'activité hélicase nécessite tous les composants du complexe CMG (Cdc45-Mcm-GINS) et que les protéines des complexes MCM et GINS présentent une localisation nucléaire tout au long du cycle cellulaire de G1 à la télophase. Seule Cdc45 possède une localisation qui diffère du reste du complexe : nucléaire en G1, enrichi dans le noyau en S/G2 et cytoplasmique durant la mitose. Cdc45 possède un signal d'export nucléaire putatif, NES²¹. En l'absence de l'identification de Cdt1 les auteurs suggèrent un rôle pour Cdc45 qui protégerait contre la re-réplication.

Les RNAi de Mcm3 et Cdc45 provoquent un drastique arrêt de la croissance et une augmentation de zoides dès J2 post-induction. Dans les expériences de RNAi des différentes protéines Mcm, ils observent des formes irrégulières du noyau, et pour GINS et Cdc45, un défaut de la ségrégation des chromosomes. Cependant ce ne sont que des observations faites en DAPI et non en FISH. Ces résultats suggèrent néanmoins que le complexe CMG interviendrait également dans le maintien de l'intégrité structurale du noyau.

Des expériences de GST-pull-down et IP ont montré que TbOrc1/Cdc6 interagissait avec Orc1b ; et TbMcm3 pouvait se lier à TbOrc1/cdc6 et TbOrc1b. Plus tard (Tiengwe *et al.*, 2012) identifieront trois nouvelles protéines interagissant avec TbOrc1/Cdc6 : Tb927.10.13380(ORC4), Tb09.160.3120 et Tb927.10.7980. Les protéines identifiées

²¹ Nuclear export signal

comme ayant un rôle dans le complexe pré-répliatif ORC de *T. brucei* sont résumées dans la Figure 2-22.

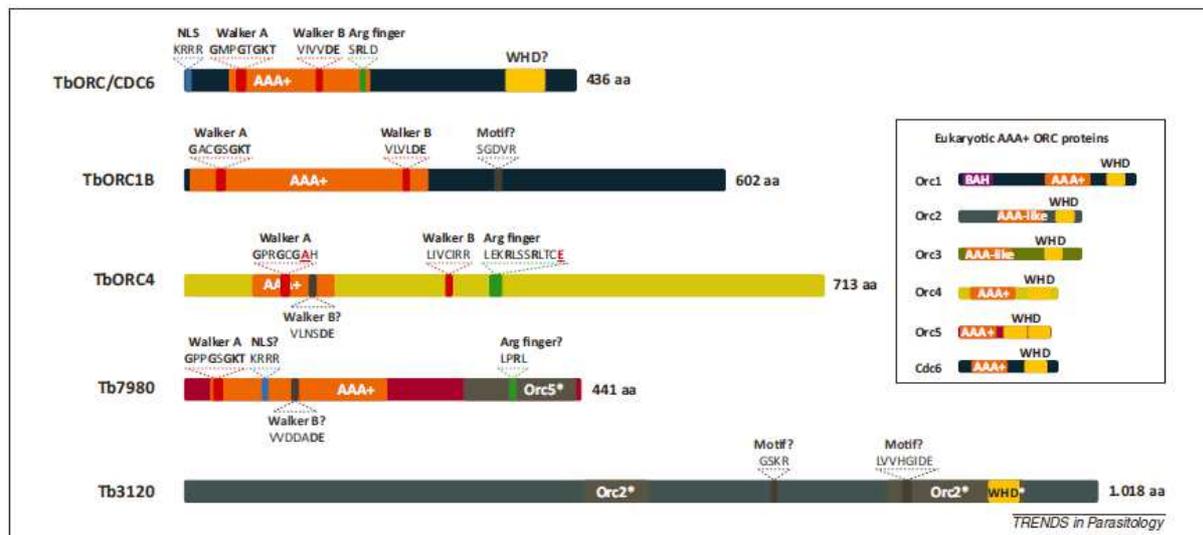


Figure 2-22 Protéines putative du complexe ORC chez *T. brucei*.

L'encart de droite représente les sous-unités du complexe ORC eucaryotes avec la représentation des motifs et des domaines putatifs. (Tiengwe *et al.*, 2014)

2.8.4.2 Les origines de réplication chez *T. brucei* :

A ce jour, seule l'équipe de Richard McCulloch s'est intéressée aux origines de réplication de *T. brucei* (Tiengwe *et al.*, 2012). Dans ce travail, ils ont déterminé les sites de liaison à l'ADN de Orc1/Cdc6 par CHIP on chip; ils ont mis en évidence 953 sites répartis sur trois grandes régions des chromosomes : dans le "cœur" (38%), les positions subtélomériques (11%) et les arrangements de gènes de VSG (51%).

Dans le "cœur" des chromosomes la majorité des sites de liaison Orc1/Cdc6 se trouvent aux frontières entre deux unités de transcription multigénique (*strand switch region* ou SSR).

Les sites de liaisons aux protéines Orc1/Cdc6 sont associés aux sites d'initiation de la transcription définis par les marques d'acétylation H4K10Ac (Siegel *et al.*, 2009) (Figure 2-23). Aucun motif consensus n'a pu être défini pour les sites de liaison d'Orc1/cdc6, que ce soit dans le "cœur" des chromosomes qu'au niveau des gènes de VSG (pourtant séquences répétées)

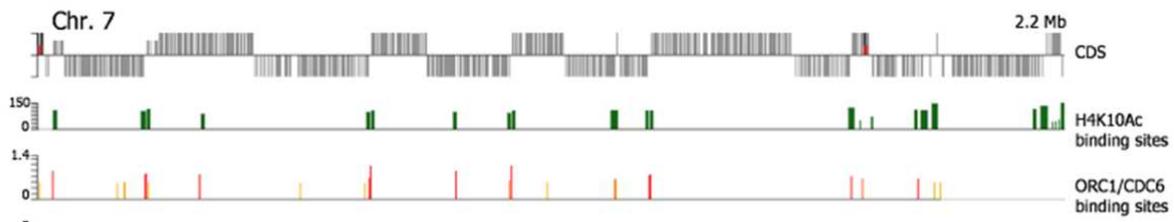


Figure 2-23 Position des H4K10Ac et des sites de liaison à l'ADN de Orc1/Cdc6 sur le chromosome 7 de *T. brucei*.

(Tiengwe *et al.*, 2012)

La technique de l'analyse de fréquence de marqueurs (Marker Frequency Analysis) (MFA) couplée au séquençage à haut débit permet de comparer deux états, ici une population de cellules en 1^{ère} moitié de phase S avec une population en G2. La comparaison du nombre de reads obtenus dans l'état S par rapport à l'état G2 permet de définir des régions sur le génome où le taux d'enrichissement de séquence est élevé. Ces régions correspondent aux origines de réplication. Dans ce travail Tiengwe et ses collaborateurs ont utilisé la MFA pour détecter les origines de réplication précoces. Ils ont pu prédire environ 100 origines de réplication pour le génome de *T. brucei* ce qui représenterait une distance inter-origines de 260 kb, ce qui s'avérera contradictoire avec nos données. Ces origines précoces se trouvent au niveau des centromères, régions connues pour être répliquées en premier. Ils ont également observé des origines asymétriques qui seraient la conséquence d'une confrontation entre la réplication et la transcription. Du fait de la forte présence d'Orc1/Cdc6 au niveau des gènes de VSG, les auteurs suggèrent qu'Orc1/Cdc6 pourrait avoir un rôle associé aux VSG situés dans une zone de non-transcription. En effet la RNAi de Orc1/Cdc6 provoque la dé-répression de l'expression des gènes de VSG (Tiengwe *et al.*, 2012) Chez *Plasmodium*, à un moment particulier du cycle, ORC1 se relocalise aux télomères, régions associées aux gènes Var. (Mancio-Silva *et al.*, 2008), suggérant un rôle de certaines protéines ORC dans la régulation de la variation antigénique responsable de l'évasion du système immunitaire. La RNAi d'Orc1/CDC6 provoque une augmentation des ARNm à proximité des SSR, où se situent les sites de liaison d'Orc1/Cdc6. Cette observation suggère que la polymérase n'a plus de signal de fin de transcription et poursuit au-delà (*readthrough*). Ceci implique qu'Orc1/Cdc6 pourrait également jouer un rôle dans la délimitation des unités de transcription (Tiengwe *et al.*, 2012).

Les protéines putatives du complexe ORC semblent être impliquées dans différents mécanismes biologiques du parasite en plus de leur rôle dans la réplication de l'ADN. Il est donc important de poursuivre l'étude de ces protéines qui peuvent être des cibles thérapeutiques et notamment chez *Leishmania* où pour l'instant peu de choses sont connues.

2.9 CRISPR/Cas9 : outil de génie génétique

2.9.1 Découverte

Le CRISPR/Cas9 a été découvert chez les bactéries et les *Archaea* (Mojica *et al.*, 2005) et (Barrangou *et al.*, 2007). Il aura fallu plusieurs années de recherche et la contribution de différentes équipes à travers le monde pour identifier et comprendre ce système qui à la base est un mécanisme de défense immunitaire des bactéries contre les virus. Si l'on regarde 20 ans en arrière, à ces balbutiements le CRISPR s'appelait SRSRs²² et ne suscitait que très peu l'intérêt des financeurs ; difficile à croire au vue de l'engouement et/ou l'inquiétude que le CRISPR a générés aujourd'hui. C'est à Alicante, en Espagne, qu'une équipe travaillant sur des bactéries a commencé à s'intéresser à la curieuse organisation d'un fragment d'ADN où se trouvaient des séquences palindromiques répétées et séparées de petites séquences variables d'ADN de 30 bases (Mojica *et al.*, 1993). Plus tard cette famille sera nommée : Cluster Regulatory interspaced Short Palindromic Repeats (CRISPR) et on trouvera des *CRISPR-associated (cas) genes* à proximité de ces séquences. La proximité de ces gènes avec les répétitions de CRISPR suggèrait que tous ces éléments étaient liés à une même fonction (Jansen *et al.*, 2002).

Même si depuis 2002, on sait qu'il y a des gènes associés aux répétitions de CRISPR, leur fonction demeurait inconnue. Pendant longtemps on a cru que le système fonctionnait comme un ARN interférant (Sorek *et al.*, 2008). Cependant deux équipes ont montré que chez les bactéries *S. epidermidis* (Marraffini & Sontheimer, 2008) et *S. thermophilus* (Garneau *et al.*, 2010), le système avait pour cible de l'ADN, bien que chez certaines espèces la cible soit de l'ARN (Hale *et al.*, 2009).

²² De l'anglais short regularly spaced repeats

2.9.2 Fonctionnement

La présence de locus CRISPR/Cas a été décrite chez de nombreuses bactéries et *Archeae* (Mojica *et al.*, 2000), parfois très différents les uns des autres en terme d'organisation ou composition mais tous servant la même fonction, la lutte contre des organismes étrangers. La Figure 2-24 décrit la classification des cinq types de systèmes CRISPR/Cas. Cette classification est basée sur la nature des protéines *cas* selon leur fonction : les protéines d'**adaptation** (utiles à l'acquisition de nouveaux *spacers*) ; celles d'**expression** (interviennent dans la maturation du crRNA et la liaison à la cible) ; d'**interférence** (coupure de la cible) et les protéines **accessoires** (rôles de régulateurs ou autre fonction associée au CRISPR) (Makarova *et al.*, 2015).

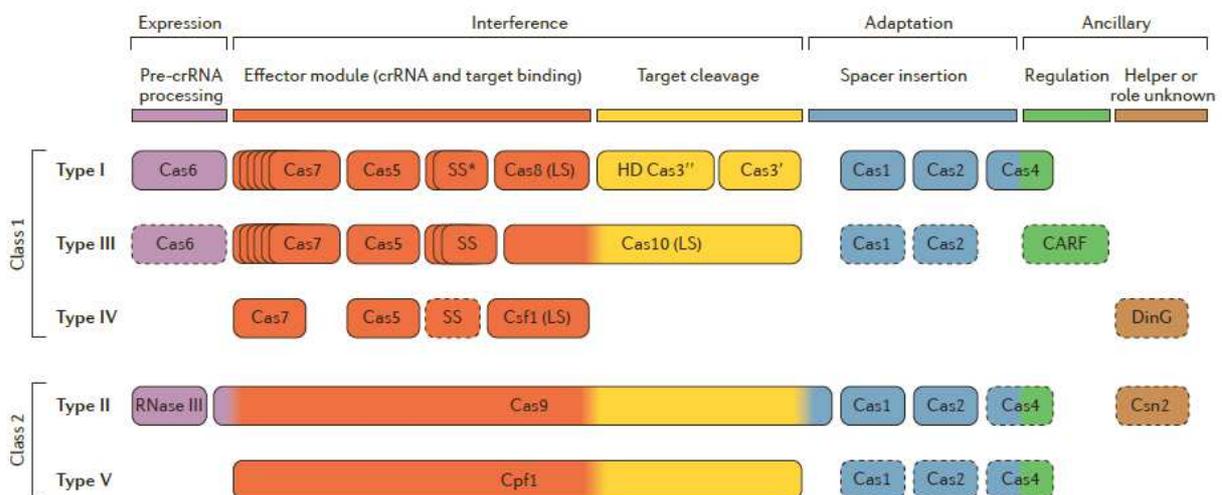


Figure 2-24 Classification des systèmes CRISPR.

D'après (Makarova *et al.*, 2015).

Chez *E. coli* on trouve un CRISPR de Type I constitué d'un complexe protéique Cas nommée cascade (Brouns *et al.*, 2008). L'inactivation indépendante de chaque gène cascade a permis de montrer que ces protéines étaient nécessaires pour la maturation d'un long ARN précurseur transcrit à partir du locus CRISPR; il sera nommé crRNAs pour CRISPR RNAs (Sorek *et al.*, 2008).

Dans cette sous-partie je vais m'intéresser plus particulièrement au système CRISPR de Classe 2 Type II retrouvé chez *S. pyogenes* (Figure 2-25). En effet le système CRISPR/Cas9 que l'on utilise aujourd'hui dérive du système découvert chez *S. pyogenes*.

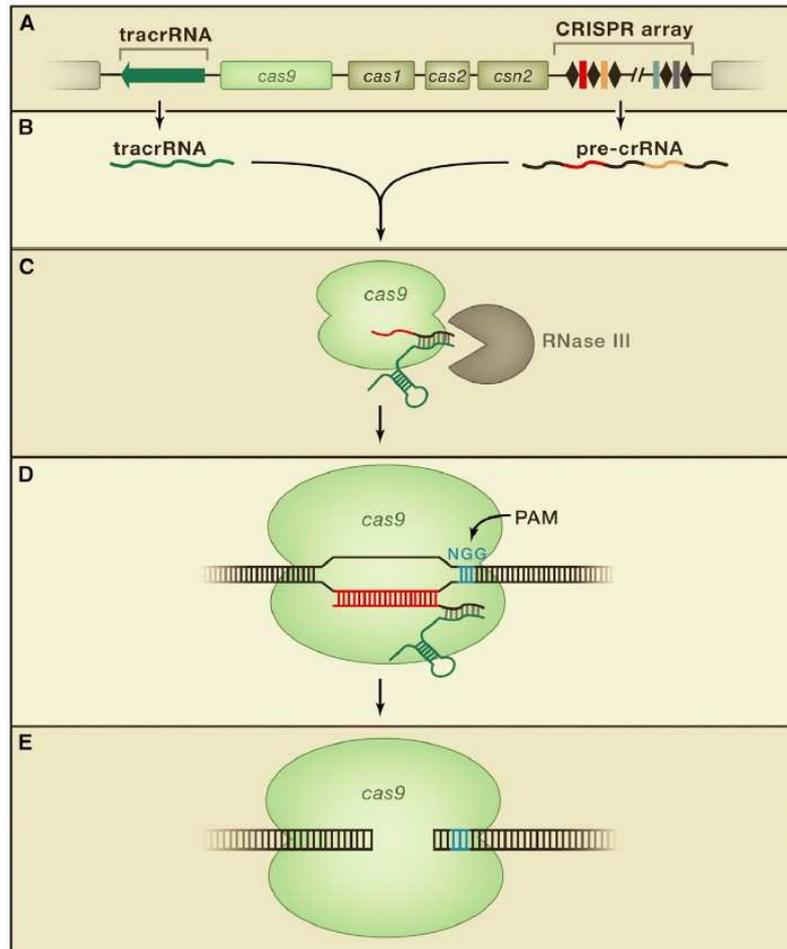


Figure 2-25. Crispr Type II chez *S. pyogenes*.

A- Organisation du locus Crispr, les rectangles de couleur dans le CRISPR array représentent les 'spacers'. B- Transcription du tracrRNA (voir texte) et du pré-CrRNA à partir du locus Crispr. C- Ces ARNs s'associent à Cas9. D- Le complexe scanne l'ADN jusqu'à trouver une séquence homologue au 'spacer' porté par le crRNA située à proximité d'un Protospacer Adjacent Motif (PAM). E- L'enzyme Cas9 provoque alors une cassure double brin (DSB) sur l'ADN. (Lander, 2016)

De la même façon que chez *E. coli* le locus CRISPR de *S. pyogenes* est composé de différents éléments (Figure 2-25A). Une partie de ces éléments sont des enzymes : cas1, cas2 et csn2 impliqués dans l'acquisition de nouveau 'spacers' provenant d'ADN étranger; et une endonucléase **cas9** qui confère l'immunité en coupant l'ADN envahisseur. L'arrangement (*array*) de CRISPR représente une forme de mémoire où les bactéries conservent une multitude de fragments d'ADN acquis de 'spacers' de virus. Les autres éléments sont des gènes codant des ARNs : le **tracrRNA** (*trans-activating* CRISPR RNA) et le **pre-crRNA** (Figure 2-25B). Ces ARNs s'associent à l'endonucléase Cas9 (Figure 2-25C), puis le complexe scanne l'ADN jusqu'à trouver une séquence

homologue au 'spacer' porté par le crRNA situé à proximité d'un **Protospacer Adjacent Motif** (PAM) Figure 2-25D. L'enzyme Cas9 provoque alors une **cassure double-brin** (DSB ²³) sur l'ADN étranger trois bases en amont du PAM (Figure 2-25E). Tout ce mécanisme va permettre l'élimination des ADN étrangers.

Le pre-crRNA est un ARN produit à partir du *CRISPR array*. Il est composé de la séquence 'spacer' encadrée en 5' et 3' de séquences répétées. La maturation du crRNA chez *S. pyogenes* est possible grâce à une endoribonucléase RNase III de l'hôte et du tracrRNA (Deltcheva *et al.*, 2011).

Grâce au travail de différentes équipes sur différents organismes, petit à petit les pièces du puzzle ont été assemblées. La logique mécanistique des différents systèmes CRISPR/cas comprise, les équipes se sont rendues compte du potentiel du CRISPR/ Cas9 en tant qu'outil génétique.

2.9.3 De la bactérie à l'outil génétique.

L'activité endonucléase spécifique de Cas9 a rappelé ce qui était utilisé pour l'édition de génome avec les ZFN²⁴ (Porteus & Carroll, 2005, Geurts *et al.*, 2009), TALEN²⁵ (Miller *et al.*, 2011) ou encore la recombinase Cre (Jullien *et al.*, 2003). Tous ces systèmes, découverts à l'origine chez des organismes vivants ont été modifiés et optimisés pour être utilisés *in vitro* ou dans des modèles hétérologues. Cependant toutes ces techniques ont montré quelques limitations, les effets 'off target', la production de nucléase à façon ou encore le fait d'être limité par la nécessité de la présence dans le génome d'une séquence spécifique pour l'édition (i.e. site LoxP). Pour la manipulation de génomes consistant en l'inactivation de gènes par interruption du cadre de lecture, insertion de mutation ou élimination de la région codante d'un gène d'intérêt (GOI), ces méthodes ont montré que pour augmenter l'efficacité de l'édition d'un génome il fallait un système capable de provoquer une cassure double brin à l'endroit désiré.

A son tour le CRISPR/cas a été optimisé et a prouvé son efficacité chez de nombreux organismes eucaryotes (Lander, 2016). Il est devenu un outil incontournable pour l'édition des génomes. Le principal système CRISPR/cas que l'on utilise aujourd'hui

²³ De l'anglais Double Strand Break

²⁴ De l'anglais Zin Finger Nuclease

²⁵ De l'anglais Transcription activator-like effector nucleases

dérive du CRISPR de Classe 2 Type II retrouvé chez *S pyogenes*. Il a été simplifié par E. Charpentier et J. Doudna qui ont joint leurs forces pour élaborer *in vitro* un “*single guide RNA*” (sgRNA) programmable pour remplacer les trois éléments, crRNA, tracrRNA et la RNaseIII qui permettaient la formation du complexe actif (Jinek *et al.*, 2012).

Au même moment l'équipe de Siksnys montrait *in vitro* que l'endonucléase crée une DSB au niveau d'un site spécifique sur l'ADN qui contient une séquence de 20 nt complémentaire à celle portée par le crRNA. Il montre également que Cas9 possède deux sites actifs distinct, RuvC et HNH, qui génèrent respectivement une coupure sur chacun des deux brins d'ADN (Gasiunas *et al.*, 2012). Ceci a montré *in vitro* que le CRISPR/Cas était un outil polyvalent et programmable pour l'édition des génomes.

En 2013 (Cong *et al.*, 2013) et (Mali *et al.*, 2013) vont ajouter des NLS et humaniser la séquence codante de Cas9 pour permettre une meilleure expression et un meilleur adressage au noyau chez les cellules humaines et de souris. Ils vont également utiliser le promoteur polymérase III du gène U6 pour exprimer le sgRNA décrit par (Jinek *et al.*, 2012). Toutes ces optimisations ont permis que le CRISPR/cas9 soit efficace chez des organismes ayant une organisation nucléaire complexe par opposition aux bactéries dans lesquelles il a été découvert.

Les cassures double-brin ainsi créées par Cas9 doivent être réparées par la cellule au risque qu'elle ne survive pas. Pour cela la cellule pourra utiliser les voies de réparation classique telle que la recombinaison homologue (HR) ou le *non homologous end joining* (NHEJ).

Cet outil est désormais largement utilisé, et a été adapté chez plusieurs parasites, *P. falciparum* (Ghorbal *et al.*, 2014), *Toxoplasma gondii* (Shen *et al.*, 2014) *T. cruzi* (Peng *et al.*, 2015), *Cryptosporidium parvum* (Vinayak *et al.*, 2015) et désormais chez *Leishmania* (Sollelis *et al.*, 2015) comme nous le verrons dans le chapitre suivant.

3 Problématique et objectifs de mon travail de thèse

Comme je l'ai précisé dans l'introduction, le laboratoire s'intéresse depuis longtemps à l'organisation des chromosomes chez *Leishmania* et plus particulièrement depuis la mise en évidence de l'aneuploïdie mosaïque. L'hypothèse que l'aneuploïdie était générée et maintenue au cours des générations suivante par une régulation non conventionnelle de la réplication suivie d'une ségrégation permissive des chromosomes a été posée. Nous nous sommes donc intéressés à la réplication de l'ADN chez ce parasite.

Leishmania et *Trypanosoma* sont deux organismes proches du point de vue de l'organisation génomique, mais les Trypanosomes sont des organismes strictement diploïdes. Il paraissait donc intéressant de comparer les paramètres de réplication de l'ADN entre ces deux organismes afin de mettre en évidence des caractéristiques propres à chaque espèce, ce qui pourrait permettre de faire la lumière sur les mécanismes qui génèrent l'aneuploïdie. Dans un premier temps, afin de déterminer les dynamiques de réplication, nous avons utilisé la technique du peignage moléculaire.

La réplication de l'ADN est rendue possible par l'activation des origines de réplication qui sont sous le contrôle de différents acteurs et marques épigénétiques. Chez les trypanosomatidés les seules études à ce sujet ont été initiées chez *T. brucei* pour identifier les protéines impliquées dans la reconnaissance des origines de réplication.

L'objectif principal de ma thèse était de cartographier les origines de réplication de *Leishmania* en utilisant deux méthodes, le ChIP-seq et le séquençage du brin naissant d'ADN.

Il était également prévu de caractériser deux protéines Orc1/Cdc6 et Orc1b décrites chez *T. brucei* comme ayant un rôle dans le complexe pré-répliatif ORC. Cependant confrontée à la faible diversité des outils génétiques disponibles chez *Leishmania* j'ai dû consacrer beaucoup de temps à la mise au point du système CRISPR/cas9 chez *Leishmania*.

J'ai également participé au travail de Maria Rosa Garcia Silvia sur la ségrégation chromosomique, pour l'identification des centromères de *Leishmania*, avec la mise au point au laboratoire du ChIP-seq et la réalisation des bibliothèques pour le séquençage.

4 Implémentation du système CRISPR-Cas9 chez *Leishmania spp*

*Dans ce chapitre je décris l'implémentation du système du CRISPR/Cas9 chez *Leishmania*. Ce sujet n'était pas inscrit dans mon programme de thèse en Décembre 2013 et je n'avais jamais entendu parler du CRISPR à cette date.*

*En Avril 2014, le Dr JJ Lopez Rubio et son équipe, récemment installé au laboratoire, étaient en pleine révision de leur manuscrit, pour le journal *Nature Methods*, dans lequel ils ont décrit l'adaptation du système CRISPR/CAS9 chez *Plasmodium falciparum* (Ghorbal et al., 2014). JJ toujours débordant d'énergie a proposé son aide pour mettre en place le système chez *Leishmania*... L'idée ne fit pas l'unanimité au laboratoire mais avec le soutien d'Yvon et l'aide de JJ et de Mehdi, post-doc, j'ai tenté l'aventure, et moins d'un an après nous soumettions dans *Cellular Microbiology* un manuscrit sur la preuve de concept du CRISPR/Cas9 chez *Leishmania* (Sollelis et al., 2015). En plus de la publication, nous avons organisé en Juillet 2015 un workshop international regroupant les laboratoires qui nous avaient souhaité se procurer les plasmides. De puis nous continuons à donner ces plasmides aux personnes intéressées par la mise en place de ce système. La mise au point du CRISPR/Cas9 chez *Leishmania* m'a permis d'obtenir des prix de poster à des conférences, et une grande expérience scientifique.*

4.1 Introduction

Le développement du système CRISPR/Cas9 nous a paru indispensable chez *Leishmania*. En effet, il y a peu d'outils génétiques disponibles pour l'étude de ce parasite en comparaison avec d'autres organismes proches tels que *Trypanosoma brucei*. Par exemple, seul *T. brucei* possède la machinerie optimale pour induire l'extinction de l'expression d'un gène par RNA interférence (RNAi). Par ailleurs, l'organisation génomique particulière de *Leishmania* et surtout sa grande plasticité génomique n'ont jamais facilité l'obtention de Knock-Out (KO) complet, c'est-à-dire le remplacement de toutes les copies du gène qui est variable en fonction de la ploïdie. En général, au laboratoire, lorsque l'équipe s'intéressait à la fonction d'un gène chez *Leishmania*, ils utilisaient souvent en première intention la RNAi chez *T. brucei* et transposaient le résultat chez *Leishmania*.

Avec le système CRISPR-Cas9 nous avons l'espoir et l'ambition de pouvoir cibler tous les chromosomes homologues en même temps et ce, quel que soit leur nombre.

Ce chapitre décrit les différentes étapes qui ont été nécessaires à la mise au point du CRISPR/Cas9 pour différentes utilisations chez *Leishmania* : le KO, l'étiquetage (« tagging ») au locus et le développement d'un système inductible.

Toutes les applications du CRISPR-Cas9 que nous avons développées sont basées sur le même principe (Figure 4-1). La première étape consiste en l'expression épisomale de la Cas9 et du sgRNA²⁶ à partir de deux plasmides pTcas9 et pLS7 respectivement. Le sgRNA contient une séquence de 20 nucléotides que nous appellerons ici le « seed » qui peut être assimilé au « spacer », il est spécifique du gène d'intérêt (GOI). Dans une seconde étape Cas9 et le sgRNA vont s'associer pour former un complexe ARN/protéine. Dans un troisième temps ce complexe scanne le génome jusqu'à trouver une séquence complémentaire au « seed » à proximité du PAM²⁷. L'endonucléase provoquera une cassure double-brin qui sera réparée par la cellule à l'aide de l'ADN donneur. Cet ADN donneur est encadré par des régions d'homologie pour permettre la recombinaison

²⁶ De l'anglais Single Guide RNA

²⁷ De l'anglais Protospacer Adjacent Motif, dont la séquence est ici NGG

homologue et peut être de différente nature : cassette de résistance, «tag²⁸», mutation ponctuelle ou autre selon la modification souhaitée.

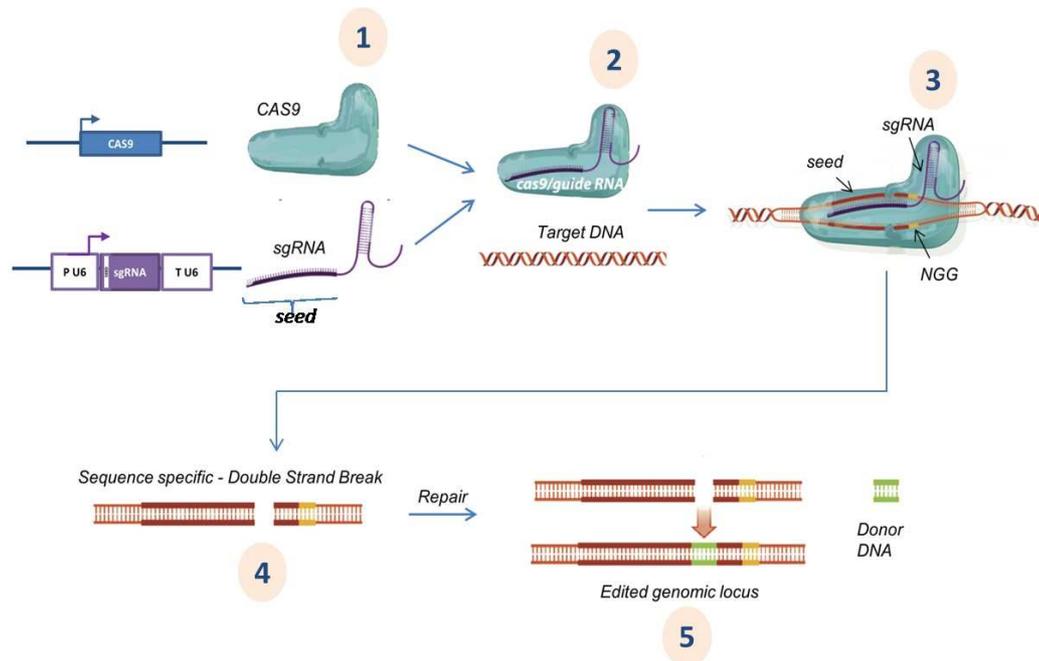


Figure 4-1 CRISPR/Cas9 in Leishmania.

Step 1: production of the endonuclease Cas9 and the sgRNA by episomal expression in a *Leishmania* strain. Step 2: formation of the Cas9/guide RNA complex. Step 3: scanning of the genomic DNA until the complex finds a complementary sequence to the seed adjacent to a PAM motif (NGG). Step 4: Cas9 creates a double strand break (DSB) and Step 5: the cell repairs the DSB with a donor DNA that we provide.

4.2 Résultats

4.2.1 Knock Out (article CRISPR-cas9, Sollelis et al. Cell Microbiol 2015)

4.2.1.1 Introduction

La première étape a été de montrer la faisabilité du système CRISPR/cas9 chez *Leishmania*. Pour cela, nous avons choisi de réaliser le KO d'un gène connu comme non essentiel qui avait déjà été réalisé chez *Leishmania*. Notre choix s'est porté sur PFR2 (*Paraflagellar protein 2*) (Santrich *et al.*, 1997). De plus, nous possédons un anticorps anti-PFR2 et avons la possibilité de faire une sonde de FISH ciblant le locus

²⁸ De l'anglais Tag : étiquette

PFR2, ce qui nous a permis de réaliser un criblage efficace de nos parasites après édition du génome par CRISPR/Cas9. Ce travail correspond à l'article N°1.

4.2.1.2 Preuve de concept du CRISPR-Cas9 (Article 1 : Sollelis et al. Cell Microbiol 2015).

Breaking report

First efficient CRISPR-Cas9-mediated genome editing in *Leishmania* parasites

Lauriane Sollelis,^{1,2} Mehdi Ghorbal,^{1,2}
Cameron Ross MacPherson,³
Rafael Miyazawa Martins,³ Nada Kuk,¹
Lucien Crobu,² Patrick Bastien,^{1,2,4} Artur Scherf,³
Jose-Juan Lopez-Rubio^{1,2} and Yvon Sterkers^{1,2,4*}

¹University of Montpellier, Faculty of Medicine,
Laboratory of Parasitology-Mycolology.

²CNRS – 5290, IRD 224 – University of Montpellier
(UMR 'MiVEGEC').

³Institut Pasteur – INSERM U1201 – CNRS ERL9195,
"Biology of Host-Parasite Interactions" Unit, Paris,
France.

⁴CHRU (Centre Hospitalier Universitaire de Montpellier),
Department of Parasitology-Mycolology, Montpellier,
France.

Summary

Protozoan pathogens that cause leishmaniasis in humans are relatively refractory to genetic manipulation. In this work, we implemented the CRISPR-Cas9 system in *Leishmania* parasites and demonstrated its efficient use for genome editing. The Cas9 endonuclease was expressed under the control of the Dihydrofolate Reductase-Thymidylate Synthase (DHFR-TS) promoter and the single guide RNA was produced under the control of the U6snRNA promoter and terminator. As a proof of concept, we chose to knockout a tandemly repeated gene family, the paraflagellar rod-2 locus. We were able to obtain null mutants in a single round of transfection. In addition, we confirmed the absence of off-target editions by whole genome sequencing of two independent clones. Our work demonstrates that CRISPR-Cas9-mediated gene knockout represents a major improvement in comparison with existing methods. Beyond gene knockout, this genome

editing tool opens avenues for a multitude of functional studies to speed up research on leishmaniasis.

Introduction

The protozoan parasite *Leishmania* causes a large spectrum of diseases in 88 countries around the world, thus representing the second cause of parasite-related deaths after malaria. The absence of effective vaccines, the extremely limited available therapeutic arsenal and the development of drug resistances are major drawbacks for clinical case management and eradication (den Boer *et al.*, 2011).

In the post-genomic era, genome editing became a major tool used in a wide range of experimental systems (Gilles and Averof, 2014). Gene knockouts were particularly used, aiming at the replacement of the targeted region by a selective marker. For this, classical molecular biology methods were first used, based on double cross-over events in homology regions (HRs) initiated by a random DNA double strand break (DSB). Linear DNA molecules bearing HRs at the ends were then used in order to increase efficiency, as the cellular machinery recognizes the DNA free ends which trigger recombination, then using the endogenous locus as a template. DSB is one of the most deleterious DNA damages that can affect cell viability and cells have developed different DNA repair pathways to overcome DSBs and survive: homologous recombination, non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). In *Leishmania*, homologous recombination is prevailing, and, in spite of the presence of Ku proteins, the NHEJ pathway seems to be absent (Passos-Silva *et al.*, 2010). More recently, it has been shown that *Trypanosoma cruzi*, a related parasite of the trypanosomatid family, can use the MMEJ pathway in the absence of template to repair DSB (Peng *et al.*, 2015).

In order to enhance the probability of a DSB and to target it on a desired region of the genome, different genome editing systems using engineered nucleases have been developed, such as Zinc-finger nucleases,

Received 30 March, 2015; revised 24 April, 2015; accepted 27 April, 2015. *For correspondence. E-mail: yvon.sterkers@univ-montp1.fr; Tel. (+33) 467 63 55 13; Fax (+33) 467 63 23 58.

© 2015 John Wiley & Sons Ltd

Talen and clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated endonuclease 9 (CRISPR-Cas9; Gaj *et al.*, 2013). None of these have been implemented so far in *Leishmania*. The CRISPR-Cas9 editing system, which was engineered from a prokaryotic immune system (Jinek *et al.*, 2012), is constituted by a single guide RNA (sgRNA) which guides the Cas9 endonuclease to cause targeted DSBs. The sgRNA carries the Cas9 binding domain and a customizable 20 nt-sequence (hereafter referred to as the seed) which gives the specificity for the target DNA site. DSB occurs three nucleotides upstream of a protospacer adjacent motif within the guide RNA target region (Jinek *et al.*, 2012). CRISPR-Cas9 system has been recently developed as a powerful technique for genome editing and as a molecular genetic tool to study gene function in various organisms, including protozoan parasites such as *Plasmodium*, *Toxoplasma* and *Trypanosoma cruzi* (Ghorbal *et al.*, 2014; Shen *et al.*, 2014; Wagner *et al.*, 2014; Peng *et al.*, 2015). The system appears particularly interesting in *Leishmania* sp. because classical genome manipulation methods experience serious limitations in this organism: (i) one of them is the difficulty in realizing gene knockouts due to its high genomic plasticity, and in particular to 'mosaic aneuploidy' (Sterkers *et al.*, 2011; Lachaud *et al.*, 2014); indeed, more than two rounds of transfection may be necessary, and in the case of essential genes, additional copies of the targeted gene/chromosome fragment/whole chromosome may arise at each round of transfection (Dubessay *et al.*, 2002); (ii) moreover, RNA interference is inefficient due to the absence of RNAi-related genes such as dicer and argonaute in most species (Lye *et al.*, 2010). Here, we set up the CRISPR-Cas9 system in *Leishmania major* using the disruption of the paraflagellar rod (PFR)2 genes as a proof of concept. In addition, we confirmed the absence of off-target edition by whole genome sequencing, as this is a major concern which is generally poorly addressed in this type of experiments.

Results and discussion

Edition of the PFR2 locus

Paraflagellar rod-2 is a major constituent of the PFR, which is a trypanosomatid-specific filamentous network attached to the axoneme of the flagellum (Maga and LeBowitz, 1999). It is encoded by three genes tandemly arrayed in a single locus on chromosome 16: LmjF16.1425 (from 591 295 to 593 094 bp), LmjF.16.1427 (594 664–596 463 bp) and LmjF.16.1430 (598 036–599 835 bp). We chose to edit the PFR2 locus because (i) it is not an essential gene as null mutants that divide normally in culture have been reported (Santrich

et al., 1997; Lye *et al.*, 2010) and (ii) specific antibodies targeting the PFR2 are available (Ismach *et al.*, 1989). To edit the PFR2 locus, we have chosen to use a two plasmids strategy: pTCAS9 carrying the Cas9 (Cong *et al.*, 2013) endonuclease gene and pLS7 which encodes the sgRNA and the donor cassette constituted by the HRs and the resistance marker (Fig. 1). We first generated *L. major* mutant parasites (termed LmjF_3171-CAS9 or M756) stably carrying the pTCAS9 episomal vector (Fig. 1A). Expression of CAS9 was checked by Western blot using an anti-flag antibody (Fig. 1B). The system requires sgRNA expression without RNA polIII-associated modifications. RNA polIII promoters such as U6 has been successfully used to implement the CRISPR-Cas9 system in other organisms, and in particular, U6 small nuclear (sn)RNA regulatory elements have been used in *Plasmodium falciparum* (Ghorbal *et al.*, 2014). The promoter and terminator of the U6snRNA (LmjF.24snRNA.01) have been reported in *Leishmania* and the minimal sequences to promote its transcription characterized (Nakaar *et al.*, 1994). The promoter sequence used here harbors sequence for the tRNA^{Gln} gene with its canonical BOX A and B, flanked by the 5'untranslated region (UTR) of U6snRNA (215 bp). The terminator comprised 120 bp of the 3'UTR from U6snRNA. In pLS7, the sgRNA was inserted between the promoter and the terminator (Fig. 1B and C). To ensure maximal specificity and efficiency of the CRISPR-Cas9 editing system, we searched for (i) two large HRs to promote efficient homologous recombination and (ii) a seed close to one of the HRs. We designed two HRs of ~ 1 kb flanking the complex locus PFR2 and a seed which targeted a 20-nt single copy sequence, preceding an NGG motif in the 5'UTR next to HR1. The HRs were cloned flanking the puromycin resistance gene (Fig. 2A and Table S1). M756 (LmjF3171-CAS9) promastigotes were transfected with the linearized pLS7 plasmid and transfectant parasites (M790) were obtained at day 11. PCR experiments showed not only the presence of edited loci but also the persistence of wild-type loci (Fig. 2B). Fluorescence *in situ* hybridization (FISH) analysis was then performed using two DNA probes from chromosome 16, one targeting the PFR2 locus and the second one another repeated gene (LmjF16.1030): PFR2-positive cells were reduced in M790 in comparison with the wild type (LmjF), whereas LmjF16.1030-positive cells were similar in both cell lines (Fig. 2C and Table S2). FISH analysis thus detected a subpopulation of parasites with no signal for PFR2 (Fig. 2C). This was confirmed by a specific immunofluorescence assay, showing both PFR2-positive and PFR2-negative cells in the M790 cell line. We then decided to clone this cell line by limiting dilution. Eight clones were screened by PCR and FISH. In two clones (M790_B4 and M790_D2), no remaining PFR2

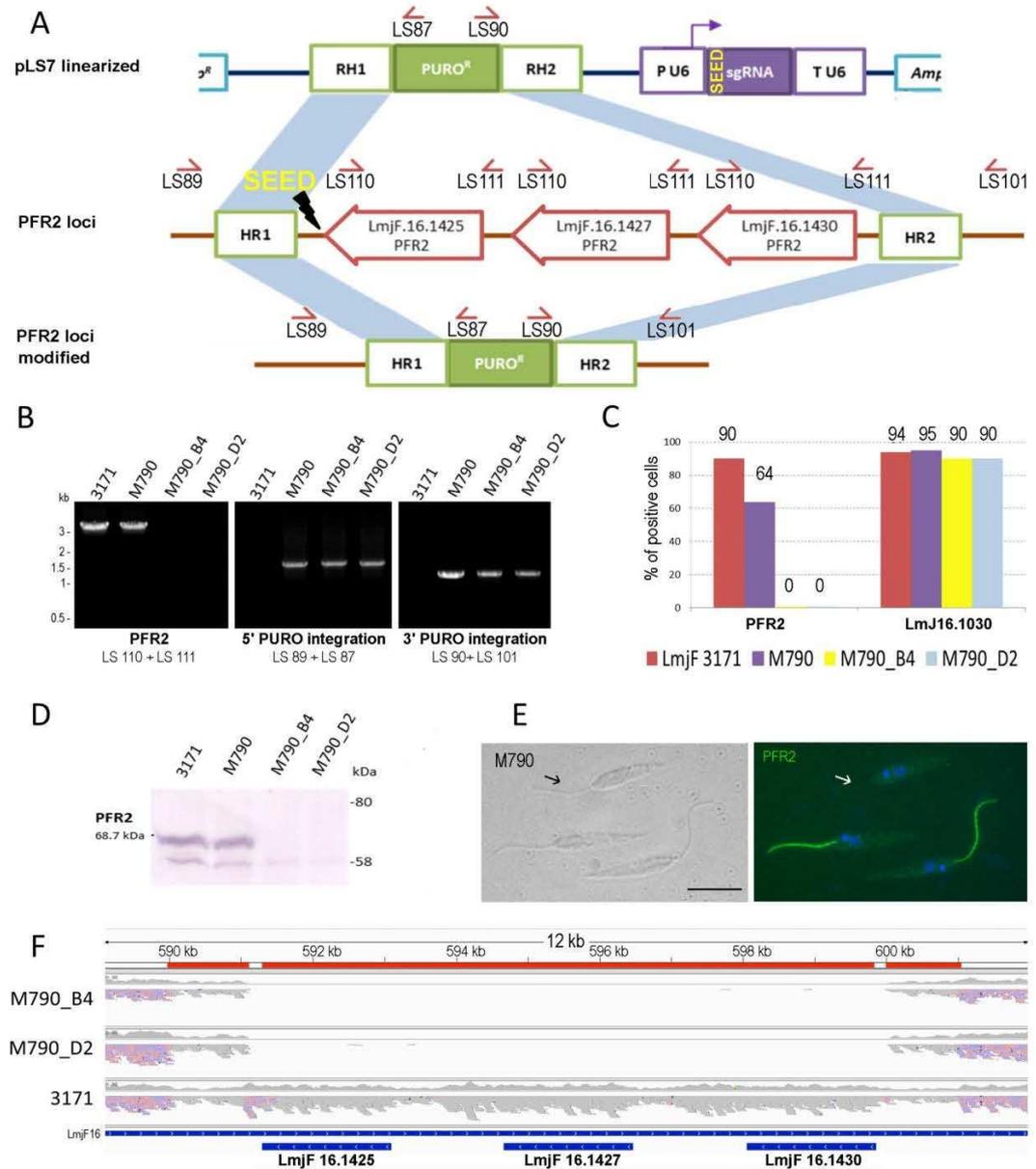


Fig. 2. Editing of the PFR2 loci using the CRISPR-Cas9 editing system in *Leishmania major*.

A. Schematic representation of the repaired locus after the double DNA strand break. pLS7 plasmid linearized by *ScaI* were positioned where the endonuclease produced the DSB. The seed targeted the 3'UTR next to HR1 (lighting). Red half-arrows (LS89 to LS91): primers used to check allele replacement in the PFR2 locus. Edition of the PFR2 locus was thoroughly verified using PCR (**B**), FISH (**C**), Western blot (**D**), an immunofluorescence assay (**E**) and whole genome sequencing (**F**).

B. PFR2 genes were PCR-amplified in the wild-type (LmjF_3171) and M790 cell lines and found absent in clones M790_B4 and _D2. PCR testing of the 5' and 3' integration was positive in the clones and absent in the two parental lines.

C. Using FISH, the PFR2 probe failed to detect chromosome 16 in the clones whereas the Lmj16.1030 probe detected it in all four cell lines.

D. No PFR2 expression was found by Western blot in the two clones.

E. Immunofluorescence using a specific Ab showed a mixture of PFR2-positive cells and edited PFR2-negative cells in M790; arrows: PFR-2 negative flagellum, bar = 5 μ m.

F. Coverage plot of the targeted region on chromosome 16 at positions 591 295 through 599 835. Mapping tracks are from B4, D2 and parental lines, from top to bottom respectively. The three red bars indicate, from left to right, HR1, the target region and HR2. The blue bars indicate gene annotations from GeneDB.

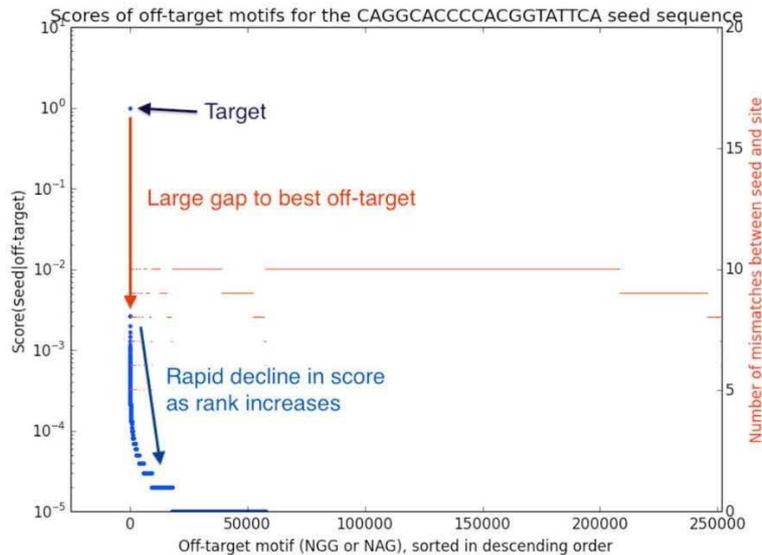


Fig. 3. Genome-wide scan of likely Cas9 targets given the seed sequence. Putative targets are ranked along the X-axis from most to least likely-to-cleave (according to Hsu *et al.*'s scoring algorithm). Y-axis 1 (blue) indicates the Hsu *et al.*'s pairwise score; Y-axis 2 (red) indicates the number of mismatches. The intended target in this figure is the top-left most point, at rank 1, as expected from any decently designed sgRNA. In this plot, the distance to the next best target and the drop-off rate is indicative of quality; here, the sgRNA sequence demonstrates a high-quality sgRNA design.

M790_B4, _D2 and LmjF_3171 respectively. M790_B4 shared 70% of its parental INDELS, leaving 401 potentially novel INDELS caused by the CRISPR protocol. Likewise, the M790_D2 line contained 868 novel INDELS. However, none of the INDELS were found within 20-nt of any putative off-target site (up to 10 mismatches) ruling out the possibility that these were caused by Cas9 off-target cleavage. All the evidences therefore point towards successful modification of the target with no off-target cleavage within detectable limits. The data obtained from whole genome sequencing of the reference strain and the two edited clones were also used to map total genomic contents to a reference-plasmid hybrid genome and thereby confirm recombination. The targeted site (Fig. 2E) appeared as a deletion in M790_B4 and _D2, but not in the wild type. In summary, the target region appears exactly as one would expect from a successful homology-directed repair between HR1 and HR2 (Fig. 2E).

Conclusions and perspectives

To our knowledge, this is the first demonstration of CRISPR-Cas9 genome editing in *Leishmania*. We report here a specific gene knockout targeting a complex locus made of three tandemly arrayed genes. We were able to obtain null parasites in a single round of transfection, that is in a short time frame (< 1 month), whereas two rounds of transfection were needed in the classical approach to knockout PFR2 (Santrich *et al.*, 1997). The CRISPR-Cas9 system may also overcome the remarkable genomic plasticity observed in *Leishmania* parasites which leads to additional copies of the targeted gene when it is essential

for viability. Recovering transfectant lines when an essential gene is targeted remains cumbersome in spite of the adaptation of CRISPR-Cas9 system. This issue could be solved by transfecting an ectopic copy of the target gene. Yet, a major improvement of the system would be to develop an inducible Cas9 gene. The CRISPR-Cas9 genome editing system in *Leishmania* sp. should open opportunities for other applications, in particular, *in situ* tagging and single nucleotide substitution marker-free approaches, which, considering the limited number of selectable markers available for *Leishmania*, would be invaluable for consecutive genome manipulations. We therefore expect that the CRISPR-Cas9 system will rapidly become a routine laboratory genome editing tool in the field of *Leishmania*.

Experimental procedures

Parasites

Promastigote forms of the *L. major* genome reference strain 'Friedlin' (MHOM/IL/80/Friedlin, LEM3171), herein termed LmjF_3171, were grown in supplemented RPMI1640 medium as described (Dubessay *et al.*, 2004). Parasites were maintained in log phase by splitting the culture every second day. For FISH analysis, in order to minimize the number of dividing cells in the population, parasites were grown to late log phase and then harvested.

Construction of vectors

All cloning experiments were performed using the In-Fusion HD Cloning Kit (Clontech); the 15 nt adaptors were added to the

primers. The endonuclease Cas9 gene together with nuclear localization signal motifs and three flag tags were amplified from pUF1-Cas9 (Ghorbal *et al.*, 2014) using the primer pair LS84/LS85. The PCR product was cloned into pTH6 (Dubessay *et al.*, 2006) which bears the hygromycin resistance gene and where Cas9 is expressed under the Dihydrofolate Reductase-Thymidylate Synthase (DHFR-TS) promoter (pTCAS9) (Fig. 1A). The design of pLS6 was based on the pL6 plasmid used for *P. falciparum* (Ghorbal *et al.*, 2014). The sgRNA cassette expression was ordered as a gene block (IDT Company®) to avoid extra bases from restriction sites; it was then inserted on a pGEMTeasy™ (Promega®) backbone between the *Apal* and *NcoI* restriction sites. We then proceeded to cloning the homology regions boxes (HR1 and HR2): single-copy sequences on each side of the three copies of PFR2 genes were amplified using the primer pairs LS75/LS76 and LS77/LS79 for HR1 and HR2 respectively (Fig. 1B). HR1 (1137 bp) and HR2 (1054 bp) were cloned between *KpnI* and *XbaI* and between *MfeI* and *HpaI*, respectively, and flanked the selection marker (puromycin).

The 20-nt of the seed surrounded by the 15-nt adaptors necessary for InFusion® cloning were ordered as two oligonucleotides (PFR2Seed3For and PFR2Seed3Rev primers in Table S4) and annealed using the following program: 95°C for 10 min; cooling from 95°C to 25°C using a $-1.5^{\circ}\text{C s}^{-1}$ ramp and 1 min pause every 10°C and keeping at 4°C. Then, the seed was cloned into the pLS6 plasmid previously digested by *BsgI* to obtain pLS7.

Transfection and cloning

Parasites were transfected as previously described by electroporation (Dubessay *et al.*, 2006). For transfections with circular DNA, 100 µg of each plasmid were used. For transfections with linear DNA, plasmids were linearized using the restriction enzyme *ScaI*, verified for linearization in agarose gel and 100 µg were subsequently used for transfection. Prior to transfection, the appropriate DNA was ethanol-precipitated and resuspended in 50 µl of water. Drug pressure (30 µg ml⁻¹ hygromycin and 30 µg ml⁻¹ puromycin) was applied 15–20 h after transfection. For cloning by limit dilution, exponential log phase culture was diluted to 5×10^3 cells ml⁻¹ and several dilutions were performed to obtain 0.5 cell per well in 96 culture plates. Positive wells were transferred into 24-well plate to expand the culture.

Western blot

Parasites were centrifuged and washed with a protease inhibitor cocktail (Sigma®). Cells were then lysed for 5 min at 75°C in loading buffer (0.5 M Tris-HCl 10% glycerol, 2% sodium dodecylsulfate (SDS), 2% dithiothreitol (DTT), 0.01% bromophenol blue). A total of 5×10^6 cell equivalents were loaded per well and the lysates were separated on a 10% SDS-polyacrylamide gel electrophoresis gel and then transferred onto a Polyvinylidene Difluoride (PVDF) membrane. Membranes were blocked using PBS, 6% skimmed milk and 0.05% Tween 20 for 1 h. For Cas9 expression, membranes were probed with a mouse anti-FLAG antibody (Sigma®) diluted at 1:100 then by alkaline phosphatase-conjugated anti-mouse secondary antibodies diluted at 1:7500 and developed with nitro-blue tetrazolium and

5-bromo-4-chloro-3'-indolylphosphate (NBT/BCIP) (Promega®). To detect PFR2 expression, the antibodies used were a mouse monoclonal anti-PFR2 antibody (generously provided by Professor Diane McMahon-Pratt, Yale) at 1:500 and an anti-mouse phosphatase alkaline conjugated (S3721, Promega®) at 1:7500.

Analysis of mutant parasite strains by PCR

In the mutants, target-gene disruption and puromycin-cassette integration (5' integration and 3' integration) were verified by performing PCR on genomic DNA (see Table S4 for primer pairs list). LmjF_3171 was used as a control. gDNA extraction was performed by DNeasy Blood and Tissue Kit (Qiagen®) according to the manufacturer's instructions.

FISH

The two DNA probes used for FISH were constituted by the PFR2 genes and the LmjF16.1030 loci which were PCR-amplified using the primer pairs LS110/LS111 and LS113/LS114 respectively (see Table S4). Probes were labelled with tetramethyl-rhodamine-5-dUTP (Roche Applied Sciences®) using the Nick Translation Mix (Roche Applied Sciences®). FISH analysis was performed as described elsewhere (Sterkers *et al.*, 2011). Slides were mounted in Vectashield with 4',6'-diamidino-2-phénylindole (DAPI) (Vector Laboratories®).

Immunofluorescence assay

Cells were fixed in 4% paraformaldehyde and air-dried on microscope immunofluorescence slides. Slides were treated with triton 0.2% in PBS 1X and saturated with fetal bovine serum (FBS) 2%. Anti-PFR2 antibody (2E10B7, kindly provided by Professor Diane McMahon-Pratt, Yale) diluted to 1:1000 was added for 1 h. After three washes with PBS 1X/2% FBS, goat anti-mouse AF488 secondary antibodies (Molecular Probes®, ref. A-11001) diluted to 1:1500 were added for 45 min. Slides were finally mounted with Vectashield (Vector Laboratories®) and Hoechst at 1:1000.

Microscopy and imaging

Leishmania cells were viewed by bright field and fluorescence was visualized using appropriate filters on a Zeiss® Axioplan 2 microscope with a 100× objective. Digital images were captured using a Photometrics CoolSnap CCD camera (Roper Scientific®) and processed with MetaView (Universal Imaging®). In order to view the entire nucleus and to count the total number of labelled chromosomes, we used a Piezo controller for Z-Stack image acquisitions of 20 planes of 0.25 µm for each analysed cell. Of note, for FISH analysis, mitotic cells were excluded from the analysis.

Whole genome sequencing

DNeasy Blood and Tissue Kit (Qiagen®) was used to prepare the genomic DNA from LmjF_3171, M790_B4 and M790_D2 cell lines. High-power sonication in a Bioruptor UCD-200 (Diagenode®), for three 10 min cycles, with 30 s ON pulses,

followed by 30 s OFF, was used to fragment genomic DNA (1.25 µg) in 125 µl of Tris-ethylenediaminetetraacetic acid (EDTA) buffer. DNA fragments (100 ng) were used to prepare indexed DNA libraries for next-generation sequencing using the TruSeq Nano DNA LT Sample Prep Kit (Illumina®) following the manufacturer's instructions. After purification, fragment sizes and library quality were estimated in a Bioanalyzer 2100 (Agilent®) and the DNA concentration was determined in a Qubit Fluorometer (Life Technologies®). Then, 100 bp reads were sequenced in a HiSeq2500 sequencer (Illumina®) in the Transcriptome and Epigenome Platform (PF2) at Institut Pasteur.

Off-target analysis

The Protopacer Workbench software suite was used to find putative off-targets up to 10 nucleotides different from the original seed, 'CAGGCACCCACGGTATTCA'. The software scored each off-target according to the algorithm published elsewhere (Hsu *et al.*, 2013). The resulting score for each off-target is proportional to the chance of off-target cleavage actually occurring. VarScan2 software was used to detect indels and indel/off-target overlap occurring in B4, D2 and LmjF_3171 in comparison with the current version of the *L. major* genome (Friedlin strain) at GeneDB database (2013-01-16). The software's suitability and parameters were previously evaluated (Ghorbal *et al.*, 2014) and showed high sensitivity and specificity at a *P*-value threshold of 0.05.

Acknowledgements

We gratefully thank Michel Pagès for fruitful discussions and for critical reading of the manuscript. We thank Professor Diane McMahon-Pratt (Yale School of Public Health) for the gift of anti-PFR2 antibodies. We also thank Odile Sismeiro, Caroline Proux and Jean-Yves Coppée from Institut Pasteur's Plateforme Transcriptome et Epigenome (PF2) for helping in the DNA sequencing. Finally, we gratefully acknowledge Vicki Diakou (Plateau MRI DBS-UM Optique, RIO Imaging Platform, Montpellier) and Laurence Berry (University of Montpellier/CNRS, UMR 5539) for expert advice and assistance in fluorescence microscopy, as well as Mauld Lamarque (University of Montpellier/CNRS, UMR 5539) for her fruitful help at a critical step. This work was supported by the Agence Nationale de la Recherche within the frame of the 'Investissements d'avenir' program (ANR-11-LABX-0024-01 'PARAFRAP'), ERC grant (PlasmoEscape 250320), ANR-13-ISV3-0003-01-NSFC (no. 81361130411), ANR 11 JSV3 004 01 PlasmoPiggyBac, the Centre National de la Recherche Scientifique, the French Ministry of Research and the Centre Hospitalier Universitaire de Montpellier. LS was funded by the French Parasitology consortium ParaFrap (ANR11-LABX0024), MG by ANR 11 JSV3 004 01 PlasmoPiggyBac, CRM and RMM were funded by ANR-13-ISV3-0003-01-NSFC. JJLR was funded by the Institut National de la Santé et de la Recherche Médicale.

References

den Boer, M., Argaw, D., Jannin, J., and Alvar, J. (2011) Leishmaniasis impact and treatment access. *Clin Microbiol Infect* **17**: 1471–1477.

- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–823.
- Dubessay, P., Ravel, C., Bastien, P., Crobu, L., Dedet, J.P., Pages, M., and Blaineau, C. (2002) The switch region on *Leishmania major* chromosome 1 is not required for mitotic stability or gene expression, but appears to be essential. *Nucleic Acids Res* **30**: 3692–3697.
- Dubessay, P., Blaineau, C., Bastien, P., and Pages, M. (2004) Chromosome fragmentation in leishmania. *Methods Mol Biol* **270**: 353–378.
- Dubessay, P., Blaineau, C., Bastien, P., Tasse, L., Van Dijk, J., Crobu, L., and Pages, M. (2006) Cell cycle-dependent expression regulation by the proteasome pathway and characterization of the nuclear targeting signal of a *Leishmania major* Kin-13 kinesin. *Mol Microbiol* **59**: 1162–1174.
- Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**: 397–405.
- Ghorbal, M., Gorman, M., Macpherson, C.R., Martins, R.M., Scherf, A., and Lopez-Rubio, J.J. (2014) Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat Biotechnol* **32**: 819–821.
- Gilles, A.F., and Averof, M. (2014) Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. *Evodevo* **5**: 43.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**: 827–832.
- Ismach, R., Cianci, C.M., Caulfield, J.P., Langer, P.J., Hein, A., and McMahon-Pratt, D. (1989) Flagellar membrane and paraxial rod proteins of *Leishmania*: characterization employing monoclonal antibodies. *J Protozool* **36**: 617–624.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821.
- Lachaud, L., Bourgeois, N., Kuk, N., Morelle, C., Crobu, L., Merlin, G., *et al.* (2014) Constitutive mosaic aneuploidy is a unique genetic feature widespread in the *Leishmania* genus. *Microbes Infect* **16**: 61–66.
- Lye, L.F., Owens, K., Shi, H., Murta, S.M., Vieira, A.C., Turco, S.J., *et al.* (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* **6**: e1001161.
- MacPherson, C.R., and Scherf, A. (2015) Protopacer Workbench: flexible guide-RNA design for CRISPR applications. *Nat Biotechnol*. In press.
- Maga, J.A., and LeBowitz, J.H. (1999) Unravelling the kinetoplastid paraflagellar rod. *Trends Cell Biol* **9**: 409–413.
- Nakaar, V., Dare, A.O., Hong, D., Ullu, E., and Tschudi, C. (1994) Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. *Mol Cell Biol* **14**: 6736–6742.
- Passos-Silva, D.G., Rajao, M.A., Nascimento de Aguiar, P.H., Vieira-da-Rocha, J.P., Machado, C.R., and Furtado, C.

- (2010) Overview of DNA repair in *Trypanosoma cruzi*, *Trypanosoma brucei*, and *Leishmania major*. *J Nucleic Acids* **2010**: 840768.
- Peng, D., Kurup, S.P., Yao, P.Y., Minning, T.A., and Tarleton, R.L. (2015) CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. *MBio* **6**: e02097-14.
- Santrich, C., Moore, L., Sherwin, T., Bastin, P., Brokaw, C., Gull, K., and LeBowitz, J.H. (1997) A motility function for the paraflagellar rod of *Leishmania parasites* revealed by PFR-2 gene knockouts. *Mol Biochem Parasitol* **90**: 95–109.
- Shen, B., Brown, K.M., Lee, T.D., and Sibley, L.D. (2014) Efficient gene disruption in diverse strains of *Toxoplasma gondii* using CRISPR/CAS9. *MBio* **5**: e01114.
- Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P., and Pagès, M. (2011) FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. *Cell Microbiol* **13**: 274–283.
- Wagner, J.C., Platt, R.J., Goldfless, S.J., Zhang, F., and Niles, J.C. (2014) Efficient CRISPR-Cas9-mediated genome editing in *Plasmodium falciparum*. *Nat Methods* **11**: 915–918.

Sequence data submission

Sequences of the plasmids pTCas9 and pLS5 have been submitted to the GenBank database. ChIP-seq data have been submitted to the "European Nucleotide Archive (ENA)", with the accession number: PRJEB9349, and accessible online through the following url: <http://www.ebi.ac.uk/ena/data/view/PRJEB9349>

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. List of essential sequences used for the construction of plasmids.

Table S2. Validation of the CRISPR-Cas9 edition of the PFR2 locus in *Leishmania major*.

Table S3. Top 10 potential off-target sites of the sgRNA in LmjF.

Table S4. Primer pairs used in the study.


```

gcacaagcgtctggggggggagagcgggctaacagcgtatagaggtgggagatggggcgtatgctcaacctcaccattccctcctctgctcgccttccgctgc
tgctgttcccttcgatacgcctctggccccccccccacacacccacatctaactgggggacagcacacgtgagcggaaaggggcaagtgaccgaggtgg
ctggtccaggtcccgggtcctccacacctgcagcaccggcgtgggtggctcgtgagccccctgtcgggaagtgccagagtcgccattccgcgtccacccc
cggatcgacgcggcactgctgagctctgcggttctcgttgcctcgtctctgagaagagttataattcaactgcgagaagaatgccgagcacagagtg
cgctctcaacggagcggcacacgcacacacagggcagatgaacagaccctcaggtgaacataaaaagacagaggagggacgggagctgtaatacagctgcg
gtcatgacatactttacatggattccacgagaccagcaaggagagacccgacaaaacagagaagcaaggcatcaacgcagcgcagagattgtaggatgcagaca
tgaacaggcacacgcgcagctgcccgcgctgcaggggaacgcagcagggcctccctccccctacacacgcgcgcgcagggatccaatgacacccagcag
acgacgcttgaagacgaaacacacagctacagggcaaaaggggtacatagaacgctacacgtgaagggaaagccaaacacagaagtgtaactcgtatgcattc
ccaccgacacagcgtgctgacagggcacacacaccacaatagtaagaagactgctgaggaggaggggggggggcactgtggcgtagcagcggcccaga
gcaatgcagccacgcggggaagaacaagagactggacgcgagctcgtcagcttcgcgaccgaaacagatgatacggggacattatccagtg

```

In the Cas 9 sequence, NLSs are highlighted in red and the three FLAG tags in green. In the 5' and 3' PURO integration sequences: plasmid sequences are in uppercase and genomic DNA sequences are highlighted in grey and lower case.

Table S3: Top-10 potential off-target sites of the sgRNA in LmjF

Location	Strand	Target	PAM	Mismatches	Score	Type
LmjF.16:591220	+	CAGGCACCCACGGTATTCA	CGG	0	1.0E+0	Intended target
LmjF.24:43866	+	CtGGCgCCgCAaGGTATTCA	CAG	4	2.6E-3	Off-target
LmjF.24:152979	+	CtGGCgCCaCAaGGTATTCA	CAG	4	2.6E-3	Off-target
LmjF.02:327010	-	CAcGtgCgtgACGGTATTCA	CGG	6	2.0E-3	Off-target
LmjF.28:274601	+	AAGGCACaCCACGGcAaTCA	CAG	3	1.7E-3	Off-target
LmjF.04:366647	-	TAtGaACCCgACGGTcTTCA	AGG	4	1.5E-3	Off-target
LmjF.14:145241	+	TAtGCACgCCACGGTATTgg	CGG	4	1.3E-3	Off-target
LmjF.34:160890	+	GctGgACggCACGGTATTcc	TGG	6	1.2E-3	Off-target
LmjF.23:506808	+	GAGatgaaaCACGGTATTCA	AGG	6	1.1E-3	Off-target
LmjF.14:201757	+	CgtcCAcTcgtgGGTATTCA	AAG	7	1.1E-3	Off-target

Table S2: Validation of the CRISP-Cas9 edition of the PFR2 locus in Leishmania major

Strain name	Description	Edited parasites	WT/PFR2 (PCR)	5' PURO integration (PCR)	3' PURO integration (PCR)	PFR2 (FISH)	LmjF16.1030 (FISH)	PFR2 (Wblot)	PFR2 (IFA)	WGS
LmjF_3171	Reference strain	NA	Pos	Neg	Neg	90%	94%	Pos	Pos	Reference sequence
M756	LmjF3171_pTCas9	NA	NA	NA	NA	NA	NA	NA	NA	NA
M809	M756_pLS6 circular	No	Pos	Neg	Neg			Pos	Pos	
M810	M756_pLS6 linear	No	Pos	Neg	Neg			Pos	Pos	
M788	M756_pLS7 circular	Yes, in combination with WT parasites	Pos	Pos	Pos			Pos	Mixed population	
M790	M756_pLS7 linear	Yes, in combination with WT parasites	Pos	Pos	Pos	64%	95%	Pos	Mixed population	
M790_B4 M790_D2	M756_pLS7 linear	Yes	Neg	Pos	Pos	0% 0%	90% 90%	Neg	Neg	Edited locus, no off target

WT: wild-type, FISH: Fluorescent in situ hybridization, Wblot: Western blot, IFA: Immunofluorescence assay, WGS: whole genome sequence, NA: not applicable, Pos: positive, Neg: negative, pLS6: control plasmid without seed, pLS7: plasmid with the seed.

Table S4: Primer pairs used in the study

Primer name	Primer sequence	Application
LS75_for	ATTCGATTCGAAGGGTACC ATGCCACACAAGAAGGG	To clone homology region 1 in pLS7
LS76_rev	GATAGGAGGG CTAGA CATGCTTGTAAGTGTGTTTGG	
LS77_for	GTGCAAGCCTCAATTG GATGAAATCTGGGAACCTCG	To clone homology region 2 in pLS7
LS79_rev	CTGCAGCTGAAGGTTAAC TCACACGGCCAGAGG	
LS84_rev	ATGGTGATGGTGGTTAAC TTACTTTTTCTTTTTGGCC	To clone CAS9 in pTH6
LS85_for	GAGCACTAGTGGTTTAAAC ATGGACTATAAGGACCAG	
PFR2Seed3For	CAGGCACCGCTGGTGGC AGGCACCCACGGTATTCA GTTTTAGAGCTAGAAAT	To clone the seed in pLS7
PFR2Seed3Rev	ATTCTAGCTCTAAAAC TGAATACCGTGGGGTGCT GCCACCAGCGGTGCCTG	
LS87_rev	TCCTCGATCTCCTCGTG	To assess PFR2 edition : 5' PURO integration PCR
LS89_for	ACAACAACCCGACGG	
LS90_for	CAGAAAGTGGTAGAGAGACTGACG	To assess PFR2 edition : 3' PURO integration PCR
LS101_rev	CTTCTTTTTGCTATTCTGCTAC	
LS110_for	GCGATCTTCACTCTTCTCTG	To assess PFR2 edition : FISH probe and wild-type locus PCR
LS111_rev	CACTGGATATGGTCCCTGT	
LS113_rev	GGGTACGTCAGGAGCAGG	To assess PFR2 edition : FISH probe (LmjF16.1030)
LS114_for	TCTACAGCTTCGTGAACGGC	

15 nt adaptators for INfusion® are in bold and restriction enzyme sites underlined

4.2.1.3 Discussion

Ce travail nous a permis de faire la preuve du concept du CRISPR/cas9 chez *Leishmania*, et diffuser nos plasmides à une trentaine de laboratoire à travers le monde. Nous sommes cependant conscients des limites du système mis en place. Jusqu'à présent nous n'avons pas obtenu une population de cellules 100% modifiée sans passer par un clonage. Ce défaut d'efficacité a été attribué au fait qu'en IFA²⁹ Cas9 était visible dans le cytoplasme de la cellule et non dans le noyau. Une des solutions pour améliorer le système était de remplacer le NLS de Cas9 par une séquence NLS spécifique de *Leishmania* décrite quelques années plutôt (Dubessay P 2006). Ce nouveau NLS devait permettre un adressage de Cas9 plus efficace au noyau de la cellule. Malheureusement, aucune amélioration de la localisation de Cas9 n'a été observée, l'endonucléase était toujours présente majoritairement dans le cytoplasme. Une faible efficacité de transcription du sgRNA par le promoteur U6 a aussi été évoquée. En effet nous avons utilisé un promoteur auquel il manquait un élément (celui présent à l'intérieur du gène U6) sur les trois nécessaires au bon fonctionnement de ce promoteur. Pour vérifier cette hypothèse, une qRT-PCR³⁰ a été réalisée, afin de comparer l'expression du messenger U6 contrôlé par son promoteur natif, au sgRNA sous le contrôle du même promoteur mais tronqué (Figure 4-2). La quantité relative de messenger obtenu avec le promoteur tronqué est supérieure de 1.5 fois au messenger U6. L'hypothèse liée au promoteur est donc caduque.

²⁹ De l'anglais immunofluorescence assay

³⁰ De l'anglais quantitative Real Time - Polymerase Chain Replication

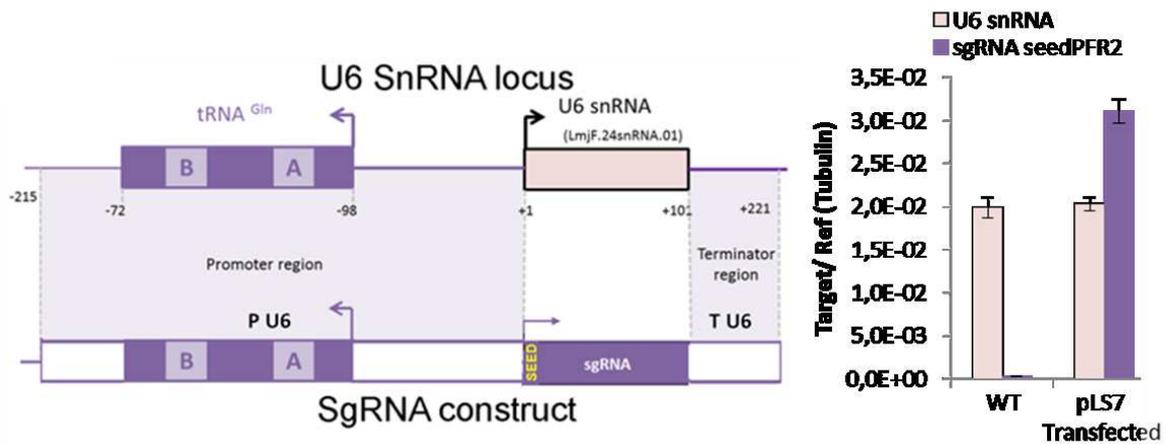


Figure 4-2 Measure of relative expression of the sgRNA and U6 promoter by qRT-PCR.

Aujourd'hui plusieurs équipes utilisent un système de transcription *in vitro* pour produire le sgRNA, et transfectent directement l'ARN guide dans les cellules (communication orale & (Peng *et al.*, 2015)). Ainsi, il est possible que la transfection d'un ARN soit plus efficace que celle d'un plasmide, ce qui augmenterait le taux d'édition. L'utilisation de la traduction *in vitro* semble aussi un atout dans le cas où nous voudrions réaliser plusieurs modifications à différents endroits dans le génome d'une même souche. A ce jour, seule une seconde étude utilisant le CRISPR/Cas9 chez *Leishmania* a été publiée (Zhang & Matlashewski, 2015). Les travaux réalisés chez *L. donovani* ont permis de confirmer la faisabilité du CRISPR chez *Leishmania*, mais ils ont surtout apporté un éclairage intéressant sur les mécanismes utilisés par la cellule pour réparer la DSB induite par l'endonucléase Cas9. En effet les principales protéines impliquées dans le mécanisme de NHEJ³¹ sont absentes du génome de *Leishmania*, ce qui laissait supposer que ce parasite utilisait principalement la recombinaison homologue. Ceci a été montré par Zhang *et al.* (Zhang & Matlashewski, 2015), qui ont également mis en évidence que *Leishmania* était capable de réparer une cassure double brin par Microhomology-Mediated End Joining (MMEJ) avec une faible efficacité. Ils ont aussi démontré que des régions d'homologie de seulement 25 pb et un oligonucléotide simple-brin comme ADN donneur sont suffisants pour initier la réparation d'une DBS, l'intégration précise d'un marqueur de résistance, d'un « tag » GFP ou d'un codon stop. En revanche leur système pour exciser un gène est basé sur la production de deux DSB

³¹ De l'anglais non homologous end joining

de part et d'autre du gène d'intérêt grâce à deux sgRNA produits en tandem à partir d'un même promoteur. Ce moyen d'expression est astucieux, mais nous avons montré que la présence d'un seul sgRNA était suffisante pour l'excision d'un locus. Cependant dans leur cas ils n'apportaient pas d'ADN donneur pour insérer un marqueur de résistance et observaient une faible efficacité de leur système. La sélection des parasites édités se faisait grâce à la résistance à la miltéfosine conférée par la perte du gène cible. Ce système n'est pas transposable pour étudier des gènes non impliqués dans la résistance aux drogues. Le fait également de ne pas fournir l'ADN donneur a montré que la DSB peut être réparée de différentes façons, et que l'on ne peut pas la contrôler. Cependant la comparaison de leur approche avec la nôtre nous a permis de faire évoluer notre stratégie dans le sens où l'utilisation de régions d'homologies inférieures à 1kb est possible. Cependant nous avons déjà commencé de réduire la taille des HRs³² de moitié (500 pb) et montré que l'édition des génomes était toujours possible (non publié). Mais surtout, avec des HRs si petites (25pb), une approche par PCR devient possible, ce qui implique de ne plus les cloner dans un vecteur, et génère un gain de temps considérable. Néanmoins, le développement de cet outil a été un avantage considérable pour le laboratoire où il est aujourd'hui utilisé en routine et nous a permis d'initier des collaborations en interne et à l'étranger.

4.2.2 « Tagging » au locus endogène par CRISPR/Cas9

4.2.2.1 Introduction

Le principal outil pour étudier la localisation d'une protéine chez *Leishmania* est l'expression épisomale d'une version marquée de la protéine. Cependant l'absence de région promotrice bien définie chez *Leishmania* ne nous permet pas de reproduire les conditions d'expression *in vivo*. De plus le « tagging » (étiquetage) au locus n'est pas quelque chose de répandu dans la communauté, il est difficile à réaliser car nos observations montrent que l'intégration d'une étiquette peut se faire aléatoirement et non à la position ciblée.

³² De l'anglais "Homology Region" pour Région d'homologie

Aujourd'hui le parti-pris est de se placer dans des conditions d'expression au plus proche de celles du gène endogène. Le système CRISPR nous donne un avantage considérable, car il permet de générer un DSB à un endroit précis et de favoriser l'intégration d'une étiquette dans le génome.

4.2.2.2 Stratégie

Au moment de choisir la stratégie à adopter pour le « tagging », deux choix étaient possibles pour l'intégration du « tag » : (i) l'intégration d'un « tag » associé à un marqueur de sélection (Figure 4-3A), ou (ii) l'intégration d'un « tag » seul au locus (marker free) (Figure 4-3B). Il est important de rappeler que chez *Leishmania* comme chez *T. brucei* ce sont les régions 5' et 3' UTR qui sont impliquées dans la stabilité des ARNm, cependant ces régions sont encore mal connues et peuvent avoir une taille variable en fonction de l'épissage.

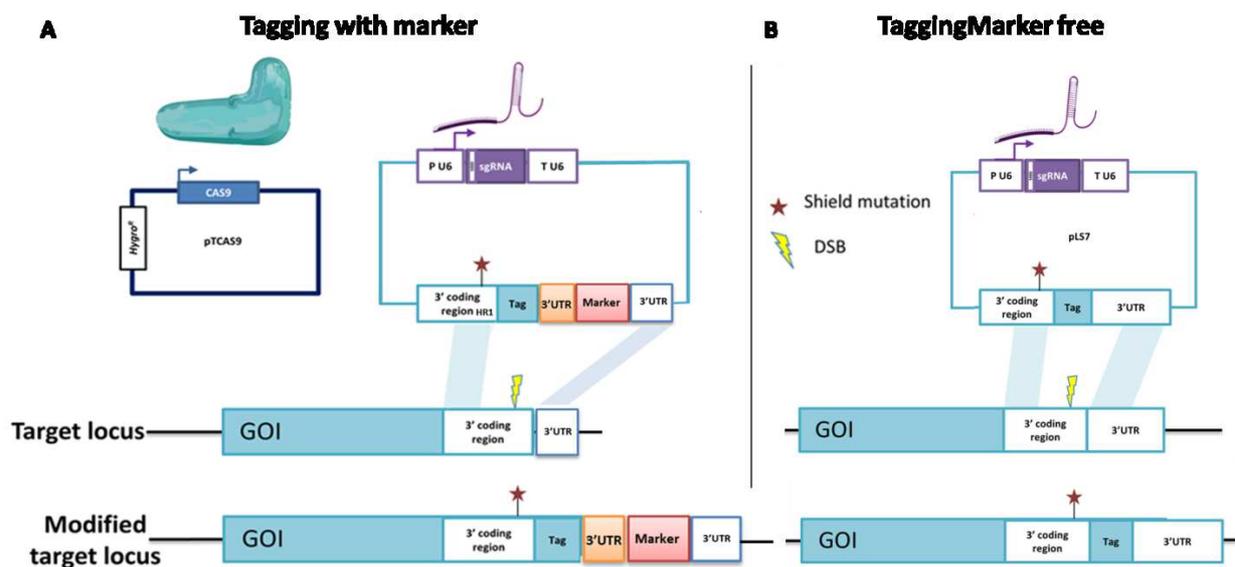


Figure 4-3. Strategy for tagging genes at the locus in *Leishmania* with CRISPR/Cas9 editing tool.

A-with a selective marker, B- marker free.

Lorsque l'intégration du « tag » est associée à un marqueur de sélection, il est sous-entendu qu'il faille modifier le 3'UTR (« Tagging » N terminal) ou le 5'UTR (« Tagging » C terminal). En effet, l'intégration d'un marqueur de sélection qui contient ses propres UTRs implique également de remplacer la région UTR endogène par une région UTR exogène (on utilise généralement celles de la DHFR-TS). L'avantage de cette

construction était de pouvoir sélectionner les parasites « édités » par un antibiotique. Ce système constitue une avancée par rapport à l'expression épisomale où aucun élément régulateur endogène n'est conservé.

La deuxième stratégie consistait en une simple intégration du « tag » en aval de l'ATG ou en amont du codon STOP du gène. Ceci permettait de conserver l'intégrité des régions UTR importantes pour la stabilité et la traduction du messager. Pour ces deux stratégies deux nouveaux vecteurs ont été construits sur la base de pLS5 décrit dans la Figure 4-3 en vue d'introduire un « tag » GFP en Cter.

Dans le cas d'un « tag » associé au marqueur de sélection la construction est basée sur le principe du polycistron : (i) le « tag » est suivi de (ii) la région 3' UTR de la DHFR-TS, puis de (iii) la cassette d'expression du marqueur de sélection qui comprend le gène de sélection encadré par les régions 5' et 3' UTR de la DHFR-TS, et pour finir (iv) des régions d'homologies permettant l'intégration.

La construction pour l'intégration du « tag » seul est beaucoup plus simple : elle comprend la séquence du « tag » encadré par les HRs. Cependant un marqueur de sélection sur le plasmide a été conservé afin d'améliorer la sélection des parasites après transfection, mais ce marqueur ne sera pas intégré. Dans le cas d'un « Tagging » en C-terminal « *marker free* », la HR1 correspond à la région 3' du GOI³³ et la HR2 à la région 3'UTR du GOI. Ces deux régions sont amplifiées par PCR à partir de l'ADN génomique et clonées dans le vecteur pLS5 pour obtenir pLS6. Entre ces deux HRs se trouvent le « tag » que l'on souhaite intégrer.

Contrairement au KO, lorsque l'on choisit une stratégie de « Tagging » (choix du *seed* et des HRs) il est important d'obtenir une DSB proche de la position où l'on envisage l'insertion du « tag ». En effet, lorsque l'ADN donneur est apporté à la cellule pour réparer la DSB, il est nécessaire d'avoir un système pour détruire la séquence reconnue par le système CRISPR/Cas9 (*Seed/PAM*). Ceci afin de protéger le locus d'une autre coupure après l'édition. Pour ce faire, la situation idéale est de pouvoir intégrer une mutation silencieuse pour faire disparaître le NGG, ou de détruire le *seed* en intégrant le « tag » au milieu de la séquence, ou encore de séparer physiquement le *seed* du PAM. Malheureusement ces situations ne sont pas toujours possibles. Dans ce cas il est

³³ De l'anglais Gene Of Interest

intéressant de pouvoir placer le *seed* au niveau d'une amorce servant à amplifier une des régions d'homologie. Cette amorce pourra porter une mutation protectrice ou « *shield* » qui sera intégrée au génome lors de la réparation par la cellule du DSB avec l'ADN donneur.

En première intention, la construction de vecteurs avec marqueur de sélection pour le « Tagging » au locus des protéines LmOrc1b et LmOrc1/Cdc6 de *L. major* a été réalisée, ces deux protéines faisant partie de mon projet de thèse initial (voir section 5.2- *Identification et caractérisation des protéines Orc1/Cdc6 et Orc1b impliquées dans la réplication de l'ADN*). Malheureusement, pour ces deux protéines nous n'avons jamais pu obtenir de cellules marquées à la GFP après transfection des parasites. Cependant, après transfection nous avons observés que les leishmanies avaient un temps de pousse très ralenti (20j au lieu d'une dizaine habituellement). L'hypothèse évoquée était qu'il s'agissait d'un KO fonctionnel : les cellules éditées produisaient une protéine non fonctionnelle qui conduisait à des cellules non viables.

Au vu de ces résultats décevants, une deuxième stratégie de « Tagging » avec la GFP en C terminal et sans marqueur de sélection a été tentée pour LmOrc1b et LmOrc1/Cdc6 et une autre protéine, la flabarine. La flabarine est une protéine du flagelle étudiée au laboratoire par Gilles Merlin (Tetaud *et al.*, 2016). Une souche exprimant la flabarine-GFP a été produite. L'intégration du « tag » au locus a été démontrée par PCR, et la présence de la protéine au niveau du flagelle par IFA avec les stratégies *marker free* et non *marker free* (Figure 4-4).

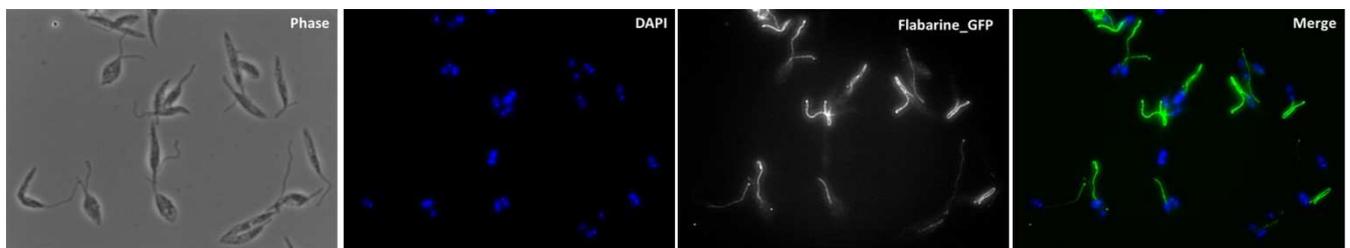


Figure 4-4 Localization of Flabarin.

IFA showing the localization of LmFlabarin-GFP (Green) at the flagellum by insertion of a GFP tag in the 3' of the flabarine gene by CRISPR/Cas9 with the marker free tagging strategy. DAPI (blue) stains the nucleus and the kinetoplast.

Après transfection nous avons obtenu une population mixte avec des cellules possédant un marquage au niveau du flagelle et d'autres cellules sans marquage. Un clonage de

cette souche a permis obtenir une population homogène avec 100% des cellules ayant un marquage au niveau du flagelle. Cependant nous n'obtenions toujours pas de marquage pour les ORC en *marker free*. Convaincu que le système fonctionnait au vu des résultats avec la Flabarine, nous avons décidé d'utiliser un « tag » plus petit (3xHA) et d'utiliser le système CRISPR/Cas9 inductible que nous développons en parallèle (*voir section 5.2.2 Résultats*). La technique du tag 3xHA avait fonctionné pour des nucléoporines (Mlp1 et Mlp2) étudiées au laboratoire (Crobu et al. Données personnelles).

4.2.3 Systèmes inductibles pour CRISPR/Cas9

Le besoin d'un système inductible est devenu une nécessité comme outil de génétique réverse. En effet, pour étudier des gènes essentiels on ne peut pas utiliser la stratégie du KO classique, puisque les cellules ne peuvent pas se développer après la perte de la fonction de certains gènes. Jusqu'alors chez *Leishmania* un seul outil de ce type est disponible mais pas utilisé par la communauté, laissant penser qu'il n'est pas assez robuste (Yan *et al.*, 2001, Yan *et al.*, 2002, Yao *et al.*, 2007). Cependant le système mis au point plus récemment chez *L. mexicana* montre un meilleur contrôle de la répression du promoteur inductible (Kraeva *et al.*, 2014). Il s'agit du système T7TR où le gène dont on souhaite contrôler l'expression est placé sous le contrôle d'un promoteur T7 polymérase inductible à la tétracycline. Plus récemment, le travail de Duncan *et al.* a utilisé le système DiCre inductible avec la rapamycine chez *Leishmania* (Duncan *et al.*, 2016). Nous nous sommes inspiré de ces deux techniques pour mettre au point un CRISPR/Cas9 inductible chez *Leishmania*.

4.2.3.1 CRISPR/Cas9 inductible grâce au système T7TR

✓ **Méthodologie**

Le système T7TR inductible à la tétracycline est constitué de quatre éléments : (i) la polymérase T7, (ii) le promoteur T7, (iii) un opérateur dépendant de la tétracycline et un (iv) répresseur (Figure 4-5).

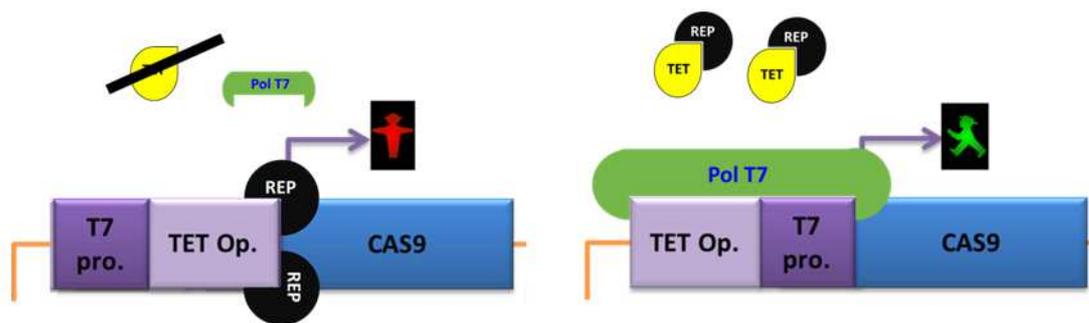


Figure 4-5. Induction of Cas9 expression by a T7TR system in *Leishmania infantum*.

In absence of tetracycline (tet) the repressor (REP) bind the tet operator (tet Op.) resulting in a null expression of Cas9. In presence of tetracycline, the repressor is trapped by the inducer (tet), Tet op. is free to welcome the polymerase T7 (Pol T7) resulting in Cas9 transcription.

Pour que le système fonctionne chez *Leishmania* il faut intégrer chacun de ces éléments dans le génome du parasite **Error! Reference source not found.** Deux souches de *Leishmania*, *L. infantum* (*L. chagasi* (MHOM/BR/00/1669)) (Yao *et al.*, 2007) et *L. mexicana* (MNYC/BZ/62/M379) (Kraeva *et al.*, 2014) avaient déjà été modifiées pour exprimer la polymérase T7 et le répresseur. Ces deux souches nous ont été données, ainsi que le plasmide permettant de placer notre gène d'intérêt sous le contrôle du promoteur inductible.

Le plasmide pTubind (Figure 4-6) se compose d'une séquence opératrice qui contrôle l'activité promotrice dépendant de la T7 pol. Nous avons choisi de placer l'endonucléase Cas9 sous le contrôle de ce promoteur inductible **Error! Reference source not found.** Ce plasmide comporte des séquences homologues au gène de la tubuline pour permettre une intégration du plasmide par recombinaison homologue. Lorsque les souches sont cultivées en absence de tétracycline (Tet) le répresseur se fixe sur l'opérateur, et la polymérase T7 ne peut pas se positionner sur son promoteur pour initier la transcription du gène. En revanche en présence de l'inducteur Tet, le répresseur sera séquestré et la T7pol pourra transcrire le gène d'intérêt (Figure 4-5).

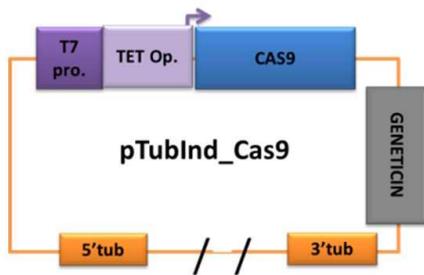


Figure 4-6 Plasmide pTubInd_Cas9.

✓ **Expression inductible de Cas9**

La séquence codante de Cas9 a été clonée dans pTubInd et le vecteur transfecté dans les souches de *L. infantum* T7TR et *L. mexicana* T7TR. Pour vérifier la robustesse du système nous avons dans un premier temps placé la GFP sous le contrôle de ce promoteur inductible et observé, sous un microscope à fluorescence, le pourcentage de cellules exprimant la GFP en présence ou absence de tétracycline. En présence de l'inducteur, à j+2 post induction toutes les cellules exprimaient la GFP. Dans la culture contrôle sans tétracycline aucune cellule n'exprimait la GFP.

Nous avons procédé à la même expérience avec Cas9 à la place de la GFP. Après induction, 90% des cellules exprimaient Cas9 pour *L. infantum* T7TR et, seulement un très faible pourcentage pour *L. mexicana* T7TR. Finalement, seule *Lin*fT7TR_Cas9 a été clonée afin d'obtenir (i) une expression plus homogène de Cas9 et, (ii) de travailler sur une souche endémique retrouvée dans le Bassin Méditerranéen. De la même façon que la souche parentale, les clones avaient une expression variable de Cas9 avec 80 à 91% de cellules marquées. Le clone ayant le taux le plus élevé de cellules marquées a été sélectionné. Cependant après quelques temps en culture, ce clone avait un comportement inhabituel : les cellules adhéraient aux parois de la flasque de culture sans jamais atteindre une concentration cellulaire élevée. Nous avons donc décidé de travailler avec la souche mère avant clonage. La quantification des cellules exprimant Cas9 en présence et absence d'inducteur nous a permis de mettre en évidence une fuite du système inductible (Figure 4-7).

CAS9 induction by 0.5 µg/mL of Tetracycline

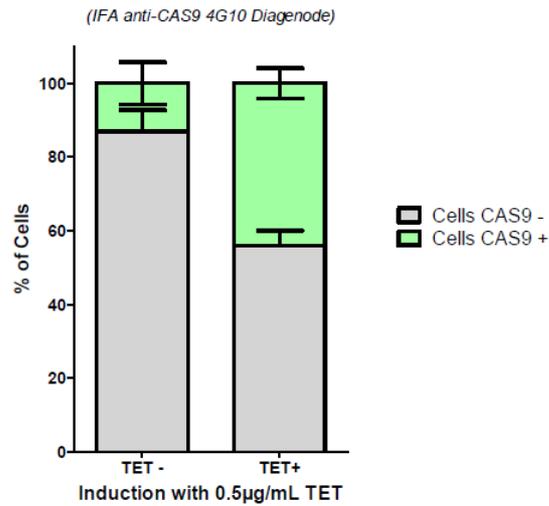


Figure 4-7 Expression of Cas9 under T7 promotor.

IFA anti-Cas9 on *L. infantum* strain which grows in presence or absence of tetracycline. Gray bars correspond to the cells which do not express Cas9 and green bars correspond to cells which expressing Cas9.

Dans un premier temps nous avons pensé que la fuite était due à la présence de traces de tétracycline présentes dans le sérum utilisé dans le milieu de culture. En effet, le sérum que nous utilisions en routine au laboratoire pouvait contenir des traces de tétracycline. Afin de réduire cette fuite nous avons cultivé les cellules avec un sérum de veau fœtal certifié sans tétracycline. La fuite a été quantifiée par dosage de Cas9 ARNm par qRT-PCR (Figure 4-8).

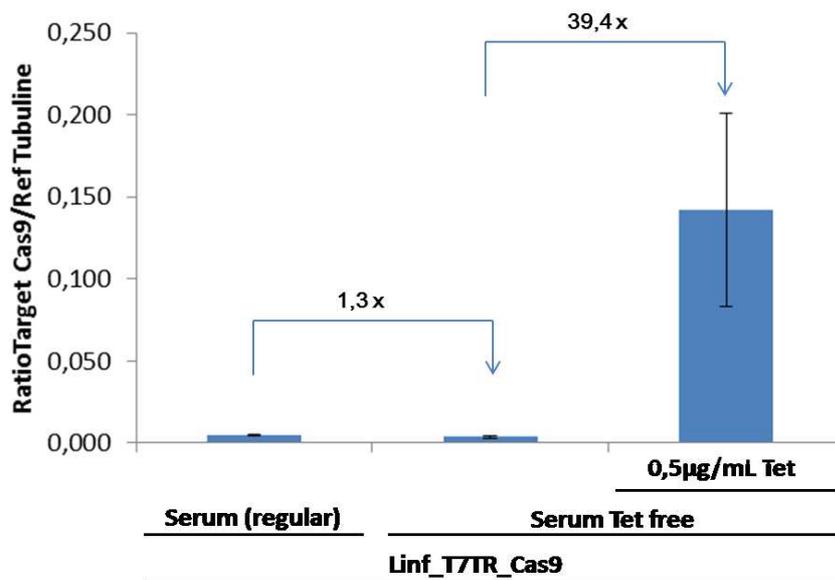


Figure 4-8 Measure of Cas9 mARN by qRT-PCR in Linf_T7TR_Cas9.

In presence of regular serum (containing traces of tetracycline) we observed 1.3 times more Cas9 mRNA than in tetracycline-free serum conditions. When we added 0.5 µg/mL of tetracycline the expression increased 39.4 times.

Finalement, la qRT-PCR nous a permis de montrer que la « fuite » du système d'expression de CAS9 n'était pas significativement différente entre les deux conditions.

Pour obtenir la souche LinfT7TR_Cas9, une transfection du plasmide pTubind_Cas9 linéarisé a été réalisée afin de l'intégrer dans le locus de la tubuline. Nous n'avons pas pu montrer l'intégration par PCR, et n'avons pas encore réalisé le PFGE³⁴ afin de vérifier cette intégration. Cependant, il est possible que le plasmide ait pu se re-circulariser pour former un épisome exprimant Cas9 ; auquel cas le nombre d'épisomes présents dans la cellule n'est pas contrôlable. On peut donc penser que la quantité de répresseur produite par un gène intégré dans le génome de *Leishmania* n'est pas suffisante pour garantir une répression de l'expression de tous les épisomes.

Afin d'améliorer ce système il convient de vérifier l'intégration de Cas9 dans le génome en une copie unique, et de s'assurer qu'il n'y ait plus d'épisome. Malgré cette fuite nous avons utilisé le système pour réaliser différents KO et « tagging ».

4.2.3.2 CRISPR/Cas9 inductible grâce au système DiCre

✓ **Introduction**

L'utilisation du système T7TR a permis de mettre en évidence que le contrôle de l'expression de gènes à partir d'un promoteur n'est pas aisé chez *Leishmania*. De plus, *Leishmania* est capable de transcrire des gènes en absence de promoteur. Ainsi le système T7TR ne semble pas la meilleure solution pour le contrôle transcriptionnel de Cas9.

Nous avons imaginé un nouveau système inductible impliquant la technologie du DiCre, système dérivé du bactériophage P1. La protéine Cre est un enzyme qui catalyse la recombinaison entre deux sites spécifiques (loxP) sur un brin d'ADN (Sternberg & Hamilton, 1981). Ces sites loxP, sont composés d'un espaceur de 8 pb encadré de deux séquences palindromique de 13 pb. La recombinaison a lieu entre deux sites loxP

³⁴ De l'anglais Pulsed Field gel electrophoresis

distincts dans le génome. Selon l'orientation de ces sites loxP dans le génome la recombinaison pourra provoquer une inversion, délétion ou translocation (Figure 4 9).

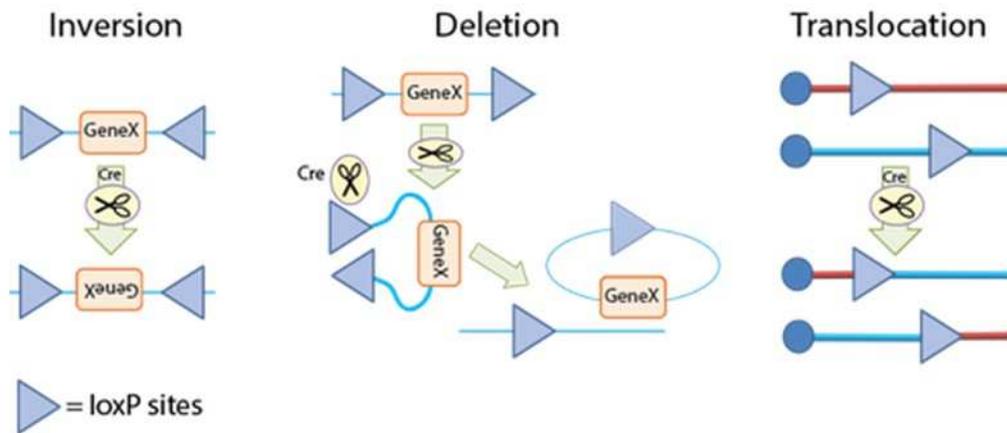


Figure 4-9. Possible Cre recombination events between two loxP sites.

Inversion: If the loxP sites are on the same DNA strand and are in opposite orientations, recombination results in an inversion and the region of DNA between the loxP sites is reversed. Deletion: If the sites are in the same orientation, the sequence between the loxP sites is excised as a circular piece of DNA (and is not maintained). Translocation: If the sites are on separate DNA molecules, a translocation event is generated at the loxP sites.

(Source <https://www.addgene.org/cre-lox/>)

Il a été possible de rendre cette activité enzymatique inductible en séparant la protéine en deux fragments sans activité enzymatique. Ces deux fragments, Nter et Cter, sont fusionnés aux protéines FKBP et FRP respectivement. En présence d'une petite molécule, la rapamycine, FKBP et FRP vont former des dimères hétérologues. Cette dimérisation conduit à l'association des deux domaines complémentaires Cre et reconstitue l'activité enzymatique Cre (Jullien *et al.*, 2003).

✓ Méthodologie

Le système DiCre a récemment été utilisé chez *T. cruzi* et *L. mexicana* pour la délétion de gènes (Kangussu-Marcolino *et al.*, 2014, Duncan *et al.*, 2016). Chez *Leishmania*, organisme aneuploïde mosaïque, le système DiCre présente un inconvénient majeur car il ne peut pas se suffire à lui-même pour générer des KO. Il faut prévoir d'abord une stratégie de KO classique pour cibler le premier allèle puis, introduire les sites loxP au

niveau du deuxième allèle à éditer. Cette stratégie peut être fastidieuse si l'on travaille sur un gène essentiel, sans oublier que le chromosome peut être majoritairement trisomique et/ou que sa « somie » peut fluctuer au fur et à mesure des cycles d'édition du génome.

Une autre possibilité, était de combiner le CRISPR/Cas9 avec le DiCre pour réguler la fonction de Cas9. L'idée est de produire une Cas9 non fonctionnelle qui pourra être « réparée » avec l'intervention de la DiCre.

Pour rendre Cas9 non fonctionnelle, nous avons introduit un insert (quatre « tag » C-myc suivie de trois codons stop) dans la séquence codante de Cas9 en créant un décalage du cadre de lecture. Cette cassette est introduite en position 1151 de la Cas9 dans un domaine non structuré de la protéine (Zetsche et al., 2015). De part et d'autre de cette séquence nous avons placé des sites loxP pour obtenir la cassette LoXP (Figure 4 10). Cette cassette devra être excisée par la Cre recombinase pour rendre Cas9 de nouveau active.

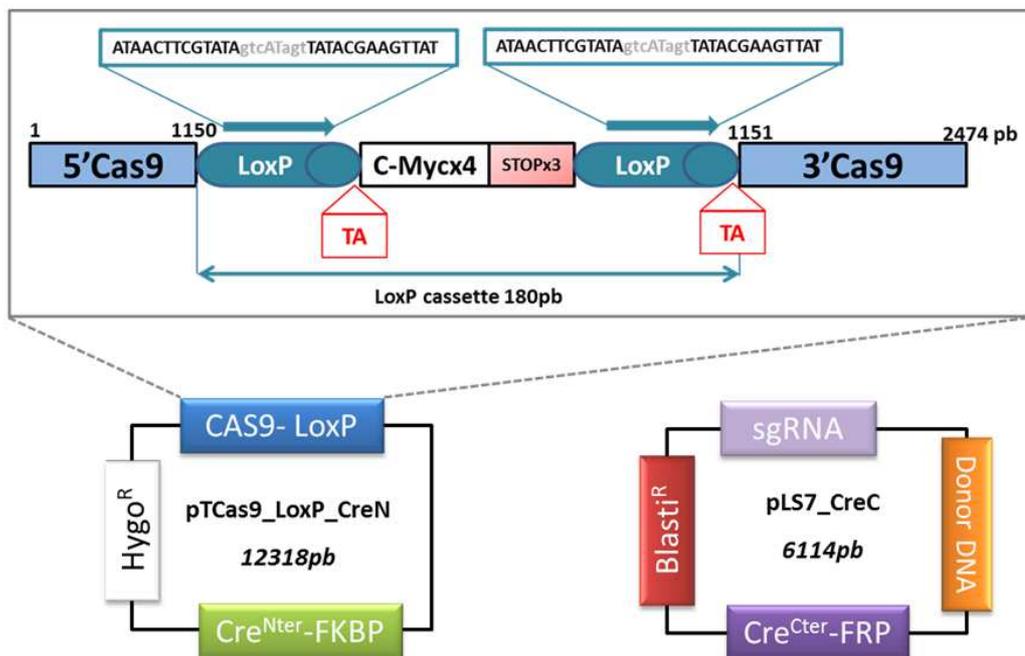


Figure 4-10 Cas9_LoxP cassette for a CRISPR/Cas9 inducible by DiCre.

A LoXP-cassette (180 pb) is inserted at position 1151 in the Cas9 open reading frame. The cassette is composed by a loxP site at each extremity, In between four c-Myc tags followed by three stop codons in frame of the Cas9 5' sequence. Two extra bases were added to restore the Cas9 open reading frame after DiCre edition.

Dans un deuxième temps il fallait exprimer les deux sous-unités Cre Nter-FKBP et Cre Cter-FRP. Elles ont été introduites dans pTcas9-LoxP et pLS7 respectivement. Ces deux plasmides (pTcas9_LoxP_CreN et pLS7_CreC) peuvent être transfectés dans n'importe quelle espèce de *Leishmania*.

Après transfection de ces deux plasmides, et en absence de rapamycine, les parasites produiront une Cas9 tronquée et les deux sous-unités Cre. Lorsque nous voudrons induire d'activité de Cas9, il suffira de supplémenter le milieu de culture avec de la rapamycine qui induira la reconstitution de l'activité de DiCre (Figure 4-11). Deux enzymes reconnaîtront indépendamment les deux sites loxP et élimineront la cassette LoxP par recombinaison homologue. Cette délétion permettra d'obtenir un gène codant une Cas9 fonctionnelle, qui en association avec le sgRNA pourra éditer le génome.

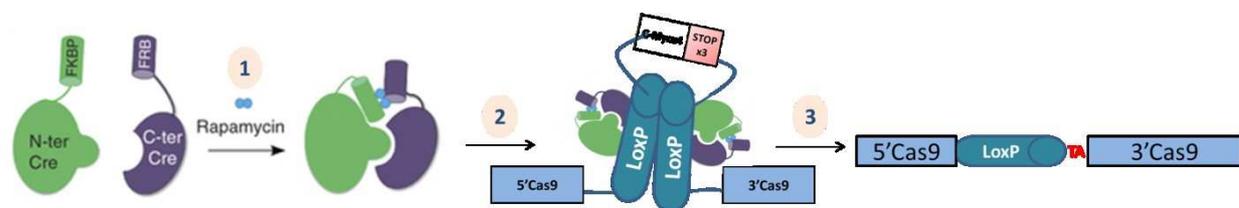


Figure 4-11 Reconstitution of a functional Cas9 by an inducible DiCre.

Step 1: Rapamycin induces the dimerization of the two Cre fragments through FKBP and FRP domains. Step2: active DiCre enzymes recognizes *loxP* sites. Step 3: recombination between two *loxP* sites leads to the deletion of the LoxP-cassette, restoring Cas9 open reading frame.

Cette nouvelle stratégie est actuellement en cours de développement au laboratoire et nous devons vérifier plusieurs choses pour valider notre système : (i) Vérifier que Cas9_LoxP n'est pas fonctionnelle en absence de Cre et que les sous-unités Cre sont bien exprimées chez *Leishmania*. (ii) Vérifier que l'on peut induire la dimérisation des deux fragments FKBP et FRP pour restaurer l'activité enzymatique et rétablir le cadre de lecture de Cas9 en présence de rapamycine au niveau ADN. (iii) S'assurer que Cas9 reconstituée est bien capable d'éditer le génome en présence d'un sgRNA. (iv) Enfin estimer si le système (DiCre+CRISPR/Cas9) ne fuit pas en absence de rapamycine.

4.3 Conclusion

Finally, nous avons démontré que l'édition du génome de *Leishmania* avec la technologie du CRISPR/ Cas9 est désormais possible, et qu'il était possible de réaliser des remplacements alléliques et des « Tagging » au locus dans différentes espèces de *Leishmania*. Le développement de cet outil a été un avantage considérable pour le laboratoire où il est aujourd'hui utilisé en routine et nous a permis d'initier des collaborations en interne et à l'étranger. La mise au point d'un système CRISPR/Cas9 inductible est plus complexe. Nous avons exploré deux pistes : le système inductible T7TR, et la technologie du DiCre. La première a montré quelques faiblesses, cependant nous avons désormais identifié des points à améliorer. Néanmoins, la deuxième stratégie qui allie le CRISPR et le DiCre semble plus prometteuse, d'autant plus qu'une stratégie similaire a fait ces preuves chez *Cryptosporidium parvum* (communication orale Boris Striepen, 2nd ParaFrap Conference). Nous espérons donc prochainement montrer la faisabilité du CRISPR/Cas9 inductible chez *Leishmania*. Cet outil nous sera indispensable pour l'étude de protéines essentiels, tels que les protéines ORCs impliquées dans la réplication de l'ADN.

5 Etude de la réplication de l'ADN chez Leishmania.

Dans ce chapitre, les travaux menés lors de ce travail de thèse pour mieux comprendre les mécanismes de réplication de l'ADN sont rapportés. La première partie sera consacrée à la technique du peignage moléculaire qui a permis de déterminer les paramètres dynamiques de la réplication de l'ADN chez trois espèces de Leishmania et T. brucei. Cette technique a été optimisée pour Leishmania au laboratoire par Slavica Stanojic ; ma contribution dans ce travail se résume à une aide technique et à la participation à la relecture du manuscrit.

La seconde partie concernera la caractérisation des protéines impliquées dans la réplication de l'ADN et tout le travail visant à cartographier des origines de réplication chez Leishmania. Ce travail a nécessité beaucoup d'optimisations et j'ai dû faire face à des difficultés, ce qui explique que ce travail soit moins abouti.

En décembre 2014, lors de mon comité de thèse en fin de première année, il m'a vivement été conseillé de reconsidérer le projet de « cartographie des origines de réplication par ChIP_seq » puisqu'un travail similaire venait d'être présenté en congrès. A partir de ce moment-là j'ai consacré beaucoup de temps au projet CRISPR-Cas9 et mis en pause l'optimisation du ChIP_seq. Cependant ce travail préliminaire a été très utile à Maria Rosa Garcia Silva pour son travail sur l'identification des centromères.

Malgré ma déception de devoir arrêter l'identification des origines, j'ai recentré mon projet en m'intéressant à la caractérisation fonctionnelle des protéines Orc1/Cdc6 et Orc1b en utilisant le CRISPR/cas9. Ce travail n'est encore que préliminaire mais j'ai décidé de l'ajouter à mon mémoire afin d'assembler le puzzle et d'identifier les pièces manquantes.

Enfin pour clore ce chapitre j'essayerai de placer et de discuter mon travail dans le contexte actuel des récentes publications sur le sujet : en Octobre 2015 le travail de C. Marques et coll. montrant par la technique du MFA³⁵_seq qu'il n'y aurait qu'une seule origine de réplication par chromosome chez Leishmania (Marques et al., 2015) alors que notre travail en peignage moléculaire a clairement montré qu'il y en a plusieurs (Stanojic et al., 2016), puis plus récemment le travail de Lombrana et coll. décrivant 5100 origines potentielles (Lombrana et al., 2016).

C'est pourquoi, au vu de ces données contradictoires de la littérature, nous avons décidé de repartir à la course aux origines de réplication avec le ChIP_seq et le séquençage du brin naissant dans la dernière année de ma thèse.

³⁵ Marker Frequency Analysis_sequencing

5.1 Dynamique de réplication chez *Leishmania* (Article 2: Stanojcic et al. Sci Rep 2016)

5.1.1 Introduction

La réplication de l'ADN est un mécanisme commun à tous les organismes. Comme nous avons pu le voir dans l'introduction, la régulation et l'initiation de la réplication sont différentes d'un organisme à l'autre. Chez les eucaryotes supérieurs la réplication est initiée à partir d'un lot d'origines pour aboutir à la complète réplication du génome en un temps donné.

Chez les trypanosomatidés le mécanisme et la dynamique de la réplication sont encore mal connus. Chez *Leishmania*, l'aneuploïdie mosaïque est maintenue grâce à un taux élevé de répartition asymétrique des chromosomes lors de la mitose, entraînant le gain ou la perte de chromosomes entiers. Il a donc été suggéré que ce parasite possédait une régulation non-conventionnelle de la réplication de ces chromosomes (Sterkers *et al.*, 2012).

Afin de pouvoir comprendre les mécanismes qui permettent de contrôler la réplication, il nous a paru judicieux de comparer la dynamique de réplication entre *Leishmania* et un organisme diploïde de la même famille, *T. brucei*. Pour cela, nous avons utilisé la technique du peignage moléculaire pour caractériser les paramètres de progression de la fourche de réplication, ainsi que les distances inter-origines chez trois espèces de *Leishmania*, *T. brucei*. Les cellules de mammifère (MEF) ont été utilisées comme contrôle.

Ce travail est présenté dans la publication ci-dessous (Stanojcic *et al.*, 2016).

5.1.2 Résultats

SCIENTIFIC REPORTS

OPEN

Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes *Leishmania* and *Trypanosoma brucei* versus mammalian cells

Received: 06 October 2015
Accepted: 17 February 2016
Published: 15 March 2016

Slavica Stanojčić¹, Lauriane Sollelis¹, Nada Kuk¹, Lucien Crobu², Yves Balard¹, Etienne Schwob⁴, Patrick Bastien^{1,2,3}, Michel Pagès² & Yvon Sterkers^{1,2,3}

Leishmania and *Trypanosoma* are unicellular parasites that possess markedly original biological features as compared to other eukaryotes. The *Leishmania* genome displays a constitutive 'mosaic aneuploidy', whereas in *Trypanosoma brucei*, the megabase-sized chromosomes are diploid. We accurately analysed DNA replication parameters in three *Leishmania* species and *Trypanosoma brucei* as well as mouse embryonic fibroblasts (MEF). Active replication origins were visualized at the single molecule level using DNA molecular combing. More than one active origin was found on most DNA fibres, showing that the chromosomes are replicated from multiple origins. Inter-origin distances (IODs) were measured and found very large in trypanosomatids: the mean IOD was 160 kb in *T. brucei* and 226 kb in *L. mexicana*. Moreover, the progression of replication forks was faster than in any other eukaryote analyzed so far (mean velocity 1.9 kb/min in *T. brucei* and 2.4–2.6 kb/min in *Leishmania*). The estimated total number of active DNA replication origins in trypanosomatids is ~170. Finally, 14.4% of unidirectional replication forks were observed in *T. brucei*, in contrast to 1.5–1.7% in *Leishmania* and 4% in MEF cells. The biological significance of these original features is discussed.

In all eukaryotes, DNA replication is a tightly regulated process that initiates from multiple replication origins that are distributed along the chromosomes. Accurate regulation of DNA replication ensures faithful transmission of eukaryotic genomes and maintenance of genome ploidy and chromatin organization¹. DNA replication is under the control of genetic and epigenetic mechanisms and is closely linked to the functional organization of the nucleus². In most eukaryotes, origins are not conserved sequences but appear to be defined by DNA structure and chromatin environment². An excess of potential replication origins is made before the S phase of the cell cycle, through the assembly of pre-replication complexes. However, only a subset of these potential origins is activated during S phase to initiate DNA synthesis. Flexibility in the choice of replication origins is a general feature of both yeast and metazoan replication origins, but the choice is surprisingly plastic and opportunistic³.

Conventional eukaryotic models are actually closely related in the evolutionary time scale⁴ and more 'divergent' eukaryotes might well have different DNA replication mechanisms. Trypanosomatids branched very early in eukaryotic history⁴ and they have attracted attention as original models for biological and comparative studies. *Trypanosoma brucei* and *Leishmania* sp. are protozoan parasites responsible for some of the most debilitating diseases in humans and livestock such as leishmaniasis, sleeping sickness and nagana. Despite the strong synteny observed between the genomes of these two trypanosomatids, their organization is very different. The genome of *T. brucei* is composed of 11 megabase-sized chromosomes ranging from 1 to 6 Mbp in length, 3–5 intermediate

¹University of Montpellier, Faculty of Medicine, Laboratory of Parasitology-Myology, Montpellier, F34090, France. ²CNRS 5290 - IRD 224 - University of Montpellier (UMR "MiVEGEC"), Montpellier, F34090, France. ³University Hospital Centre (CHU), Department of Parasitology-Myology, Montpellier, F34090, France. ⁴Institute of Molecular Genetics, CNRS UMR5535 & University of Montpellier, Montpellier, F34293, France. Correspondence and requests for materials should be addressed to M.P. (email: gpp@univ-montpl.fr) or Y.S. (email: yvon.sterkers@univ-montpl.fr)

chromosomes of 200–500 kb and ~100 mini-chromosomes ranging from 30 to 150 kb^{5,6}. In contrast, the genome of *Leishmania* comprises 36 chromosomes for the Old World species, hereafter *L. major* and *L. donovani*, and 34 for the New World species, hereafter *L. mexicana*^{7,8}, ranging from 0.3 to 3 Mb in size. In all trypanosomatids, genes are organized in large unidirectional polycistronic units (so-called “directional gene clusters” or DGCs), made of many functionally non-related genes present on the same DNA strand. Neighbouring transcription units can be either convergent or divergent, and the regions between them are known as strand switch regions (SSRs). Divergent SSRs have been associated with the initiation and convergent ones with the termination of transcription^{9,10}. Both types of SSRs have also been shown to contain origins of replication^{11,12} as well as centromeric regions¹³ in *T. brucei* but not *Leishmania*. Another unique feature is the constitutive ‘mosaic aneuploidy’ demonstrated by Fluorescent *in situ* hybridization (FISH) data in *Leishmania* strains^{14,15}, contrasting with the diploidy of *T. brucei* established for its 11 ‘megabase’ chromosomes¹⁶. The biological significance of mosaic aneuploidy in *Leishmania sp.* remains ill defined, but some insights were obtained by studying the distribution of chromosomes during mitosis. FISH monitoring of the sister chromosomes segregation during mitotic division showed that *Leishmania* display high rates of asymmetric chromosome allotments, and that all dividing nuclei totalized an odd number of chromosome homologues^{14,17}. An even number of chromosome homologues in dividing cells indicates chromosome segregation defects^{18,19}, while an odd number rather suggests an unconventional regulation of DNA replication; this would allow over- and under-replication events leading to either the complete re-duplication or the absence of a given chromosome at the end of the cell cycle. The regulation of DNA replication is thought to be essential for the occurrence and persistence of mosaic aneuploidy in *Leishmania*^{14,17} and might also be responsible for a laboratory-induced, well tolerated aneuploidy in *T. brucei*²⁰. Here, we analysed DNA replication in three *Leishmania* species (promastigote forms of *L. donovani*, *L. major* and *L. mexicana*) and in the related *T. brucei* (procyclic forms) at the single molecule level, using state-of-the-art DNA molecular combing. This technique allows uniform stretching of DNA fibres, hence accurate measurements of inter-origin distances (IODs) as well as of the symmetry and velocity of progressing replication forks²¹. We show here that DNA replication in these parasites exhibit highly original features. The inter-origin distances are very large and DNA replication progresses more rapidly, particularly in *Leishmania*, than in any other eukaryotic cells analysed so far. The estimated number of active origins was similar in all trypanosomatids (150–180). The major difference in DNA replication dynamics between these parasites is the high occurrence of unidirectional replication forks in *T. brucei*. To our knowledge, this is the first study of replication dynamics in *Leishmania*, the first one to use DNA molecular combing in trypanosomatids, and the first one to compare the replication dynamics of so-called ‘divergent’ eukaryotes and higher eukaryotes in the same experiments.

Results

Setting up the conditions for DNA replication analysis in trypanosomatids by DNA molecular combing. Exponential cultures of asynchronously growing cell populations of each trypanosomatid were labelled with two consecutive pulses of IdU and CldU, DNA fibres were then purified and combed onto silanized coverslips (Fig. 1A; see Material and Methods). This procedure generated long, parallel DNA fibres, with a uniform extension of 2 kb/ μm ²². The incorporated modified nucleosides were immunodetected and DNA was also systematically counterstained to ensure that (i) replication signals belonged to the same fibre, (ii) two fibres were not overlapping and (iii) the signals were not discontinuous (see Material and Methods). After incorporation of two modified nucleotides into newly replicated DNA, at least three different signal types can be expected (Fig. 1B). Several concentrations of modified nucleosides, iodo-deoxyuridine (IdU) and chloro-deoxyuridine (CldU), were tested for labelling trypanosomatids, and a minimal concentration of 300 μM was found necessary to obtain satisfactory signals of newly replicated DNA fibres. Since these concentrations were 10 \times higher than those usually used to label cells for DNA combing analysis, we used mouse embryonic fibroblasts (MEF cells), labelled under exactly the same conditions as a control. The analysis of DNA replication in MEF cells by molecular combing has been detailed elsewhere^{23,24}. This allowed us to check that speed and other parameters of DNA replication were not altered by the high concentrations of modified nucleosides.

Inter-origin distances in trypanosomatids. Inter-origin distance (IOD) is defined as the length between two adjacent replication initiation sites and can be determined by measuring centre-to-centre distances between two adjacent progressing forks (Fig. 1B). Representative chromosome fibres from this analysis are shown in Fig. 2A, with the position of the presumed DNA replication origins. The distributions of all IODs measured in MEF cells and four trypanosomatids are presented in Fig. 2B. More than 100 IODs were measured for each of the five cell lines studied (Table 1). The mean IOD was calculated as 136 kb (median 125 kb, see Table 1) in MEF cells, which is in agreement with the previously published data^{23,24}. The mean IOD in *T. brucei* was 160 kb (median 149 kb). The mean IODs in *L. donovani*, *L. major* and *L. mexicana* were 189 kb (median 180 kb), 195 kb (median 193 kb) and 226 kb (median 203 kb), respectively (Table 1). The range of IODs was particularly large in *Leishmania* species where they fluctuated between 30–628 kb, 48–625 kb, 49–628 kb in *L. donovani*, *L. major* and *L. mexicana*, respectively (Fig. 2B, Table 1). In *T. brucei* the range of IOD variations was smaller than in *Leishmania* (37–422 kb) but greater than in MEF cells (27–372 kb). Sample size, fibre length and DNA counterstaining are crucial parameters to gain robust information about replication dynamics; in particular IOD measurements can be biased by DNA fibre length²⁴. We therefore verified the length classes of combed DNA fibres for MEF cells, *T. brucei* and the three *Leishmania* species. The distribution of DNA fibre lengths was very similar in all five samples (Supplemental Fig. 1), showing that IOD comparisons among trypanosomatids and MEF cells is not biased by DNA fibres length. Statistical analysis of these measurements confirmed that IODs are significantly larger in *Leishmania* than in other cell types, but also larger in *T. brucei* than in MEF cells (Fig. 2C).

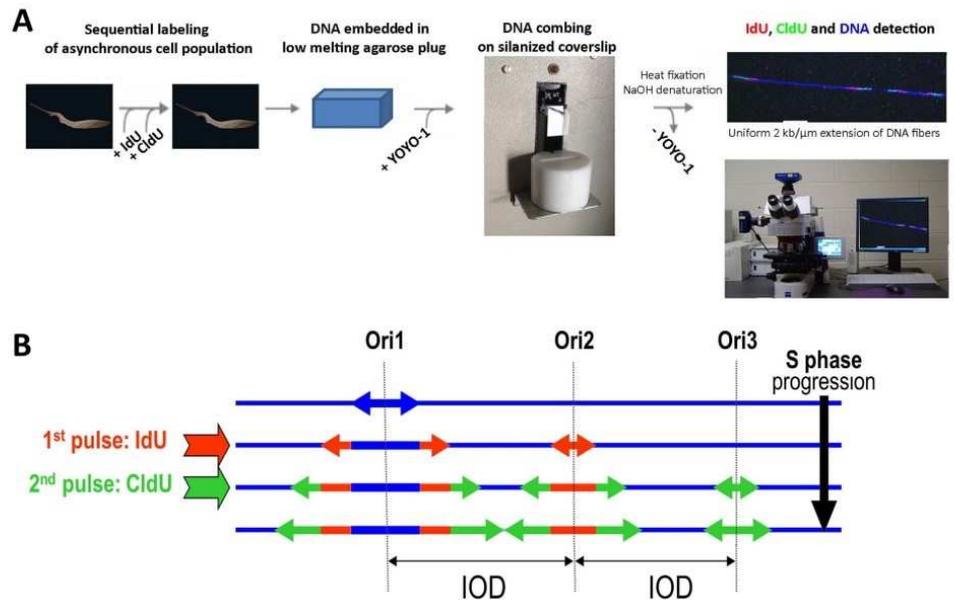


Figure 1. Schematic representation of the experimental approach and different signal patterns observed after two-pulse labelling of replicating DNA. (A) Asynchronous cells were successively labelled with equal pulses of iodo-deoxyuridine (IdU) and chloro-deoxyuridine (CldU) and chromosomal DNA combed on silanized coverslips. (B) Three major types of signal patterns may be observed after molecular combing and immuno-detection of the modified nucleosides, depending whether DNA replication initiates before the first pulse (Ori 1), during the first pulse (Ori 2) or during the second pulse (Ori 3). DNA fibres are counterstained in blue; IdU tracks, corresponding to the incorporation of the modified nucleotide during the first pulse, are revealed in red, whereas CldU tracks corresponding to the second pulse are in green. Arrows represent the direction of the replication forks progression, which is visually detected as a transition from red to green labeling. DNA replication initiation sites (active origins or Ori) stand in the centre of the bidirectional replication fork. The inter-origin distance (IOD) was determined by measuring centre to centre distance between two adjacent progressing bidirectional forks.

Velocity of replication forks in trypanosomatids. To calculate and compare the velocities of replication forks in these five eukaryotic cell types, we divided the length of the second (CldU) labelling track by the duration of the second nucleoside pulse (see Fig. 1B). Representative chromosome fibres used for this analysis are shown in Fig. 3A, where only signals from progressing bidirectional forks are presented (IdU and CldU labeling). The distributions of >100 fork velocity measurements for all five cell types are presented in Fig. 3B and Table 1. In MEF cells, labelled under exactly the same conditions as trypanosomatid cells, the detected mean velocity was 1.36 kb/min (median 1.06 kb/min), which is in agreement with previous reports using lower concentrations of modified nucleosides^{23,24}. In *T. brucei*, the mean replication fork velocity was 1.87 kb/min (median 1.83 kb/min), with extremes ranging from 0.34–6.09 kb/min. In *Leishmania*, these figures were 2.43 (range 0.88–4.65) kb/min, 2.45 (range 0.62–4.97) kb/min and 2.59 (range 1.11–6.03) kb/min in *L. major*, *L. donovani* and *L. mexicana* respectively. Comparative statistical analysis of fork velocities in the four trypanosomatids and MEF cells confirmed that DNA replication progressed at a similar velocity in the three *Leishmania* strains, and significantly faster in *Leishmania* than in *T. brucei* and MEF cells (Fig. 3C).

The interdependence of IODs and fork velocities was also analysed in these five eukaryotic cell types (Fig. 4). The mean fork velocities and IOD values tend to be positively correlated (coefficient of determination $R^2 = 0.903$): the longer IODs observed in *Leishmania* cells result in faster fork progression. Similarly, the IODs were significantly shorter in *T. brucei* than in *Leishmania*, and the progression of replication forks was slower. The IODs were the shortest, and the progression of replication forks the slowest, in MEF cells.

DNA replication dynamics in trypanosomatids. To further describe the replication dynamics, we analysed (i) the temporal order of adjacent origin firing, (ii) the symmetry of bidirectional replication forks' progression and (iii) the presence of unidirectional replication forks. Temporal regulation of origin activation allows adjacent DNA replication origins firing either synchronously, when two or more neighbouring origins fire at the same time, or asynchronously when adjacent origins fire at different time points. One potential role for this regulation is to minimize fluctuations in S phase duration and to avoid persistence of unreplicated DNA in mitosis²⁵. We found both synchronous and asynchronous firing of adjacent DNA replication origins (Fig. 5A). The replication origins can be activated in an asynchronous manner (Fig. 5A, lower panel) when one replication origin is activated in between two previously activated origins. Asynchronous firing of adjacent origin was observed more

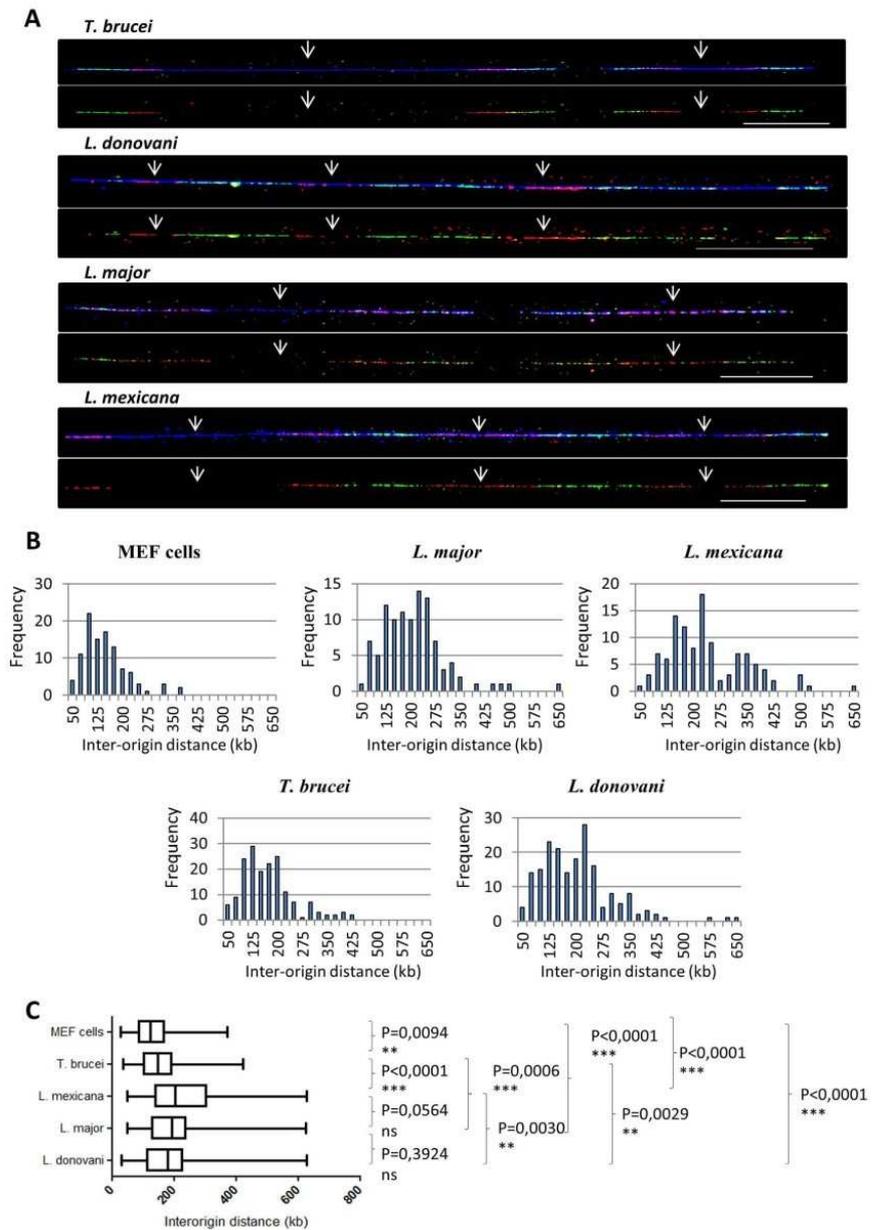


Figure 2. Single-molecule analysis of the inter-origin distances in trypanosomatids using molecular combing. (A) Representative chromosome fibres from trypanosomatids used for the analysis of inter-origin distances (IOD). Positions of the presumed origins are indicated with arrows. Scale bar: 50 kb. For each species, upper panels: counterstaining of DNA fibres is in blue, first pulse nucleosides (IdU) in red and second pulse nucleosides (CldU) in green; lower panels show the tracks from the first and second pulse nucleosides extracted from the upper panel. (B) Statistical distribution (frequency) of IODs in MEF cells and four trypanosomatids. (C) Comparative analysis of IODs in MEF cells and four trypanosomatids obtained after one-way ANOVA analysis of variance. Boxes present 25–75% range. Whiskers present minimum–maximum range. The two-tailed Mann-Whitney test was used to compare the median values of IOD between each two samples and to calculate the corresponding P values ($P < 0.05$ was taken as significant).

frequently in trypanosomatids than in MEF cells (Fig. 5B) and among the trypanosomatids, more frequently in *Leishmania* species than in *T. brucei*. In *Leishmania* species, asynchronous adjacent origin firing was three times more frequent than synchronous adjacent origin firing. The symmetry of progressing replication forks, a

	<i>L. donovani</i>	<i>L. major</i>	<i>L. mexicana</i>	<i>T. brucei</i>	MEF cells
Inter-origin distance (kb)					
Median (kb)	179.7	192.7	203.3	148.8	124.6
Mean (kb)	188.5	194.8	226	159.9	135.8
Minimum (kb)	29.5	48.6	49.3	36.7	27.4
Maximum (kb)	628.1	625.4	628.1	421.6	372.3
25% Percentile (kb)	114.2	128.2	140.7	103.8	86.37
75% Percentile (kb)	228.3	237.5	306.4	192.4	168.1
Number of values measured	189	130	118	172	140
Velocity of replication forks (kb/min)					
Median (kb/min)	2.37	2.45	2.48	1.84	1.16
Mean (kb/min)	2.45	2.44	2.59	1.87	1.36
Minimum (kb/min)	0.62	0.88	1.12	0.34	0.25
Maximum (kb/min)	4.97	4.65	6.03	6.10	5.42
25% Percentile (kb/min)	2.03	2.01	2.18	1.53	0.80
75% Percentile (kb/min)	2.82	2.82	2.94	2.17	1.65
Number of values measured	238	134	273	583	175
Asymmetry of replication forks (ratio long/short forks)					
Median (ratio long/short forks)	1.03	1.03	1.04	1.07	1.10
Mean (ratio long/short forks)	1.07	1.12	1.08	1.19	1.22
Minimum (ratio long/short forks)	1	1	1	1	1
Maximum (ratio long/short forks)	2.03	2.34	1.56	3.77	3.13
25% Percentile (ratio long/short forks)	1.00	1.01	1.01	1.02	1.04
75% Percentile (ratio long/short forks)	1.08	1.10	1.10	1.18	1.24
Number of values measured	104	72	85	112	126
Number of active replication origins/ haploid genome					
Size of the haploid genome sequenced*	32.4Mb	32.8Mb	32Mb	26Mb	2.8 10 ³ Mb
haploid genome/median IOD	180	170	157	175	22.5 10 ³
haploid genome/mean IOD	172	168	142	163	20.6 10 ³

Table 1. DNA replication parameters in three *Leishmania* species, *T. brucei* and MEF cells. *References for genome sequences are Downing *et al.* 2011 for *L. donovani*, Ivens *et al.* 2005 for *L. major*, GeneDB for *L. mexicana*, Berriman *et al.* 2005 for *T. brucei* and NCBI resources for MEF cells. The figures given for measured values were collected from three or more independent biological experiments.

parameter used to estimate the occurrence of fork stalling, was also studied. Here again, MEF cells were used as a control since the analysis of replication fork symmetry in these cells was published previously²⁴. The forks symmetry can be estimated as a ratio of the lengths of the longer over the shorter progressing forks that are emanating from a given origin. Progressing forks are symmetric when this ratio is equal to 1, and asymmetric when it is > 1 (Fig. 6A and Methods). The distribution of these ratios in all cell types studied is shown in Fig. 6B. Asymmetric replication forks were present in MEF cells as reported previously²⁴, with a median ratio of 1.10 (Fig. 6B and Table 1). Fork asymmetry values were uniform in *Leishmania* species, with median values of 1.02, 1.03 and 1.04 for *L. donovani*, *L. major* and *L. mexicana*, respectively. On the other hand, *T. brucei* was significantly more enriched in asymmetric forks (median 1.07) than *Leishmania* (Fig. 6B). Some examples of progressing replication bidirectional forks with high asymmetry, observed in *T. brucei*, are shown in Fig. 6C. The presence of asymmetric or unidirectional replication forks has to be analysed separately because the latter were excluded from the calculation of the progressing forks ratio. A schematic presentation of the analysed unidirectional replication forks is shown in Fig. 6D. The total number of unidirectional replication forks was counted, avoiding the artefacts of broken DNA fibres (Fig. 6D). The percentage of unidirectional forks over the total number of replication forks analysed was calculated in all five cell types (Fig. 6E). This was highest (14.4%) in *T. brucei*, as opposed to MEF cells, where only 4% of unidirectional replication forks were observed, and *Leishmania*, where unidirectional replication forks were rarely seen (1.7%, 1.5%, and 1.5% for *L. donovani*, *L. major* and *L. mexicana*, respectively). Some examples of unidirectional replication forks in *T. brucei* are presented in Fig. 6F.

Discussion

In the present study, we examined the DNA replication dynamics in three *Leishmania* species and *T. brucei* in comparison with mouse embryonic fibroblasts (MEF). For this purpose, we used DNA molecular combing on asynchronously cultivated cells which allowed us to follow the DNA replication dynamics in cells at different time points of the S phase. Molecular combing allows monitoring the replication origins activation and the progression of replication forks, hence studying the real-time dynamics of genome replication, at a resolution of 2–5 kb and with the possibility to study > 1 Mb-long DNA fibres²². It investigates the dynamics of DNA replication at the level of individual molecules, as opposed to high-throughput genomic approaches, which provide compiled replication patterns of the whole cell population.

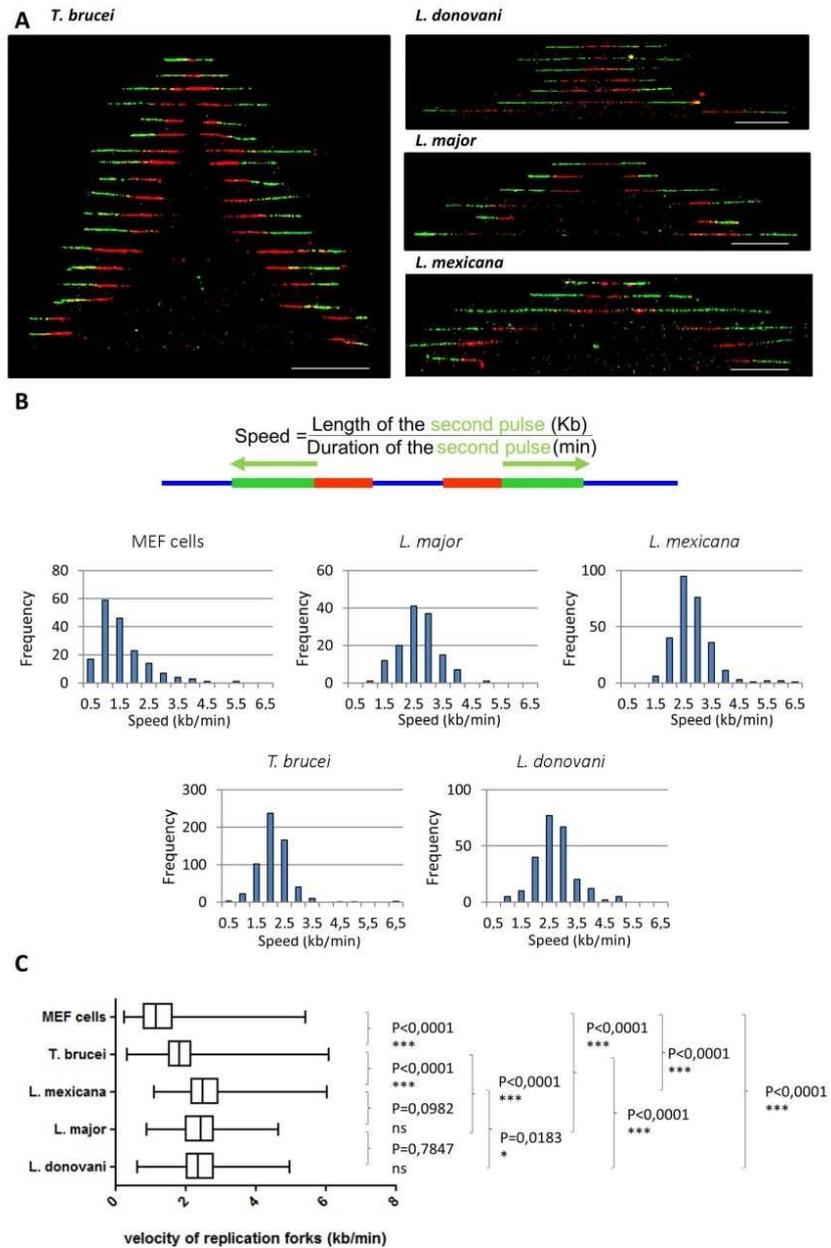


Figure 3. Single-molecule analysis of the velocity of replication forks in MEF cells and four trypanosomatids using DNA molecular combing. (A) Representative bidirectional replication forks taken from different microscopic fields were artificially assembled and centred on the position of the presumed origins. Red tracks: IdU, green tracks: CldU. Counterstaining of DNA was performed but is not shown here for the sake of clarity. Scale bar: 50 kb. (B) Statistical distribution of replication forks velocity measurements in MEF cells and four trypanosomatids. The velocity of the replication fork was calculated by dividing the length of the 'green' tracks with the duration of the second pulse. (C) Comparative analysis of the velocity of replication forks in the five cell lines obtained after one-way ANOVA analysis of variances. Boxes present 25–75% range. Whiskers present minimum–maximum range. The two-tailed Mann-Whitney test was used to compare the median values of the velocity of replication forks between each two samples and to calculate the corresponding P values ($P < 0.05$ was taken as significant).

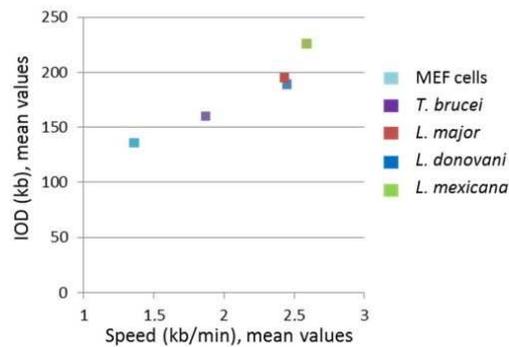


Figure 4. Positive correlation between mean inter-origin distances (IODs) and fork velocities. 2-way representation of mean values of fork velocities and IODs in MEF cells, *T. brucei* and three *Leishmania* species.

This study showed that DNA replication forks are moving fast in trypanosomatids, especially in *Leishmania* species, compared to other eukaryotes (Fig. 3B and Table 1). Indeed, most reported mean velocities of replication forks in eukaryotes vary from 0.8 kb/min in *Drosophila* Kc cells²³ to 1.4–1.5 kb/min in mouse embryonic stem cells, human primary normal keratinocytes, and Chinese hamster embryonic fibroblasts (CHEF); and the highest replication fork speed ever reported was 1.9 kb/min in transformed JEFF lymphoblastoid cells²⁴. We found a DNA replication velocity of 2.4–2.6 kb/min in the three *Leishmania* species and 1.9 kb/min in *T. brucei*. Several factors play important roles in fork progression rates, among which the intrinsic properties of the DNA polymerase²⁶, the chromatin structure and condensation state²⁷, and the concentration of the dNTPs^{28,29}. In trypanosomatids, all canonical histones homologues can be found in the genome^{30–32}, but the patterns and types of histone modifications show substantial differences when compared to higher eukaryotes^{33–36,37}. Chromatin is classically packaged around histones, but the chromosome condensation state is more relaxed in these organisms than in higher eukaryotes. A recent study of replication velocity in *T. brucei* using another technique named ‘single molecule analysis of replicated DNA or SMARD’, yielding spread DNA rather than combed fibres, found even higher replication velocities in procyclic and bloodstream forms (mean 3.7 ± 0.1 and 4.4 ± 0.1 kb/min, respectively)³⁸. We consider our data more accurate, because (i) DNA molecular combing enables the linearization of long DNA fibres and their uniform extension at 2 kb/ μ m; (ii) the DNA fragments obtained using SMARD are much smaller, which did not allow measuring IODs, (iii) no control cells were used in the SMARD study. In addition, the authors did not specify that they put the cells on ice after labelling, and we experimentally determined that, if this is not done, DNA replication continues, yielding artificially high replication fork velocities, e.g. 3.8 kb/min in *L. donovani* (data not shown).

Molecular combing also allows measuring the distances between active origins, or IODs, which have never been reported for divergent eukaryotes. IODs in *T. brucei* (mean 160 kb) and *Leishmania* (mean 226 kb for *L. mexicana*) were significantly larger than those reported in the literature for mammalian cells, e.g. 111 kb, 136 kb, 139 kb and 147 kb, in human primary keratinocytes, MEF cells, mouse embryonic stem cells and JEFF cells, respectively^{23,24,39}. IODs in trypanosomatids were also much larger than those observed in insect cells (*Drosophila* Kc cells)²³, *Arabidopsis thaliana*⁴⁰ and budding yeast^{41,42} where the average IODs are 73 kb, 77 kb and 46 kb, respectively. One explanation for this uncommon feature might be the highly specific gene organization in trypanosomatids as large directional gene clusters. In addition, we observed a positive correlation between fork velocity and IODs in the four trypanosomatids and MEF cells (Fig. 4D): the larger was the mean IOD in one cell type, the faster was the progression of replication forks. A similar correlation was previously described for lymphoblastoid cells and fibroblasts⁴³ and it can be also inferred from previously published data in other cell types^{22–24}. This tendency suggests that fork velocity and IOD are co-regulated by a mechanism that ensures the replication of the entire genome during the cell cycle. One may consider that this mechanism is universal and conserved throughout evolution.

It is possible to infer the number of active origins per haploid genome from our data (Table 1). Dividing the size of the haploid genome of *T. brucei*, i.e. 26 Mb³⁰, by the mean IOD of 160 kb (or the median 149 kb), we inferred the presence of ~170 (163–175) active origins per haploid genome in this organism. The same calculation applied to *L. major* and *L. donovani* (genome size 32.8 Mb and 32.4 Mb, respectively^{44,45}) yields an estimate of 168–180 active origins per haploid genome; and a slightly smaller number of active replication origins was computed for *L. mexicana* (around 150). These figures have to be compared with the 42 early firing DNA replication origins that were recently mapped in the *T. brucei* genome using genome-wide marker frequency analysis coupled with deep sequencing (MFaseq), which compares DNA read depth in replicating cells versus non-replicating cells⁴⁶, and with the ~100 active origins predicted by these authors by extrapolation⁴¹. The discrepancy between these two estimates of active origin numbers still appears paradoxical because genome-wide analysis should overestimate the number of active origins as it compiles all origins from the cell pool (reviewed in⁴⁷). The SMARD analysis³⁸ supported the presence of more, unmapped, replication origins in *T. brucei* but without estimation of a final origin number. Therefore, our results bring a more precise insight on the number of active origins in trypanosomatids.

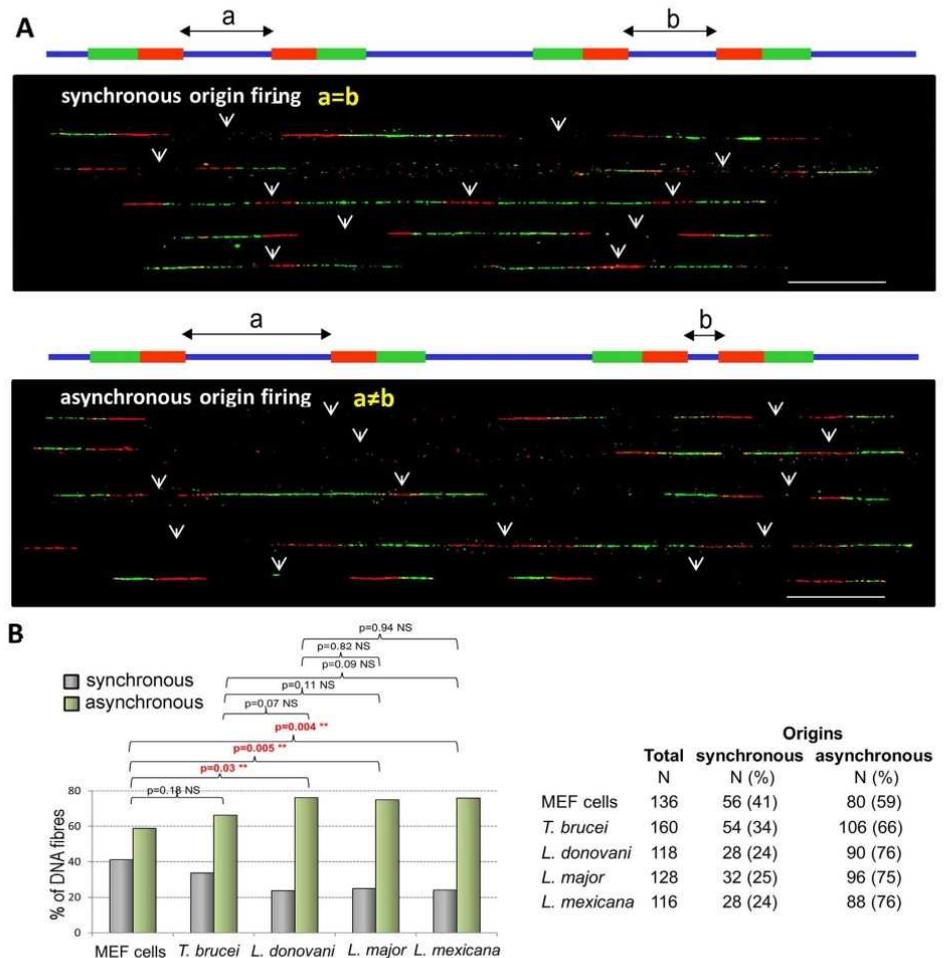


Figure 5. Analysis of synchronous and asynchronous adjacent origin firing. (A) Representative chromosome fibres used for the analysis of DNA replication dynamics showing synchronous (upper picture) and asynchronous (lower picture) firing of adjacent origins (arrows). Diagrams: “a” and “b” represent the lengths of the tracks of non-labelled DNA in two adjacent origins. They are used to discriminate between synchronous and asynchronous origins: if “ $a = b$ ” the two adjacent origins are synchronous, if “ $a \neq b$ ”, they are asynchronous. (B) The number of DNA fibres bearing synchronous and asynchronous adjacent origin firing was counted in the five cell types (MEF cells, *T. brucei* and three *Leishmania* species). Numbers and percentages are presented in the table (right) and the percentages of each type of firing represented as histograms (left). The Chi-square test was used to calculate the P values ($P < 0.05$ was taken as significant).

One of the major findings of our study is the presence of multiple active origins on the chromosomes of *Leishmania*. Indeed, MFAseq has also recently been applied to two *Leishmania* species (*L. major* and *L. mexicana*)⁴⁸ and found that each chromosome of *Leishmania* appears to be replicated from a single origin. Our data contradict this statement. We were able to see more than one active origin on most DNA fibres in all three analyzed *Leishmania* species, like in *T. brucei* (Figs 2A and 5A) and any other eukaryote analyzed so far. Incidentally, in their discussion, the authors raised the issue that their data may not be relevant for *Leishmania* biology and that a single-origin chromosome is not consistent with the duration of the S phase estimated at 2.9 h in *L. mexicana*⁴⁹. Indeed, considering the mean fork velocity of 2.6 kb/min found here for *L. mexicana* (Table 1), the replication of the largest chromosomes (1–3 Mb) from a single origin would take 6–20 hours. The robustness of our approach clearly shows that the *Leishmania* chromosomes, at least the largest ones, are replicated from multiple origins.

The most distinguishing feature of DNA replication in *T. brucei* as compared to *Leishmania* and MEF cells is the high frequency of unidirectional replication forks. The reason for such a discrepancy between the two parasites is unknown. However, it is tempting to implicate the numerous ‘mini-’ and ‘intermediate’ chromosomes of *T. brucei*, since these are specific of *T. brucei* and built around repetitive palindromes⁵⁰; the latter have been proposed to adopt hairpin structures during lagging-strand synthesis, which represent obstacles to fork

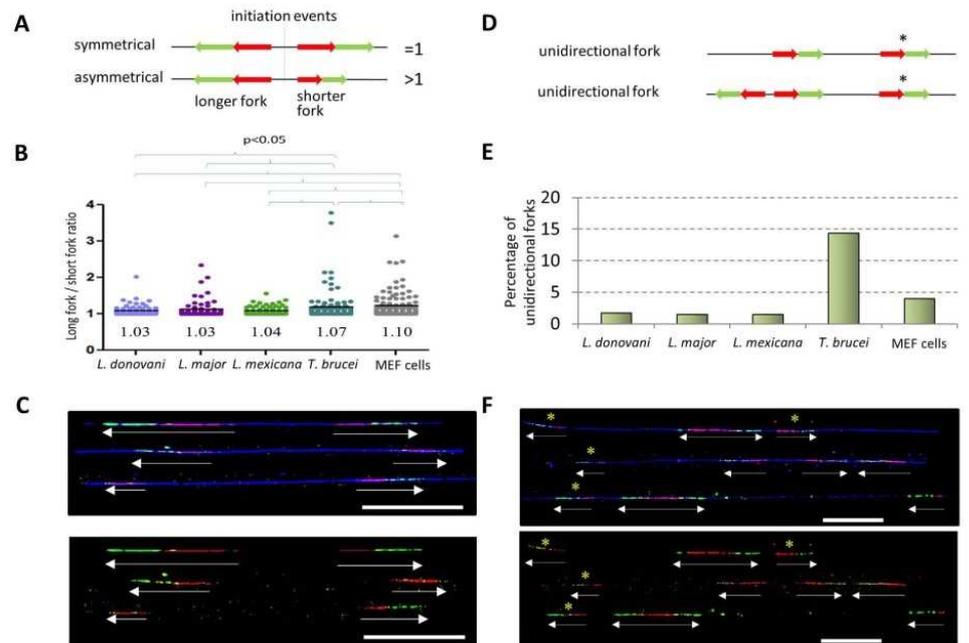


Figure 6. Analysis of the symmetry of progressing bidirectional replication forks. (A) Presentation of the concept of long fork/short fork ratios. The long fork/short fork ratio corresponds to the ratio of the longer IdU + CldU signal over the shorter IdU + CldU signal of bidirectional replication forks. A ratio = 1 indicates symmetry (see Materials & Methods). A ratio > 1 indicates fork asymmetry, while a ratio = 1 indicates symmetry (see Materials & Methods). (B) Distribution of long fork/short fork ratios in MEF cells, *T. brucei* and three *Leishmania* species. Median (black bar) values are indicated below the distribution clouds. Statistically significant differences are indicated; p values were calculated using the two-tailed Mann-Whitney test. (C) Representative pairs of asymmetric replication forks from *T. brucei*. The forks were assembled from different microscope fields. Arrows indicate the lengths of the bidirectional forks analysed. Red tracks: IdU, green tracks: CldU, blue tracks: DNA. DNA counterstaining is not shown in the bottom panel. Scale bar: 50 kb. (D) Schematic presentation of the unidirectional replication forks that were analysed. Forks indicated by asterisks are counted as unidirectional forks and used in the calculation of percentage of unidirectional forks. In the top fibre, the left fork was not included for calculation as one cannot distinguish between a unidirectional fork and a fibre broken between the bidirectional forks. (E) Histograms showing the percentage of unidirectional replication forks over the total number of analysed forks in MEF cells, *T. brucei* and three *Leishmania* species. (F) Representative DNA fibres from *T. brucei* showing unidirectional forks. The fibres were assembled from different fields. Arrows indicate the direction of replication forks. Unidirectional forks indicated by asterisks were used for this analysis. Same color code and legend as in (C).

progression⁵¹. Similarly, collisions between transcription and replication can also cause replication fork stalling. Using MFA seq, Tiengwe *et al.*¹¹ observed asymmetric replication peaks within directional gene clusters (DGCs) on megabase-sized chromosomes and noted that replication was slower when transcription was progressing in the direction opposite to DNA replication. However the organization of the genome in DGCs is not a specificity of *T. brucei* but a feature shared by all trypanosomatids. Finally, they did not observe unidirectional replication peaks.

In conclusion, DNA molecular combing here revealed and clarified original features of DNA replication in trypanosomatids in comparison with other eukaryotic model organisms, such as large IODs, fast replication fork progression and, in *T. brucei*, the elevated occurrence of unidirectional replication forks. The characteristics of DNA replication in these either aneuploid or diploid organisms are often similar but also show differences. Yet, these are not sufficient to explain their differences in ploidy patterns. Therefore, rather than to different replication parameters, it is probable that the mechanisms responsible for the unique mosaic aneuploidy of *Leishmania* are related to a 'leaky' regulation of DNA replication, still to be characterized, and most likely at the level of the licensing or re-licensing of DNA replication origins.

Methods

Parasites and in vitro culture. Promastigotes of *Leishmania major* 'Friedlin' (MHOM/IL/81/Friedlin), *L. mexicana* (MHOM/GT/2001/U1103, clone 25), and *Leishmania donovani* (LD1S) were grown at 26 °C in RPMI1640 (Gibco BRL) supplemented with 10% fetal bovine serum (FBS)⁵². Procyclic forms of the Lister 427 wild type strain and Lister 427 29–13 cell line of *T. brucei* were grown at 27 °C in SDM-79 (PAA Laboratories)

supplemented with 10% FBS and 7 µg/ml hemin, and with 30 µg/ml of hygromycin and 10 µg/ml of geneticin for the 29–13 line⁵³.

DNA molecular combing. Asynchronous insect form cell populations of *T. brucei* and three *Leishmania* species (*L. donovani* Ld1S, *L. major* Friedlin 3171 and *L. mexicana* U1103) were grown to the density of 4.10⁶ cells/mL and sequentially labelled with two modified nucleosides, iodo-deoxyuridine (IdU, Sigma) and chloro-deoxyuridine (CldU, Sigma). Cells were sequentially labelled with 300 µM IdU and 300 µM CldU for 20 (or 15) minutes each, without intermediate wash. After labeling, the cells were immediately placed on ice to stop DNA replication. MEF cells derived from 13.5-d mouse embryos were cultured as previously described⁵⁴ and used at passage 4–5. They were labelled with IdU and CldU in the same manner and with the same concentrations of modified nucleosides as parasite cells. Cells were then centrifuged (3000 g for 5 min at +4 °C) and washed once with 1 × phosphate-buffered saline (PBS); 1.10⁸ cells were resuspended in 100 µl of 1 × PBS with 1% low-melting agarose in order to embed cells in agarose plugs. Plugs were incubated in 0.5 mL of 0.5 M EDTA with 1% N-lauryl-sarcosyl and 2 mg/mL proteinase K and incubated at 45 °C for 2 days (fresh solution was added on the second day). Complete removal of digested proteins and other degradation products was performed by washing the plugs in 0.5 M EDTA and TE buffer several times. Protein-free DNA plugs were then stored in 0.5 M EDTA at 4 °C or used immediately for combing. Agarose plugs were stained with YOYO-1 fluorescent dye (Molecular Probes) in TE buffer for 1 h, washed with TE buffer, resuspended in 100 µl of TE buffer and melted at 65 °C for 15 minutes. The solution was maintained at 42 °C for 15 minutes and treated overnight with β agarase (Sigma Aldrich). After digestion, 4 ml of 50 mM MES (2-(N-morpholino) ethanesulfonic acid, pH 5.7) were added very gently to the DNA solution and then DNA fibres were combed and regularly stretched (2 kb/µm) on silanized coverslips as described previously²². Combed DNA was fixed at 65 °C for 2 hours, denatured in 1N NaOH for 20 minutes and washed several times in PBS. After denaturing, silanized coverslips bearing the DNA fibres were blocked with 1% BSA and 0.1% Triton X100 in PBS. Immuno-detection was done with antibodies diluted in 1 × PBS, 0.1% Triton X100, 1% BSA and incubated at 37 °C in a humid chamber for 60 min. Each step of incubation with antibodies was followed by extensive washes with 1 × PBS. Immuno-detection was done with anti-ssDNA antibody (1/100 dilution, Chemicon), the mouse anti-BrdU antibody (1/20 dilution, Becton Dickinson) and a rat anti-BrdU antibody (1/20 dilution, Sera Lab) that recognize the IdU and CldU tracks, respectively. The mouse anti-BrdU, clone B44, is derived from hybridization of mouse Sp2/0-Ag14 myeloma cells with spleen cells from BALB/c mice immunized with iodouridine-conjugated ovalbumin. It reacts with iodouridine and BrdU^{55,56}. The rat anti-BrdU antibody, clone BU1/75 (ICR1), cross reacts with chlorodeoxyuridine (CldU) but does not cross react with thymidine or iododeoxyuridine⁵⁶. The secondary antibodies were goat anti-rat antibody coupled to Alexa 488 (1/50 dilution, Molecular Probes), goat anti-mouse IgG2a coupled to Alexa 647 (1/100 dilution, Molecular Probes). Coverslips were mounted with 20 µl of Prolong Gold Antifade (Molecular Probes), dried at room temperature for 2 hrs and processed for image acquisition using a fully motorized Leica DM6000 microscope equipped with a CoolSNAP HQ2 1 CCD camera and controlled by MetaMorph (Roper Scientific). Images were acquired with a 40 × objective, where 1 pixel corresponds to 322 bp. Thus, at a magnification of 40 ×, one microscope field corresponds to ~450 kb. Observation of longer DNA fibres therefore requires the capture of adjacent fields or the use of a fluorescence microscope equipped with a motorized stage that enables scanning of slides with high precision. Fibres < 150 kb were excluded from the analysis. Inter-origin distances and the velocity of replication forks were measured manually using the MetaMorph software. Statistical analyses of inter-origin distances and velocities of replication forks were performed using Prism 5.0 (GraphPad). At least three independent cell labeling and combing experiments were performed for each analysis presented here.

Statistical analysis. Replication fork speed was estimated on individual forks displaying an IdU track flanked by a CldU track. Only intact forks were recorded as ascertain by DNA counterstaining. The estimations of fork asymmetry were calculated as the ratio of the longer track over the shorter track in progressing bidirectional forks. A longer fork/shorter fork ratio > 1 indicates asymmetry. Inter-origin distances were measured as the distance (kb) between the centers of two adjacent progressing forks located on the same DNA fibre. The GraphPad Prism (GraphPad Software) was used to generate graphs and perform statistical analysis. DNA replication parameters generally do not display a Gaussian distribution⁵⁷. Statistical comparisons of the distributions were therefore assessed using the nonparametric Mann–Whitney rank sum test. Two-tailed tests were systematically used. The number of synchronous and asynchronous origins in the different cell types were compared using Chi-square tests. Statistical significance was set at $P \leq 0.05$. The correlation between fork speed and IODs was analysed using linear regression and the coefficient of determination R^2 .

References

- Mechali, M. DNA replication origins: from sequence specificity to epigenetics. *Nat Rev Genet.* **2**, 640–645 (2001).
- Lemaitre J. M., Vassetzky Y. & Méchali M. Analysis of Chromatin Assembly, Chromatin Domains, and DNA Replication Using Xenopus Systems. In, Mapping Protein/DNA Interactions by Cross-Linking [Internet]. Paris: Institut national de la santé et de la recherche médicale; 2001. Available from, <http://www.ncbi.nlm.nih.gov/gate1.inist.fr/books/NBK7096/> (Accessed: 8th February 2016).
- Mechali, M., Yoshida, K., Coulombe, P. & Pasero, P. Genetic and epigenetic determinants of DNA replication origins, position and activation. *Curr Opin Genet Dev.* **23**, 124–131 (2013).
- Baldauf, S. L. The deep roots of eukaryotes. *Science* **300**, 1703–1706 (2003).
- Daniels, J. P., Gull, K. & Wickstead, B. Cell biology of the trypanosome genome. *Microbiol Mol Biol Rev.* **74**, 552–569 (2010).
- Ersfeld, K. Nuclear architecture, genome and chromatin organisation in *Trypanosoma brucei*. *Res Microbiol.* **162**, 626–636 (2011).
- Wincker, P. et al. The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res.* **24**, 1688–1694 (1996).

8. Britto, C. *et al.* Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene* **222**, 107–117 (1998).
9. Wong, A. K., Curotto de Lafaille, M. A. & Wirth, D. F. Identification of a cis-acting gene regulatory element from the *lemdr1* locus of *Leishmania enriettii*. *J Biol Chem.* **269**, 26497–26502 (1994).
10. Martínez-Calvillo, S., Nguyen, D., Stuart, K. & Myler, P. J. Transcription initiation and termination on *Leishmania* major chromosome 3. *Eukaryot Cell* **3**, 506–517 (2004).
11. Tiengwe, C. *et al.* Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell Rep.* **2**, 185–197 (2012).
12. Tiengwe, C. *et al.* Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLoS One* **7**, e32674 (2012).
13. Obado, S. O., Bot, C., Nilsson, D., Andersson, B. & Kelly, J. M. Repetitive DNA is associated with centromeric domains in *Trypanosoma brucei* but not *Trypanosoma cruzi*. *Genome Biol.* **8**, R37 (2007).
14. Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P. & Pagès, M. FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. *Cell Microbiol.* **13**, 274–283 (2011).
15. Lachaud, L. *et al.* Constitutive mosaic aneuploidy is a unique genetic feature widespread in the *Leishmania* genus. *Microbes Infect.* **16**, 61–66 (2014).
16. Melville, S. E., Leech, V., Gerrard, C. S., Tait, A. & Blackwell, J. M. The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Mol Biochem Parasitol.* **94**, 155–173 (1998).
17. Sterkers, Y. *et al.* Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. *Mol Microbiol.* **86**, 15–23 (2012).
18. Bessat, M. & Ersfeld, K. Functional characterization of cohesin SMC3 and segregation of large and minichromosomes in *Trypanosoma brucei*. *Mol Microbiol.* **71**, 1371–1385 (2009).
19. Akiyoshi, B. & Gull, K. Discovery of unconventional kinetochores in kinetoplastids. *Cell* **156**, 1247–1258 (2014).
20. Morelle, C. *et al.* The nucleoporin Mlp2 is involved in chromosomal distribution during mitosis in trypanosomatids. *Nucleic Acids Res.* **43**, 4013–4027 (2015).
21. Schwob, E. *et al.* Use of DNA combing for studying DNA replication *in vivo* in yeast and mammalian cells. *Methods Mol Biol.* **521**, 673–687 (2009).
22. Michalet, X. *et al.* Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* **277**, 1518–1523 (1997).
23. Cayrou, C. *et al.* Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* **21**, 1438–1449 (2011).
24. Techer, H. *et al.* Replication dynamics: biases and robustness of DNA fiber analysis. *J Mol Biol.* **425**, 4845–4855 (2013).
25. Raghuraman, M. K. *et al.* Replication dynamics of the yeast genome. *Science* **294**, 115–121 (2001).
26. van Oijen, A. M. & Loparo, J. J. Single-molecule studies of the replisome. *Annu Rev Biophys.* **39**, 429–448 (2010).
27. Conti, C. *et al.* Inhibition of histone deacetylase in cancer cells slows down replication forks, activates dormant origins, and induces DNA damage. *Cancer Res.* **70**, 4470–4480 (2010).
28. Anglana, M., Apiou, E., Bensimon, A. & Debatisse, M. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* **114**, 385–394 (2003).
29. Poli, J. *et al.* dNTP pools determine fork progression and origin usage under replication stress. *Embo J.* **31**, 883–894 (2012).
30. Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–422 (2005).
31. Burri, M., Schlimme, W., Betschart, B. & Hecker, H. Characterization of the histones of *Trypanosoma brucei* blood stream forms. *Acta Trop.* **58**, 291–305 (1994).
32. Schlimme, W., Burri, M., Betschart, B. & Hecker, H. Properties of the histones and functional aspects of the soluble chromatin of epimastigote *Trypanosoma cruzi*. *Acta Trop.* **60**, 141–154 (1995).
33. Alsford, S. & Horn, D. Trypanosomatid histones. *Mol Microbiol.* **53**, 365–372 (2004).
34. Figueiredo, L. M., Cross, G. A. & Janzen, C. J. Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat Rev Microbiol.* **7**, 504–513 (2009).
35. Janzen, C. J. *et al.* Unusual histone modifications in *Trypanosoma brucei*. *FEBS Lett.* **580**, 2306–2310 (2006).
36. Mandava, V. *et al.* Histone modifications in *Trypanosoma brucei*. *Mol Biochem Parasitol.* **156**, 41–50 (2007).
37. Cayrou, C. *et al.* The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res.* **25**, 1873–85 (2015).
38. Calderano, S. G. *et al.* Single molecule analysis of *Trypanosoma brucei* DNA replication dynamics. *Nucleic Acids Res.* **43**, 2655–2665 (2015).
39. Conti, C. *et al.* Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol Biol Cell* **18**, 3059–3067 (2007).
40. Costas, C. *et al.* Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks. *Nat Struct Mol Biol.* **18**, 395–400 (2011).
41. Lengronne, A., Pasero, P., Bensimon, A. & Schwob, E. Monitoring S phase progression globally and locally using BrdU incorporation in TK(+) yeast strains. *Nucleic Acids Res.* **29**, 1433–1442 (2001).
42. Yabuki, N., Terashima, H. & Kitada, K. Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7**, 781–789 (2002).
43. Palumbo, E., Tosoni, E. & Russo, A. General and specific replication profiles are detected in normal human cells by genome-wide and single-locus molecular combing. *Exp Cell Res.* **319**, 3081–3093 (2013).
44. Ivens, A. C. *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**, 436–442 (2005).
45. Downing, T. *et al.* Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* **21**, 2143–2156 (2011).
46. Muller, C. A. *et al.* The dynamics of genome replication using deep sequencing. *Nucleic Acids Res.* **42**, e3 (2014).
47. Fragkos, M., Ganier, O., Coulombe, P. & Mechali, M. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol.* **16**, 360–374 (2015).
48. Marques, C. A., Dickens, N. J., Paape, D., Campbell, S. J. & McCulloch, R. Genome-wide mapping reveals single-origin chromosome replication in *Leishmania*, a eukaryotic microbe. *Genome Biol.* **16**, 230 (2015).
49. Wheeler, R. J., Gluenz, E. & Gull, K. The cell cycle of *Leishmania*: morphogenetic events and their implications for parasite biology. *Mol Microbiol.* **79**, 647–662 (2011).
50. Wickstead, B., Ersfeld, K. & Gull, K. The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes. *Genome Res.* **14**, 1014–1024 (2004).
51. Lebofsky, R. & Bensimon, A. DNA replication origin plasticity and perturbed fork progression in human inverted repeats. *Mol Cell Biol.* **25**, 6789–6797 (2005).
52. Dubessay, P., Blaineau, C., Bastien, P. & Pages, M. Chromosome fragmentation in *Leishmania*. *Methods Mol Biol.* **270**, 353–378 (2004).
53. Wirtz, E., Leal, S., Ochatt, C. & Cross, G. A. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol.* **99**, 89–101 (1999).

54. Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).
55. Gratzner, H. G. Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: A new reagent for detection of DNA replication. *Science* **218**, 474–475 (1982).
56. Bakker, P. I., Aten, J. A., Tukker, C. I., Barendsen, G. W. & Veenhof, C. H. Flow cytometric analysis of experimental parameters for the immunofluorescent labeling of BrdUrd in various tumour cell lines. *Histochemistry* **91**, 425–429 (1989).
57. Bianco, J. N. *et al.* Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods* **57**, 149–157 (2012).

Acknowledgements

We acknowledge Marjorie Drac from the Montpellier DNA Combing Facility for providing silanized cover slips. We thank Virginie Georget and Julien Cau from the Montpellier RIO Imaging Facility (MRI) for their help and microscopy facilities. We are very grateful to Jeremy Mottram and Gilles Merlin for providing the *L. mexicana* strain, Gerald Spaeth (Institut Pasteur) for providing the *L. donovani* strain, Derrick Robinson and Frédéric Bringaud (University Victor Segalen Bordeaux 2) for providing the *T. brucei* wild-type strain and the *T. brucei* 29–13 cell line, respectively, as well as Marta Bialic (IGMM, Montpellier) for the MEF cells. This study was supported by the Agence Nationale de la Recherche within the frame of the “Investissements d’avenir” program (ANR-11-LABX-0024-01 “PARAFRAP”), the Institut Pasteur *Leishmania* Consortium “LeISHield”, the Centre National de la Recherche Scientifique (CNRS), the French Ministry of Research and the Centre Hospitalier Universitaire of Montpellier.

Author Contributions

S.S., L.S., N.K. and L.C. conducted the experiments, Y.B. built the molecular combing device; S.S., M.P., Y.S. and E.S. designed the experiments and M.P., S.S., Y.S. and P.B. wrote the paper with feedback from all authors.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Stanojic, S. *et al.* Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes *Leishmania* and *Trypanosoma brucei* versus mammalian cells. *Sci. Rep.* **6**, 23142; doi: 10.1038/srep23142 (2016).

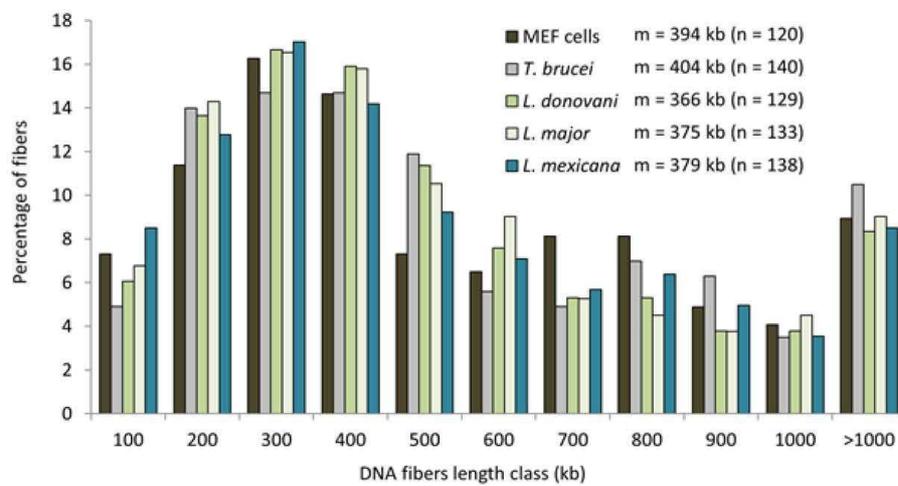


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

SUPPLEMENTAL MATERIAL

Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes *Leishmania* and *Trypanosoma brucei* versus mammalian cells

Slavica Stanojic, Lauriane Sollelis, Nada Kuk, Lucien Crobu, Yves Balard, Etienne Schwob, Patrick Bastien, Michel Pagès and Yvon Sterkers



Supplemental Fig 1. Distribution of length classes of DNA fibers in MEF cells, *T. brucei* and three *Leishmania* species.

DNA fibers combed on the coverslips were measured. The median fiber length in kb (m) and the number of measures (n) are indicated for each sample.

5.1.3 Conclusion

Ce travail a permis de mettre en évidence plusieurs aspects originaux de la dynamique de réplication chez les trypanosomatidés. Un résultat majeur de cette étude montre que *Leishmania* possède les plus grandes distances inter-origines et la plus grande vitesse de réplication observées parmi les autres eucaryotes déjà étudiés. De plus, nous avons pu estimer que le génome de *Leishmania* possède environ 168 origines de réplication.

L'une des hypothèses pour expliquer l'aneuploïdie chez *Leishmania* était qu'un double allumage (*firing*) (double activation des origines) de certaines origines expliquait la présence de chromosomes surnuméraires, et l'absence d'allumage provoquait la perte de chromosomes. Cependant, la technique du peignage n'a pas mis en évidence d'évènements de double allumage. Il est cependant possible que ce double évènement d'activation ait lieu hors du laps de temps défini par la durée du temps de marquage qui est de 10 min. Une solution serait de faire des gels bidimensionnels, associés au « bubble-seq », qui permettraient de visualiser ce type d'évènement.

Les origines de réplication actives chez *T. brucei* étudiées, avec à la méthode du MFA, (Tiengwe *et al.*, 2012) ont permis de mettre en évidence des origines de réplication symétriques et asymétriques. Les origines de réplication asymétriques seraient la conséquence d'une compétition entre les machineries de transcription et de réplication, la première bloquant la progression de la fourche de réplication. En effet ces situations sont toujours retrouvées à l'intérieur des SSR au niveau de sites où la transcription s'arrête et repart (*divergent SSR*). En peignage moléculaire nous avons également observé chez *T. brucei* un taux élevé de fourches asymétriques qui corrobore bien les observations précédentes (Tiengwe *et al.*, 2012).

Le peignage a permis de confirmer les données de MFA chez *T. brucei* où 14,4 % des fourches de réplication observées sont unidirectionnelles contre 1,5% chez *L. major*. Les fourches de réplication unidirectionnelles pourraient être également une conséquence du couplage de la réplication à la transcription. Ce phénomène se retrouverait chez *T. brucei* en particulier dans les régions contenant les VSG et serait absent de chez *Leishmania*.

Nous avons observé d'autres différences entre *Leishmania* et *T. brucei*, notamment sur la grande variabilité des distances inter-origines. Ces valeurs vont de 48,6 kb à 625,4 kb pour *L. mexicana*, une telle amplitude n'a pas été observée ni chez *T. brucei* ni chez les

cellules MEF. Ceci peut suggérer une flexibilité dans l'activation des origines. Toujours chez *Leishmania*, il a été noté une proportion très importante d'origines asynchrones (75% vs 66% chez *T. brucei* et 59% chez les cellules MEF). Ce constat peut faire penser à l'existence, comme chez les autres eucaryotes, d'origines de réplication qui s'activent tardivement et d'autres précocement. Cette flexibilité du « *licensing* » ouvre la possibilité de l'existence de double allumage qui serait à l'origine de la synthèse de chromosomes surnuméraires. Finalement, le nombre d'origines de réplication estimé par génome haploïde ($n=157$) et la vitesse moyenne de réplication mesurée (2.48 kb/min), permettraient à la cellule de répliquer son génome en 1,4 h. Ainsi, *L. mexicana* aurait le temps de répliquer deux fois la totalité de son génome en 2,9 h, et donc le temps de produire des chromosomes surnuméraires caractéristiques de l'aneuploïdie mosaïque.

5.2 Identification et caractérisation des protéines Orc1/Cdc6 et Orc1b impliquées dans la réplication de l'ADN.

5.2.1 Introduction

Dans la continuité du travail sur la réplication en peignage moléculaire, il était primordial de s'intéresser aux mécanismes qui contrôlent l'activation des origines de réplication. Quels sont les acteurs, à quel moment interviennent-ils, et quels sont les moyens de contrôle ?

La réplication de l'ADN des eucaryotes implique le recrutement de différents facteurs. La réplication est initiée par le recrutement du complexe protéique ORC composé de six sous unités (Orc1 à Orc6). Chez les trypanosomatidés, la réplication de l'ADN nucléaire est encore mal connue. Les seules données que nous avons jusqu'à présent concernaient *T. brucei*, organisme diploïde (Tait, 1980) Une des hypothèses était que l'aneuploïdie mosaïque de *Leishmania* est la conséquence d'une régulation non-conventionnelle de la réplication. Pour comprendre comment *Leishmania* régule la réplication de son génome nous avons étudié la fonction de deux protéines orthologues de *T. brucei* LmOrc1 et LmOrc1/Cdc6. Pour cela nous avons utilisé l'expression épisomale et le CRISPR/Cas9 pour la localisation des protéines. L'étude de la fonction de ces protéines a été réalisée chez *Leishmania* à partir de KO obtenus avec la technique du CRISPR-Cas9, et chez *T. brucei* par la complémentation fonctionnelle des RNAi.

5.2.2 Résultats

5.2.2.1 Identification et localisation des protéines ORC.

Leishmania et *Trypanosoma* sont des parasites qui présentent une forte synténie de leur organisation génomique. Les précédentes études (Godoy *et al.*, 2009, Dang & Li, 2011) menées chez *T. brucei* ont permis d'identifier plusieurs gènes associés à la réplication de l'ADN, dont tous les orthologues sont présents chez *L. major* (Tableau 5-1.). Pour ce travail, nous nous sommes concentrés sur deux facteurs LmOrc1/Cdc6 (LmjF.28.0030) et LmORC1b (LmjF.26.2210).

Tableau 5-1 Replication factors

Eukaryotes factors	Tryps Factors	<i>Trypanosoma brucei</i>	<i>Leishmania major</i>
Orc1	ORC1/CDC6	Tb11.02.5110 Tb927.11.7216.1	LmjF.28.0030
	ORC1B	Tb927.9.2030	LmjF.26.2210
Orc2	Not found		
Orc3	Not found		
Orc4	ORC4	Tb927.10.13380	LmjF.18.0720
Orc5	Not found		
Orc6	Not found		
Cdc6	Not found		
	new	Tb09.160.3120	LmjF.36.6700
	new	Tb927.10.7980	LmjF.01.0660
Cdt1	Not found	Not found	Not found
MCM2	MCM2	Tb927.11.7860	LmjF.28.0850
MCM3	MCM3	Tb927.2.3930	LmjF.33.2700
MCM4	MCM4	Tb927.11.12250	LmjF.09.0250
MCM5	MCM5	Tb927.11.5570	LmjF.24.0910
MCM6	MCM6	Tb927.11.11640	LmjF.28.2385
MCM7	MCM7	Tb927.11.16140	LmjF.32.2960
Cdc45	TbCDC45	Tb927.11.2710	LmjF.33.2320
GIN5	Sld5 TbSld5	Tb927.3.4810	LmjF.29.2260
Pfs1	TbPfs1	Tb927.9.4990	LmjF.15.1510
Pfs2	TbPfs2	Tb927.11.10460	LmjF.36.4190
Pfs3	TbPfs3	Tb927.4.4680	LmjF.31.1660
MCM10	TbMCM10	Tb927.9.1730	LmjF.26.2390
MCM8	TbMCM8	Tb927.10.10410	LmjF.05.0330
MCM9	TbMCM9	Tb927.9.9600	LmjF.35.4910
MCM-BP	TbMCM-BP	Tb927.7.1770	LmjF.26.0150
PCNA	TbPCNA	Tb927.9.5190	LmjF.15.1450

La localisation des protéines a dans un premier temps été réalisée à partir d'une expression épisomale (utilisant le vecteur d'expression pTH6) pour LmOrc1/Cdc6, LmORC1b et LmORC4 chez *L. major*. Nous avons utilisé différents « tags » GFP et 12x C-myc pour vérifier qu'il n'y avait pas d'effet du « tag » sur la localisation des protéines. Le « tag » C-Myc avait été utilisé dans l'optique des expériences de ChIP. Ces « tags » ont été placés en Cter ou Nter afin de vérifier si leurs positions ne provoquaient pas une relocalisation des protéines. Pour ce type de protéines, nous nous attendions à une localisation nucléaire. Les localisations obtenues sont résumées dans le Tableau 5-2.

Tableau 5-2: Replication factors localization using different tag.

LmjF proteins	GFP		1 C-myc		12 C-myc	
	Cter	Nter	Cter	Nter	Cter	Nter
Orc1/cdc6	nuclear	nuclear	no signal	cytoplasmique	cytoplasmique	cytoplasmique
Orc1b	nuclear	nuclear	-	-	nuclear	nuclear
Orc4 S*	cytoplasmique	cytoplasmique	-	-	cytoplasmique	cytoplasmique
Orc4 L*	no signal	nuclear	nuclear	no signal	cytoplasmique	cytoplasmique

*The genomic sequence of ORC4S is the sequence annotated in TryTrypDB. This sequence does not comprise an NLS. So we checked the upstream region of the ATG and found an NLS motif and a second ATG. We called 'ORC4L' this new ORF. The overexpression of either LmORC4S or L reduced the multiplication rate of the parasites, and the percentage of dividing cells increased. More experiments are needed to better characterize this phenotype.

Le « tag » GFP, quel que soit sa position, permet une localisation correcte des protéines, sauf pour la protéine ORC4. En revanche, le «tag» C-Myc ne permet pas d'obtenir une localisation nucléaire, contrairement à ce qu'avait obtenu Tiengwe avec TbOrc1/Cdc6 (Tiengwe *et al.*, 2012). D'après ces premières expériences, il semble que la nature du «tag» et sa position aient une grande influence sur la correcte localisation des protéines.

La suite de notre étude a porté sur LmOrc1/Cdc6 marquée avec la GFP (Figure 5-1), et LmOrc1b marquée avec le «tag» C-Myc (Figure 5-2). Le marquage de LmOrc1b a été mis en évidence par immunofluorescence à l'aide d'un anticorps anti-C-myc (IFA), et une visualisation directe après fixation pour les constructions GFP.LmOrc1/Cdc6. Dans les cellules en interphase, une localisation nucléaire proche du nucléole a été mise en évidence (Figure 5-1A-B top panel). En revanche, dans les cellules en télophase, la protéine semble suivre le fuseau mitotique si on considère le pont entre les deux noyaux en division (Figure 5-1A-B bottom panel). Nos résultats montrent que la position du « tag » GFP ne perturbe pas la localisation de la protéine, et que l'expression épisomale a permis de localiser la protéine dans des cellules viables, contrairement à l'étude de Kumar *et al.* qui ont seulement pu obtenir leur localisation en expression transitoire (Kumar *et al.*, 2008).

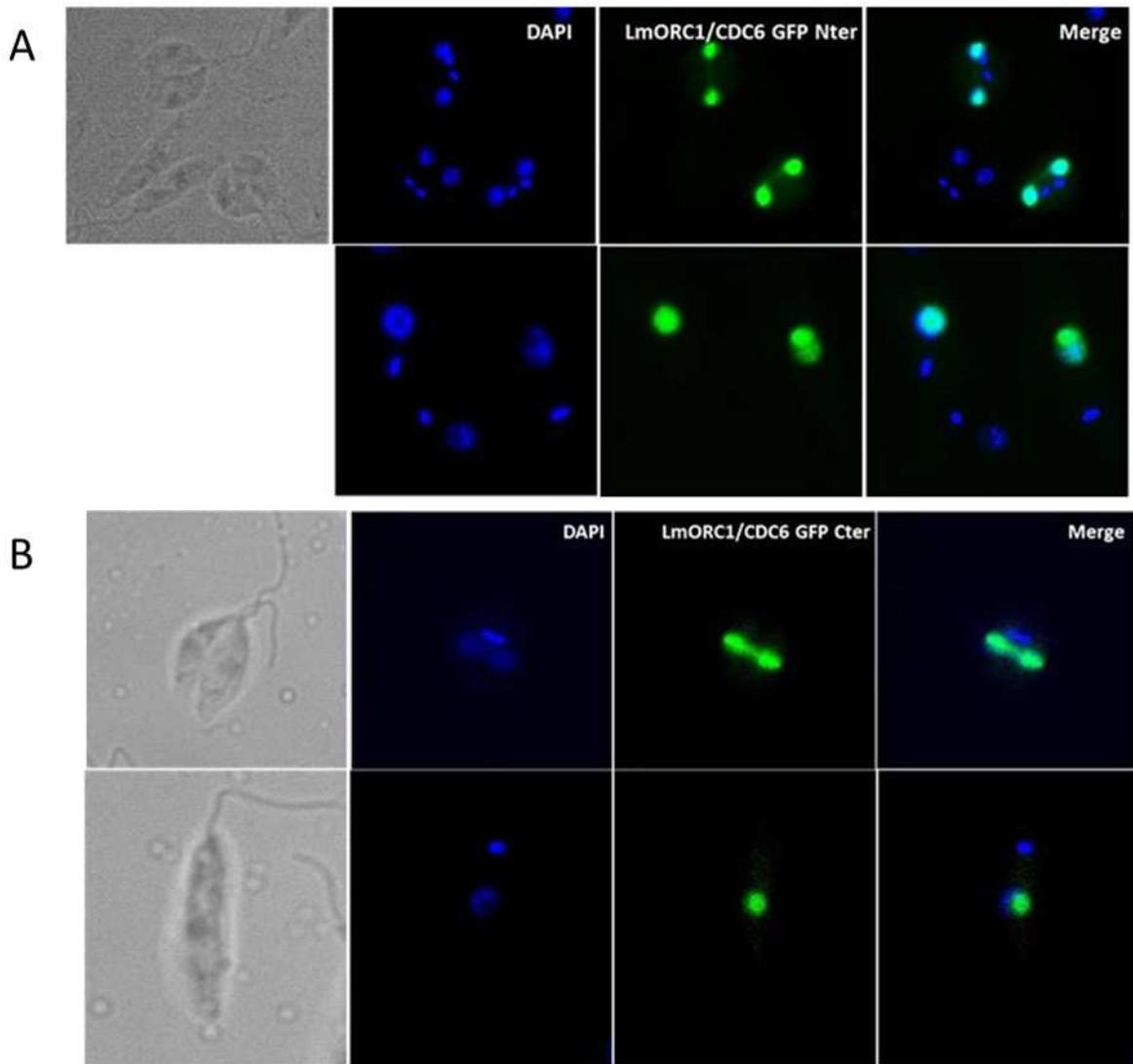


Figure 5-1 Localization of LmOrc1/Cdc6 in *L. major*.

A- GFP_Nter, B- GFP_Cter. LmOrc1/Cdc6_GFP signals locate in the nucleus. In addition, this pattern seems to be modified during the cell cycle progress.

La protéine LmOrc1b est elle aussi nucléaire comme LmOrc1/Cdc6, mais sa localisation dans le noyau semble un peu différente. On peut voir sur la Figure 5-2 que LmOrc1b forme un pattern "en nid d'abeille" dans le noyau, et semble localisée dans les zones où le DAPI apparaît moins dense suggérant une association préférentielle de la protéine avec l'euchromatine.

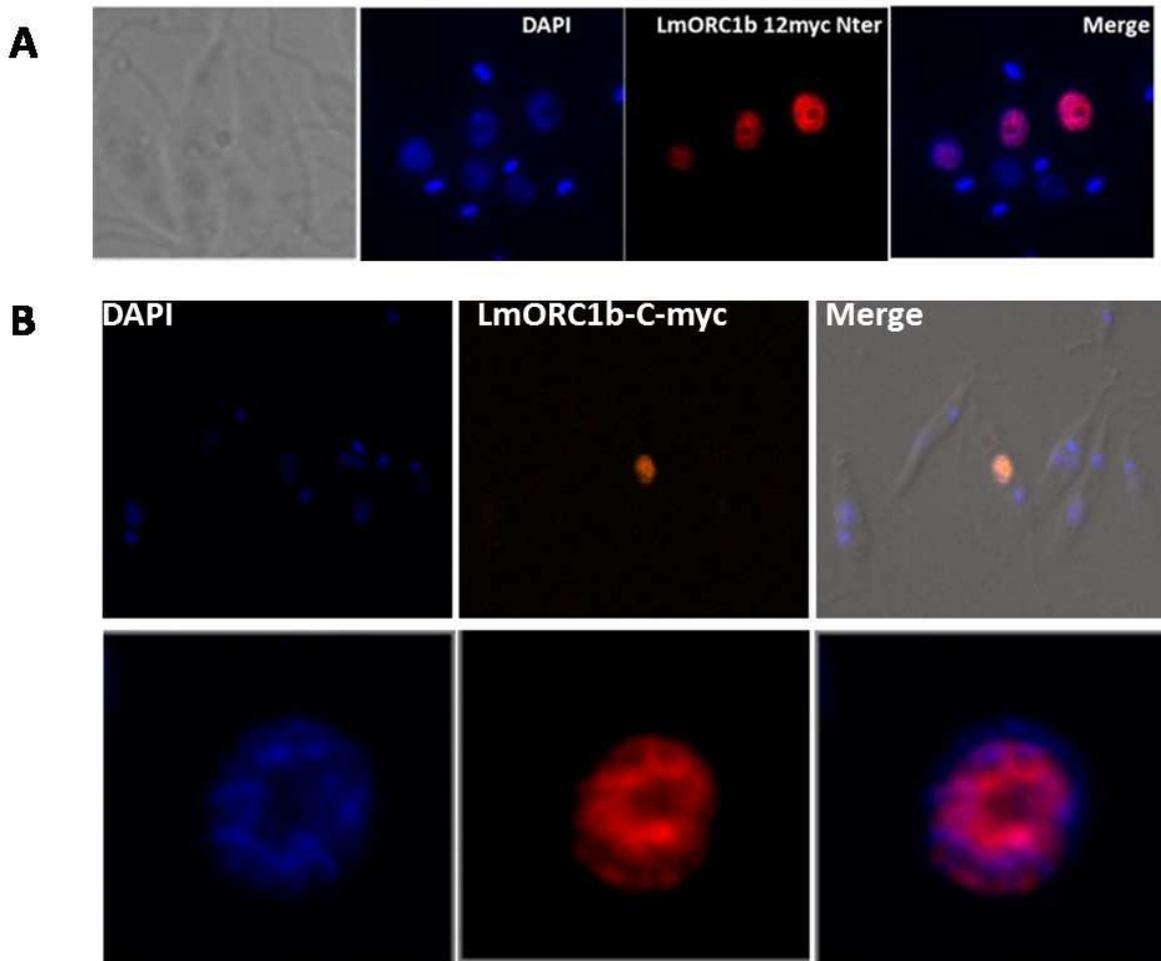


Figure 5-2 Localization of LmOrc1b in *L. major*.

A- LmORC1b 12C-myc_Nter; B- LmORC1b 12C-myc_Cter. We observed a nuclear localization for all constructs with a particular pattern in honeycomb (bottom panel)

Ainsi, il semble que certains acteurs de la réplication peuvent avoir une localisation cellulaire différente en fonction du stade du cycle cellulaire. Pour LmOrc1b et LmOrc1/Cdc6, nous avons envisagé de suivre leur localisation au cours du cycle cellulaire en couplant le marquage des protéines avec une incorporation à l'EdU³⁶, un analogue de la thymidine, pour identifier les cellules en réplication. Il fallait cependant prendre en compte deux problèmes: premièrement, toutes les cellules n'exprimaient pas la protéine tagguée, et celles qui l'exprimaient n'avaient pas la même intensité d'expression en IFA ; deuxièmement, l'expression épisomale n'est pas sous le contrôle des régions régulatrices endogènes ce qui pouvait fausser les interprétations.

³⁶ 5-ethynyl-2'-deoxyuridine

Dans le souci de conserver une régulation de l'expression des protéines dans les conditions endogènes nous avons utilisé le CRISPR/Cas9 pour réaliser un « tagging » au locus comme expliqué dans la section (4.2.2 « *Tagging* » au locus endogène par CRISPR/Cas9).

Nous avons obtenu seulement une localisation pour Orc1b avec un « tag » 3xHA dans la souche de Linf_T7TR (Figure 5-3B). Comme attendu, la localisation est nucléaire mais toutes les cellules n'expriment pas la protéine marquée. De plus, comme le montre les résultats de PCR, toutes les cellules ne sont pas éditées (Figure 5-3A). Le clonage des cellules est en cours afin d'obtenir une population homogène. Une fois cette population obtenue (toutes les cellules éditées), il sera intéressant, avec une qRT-PCR, de comparer l'expression du messenger LinfOrc1b dans Linf_T7TR à la population éditée LinfORC1b_3HA; ceci pour montrer que le « tagging » n'affecte pas la quantité d'ARN transcrit.

En ce qui concerne Orc1/Cdc6 nous n'avons toujours pas obtenu de parasite édité, nous souhaitons désormais faire une tentative en plaçant le « tag » en N-terminal de la protéine.

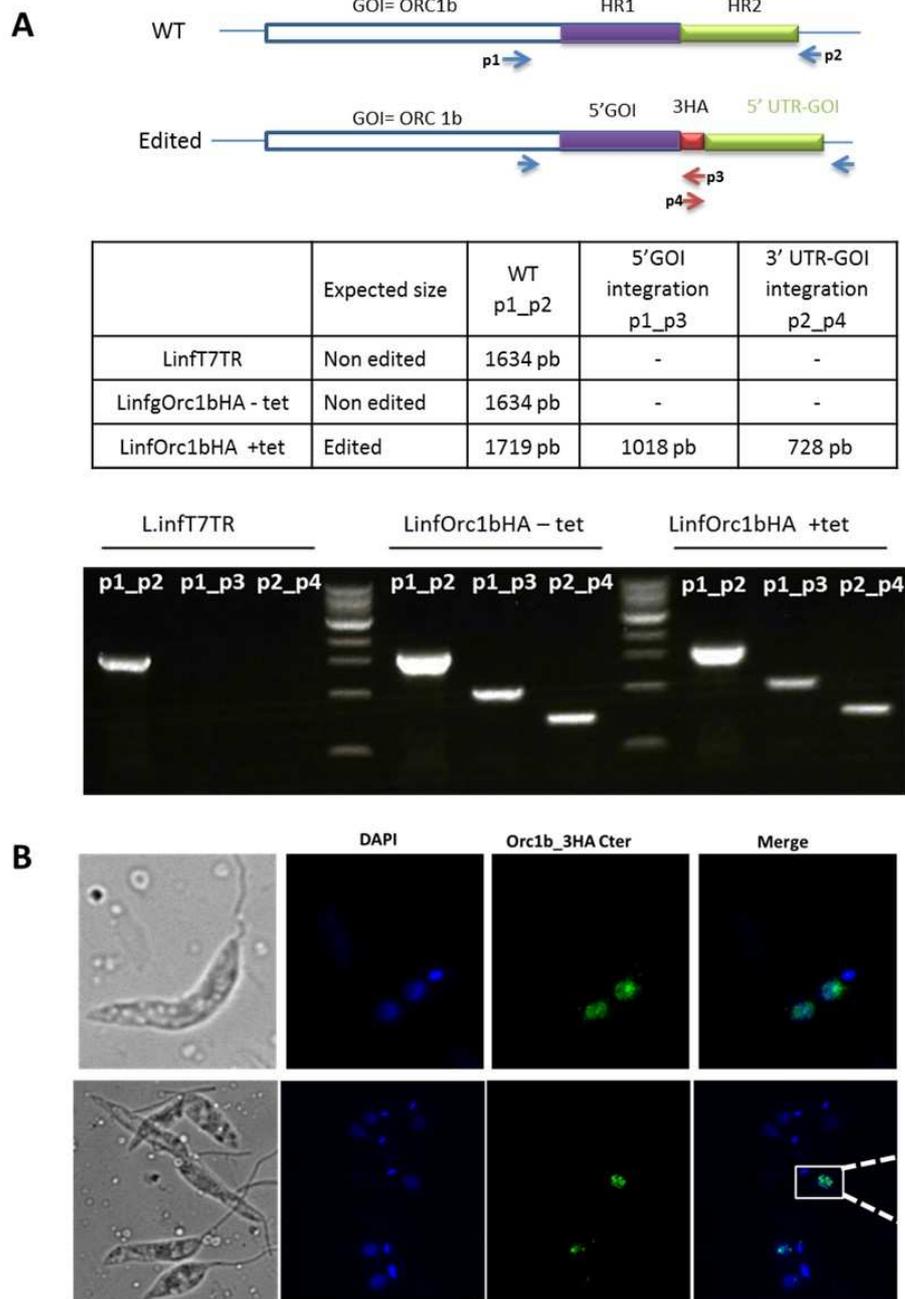


Figure 5-3 Tagging marker free of Linf_Orc1b_3HA_Cter with CRISPR/Cas9.

A- Primers used to check by PCR the integration of the 3HA tag at the 5' end of Linf_Orc1b (p1_p3 and p2_p4) and the presence of the WT locus (p1_p2) in LinfT7TR and Linf_Orc1b_3HA strains, in presence or absence of tetracycline. B-immunofluorescence anti-HA of LinfOrc1b, showing a nuclear localization in dividing (top panel) and non dividing cells (bottom panel).

5.2.2.2 Etude fonctionnelle des protéines ORCs par RNAi.

TbOrc1b et TbOrc1/Cdc6 sont connues pour être impliquées dans la réplication de *T. brucei*. Seule l'inactivation de Tb Orc1/Cdc6 par RNA interférence avait été réalisée

(Godoy *et al.*, 2009). Nous avons répété et complété ces données par des expériences de RNAi de TbOrc1b et de double RNAi de TbOrc1b et TbOc1/Cdc6. Dans un deuxième temps, nous avons réalisé la complémentation de ces RNAi avec les protéines orthologues de *Leishmania*, dont les séquences sont suffisamment différentes pour ne pas être sensibles au mécanisme d'interférence. Les expériences de RNAi ont été réalisées par des stagiaires de master que j'ai encadrés pour ce travail.

Le suivi de la croissance des cellules après RNAi de TbOrc1b a montré une réduction de la croissance dès J+3 post induction (Figure 5-4A). L'efficacité du RNAi a été contrôlée par qRT-PCR : nous avons observé une diminution de la quantité de messenger de 2,6 fois (Figure 5-4B). Nous avons observé la présence de cellules sans noyau (zoïdes) dans une proportion d'environ 17% à J+7 post induction. Ce résultat a été confirmé par FACS où nous retrouvons 10,5% de zoïdes à J+5. Le FACS nous a permis également de voir une augmentation des cellules en phase G1/S témoignant de cellules bloquées G1 et début de phase S (Figure 5-4C). Tous ces résultats suggèrent l'implication d'Orc1b dans la réplication.

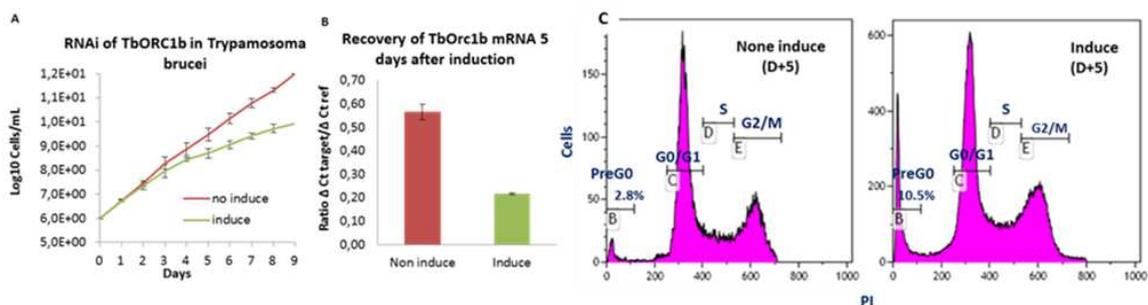


Figure 5-4 RNA interference of TbORC1b in procyclic forms of *T. brucei*.

(A) Cell growth reduction. (B) Diminution of the mRNA expression of TbORC1b was measured by qRT PCR. (C) Flow cytometry: a pre-G0 peak was observed in induced cells, which is compatible with the presence of zoids detected by DAPI staining.

Par la suite nous avons éteint l'expression de TbOrc1b et TbOrc1/Cdc6 par une double RNAi (Figure 5-5). Dès J+2 post-induction nous observons un défaut de croissance plus important que pour les simples RNAi. Pour cette expérience de double RNAi, il sera nécessaire de quantifier l'efficacité du RNAi par qRT-PCR.

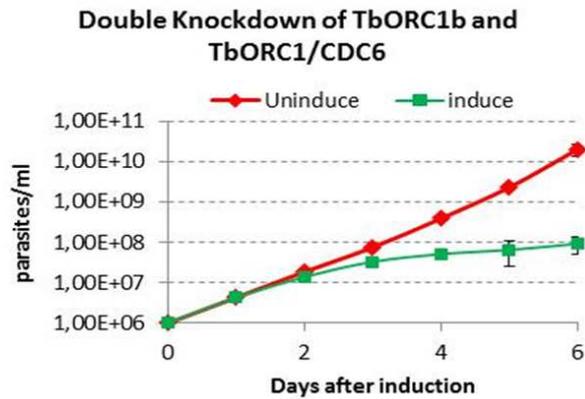
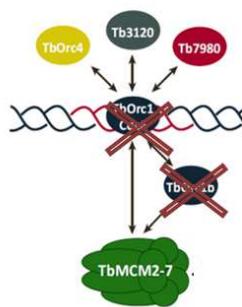


Figure 5-5. Double Knockdown of TbOrc1b and TbOrc1/Cdc6 by RNA interference. After day +2 post induction we observed a severe growth defect.

Avant la mise en place du CRISPR/Cas9 inducible chez *Leishmania* nous n'avions pas d'outil permettant l'étude fonctionnelle des gènes essentiels. Nous avons donc, dans une première approche, testé si les protéines LmOrc1b et LmOrc1/Cdc6 de *Leishmania* étaient capable de compléter les fonctions de TbOrc1b et TbOrc1/Cdc6 dans les RNAi.

Au vu des courbes de croissance nous avons observé que LmOrc1/Cdc6_GFP n'était pas capable de compléter la fonction TbOrc1/Cdc6. En revanche la version sans GFP de la protéine permet de restaurer la croissance des parasites (Figure 5-6). Dans cette expérience il reste à vérifier l'efficacité de la RNAi et le dosage de l'expression du gène hétérologue par qRT-PCR. Néanmoins nous avons montré par fluorescence que dans cette condition LmOrc1/Cdc6_GFP est exprimée et localisée au noyau chez *T. brucei*.

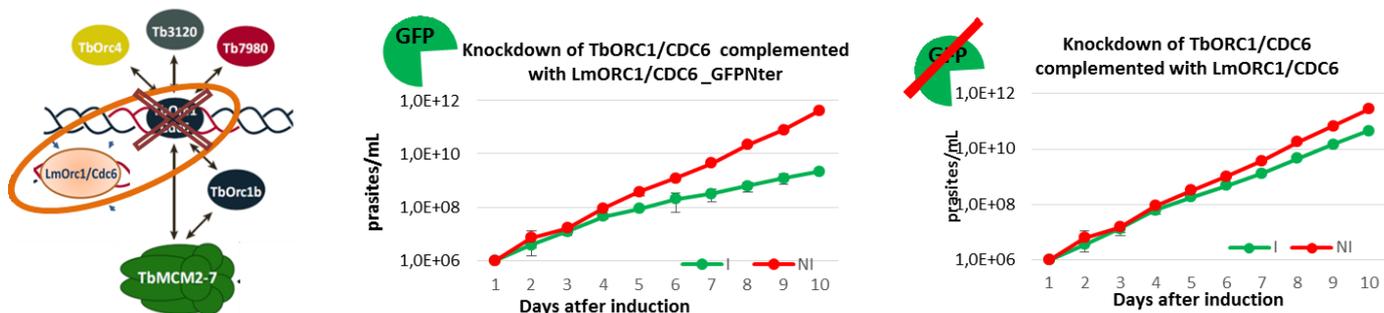


Figure 5-6. Complementation of TbOrc1/Cdc6 RNA interference by LmOrc1/Cdc6. Cells growth was monitored in absence (red line) and in presence of tetracycline (green line). Tetracycline induces at the same time the knockdown of the Tb gene and the expression of the Lm gene. When LmOrc1/Cdc6 was fused to a GFP tag, we did not

observe any complementation (left graph) but when LmORC1/CDC6 was not fused to GFP, we observed a complementation (right graph).

En revanche, la fonction de TbORC1b n'est pas complétée pas la version avec ou sans GFP de LmORC1b (Figure 5-7).

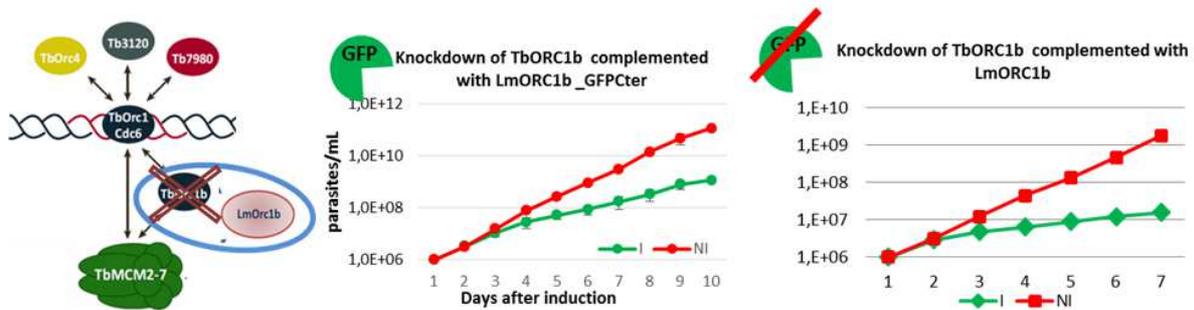


Figure 5-7. Complementation of TbOrc1b RNA interference by LmOrc1b.

Cells growth was monitored in absence (red line) and in presence of tetracycline (green line). Tetracycline induces at the same time the knockdown of the Tb gene and the expression of the Lm gene. Whether LmOrc1b is fused to a GFP tag (left graph) or not (right graph) we did not observe any complementation.

Cette donnée est très intéressante, car nous avons peut être mis en évidence une protéine qui n'aurait pas la même fonction dans les deux organismes. La comparaison des séquences protéiques ORC1b des principaux trypanosomatidés a mis en évidence que la protéine de *Leishmania* a seulement 33% d'homologie avec *T. brucei* (Figure 5-9), alors que LmOrc1/Cdc6 possède 55% d'homologie avec TbOrc1/Cdc6 (Figure 5-8). Malgré la conservation des motifs Walker A et B, LmOrc1b (870 aa) comporte plusieurs insertions dans sa séquence par rapport à *T. brucei* (602 aa). Ces observations suggèrent qu'Orc1b pourrait avoir un rôle différent chez *Leishmania* de celui mis en évidence chez *T. brucei*, ou avoir des fonctions supplémentaires et interagir avec d'autres partenaires. LmOrc1b est un candidat idéal pour réaliser une immuno-précipitation afin d'identifier de nouveaux partenaires spécifiques à *Leishmania*. Chez *T. brucei* TbOrc1b_3HA a été capable d'immuno-précipiter Tb Orc1/Cdc6 et TbMcm3 (Dang & Li, 2011).

CLUSTAL O(1.2.2) multiple sequence alignment

```

                                NLS
LmjF.28.0030  -----MKRSRRATQQEDVMRALKEGVQALSVSSALSTKDLVCRGDHARAIQQF
Tb11.02.5110b -----MKRRRDSNGKSI AALRAGVAALSVSSNIASRELTCDRDSHVKAILD F
TcCLB.508239.10 MNSRVILVFGENC MKRRRADETKKI AALRAGVAALSVSSTA AAKELTCRDAHLSAIVEF
                                ** . : *** : * * ***** : : : * . * * * * : *
                                Walker A
LmjF.28.0030  LEDEKHHTMQIF SMPGTGKT ATVNFALAQLV--SRRGTKPTAVFLNGFVVQKSSDIYYTL
Tb11.02.5110b LNDKVHPVMQVF SMPGTGKT ASVNHALALLASSPAGSKPTAVFLNGYIIQKTSDIYWTL
TcCLB.508239.10 LNDSVHPVMQIF SMPGTGKT ASVNHALTLLAQSAPPGRKPTAVFLNGYVIQKNSDIYWTL
*: * . * . * : * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                Walker B
LmjF.28.0030  HHHLTKARLGAVEPCPVAQCASRIEKRFRRHGWWGKPPALC VIVVDI VDKILEKHSKGFFK
Tb11.02.5110b NSHLSKTRLKHAENCLPEQCPALIEKRFRKQGWGSSTTPLC VIVVDI VDKVLKKNKAFFR
TcCLB.508239.10 NSHLSKARLGHTECLPDQCAALIEKRFRQGWGGASTPLC VIVVDI VDKVLKRHNKAFFR
. * * : * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                Arg finger
LmjF.28.0030  VVDWLTLPYANCKLITISNSMELQLDAKTK SRL GVVNQLVFSSYGTQELREILVHRVGAI
Tb11.02.5110b IVDWLSLPYAFCKLITISNSMELAADAKTR SRL DITKRLVFEPYSLPELKEIILRRVSHI
TcCLB.508239.10 IVDWLSFPFAFCKLVTISNSMELAADAKTR SRL DITKRLVFEPYSFSELKEIILRRVSGKI
: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
LmjF.28.0030  EPKLFADQAVNQLCTQTASHYGDVRRLLQSASAAICGLVLMRMRDDAVDVSSTDGIITLRE
Tb11.02.5110b KPTLFAEKAINYLCNQ TASHYGDVRRLLQSASSAICGLMMRIEEGYKLPKHDGLLTVKD
TcCLB.508239.10 KPTLFAEKAINYLCNQ TASHYGDVRRLLQSASSAVCGLMMKLEEGYRVPEAQDVLTVKD
: * . * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
LmjF.28.0030  IHTVVRQIFHDRFVEFIT TMRMPALFITVAVLGKETEALIRKREVDRIPLERLLAITQQ
Tb11.02.5110b VHSVVRQIFHDRFVEFIQTIRLPVVFISVAVIAVETARLFRANCDERSRLPIDSLFTATKR
TcCLB.508239.10 VHAVVRQIFHDRFVEFIQTIRLPILFISVAVVAHETNKLFRSNPGDCRLSLDGLFLSTKR
: * . * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
LmjF.28.0030  AQRKCMNVL R-----PINRAAFMEEIELLRQVSLIEVCIGEDRIPIRCVDELLDAKEEVF
Tb11.02.5110b AQERFGSVFADLHAVTLNYGAYLEIVEMLREVALIDVSVGEERIPVKTVQSLEATERAH
TcCLB.508239.10 AQQIYSASLGE PHTVDLTYGAYLDIVEMLREVALIDVSI GEERIPVKTTQNLEATEKAY
** . : . * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
LmjF.28.0030  VLLQPYQM VVDS CRLHDAFGATYGAALHL
Tb11.02.5110b ASMLQPFQT VVDACKLHDDFGTGICPLFSI
TcCLB.508239.10 VSMLQFPFPTVLDACQLHDVFGEGINPLFKT
. * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

Figure 5-8 . Clustal omega alignments of *L. major*, *T. brucei* and *T. cruzi* Orc1/Cdc6 proteins sequences.

In grey NLS, in red the Walker A and B and in blue Arginine finger motifs

```

LmjFORC1b      MPRSAVTATPRARRAGGGSGSSPTQRPSSRRSGKADTHASVPDAVFQVETLLEEAPHAGRE
TcCLB.508119.150 -----MEKAPLLERE
Tb927.9.2030 -----MMENTPLLGRE
                  Walker A
                  :*: * **

LmjFORC1b      AEYESVVAFVSRRLSHEHTNTSLFVCGCGCGKSTVKRALRAISARVLPDCAKARSLIKR
TcCLB.508119.150 QEYNAVHAFIEERLRMSHCRSLFVFGACGGKSTILRAMRCVSHRPMACDKVVGRTSAG
Tb927.9.2030 QEYNDVLRFIEERLLIQHCKSLFVFGACGGKSTIIRAMRRVTHRVMTYDKAAGRSSTG
                  **: * *:.** . **** *..*****: **: * : * .

LmjFORC1b      STLEECGWQTCVEDGDDAASVTADATCRTRKALLTPTSTPRSSRQLSAEGRSPSDANG
TcCLB.508119.150 GR-GSCGS-----GRSKSHS-----SHDNSPSPA--
Tb927.9.2030 SN-SSNGG-----GGLS-----DSPPPAAG
                  . * . ** *

LmjFORC1b      LTPSSSASLAAALSHTTSPDTSRTSASSLAAATATRSARLALHSPESSR-PSDSSVNSTA
TcCLB.508119.150 -----IVKRKPKVKKRRRGENESEDVAVD
Tb927.9.2030 -----RKQOK-----QMERQRP TKRQRGSETE-----
                  : : * . : . : .

LmjFORC1b      TLEEAASPACGVKRGRAADEWEENSIRSAELDRQATAVLTREGAWRQAQGLEQEVPRH
TcCLB.508119.150 TLE-----GAANGRISVGAR-----DASPEGDFVSSSS-----
Tb927.9.2030 -----GNAAGAAAVSA-----DNVKDGHFRSIGN-----
                  * * : * : .

LmjFORC1b      QPEPYSLAYFSMRYPELFSDDGTLATASSSPATSQFRQIFGHYVNCADVSGPLLVEAVC
TcCLB.508119.150 ---P-----LPAATLRL-----EADRRIQCHYVNCADMTSQQLCDAIA
Tb927.9.2030 ---AATPGDPTVLFPHLF-----GKGRRVQCHYVNCADMTAQQLCDSIS
                  . *: * : *****:.. * : : .

LmjFORC1b      DSIRATCSRTDGATQLLQLWLASIGKATSSRAQAIGGAAES-----HAASTSA
TcCLB.508119.150 SCFRVAQPRLDASSRNLLSEIEALPRHVTAYKRR--GVAAEAARKLLLRDSAVEHVTEPT
Tb927.9.2030 ETFCSSQVRLDAQTRTLVNAVGGIPKRVSAMRRR--SRELLKH----YDQQPIEGLPTA
                  . : : * Walker B . : : . : . : .
                  Walker B
                  . : : . : . : .

LmjFORC1b      ASKRASPPALHVALDFVEYLRGGGAKMLTLLAALAFHHPAQLALIFISNQRAVHVHPQ
TcCLB.508119.150 ALKKCVLKVPLHVLALDEVEYVRAGARGLLTALAALSVEHASQLALVIFISNQRELHVHPY
Tb927.9.2030 GLTKTTVKTPLHVLVLDVEVEYVRAAARGALASLAELSAEYASQLALVIFISNQREFVHVHP
                  . : . * : .*****:.. * : * * : . : *****:***** :****

LmjFORC1b      LMLELLLFQPYSEAQLREIGASATNAELQRCEQLVGAKVRKTAARSSAAGHKRLRASDVD
TcCLB.508119.150 MLLKELPFEAYSAQTLERICARIVAN-----VDFTEKERMVQVQ
Tb927.9.2030 MLLKELPFEAYSAEMLESIGNHIVAA-----AHFTRKEKEQVQ
                  :*: * * : * * * . . . . :*:

LmjFORC1b      IKPRLYDYIARKALLEFSGDVRQVIAMCHRVSVAWREVAEAKLEAAAAGAATAATSTPP
TcCLB.508119.150 LSPALIAYIAKKAVHEYSGDVRQVAMSRRVLVHTAQOEARDEAIK---RGDVRNDETPO-
Tb927.9.2030 LSPGLIRYIARKALVEFSGDVRQVAVMCRRLIHTAVEEQKKSVP---KGRVSTEGGET-
                  . * * * : * : ***** * : * : . * * . * .

LmjFORC1b      AMARSDSSSTAVVRNTSVADFSSKLEAAVEAYRVDKCAAATTQVSTTFSGTSQSSSDAAG
TcCLB.508119.150 -----RRSGS-----G
Tb927.9.2030 -----P-RGS-----R
                  * :

LmjFORC1b      IKDTSSTRPSGAAAASCASITAAAQLAPVPSSWTPSSGVMTLAKSVRVLQSNQV-ESDID
TcCLB.508119.150 I-----TRAGAVPGNTSPASPAAVMTLAQSQKVLRGTAVVSDRVF
Tb927.9.2030 E-----LAGATAPGGNSPGVPAALITLAQSQKILRGPDSVGDVRF
                  . * * * : : : * : * : . : . :

LmjFORC1b      KHVGTLPQTLVYLSCIVVLTLRKQEERTYASTSVGGRIGTRTPGAG-----RTA
TcCLB.508119.150 EYVAGMPEQMLYVLSCLVVLALRQQRDKSLAQSNVAGVSTTRRV-----RTDA
Tb927.9.2030 EYVSAMPEQMLYVLSCLVVLMLRQQRDSQIAKVISIGTSAVGRRONKTAIGSSFGSYSAA
                  : * . : * * * * * * * * * * * : * . . . . * *

LmjFORC1b      AQRAPTLSLKMHEVHQLYTRLMAKLHFPAMRADGFVQVECLADFGVLTQRLRGTDRV
TcCLB.508119.150 AAGKTEPGTFAMREVHRMYTALMGQRFPMNAAGIASAIDSFADLGIISRPOQRNEEL
Tb927.9.2030 EPGPVVPGTFTTRVVHRLYTALMGQRFPMNAAGISSAIDGFADIGIISRPORRGNEEV
                  . : : * : * * * * : * * * * * : : * : * : * : * : * : * :

LmjFORC1b      FSLNGTWTLPAMQAALVKRGEAIKQDRTTAGAGAMMENRFEEVLRANLVK-
TcCLB.508119.150 FSNFGTWSLESMSQALTRGELLRQERIEFGLDG-SENRFSEVLRKLIAGL
Tb927.9.2030 FSNFGTWTLESMSQAALTARGEALRQERVDCGLDS-AENRFEEVLRKLGILSL
                  **: * : * : * * * * * * * * * * * * * * * * :

```

Figure 5-9 Clustal omega alignment of *L. major*, *T. brucei* and *T. cruzi* Orc1b proteins sequences.

Highlighted in yellow extra amino-acids present in *Leishmania* but absent in *T. brucei* and *T. cruzi*. In red the Walker A and B conserved domains

5.2.2.3 Etude fonctionnelle des protéines ORCs par KO.

Les KO inductibles de LinfOrc1b et LinfOrc1/Cdc6 ont été réalisés chez *L. infantum* à l'aide du système inductible T7TR CRISPR/Cas9 KO bien que nous étions conscients que le système comportait des fuites (voir 4.2.3.1 CRISPR/Cas9 inductible grâce au système T7TR). Deux semaines après la transfection, nous avons effectué un suivi de croissance des lignées Linf_T7TR_Cas9, Linf_T7TR_Cas9_pL7Orc1b et Linf_T7TR_Cas9_pL7Orc1/Cdc6 en présence et en absence de tétracycline (Figure 5-10). Les courbes de croissance confirment la fuite du système inductible. En effet, nous n'observons pas de différence de croissance entre les lignées Linf_T7TR_Cas9_pL7 induites et non induites. En revanche nous observons une altération de la croissance de la lignée Linf_T7TR_Cas9_pL7Orc1/Cdc6 et une légère diminution de la croissance de Linf_T7TR_Cas9_pL7Orc1b par rapport à la lignée Linf_T7TR_Cas9.

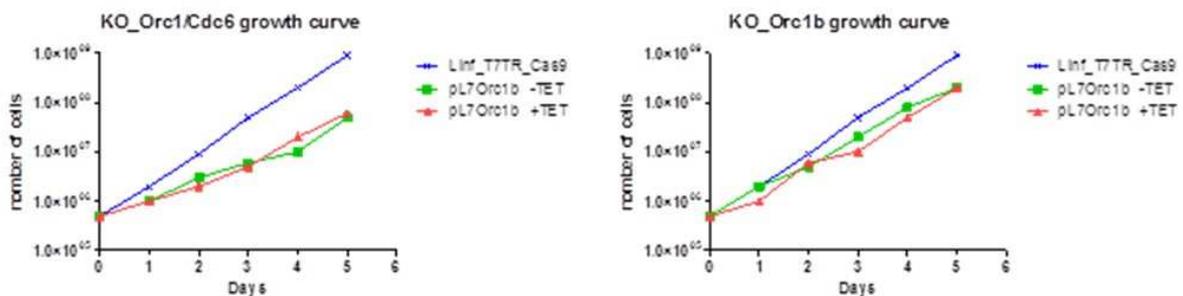


Figure 5-10 Growth defect after an induced KO of Orc1/Cdc6 and Orc1b in *L. infantum*.

Growth curves of Linf_T7TR_Cas9, Linf_T7TR_Cas9_pL7Orc1b and Linf_T7TR_Cas9_pL7Orc1/Cdc6 in presence or absence of tetracycline. The depletion of either ORC1b or ORC1 CDC6 leads to a slight growth reduction in the parasites.

Nous avons réalisé des PCR pour confirmer la présence des cellules éditées par CRISPR/Cas9. Le couple d'amorces p1-p2 permet d'amplifier le gène LinfORC1b (2621pb) et le couple p3-p4 permet de faire la distinction entre le locus sauvage (4013pb) et le locus modifié après intégration de la cassette de résistance (2557pb)

(Figure 5-11A). Nous avons extrait de l'ADN génomique de chaque souche pour effectuer des PCR afin de confirmer si nous avons ou non un remplacement du gène cible par la cassette de résistance. Les PCR de contrôle sur l'ADNg de LinfT7TR_Cas9 confirment la présence du gène Orc1b (p1_p2) et du locus sauvage (p3_p4) (Figure 5-11B left panel). Sur cette même figure, en présence (right panel) ou absence (middle panel) de tétracycline nous observons le même profil. Premièrement, nous avons mis en évidence la présence du gène Orc1b avec le couple d'amorces p1-p2 et celle du locus sauvage avec les amorces p3-p4. Deuxièmement nous avons obtenu une bande à 2557 pb avec les amorces p3-p4 ; cette bande correspond au locus édité. Ces résultats de PCR montrent que nous avons une population hétérogène de parasites édités et non édités. Ces résultats montrent également la présence d'une fuite dans le système d'induction. Ceci explique pourquoi nous ne voyons pas de différence sur les courbes de croissance entre les cellules « induites » et « non induites ». Concernant la construction ciblant Orc1/Cdc6 nous n'avons pas obtenu de PCR satisfaisantes, il a donc été décidé de redessiner de nouvelles amorces. Nous avons également tenté de cloner ces souches (Linf_T7TR_Cas9_pL7Orc1b et Linf_T7TR_Cas9_pL7Orc1/Cdc6) par la technique de dilution limite, mais sans succès.

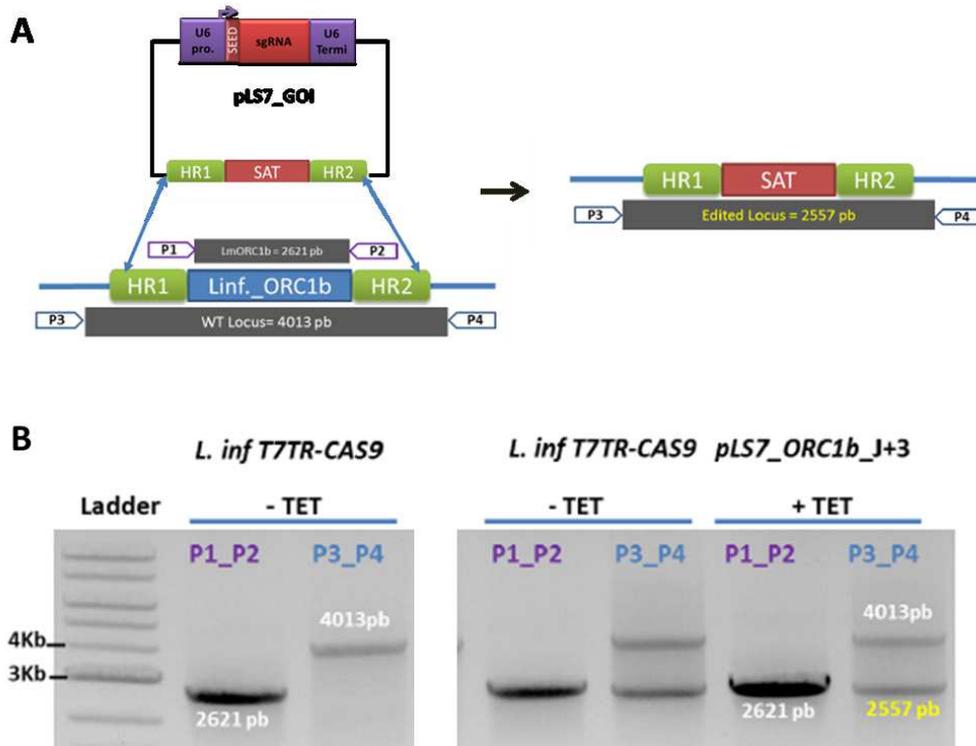


Figure 5-11 Genome edition of ORC1b was verified by PCR.

(A) Schematic representation of genome edition of LinfOrc1b locus with CRISPR/Cas9. (B) Two sets of primers were used: p1_p2 to amplify LinfORC1b (2621pb) and p3_p4 specific for the whole locus (either 4013 pb or 2557pb after edition). With p3-p4 we were able to detect the 2557pb band which testifies the resistance cassette integration and indeed the absence of LinfORC1b.

Pour comprendre ce qui se passe au niveau cellulaire dans ces souches KO, nous avons suivi la progression du cycle cellulaire en évaluant le nombre et la morphologie des noyaux (N) et des kinétoplastes (K). Cette méthode nous permet de classer individuellement chaque cellule en fonction de son stade dans le cycle cellulaire. Pour ce faire, les cellules ont été observées au microscope à fluorescence après fixation au PFA³⁷ 4% et coloration des noyaux par le DAPI. Nous avons observé que, contrairement à *L. major*, *L. infantum* (MHOM/BR/00/1669) duplique son kinétoplaste en même temps que le noyau. La Figure 5-12 résume les différents types de cellules retrouvées pour chaque condition après 5 jours d'induction à la tétracycline. Pour la condition témoin LinfT7TR_Cas9, (les parasites possèdent uniquement Cas9 et ne possèdent pas le plasmide pLS7), nous observons une majorité de cellules 1N1K en phase G1 (70%),

³⁷ paraformaldehyde

environ 20% des cellules en phase S (1eK1eN)³⁸ et des cellules 2N2K en mitose (8%). Nous verrons que ces pourcentages sont différents dans les souches « induites » et « non induites » après les KO (Linf_T7TR_Cas9_pL7Orc1b et Linf_T7TR_Cas9_pL7Orc1/Cdc6).

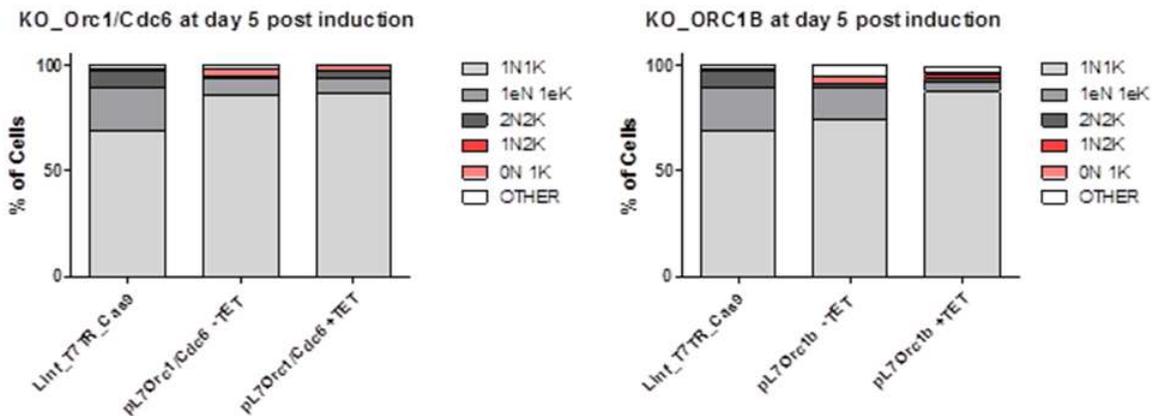


Figure 5-12 Effect of the depletion of LinfOrc1/Cdc6 and LinfOrc1b in *L. infantum* promastigote.

Quantification of the proportion of different cell types after KO induction by CRISPR/Cas9 based on the number of nuclei (N) and kinetoplasts (K) detected in individual cells stained with DAPI. Three types of population were tested: Linf_T7TR_Cas9 (control, without tetracycline (tet)), Linf_T7TR_Cas9_pLS7Orc1/Cdc6 and Linf_T7TR_Cas9_pLS7Orc1b (with (+) and without (-) tet).

Cet équilibre se trouve modifié pour les souches KO Linf_T7TR_Cas9_pL7Orc1b et Linf_T7TR_Cas9_pL7Orc1/Cdc6. Nous observons une augmentation du pourcentage des cellules 1N1K associée à une diminution de la proportion des cellules en répliation et en mitose. Enfin une proportion non négligeable de cellules anormales est retrouvée : 2% de cellules 1N2K et 4% de zoïdes 0N1K (Figure 5-13).

³⁸ e= elongated, noyau et kinétoplaste allongés.

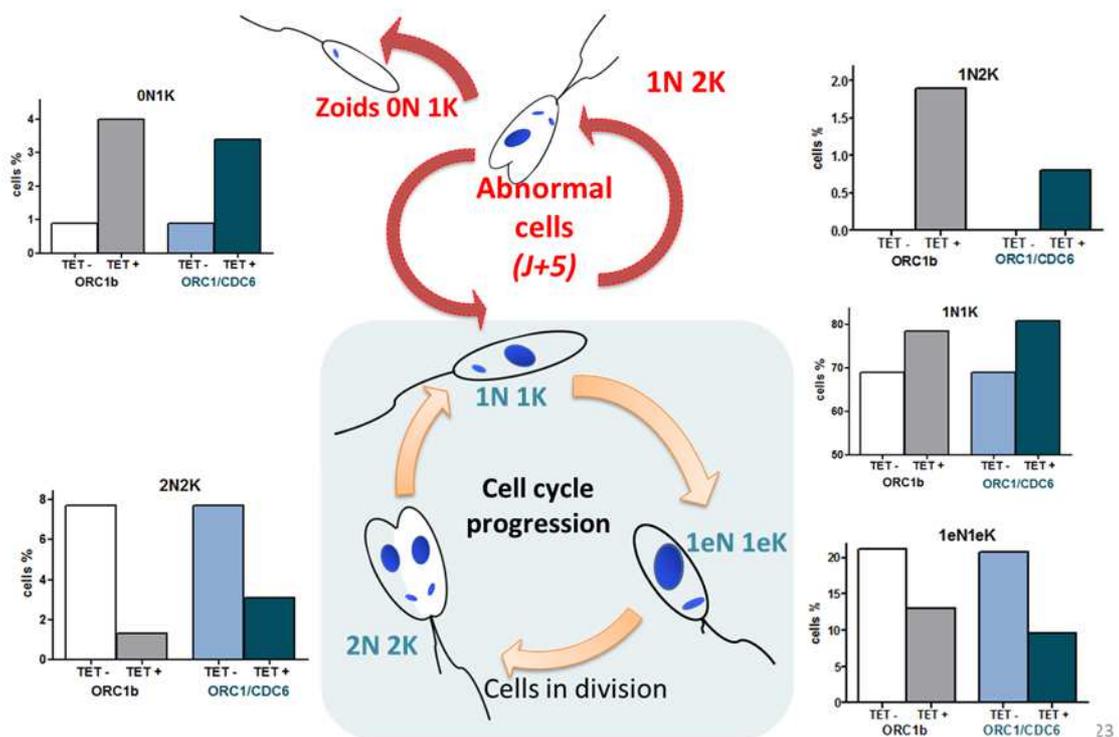


Figure 5-13 Functional study of *LinfOrc1/Cdc6* and *LinfOrc1b* inducible KO using the CRISPR/Cas9 system.

The endonuclease Cas9 is under a T7 promoter, inducible with the addition of tetracycline. The depletion of either *ORC1b* or *ORC1-CDC6* leads to a growth reduction in the cell line. The normal cell cycle, *L. infantum* (blue) shows that the parasite begins to divide its kinetoplast almost simultaneously with the nucleus. This leads to the doubling of the population every 8 hours. After the depletion of *ORC1b* or *ORC1/CDC6* we observed the appearance of abnormal cells (red arrows) : Zoids (cells without nucleus) and 2K1N cells associated with a decrease in the dividing cells : 1K2N and 2N2K. The histograms present the percentage of cell types in each step of the cell cycle in the *Linf_T7TR* – tet and *LinfOrc1b* and *LinfOrc1/Cdc6* + tet strains. This observation strongly suggests a defect in DNA replication in the absence of *ORC1b* and *ORC1cdc6*.

Ces résultats encore préliminaires nécessitent d'être répétés et confirmés, cependant ils tendent à montrer qu'en absence d'*Orc1b* et *Orc1/Cdc6* les cellules restent bloquées en G1 et ne rentrent pas en phase S.

5.2.3 Conclusion

Dans ce travail préliminaire nous avons obtenu des informations fondamentales qui vont nous permettre d'orienter nos recherches concernant le rôle que pourraient avoir *Orc1b* et *Orc1/Cdc6* chez *Leishmania*. Nous avons confirmé la localisation de ces deux

protéines au noyau et mis en évidence que le type de « tagging » utilisé pouvait affecter une correcte localisation. D'autre part les KO inductibles par CRISPR/Cas9 ont montré que l'absence de LmOrc1/Cdc6 et LmOrc1 provoque un défaut de réplication au vu de l'apparition de cellules anormales telle que les zoïdes. De plus, Orc1b pourrait avoir une fonction différente ou du moins de ne pas avoir les mêmes partenaires chez *Leishmania* et chez *T. brucei*. Nous avons désormais développé un outil de génétique réverse chez *Leishmania* qui va nous permettre de comprendre et identifier de nouveaux partenaires impliqués dans la réplication de l'ADN. En effet l'obtention laborieuse de la souche Linf_Orc1_3HA va désormais nous ouvrir plus de possibilités pour comprendre l'organisation et la régulation du complexe pré-réplicatif chez *Leishmania*. On sait désormais que le « tag » 3HA ne nuit pas à la localisation de la protéine, on pense ainsi réaliser le « tagging » d'une souche de *L. infantum* JPCM5 (non T7TR) pour poursuivre notre étude.

5.3 Identification des origines de réplication chez *Leishmania*

5.3.1 Introduction

Les études menées chez les eucaryotes sur l'initiation de la réplication ont montré une différence entre le nombre potentiel d'origines de réplication et celles qui sont actives. Les origines de réplifications sont définies par un complexe pré-réplicatif constitué des protéines ORC. Ces origines sont réparties en cluster sur les chromosomes et, seul un nombre limité d'entre elles seront activées et pourront s'associer au complexe hélicase (Cayrou *et al.*, 2015).

Un des moyens utilisé pour identifier les sites de liaison des protéines ORC sur la chromatine est la méthode de ChIP-seq³⁹. Cette technique ne permet cependant pas de distinguer les origines actives des non actives. Un des procédés utilisé pour caractériser les origines actives est l'isolement et le séquençage du brin naissant (Cayrou *et al.*, 2012). Grâce à cette méthode il est possible de connaître avec précision la position des origines actives dans une population de cellules non synchronisées.

³⁹ Chromatine Immuno Precipitation -sequencing.

A ce jour, nous avons très peu d'information sur l'organisation du complexe pré-réplicatif et sur la position et l'utilisation des origines de réplication chez *Leishmania*. Au laboratoire nous avons mis en place la technique du ChIP-seq et le séquençage du brin naissant (SNS-seq) pour étudier l'organisation et l'activation des origines de réplication.

5.3.2 ChIP-seq

Pour mettre au point cette technique du ChiP nous nous sommes inspirés du travail de (Tiengwe *et al.*, 2012) dans lequel ils identifient, chez *T. brucei*, les sites de liaison de TbOrc1/Cdc6 sur la chromatine par ChIP-on-Chip. Les auteurs utilisent une souche de *T. brucei* exprimant une protéine TbOrc1/Cdc6 fusionnée à une étiquette 12xc-Myc au locus, l'autre allèle ayant été inactivé par un classique knock-out. Cette protéine chimère leur a permis d'immuno-précipiter spécifiquement à l'aide d'un anticorps anti-Myc la chromatine à laquelle elle est associée.

Pour réaliser ce travail chez *Leishmania*, nous avons dans un premier temps identifié des protéines candidates (LmOrc1/Cdc6 et LmOrc1b) capables de faire partie du complexe pré-réplicatif chez *Leishmania major* (voir 5.2 *Identification et caractérisation des protéines Orc1/Cdc6 et Orc1b impliquées dans la réplication de l'ADN.*).

Au moment de ce travail nous ne possédions pas encore l'outil du CRISPR/Cas9 pour réaliser le « tagging » au locus de LmOrc1b et LmOrc1/Cdc6. Ces dernières ont donc été fusionnées à une étiquette GFP ou 12xc-Myc en position N- ou C-terminal. Ces constructions ont été exprimées chez *L. major Friedllin* (LmjF3171) de façon épisomale à l'aide du plasmide pTH6.

La localisation nucléaire de LmOrc1/Cdc6 et de LmOrc1b, associée au fait que ces protéines soient les orthologues de *T. brucei*, suggérait que nous avions deux bons candidats pour réaliser l'immuno-précipitation de la chromatine (ChIP).

Le ChIP_seq se décompose en plusieurs étapes :

- (i) l'étape de crosslink au formaldéhyde qui va fixer les interactions protéines-protéines associées à l'ADN ;
- (ii) l'étape de fragmentation de la chromatine ;
- (iii) l'immuno-précipitation (IP) de la chromatine à l'aide d'un anticorps spécifique ;
- (iv) la réalisation d'une librairie d'ADN pour le séquençage à haut débit Illumina et,
- (v) l'alignement des séquences obtenues avec le génome de référence.

Les détails de ce protocole après optimisation sont décrits dans l'annexe II.

Tout le travail relatif à la préparation des librairies (Diagenode) et au séquençage (Illumina) a été réalisé en collaboration avec le laboratoire BIHP d'Artur Scherf à l'Institut Pasteur. A ce jour nous avons réalisé l'immuno-précipitation (IP) d'Orc1/Cdc6_GFPNter et Cter, d'Orc1b_GFP et d'ORC1b_12xCmyc.

Seules les ChIP réalisées avec Orc1b_GFP et ORC1b_12xCmyc ont été séquencées et l'analyse bioinformatique n'a pas montré une différence d'enrichissement de séquence entre la chromatine avant IP (input) et la chromatine après IP (ChIP) (Figure 5-14). Nous avons choisi de réaliser en parallèle le ChIP avec un anticorps anti-histone acétylé comme contrôle interne du bon déroulement du ChIP (non montré).

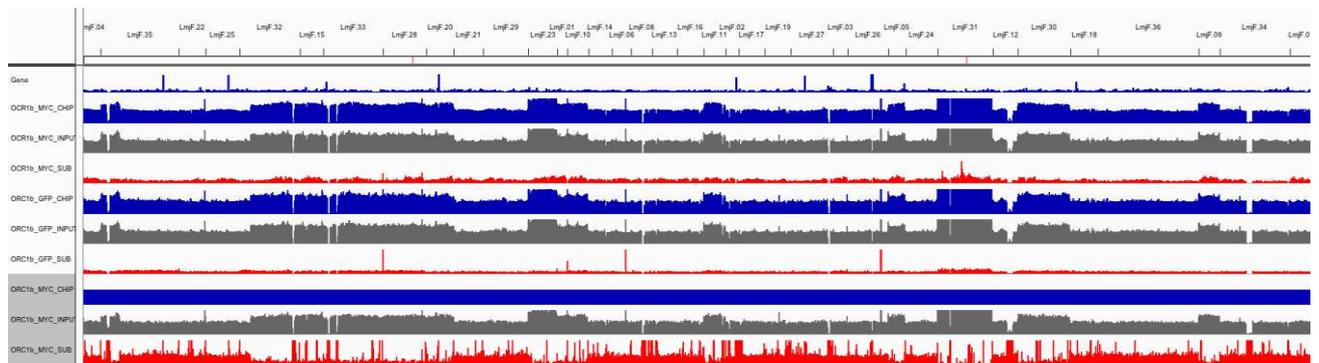


Figure 5-14 Results of ChIP-Seq experiment.

Whole genome coverage of ChIP-seq result for LmORC1b_12xCmyc (top tracks) and LmOrc1b_GFP (bottom tracks). The reads obtained after sequencing are aligned on the whole L.mjfF3171 reference genome. The ChIPs are the blue tracks, the inputs are the grey tracks and the difference between the two is the red tracks. The ChIP samples are very similar to the inputs for OCR1b_12xC-MYC and ORC1b_GFP.

Cette expérience n'ayant été réalisée qu'une fois, il est difficile de conclure. Cependant plusieurs hypothèses peuvent expliquer ce résultat : (i) nous avons réalisé l'IP avec des parasites qui sur-exprimaient notre protéine d'intérêt à partir d'un épisome. Il est possible que cette protéine produite en grande quantité, non physiologique, se lie de manière non spécifique à l'ADN ce qui expliquerait qu'on n'obtienne pas d'enrichissement de séquence spécifique après IP, surtout que la production de protéine endogène non marquée persiste. (ii) Il est possible que LmOrc1b n'ait pas un lien direct dans la reconnaissance des origines de réplication c'est-à-dire que cette protéine ne

serait pas en contact direct avec l'ADN mais en interaction fugace avec un complexe protéique.

Néanmoins, il faudrait revenir sur les étapes d'optimisation de lavages et d'éluion de l'IP pour s'affranchir des fragments de chromatine liés non-spécifiquement aux anticorps et ainsi augmenter la spécificité de notre IP. Maintenant que nous possédons l'outil CRISPR/Cas9 nous avons réalisé le « tagging » au locus de *L.infORC1b* avec 3xHA qui devrait nous permettre de réaliser un CHIP plus spécifique.

5.3.3 Séquençage du brin naissant

L'isolement du brin naissant (SNS⁴⁰) est une technique bien établie pour cartographier les origines de réplication (Cayrou *et al.*, 2012). Cette technique est basée sur le fait qu'il est possible de différencier le brin naissant précoce du reste de l'ADN génomique grâce à la présence d'une amorce ARN en 5'. Cette amorce, synthétisée par une primase au niveau de chaque origine de réplication active, permet l'amorçage de l'ADN polymérase. Un gradient de sucrose permettra de séparer les SNS des fragments d'Okazaki en fonction de leur taille (Figure 5-15).

Il s'agit d'une technique très performante qui permet de détecter des événements rares. L'abondance des SNS issus d'une origine de réplication donnée reflète l'efficacité avec laquelle cette origine est activée au sein d'une population cellulaire. Cependant il est difficile de séparer les SNS du reste de l'ADN génomique, d'où un rendement faible qui est un inconvénient pour le séquençage.

Nous avons purifié les SNS à partir d'une culture de *L. major* et de *T. brucei* en phase exponentielle de croissance, et nous avons utilisé en parallèle une culture en phase stationnaire comme contrôle. Le protocole est décrit dans l'Annexe III.

⁴⁰ De l'anglais small nascent strand

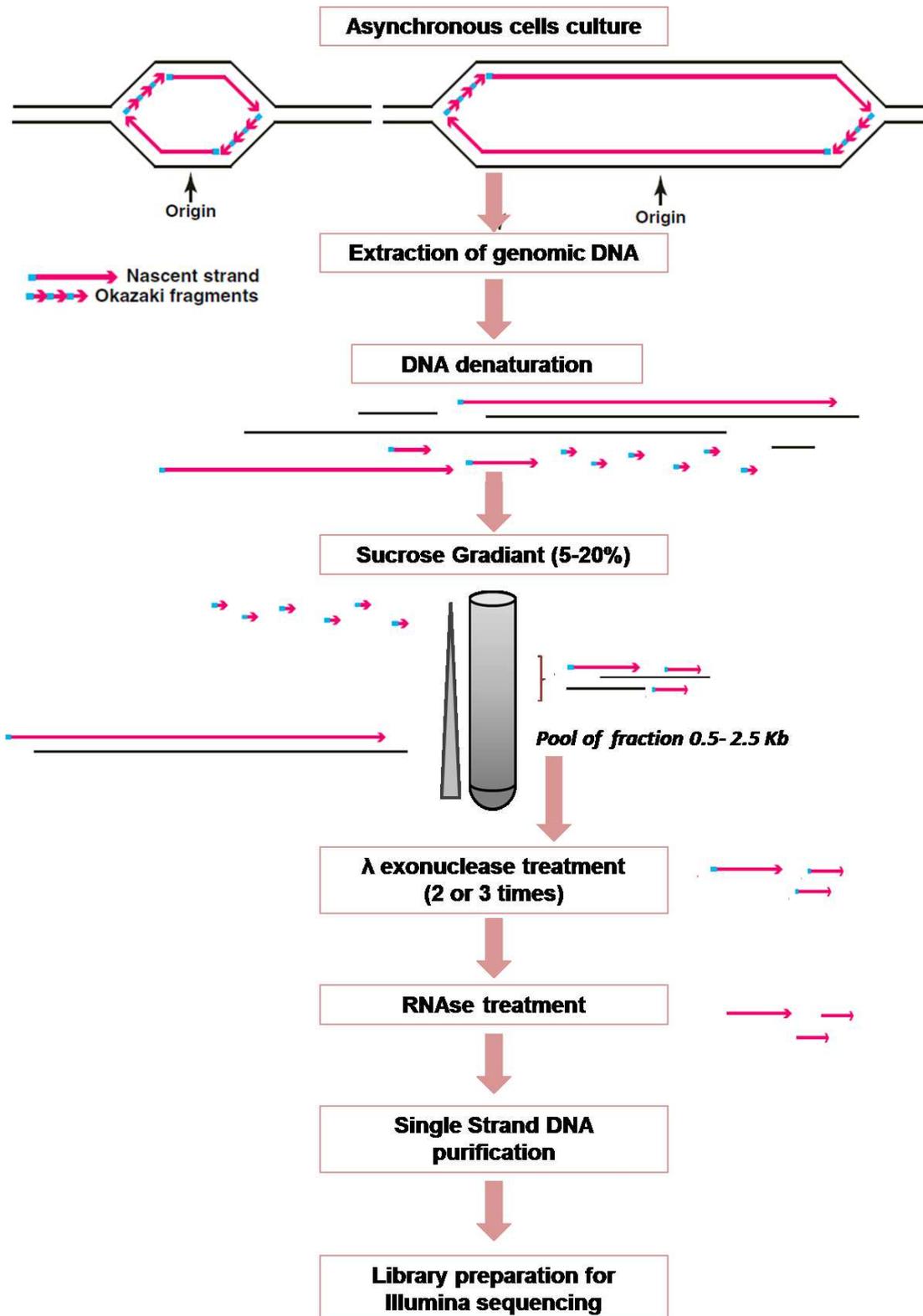


Figure 5-15 Purification procedure of small nascent DNA strands.

Small nascent strand mapping technique is based on the purification and analysis of small nascent single-stranded DNA fragments that are specifically found in small replication bubbles. Neosynthesized DNA (pink) present in replication bubbles is extruded upon denaturation of genomic DNA, and fractionated by size on a sucrose gradient to purify SNSs from large DNA fragments and Okazaki fragments (<350 nt). The RNA primed (blue) nascent strands are protected from the λ-exonuclease treatment

which reduces contamination by random broken DNA fragments of similar size. A last treatment with RNase is required to get rid of the RNA primers.

L'isolement du brin naissant a été réalisé à partir d'une extraction d'ADN génomique de cultures de parasites, puis dénaturé et déposé sur un gradient de sucrose (Annexe III). L'ultra-centrifugation a permis de séparer les fragments d'ADN selon leur taille et de sélectionner les fractions contenant des fragments d'ADN d'une taille de 500 à 2500 bases correspondant au SNS (Figure 5-16). A noter que les fragments de taille inférieure correspondent aux fragments d'Okazaki.

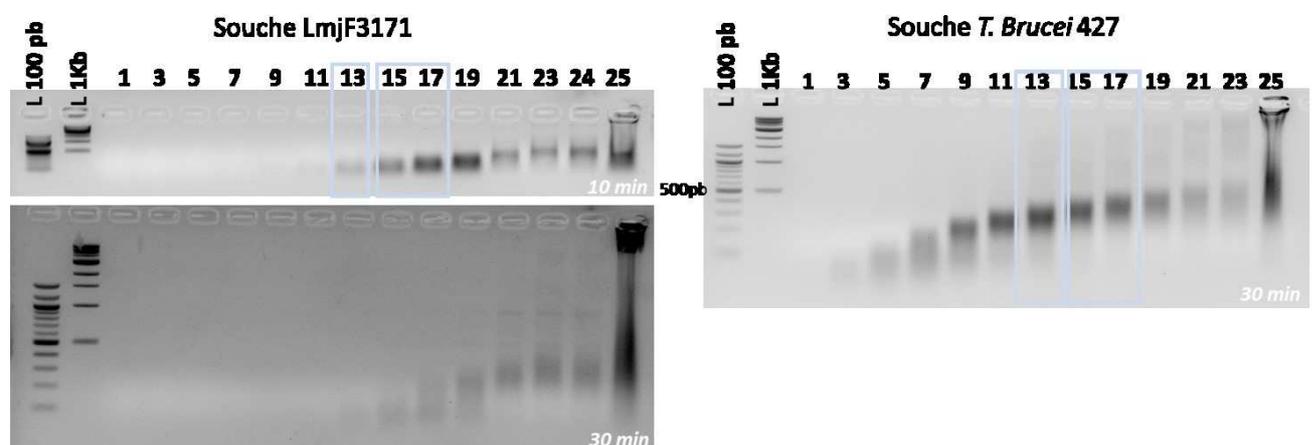


Figure 5-16 Small Nascent Strand isolation after ultracentrifugation.

All sucrose gradient fractions are run on a 1% agarose gel after ultra-centrifugation. DNA from LmjF3171 is on the left and DNA from *T. brucei* is on the right. Blue boxes correspond to the selected fractions for SNS isolation (13-14 and 15-18).

La seconde étape a consisté à traiter les fractions sélectionnées avec l'exonucléase λ (λ exo) qui a la particularité de dégrader l'ADN de 5' vers 3'. Cette étape visait à dégrader les fragments d'ADN génomique contenus dans ces fractions. Le SNS est protégé de cette activité enzymatique par l'amorce ARN situé à son extrémité 5'. Jusqu'à trois cycles de traitement λ exo pouvaient être nécessaires pour éliminer les fragments d'ADN génomique contaminants. Enfin, après un traitement final avec la RNase qui éliminait la séquence ARN nous avons obtenu une molécule d'ADN simple brin correspondant au brin naissant.

Nous avons traité en parallèle un échantillon de LmjF3171 et *T. brucei* que nous avons utilisé comme témoin positif. En effet, la connaissance de la position de certaines origines de réplication chez *T. brucei* était un bon moyen de contrôler le déroulement de

la purification du brin naissant. A chaque cycle de traitement λ exo, nous avons réalisé des qRT-PCR pour vérifier que nous avons bien un enrichissement de matériel au niveau des origines de réplication décrites chez *T. brucei*. Les amorces spécifiques utilisées se trouvent dans des régions du génome contenant ou non des origines de réplication (Tiengwe *et al.*, 2012).

Pour séquencer ces fragments nous avons constitué des librairies orientées et non orientées (Figure 5-17). Les librairies non-orientées ou librairies classiques ont été réalisées en deux temps : d'abord une fragmentation chimique des SNS pour obtenir des molécules d'une taille homogène d'environ 350 pb, suivie de la synthèse du brin complémentaire à l'aide du kit Invitrogen WGA2 ; puis la ligature des adaptateurs et « index » nécessaires à l'identification des fragments après le séquençage à haut débit Illumina en suivant les indications du kit Diagenode MicroPlex v2.

En parallèle nous avons réalisé des librairies orientées ou “stranded” (Figure 5-18) avec le kit Accel-NGS 1S Plus DNA Library Kit de chez Swift Biosciences, en suivant le protocole du fournisseur. Nous avons utilisé le sonicateur bioruptor pico de Diagenode (9 cycles, 5sec ON- 60sec OFF) pour fragmenter les SNS. Ce kit nous permet de confectionner des librairies à partir de seulement 10pg d'ADN. L'avantage de telles librairies est qu'elles permettent de garder l'orientation des brins naissants en ligant une « queue », en plus de l'adaptateur, en position 5' de chaque fragment d'ADN. Après le séquençage, cette petite « queue » va nous permettre de distinguer de quel brin matrice est issu le fragment de SNS. Grâce à ce procédé nous pourrions définir avec précision la position des origines de réplication et éventuellement des clusters.

Cette idée de réaliser des librairies orientées est née d'une discussion avec Cédric Notredame, bio-informaticien CNRS du CRG de Barcelone, qui soupçonne que nous pourrions avoir des difficultés à interpréter les données de brin naissant en utilisant des librairies non orientées. En particulier, lorsque les origines de réplication activées sont proches les unes des autres, nous serons dans l'incapacité de définir précisément leurs positions comme l'explique la Figure 5-17. :

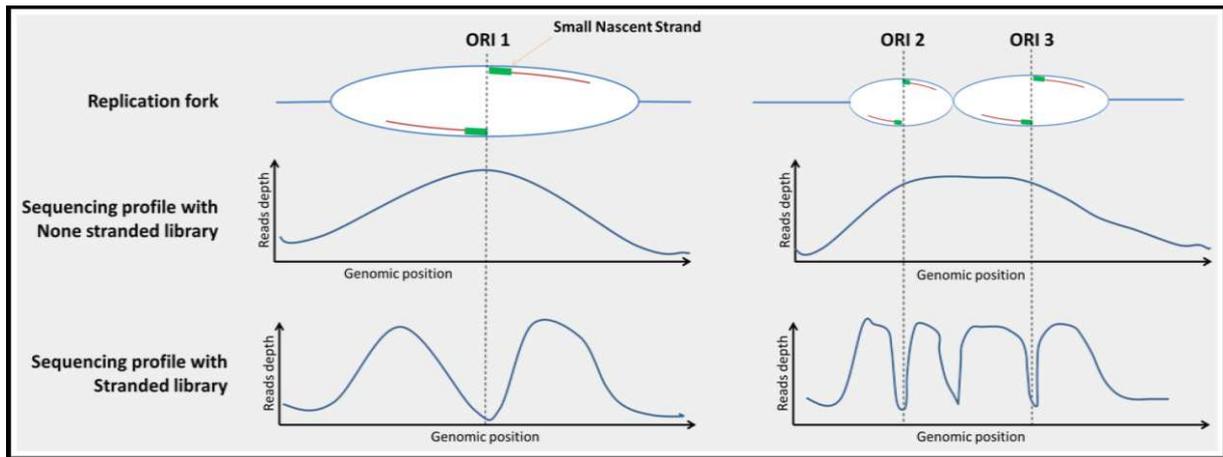


Figure 5-17. Two types of library for WGS.

In the top panel three firing origins (ORI 1-3) and their small nascent strands are represented. The middle panel corresponded to a schematic sequencing profiles obtained with a non-stranded library and the bottom panel shows profiles obtained with a stranded library.

Nous sommes désormais prêts à séquencer les échantillons. La taille et la qualité des librairies sont contrôlées sur Bioanalyser (High Sensitivity DNA Kit) de chez Agilent. Le multiplexage des échantillons a été réalisé pour que tous les échantillons puissent être séquencés sur la même « flowcell⁴¹ » sur le séquenceur Illumina NextSeq500 avec 300 cycles de séquençages à chaque extrémité dit « pair end ».

⁴¹ Cellule à flux continu, nom anglais donné à la puce sur laquelle sont déposés les échantillons à séquencer

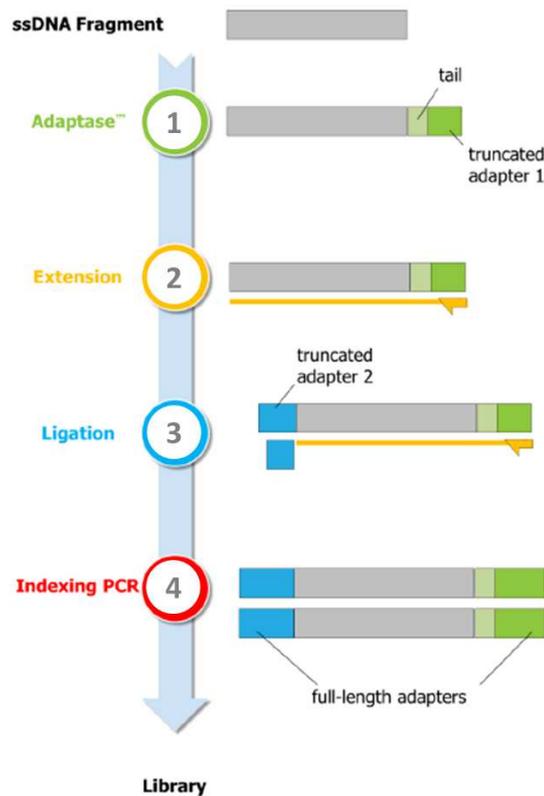


Figure 5-18 Preparation of stranded library with Accel-NGS 1S Plus DNA Library Kit from Swift Biosciences.

Step 1: The Adaptase step is a reaction that simultaneously performs end repair, tailing of 3' ends, and ligation of the first truncated adapter to 3' ends. Step2: The Extension step is used to facilitate ligation of the second truncated adapter. The synthesized strand does not get sequenced. And a bead clean-up is used to remove oligonucleotides and small fragments, and to change enzymatic buffer composition. Step 3: The Ligation reaction is used to add the second truncated adapter to the 5' ends. Step 4: The PCR step is used to increase yield and add the indexed adapter sequence.

5.4 Discussion générale sur la réplication du parasite *Leishmania*

Durant ma thèse, nous avons d'une part étudié la dynamique de réplication des *Leishmania* en la comparant à celles de *T. brucei* et des cellules MEF ; et d'autre part mis en place les techniques de ChIP-seq et de séquençage du brin naissant afin de réaliser une cartographie fine des origines de réplication. Cependant, nous devons replacer notre travail dans le contexte bibliographique actuel. En effet, les données publiées récemment proposent une cartographie des origines de réplication chez *Leishmania* après utilisation de MFA-seq (Tiengwe *et al.*, 2012, Marques *et al.*, 2015), et du séquençage de brins naissants (SNS-seq) (Lombrana *et al.*, 2016). A noter que Lombrana *et al.* ont réanalysé les données du MFA-seq de Marques *et al.* Les données obtenues par

ces deux équipes se sont finalement révélées contradictoires. Elles ne permettent pas de dégager une image claire de la réplication chez *Leishmania*. Une analyse critique de ces deux publications a été réalisée à la lumière de nos propres travaux.

Dans un souci de clarté, un tableau qui résume les principaux résultats de nos travaux en peignage moléculaire est présenté ci-dessous (Tableau 5-3).

Tableau 5-3 Résumé des principaux résultats de peignage moléculaire.

	<i>L. donovani</i>	<i>L. major</i>	<i>L. mexicana</i>	<i>T. brucei</i>	MEF
Genome size	32.4Mb	32.8 Mb	32 Mb	26 Mb	2.8.10 ³ Mb
Mean speed (kb/min)	2.45	2.44	2.59	1.87	1.16
Mean IOD (kb)	188.5	194.8	226	159.9	135.8
Total number of ORIs	172	168	142	136	20.6 .10 ³
Unidirectional forks (%)	1.7	1.5	1.5	14.4	4
Asynchronous forks (%)	76	75	76	66	59

La technique de MFA-seq, peu connue, a été utilisée pour l'étude d'organismes comme la Levure (Muller *et al.*, 2014) et les *Archeae* (Pelve et al., 2012). Son principe s'appuie sur la comparaison des profondeurs de « reads⁴² » obtenues entre une population de cellule en phase répliative (phase S) et, la même population en phase non répliative soit G1 ou G2. L'enrichissement en séquences est représenté par un score qui varie entre 1 et 2 et correspond au rapport entre les deux conditions (ex. cellules en phase S par rapport aux cellules en phase G2). La valeur 1 représente les régions non répliquées et la valeur 2 correspond aux régions répliquées du génome. Le positionnement de ces scores sur le génome donne la position des origines de réplication. La hauteur des pics de MFA-seq est également un indicateur de la fréquence d'activation et d'utilisation des origines de réplication dans le génome à l'échelle de la population. Un pic élevé est caractéristique d'une activation fréquente de l'origine. Cette technique peut être également associée à un tri par FACS des cellules pour sélectionner les populations d'intérêt (Muller *et al.*, 2014).

⁴² De l'anglais lecture, terme donné aux séquences

La technique du séquençage des brins naissants utilisée par Lombrana et al. est proche de celle que nous souhaitons appliquer et que j'ai décrite ci-dessus (Figure 5-15). Les deux principales différences avec notre travail sont que ces auteurs n'ont pas fragmenté l'ADN avant de réaliser les librairies et que les librairies réalisées ne sont pas orientées, alors que nous avons préparé des librairies non orientées et orientées.

5.4.1 Analyse critique des données de MFA-seq chez *Leishmania*

L'équipe de McCulloch a utilisé la technique du MFA-seq à la fois chez *T. brucei* (Tiengwe et al., 2012, Devlin et al., 2016) et chez *Leishmania* (Tiengwe et al., 2012, Marques et al., 2015, Devlin et al., 2016). Dans le travail initial chez *T. brucei*, cette approche a été utilisée en complément du ChiP-on-chip qui a permis d'identifier 953 sites de liaison à l'ADN de TbOrc1/Cdc6_12myc, dont 51% sont situés dans les VSG. Ils n'ont pas pu identifier de motif sur l'ADN caractérisant les sites de liaisons d'Orc1/Cdc6 ou encore au niveau des origines actives révélées par le MFA-seq. Néanmoins, ce travail (Tiengwe et al., 2012) montre que *T. brucei* possède 45 origines de réplication actives précoces réparties sur les chromosomes à des positions non canoniques. Une seconde étude sur l'implication de la réplication dans l'expression des VSG, chez *T. brucei* (Devlin et al., 2016) a permis de détecter, toujours en MFA, cinq origines de réplication supplémentaires par rapport à 47 le nombre d'origines actives en première partie de phase S. Par contre, la MFA n'a pas permis de révéler de nouvelles origines en phase S tardive. Il semble que la technique de MFA n'a pas des performances suffisantes pour détecter les origines qui sont utilisées à plus faible fréquence que d'autre utilisées majoritairement. Il faut également noter que les deux formes du parasite (BSF⁴³, chez l'Homme et PCF⁴⁴, chez le vecteur) utiliseraient les mêmes origines de réplication d'après Devlin et al., 2016.

Pendant le processus de révision de l'article (Stanojcic et al., 2016), l'équipe de McCulloch a publié un travail qui ne retrouve qu'une seule origine de réplication par chromosome chez *Leishmania*, avec la technique du MFA qu'il avait précédemment utilisé chez *T. brucei*. Ce résultat, n'est pas cohérent avec les données de peignage

⁴³ Blood stream form

⁴⁴ Procyclic form

moléculaire, ni avec la taille des grands chromosomes, la vitesse de réplication et la durée de la phase S. L'origine unique est utilisée chez les procaryotes qui possèdent de petits génomes circulaires. Ce petit nombre d'origines ne semble pas approprié pour un eucaryote possédant des chromosomes entre 0,3 et 3 Mb.

D'ailleurs, bien que les auteurs annoncent dans le titre de leur article la présence d'une seule origine par chromosome chez *Leishmania* : « *Genome-wide mapping reveals single-origin chromosome replication in Leishmania, an eukaryotic microbe* », ils admettent dans la discussion que ce n'est pas possible ⁴⁵! Il apparaît donc que la technique du MFA qui avait pourtant donné de bons résultats chez d'autres organismes (Masai *et al.*, 2010) y compris chez *T. brucei* (Tiengwe *et al.*, 2012) semble peu performante chez *Leishmania*.

Par ailleurs, chez *Leishmania*, si l'on considère que la technique du MFA manque de sensibilité et non de spécificité et qu'elle a permis de détecter certaines origines, on peut constater que 30 des 36 pics de MFA localisent avec les marques H3Ac qui correspondent aux sites d'initiation de la transcription (Thomas *et al.*, 2009), et que seulement 21% des SSR (*strand switch region*) contiennent une origine (dont 5 origines dans des SSR convergentes). Il était intéressant de constater que la position des pics obtenus en MFA est conservée entre les deux espèces *L. major* et *L. mexicana*.

Marques *et al.* suggèrent que la position des origines de réplication serait conservée chez les Kinetoplastida. Ceci serait dû aux unités de transcription, qui imposeraient une contrainte pour le déplacement des machineries de réplication et de transcription. Ils supposent aussi que les pics trouvés en MFA chez *Leishmania* pourraient correspondre à la position des centromères. En effet, cette zone du chromosome est la première répliquée, et des complexes protéiques impliqués dans la ségrégation des chromosomes viennent s'y fixer, provoquant un obstacle sur la fourche de réplication. Gêné dans sa progression, le complexe réplicatif serait contraint de se dissocier et de se réassocier en fonction de ces contraintes. Aujourd'hui nous pouvons confirmer cette hypothèse, puisque nous avons pu identifier les centromères grâce aux protéines des kinétochores. La position des centromères chez *Leishmania* et la majorité des pics trouvés en MFA-seq

⁴⁵ « *Nonetheless, if the replication rate predictions are accurate, ~50 % of Leishmania chromosomes might complete replication from one origin, but the rest could not. How, then, can Leishmania replicate their entire genome?* » (Marques *et al.*, 2015)

correspondent effectivement aux centromères de *Leishmania*. (Garcia-Silva *et al* submitted).

Deux autres critiques peuvent être faites sur l'analyse des données de l'article de Marques *et al.* 2016. La première porte sur l'analyse bio-informatique, qui a été réalisée en utilisant une fenêtre de 2500 bp; or nous verrons ci-dessous, qu'une analyse réalisée avec une fenêtre de 1000 bp donne des résultats très différents.

La dernière critique concerne la synchronisation des cellules. Les auteurs prétendent travailler sur des populations de cellules synchronisées après les avoir triées par FACS. Or du fait de l'aneuploïdie mosaïque, le contenu en ADN des leishmanies ne correspond pas strictement à 1C et 2C, certaines cellules ont un contenu intermédiaire en ADN alors qu'elles ne sont pas en division mais toujours en interphase. On ne peut donc prétendre ni trier les cellules en FACS, ni distinguer la phase S précoce de la phase S tardive. D'ailleurs dans leur étude, les auteurs n'observent pas de différence entre les cellules triées et non triées, ni entre « phase S précoce » et « phase S tardive ». Nous proposons de coupler le tri en FACS à un marquage EdU⁴⁶ des parasites pour différencier les cellules en réplication des cellules en phase non répliative. A partir de ce tri, la détection des origines de réplication MFA-seq pourrait être moins biaisée.

5.4.2 Analyse critique des données de séquençage du brin naissant

Récemment Lombrana *et al.*, 2016 se sont intéressés à l'organisation des nucléosomes (*chromatin compaction*), et aux origines de réplication chez *Leishmania* en utilisant du MN-seq et du SNS-seq respectivement (Lombrana *et al.*, 2016). L'isolement et le séquençage du brin naissant est une technique largement utilisée pour identifier toutes les origines de réplication des cellules eucaryotes d'une population (Cayrou *et al.*, 2012). Cette technique a permis de révéler 5100 origines réparties sur tous les chromosomes. Ce nombre est particulièrement important, au regard de l'origine unique décrite dans Marques *et al.* (année) et au regard de nos résultats de peignage (Stanojic *et al.*, 2016). Si ce nombre d'origines s'avère exact chez *Leishmania*, cela signifierait que *Leishmania* à l'échelle de la cellule utiliserait en moyenne 3,39% de ses origines (168 origines par génome haploïde (Stanojic *et al.*, 2016)). Avec les techniques de séquençage à haut

⁴⁶ 5-ethynyl-2'-deoxyuridine

débit il est tout à fait normal d'obtenir un nombre supérieur d'origines par rapport à une technique comme le peignage moléculaire. Mais ce chiffre est peut être surestimé; en effet, en comparant le résultat obtenu pour les cellules MEF par la même technique (Cayrou *et al.*, 2011), les auteurs ont détecté 2231 potentielles origines par 60Mb (1115ori/30Mb). Avec la technique du peignage nous avons dénombré $20,6 \cdot 10^3$ origines actives pour les MEF, ainsi, les MEF utiliseraient en moyenne 20% de la totalité de leurs origines pour répliquer leur génome et *Leishmania* aurait donc 5 fois plus d'origines de réplication potentielles que les eucaryotes supérieurs. Cependant ce nombre important d'origines peut être un avantage pour le parasite et lui offre plus de flexibilité dans le choix des origines pour répliquer son génome. Les données de peignage montrant une forte proportion d'origines asynchrones (75%), et la large amplitude de vitesse de réplication et d'IOD⁴⁷s (Stanojic *et al.*, 2016) sont également des arguments en faveur d'une grande flexibilité de *Leishmania* dans la gestion de la réplication de son génome.

Lombrana *et al.* (2016) compare aussi la position des origines trouvées par SNS-seq aux données des différentes études déjà disponibles, telles que la position des Ach3, des Bases J, ou encore l'abondance de messagers. Bien que les positions d'enrichissement de séquences obtenues avec le SNS-seq coïncident avec les différentes marques : position des Ach3, Base J, abondance de messagers, prédiction de G-quadruplexes (G-4) qui sont des régions GC riches (Lombrana *et al.*, 2016), elles n'ont pas permis d'établir un lien direct entre la position de ces marques et la position des origines de réplication, ni de décrire un motif canonique qui permettrait de prédire la position des origines de réplication. Chez *Leishmania* 74% des origines contiendraient un motif GC-riche (Lombrana *et al.*, 2016). Ceci a également été observé chez des cellules de drosophile, de souris et humaines pour lesquelles Cayrou *et al.* ont pu identifier des motifs GC riches au niveau des origines de réplication (Cayrou *et al.*, 2011).

En revanche, les auteurs (Lombrana *et al.*, 2016) observent une corrélation entre la faible occupation des nucléosomes et un fort enrichissement en brin naissant. En effet, il est connu que la réplication de l'ADN est devancée par le complexe hélicase qui relaxe l'ADN et l'initiation serait ainsi favorisée dans des régions ouvertes de la chromatine (Cayrou *et al.*, 2015).

⁴⁷ Inter-origine distance

Enfin, Lombrana *et al.* (Lombrana *et al.*, 2016) ont ré-analysé les données de MFA-seq obtenues par Marques *et al.* (Marques *et al.*, 2015) en changeant les paramètres d'analyses. Ils utilisent une fenêtre de 1kb contre 2,5kb pour normaliser les données de MFA. Avec cette nouvelle analyse et un format de représentation différente des données de MFA-seq, nous suspectons que le MFA-seq (Figure 5-19 top and bottom panel) détecte plus d'origines que le SNS-seq (Figure 5-19 middle tracks). Il semblerait qu'en plus du biais du tri par FACS la méthode d'analyse bio-informatique soit, elle aussi, un biais d'analyse.

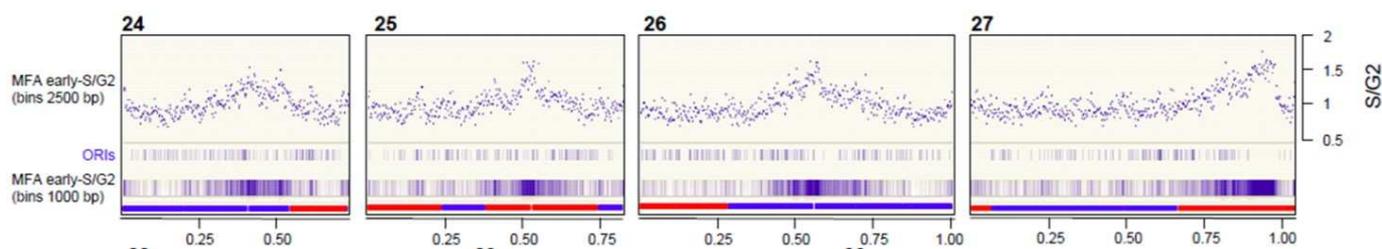


Figure 5-19 Higher ORI density at early replicating regions in *L. major* chromosomes.

DNA enrichments in early-S relative to G2 computed in 2,500 bp bins (MFA analysis, Marques *et al.*, 2015), middle tracks common ORI summits obtained by SNS-seq and bottom tracks heatmap of MFA early-S/G2 in sliding 1,000 bp bins, at 24/25/26/27 *L. major* chromosomes. Blue and red PTUs represent RNAPII transcription from the + or - strand, respectively. Chromosome sizes are denoted in intervals of 0.25 Mb.

5.4.3 Synthèse des données publiées

En résumé, il apparaît que l'on ne peut pas établir de règle pour définir la position des origines de réplication chez *Leishmania*. Les données de peignage associé au SNS-seq reflètent la grande flexibilité qu'a ce parasite pour contrôler la réplication de son génome.

L'analyse des données de ces trois articles ne nous permet pas d'avoir une image claire de la réplication chez *Leishmania*, et encore moins de comprendre les mécanismes qui contrôlent l'aneuploïdie mosaïque. Il faut compter sur le séquençage de nos échantillons de SNS stranded de *L. major* et *T. brucei* en parallèle pour nous aider à comprendre l'utilisation des origines de réplication chez *Leishmania*. Cette méthode va nous permettre d'apporter une meilleure spécificité et une meilleure définition des origines de réplication dans le cas où l'activation de ces dernières soit proche (Figure 5-17). De

plus, la cartographie des sites de liaison des protéines ORC à la chromatine en CHIP-seq et la confrontation de ces données à celles du séquençage de brins naissants sera très intéressante. Enfin ces techniques pourront généralement être très importantes dans l'étude des mutants de la réplication obtenus par la méthode du CRISPR/Cas9.

6 Conclusion & Perspectives

Au cours de ces trois années de thèse, nos travaux de peignage nous ont permis de poser une base solide pour comprendre la dynamique de réplication chez *Leishmania*. La compréhension de la dynamique de la réplication et l'identification des origines de réplication s'inscrivent dans l'optique d'élucider pourquoi et comment *Leishmania* gère l'aneuploïdie mosaïque qui est si souvent délétère dans d'autres organismes. Il faut rappeler que notre hypothèse suggère que cette aneuploïdie mosaïque résulte d'une réplication non conventionnelle de l'ADN.

A ce jour, la seule information qui ressort des différentes études chez *Leishmania* est la flexibilité de la réplication de l'ADN. La vitesse élevée de la réplication, la distance inter-origines variables, la possibilité d'initier la réplication à de nombreux endroits sur le génome et à des moments différents (origines asynchrones) sont des caractéristiques très différentes de celles retrouvées dans les autres organismes eucaryotes, et vont dans le sens de notre hypothèse. Ainsi le mode original de réplication de l'ADN chez *Leishmania* pourrait jouer un rôle dans l'émergence et le maintien de l'aneuploïdie mosaïque de *Leishmania*. L'aneuploïdie est une particularité génomique partagée par toute les espèces de *Leishmania* (Lachaud et al., 2013) qui, pour rappel, ne possèdent pas ou très peu de gènes spécifiques d'espèces bien que responsables d'un large éventail de maladies. Nous pensons donc que l'aneuploïdie mosaïque peut être considérée comme un facteur de virulence et être impliquée dans la pathogénicité du parasite.

Désormais, c'est tout le travail préliminaire effectué sur les protéines ORC qui va nous permettre de comprendre comment la réplication est régulée dans le contexte d'un organisme aneuploïde comme *Leishmania*. Par exemple Cdt1 paraît être un acteur essentiel dans le « licensing » des origines de réplication chez les eucaryotes supérieurs. Mais à ce jour aucun homologue n'a été retrouvé dans le génome des Trypanosomatidés, on peut donc se demander comment est effectué le contrôle de la réplication au cours de la phase S du cycle cellulaire.

Il est maintenant essentiel d'utiliser LmOrc1b et LmOrc1/Cdc6 pour identifier de nouveaux partenaires impliqués dans la réplication, et de s'intéresser à d'autres protéines telles que Cdc45 (qui présente 43% d'identité avec *T. brucei*) ou le complexe

hélicase MCM. Il sera ensuite important de comprendre comment ces partenaires interagissent entre eux et quel est leur rôle dans la réplication de l'ADN.

Le développement du CRISPR/CAS9, la mise en place de la technique du ChIP-seq et le séquençage du brin naissant sont, sans nul doute, des apports et des atouts considérables pour l'identification et la caractérisation des protéines impliquées dans la genèse de l'aneuploïdie mosaïque chez *Leishmania*.

7 Annexes

7.1 Annexe I :CRISPR-Cas9 for *Leishmania major*

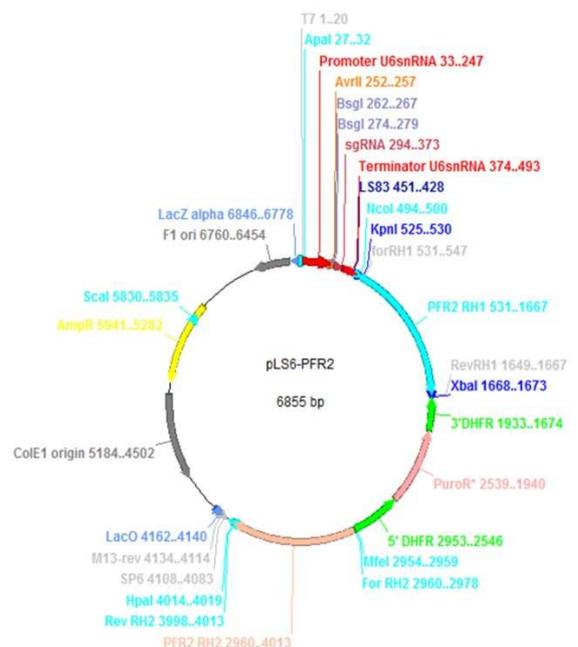
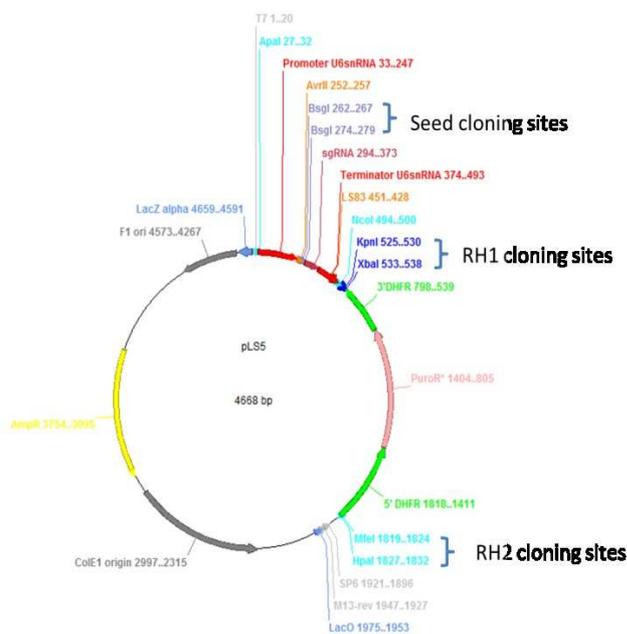
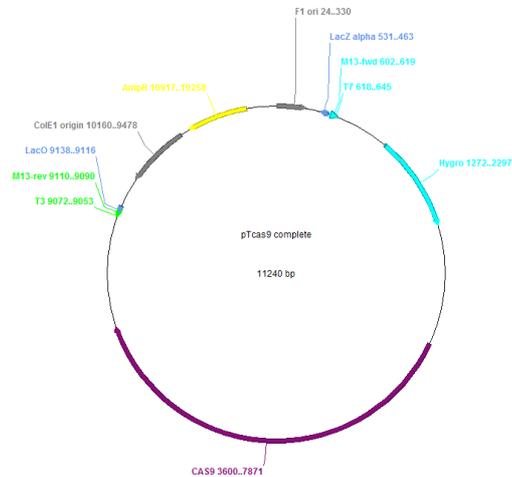
[First efficient CRISPR-Cas9-mediated genome editing in Leishmania parasites.](#) (Sollelis *et al* 2015)

Goal: How to prepare your plasmid for use in CRISPR/Cas9 approach?

What you should know:

This protocol is based on a 2 plasmids strategy:

- **pT-Cas9:** for the Cas9 nuclease expression, the same plasmid is used for all targets (you do not need to modify it)
- **pLS6-PFR2** (harboring homology regions of PFR2 but without the seed), so you have to adapt it to your GOI; 2 adaptations are needed:
 - i- replace PFR2 homology regions with your homology boxes (same as classical approach) to obtain pLS6-GOI
 - ii- add your sgRNA to pLS6-GOI to obtain pLS7-GOI



A- Replace PFR2 homology regions in pL6-PFR2 with your homology boxes (same as classical approach) to obtain pLS6-GOI

Steps:

Clone the homology regions (HRs) of your GOI between XbaI/KpnI (HR1) and HpaI/MfeI (HR2) following your favorite approach (digestion and either ligase-based either InFusion-based). Take care of your gene orientation, it has to be in "frame" with the resistance cassette.

**We have designed 500bp length HR. But we know that in Plasmodium falciparum we can drop to 250bp.*

For the cloning of the HRs we have used the InFusion Kit from Clontech.

Example of primers we designed with the **15nt to allow the infusion cloning**.

Primer name	Primer sequence	Application
LS75_for	ATTCGATTCTGAAGGGTACCC CATGCCACACAAGAAGGG	To clone homology region 1 PFR2 in pLS6-GOI
LS76_rev	GATAGGGAGGTCTAGA CATGCTTGTAGTGATGTTTTGG	
LS77_for	GTGCAAGCCTCAATTG GATGAAATCTGGGAACTCG	To clone homology region 2 PFR2 in pLS6-GOI
LS79_rev	CTGCAGCTGAAGGTTAACT CACACGGCCAGAGG	

A.1. Cloning homology region 1 of your GOI:

- i- Digest pLS6-PFR2 by XbaI/KpnI to remove HR2-PFR2 that is 1054pb long
- ii- Purify the cut DNA on PCR clean up kit Macherey (optional)
- iii- Amplify your HR1 by PCR
- iv- Do the InFusion reaction using the User Manual or [online tools](#).

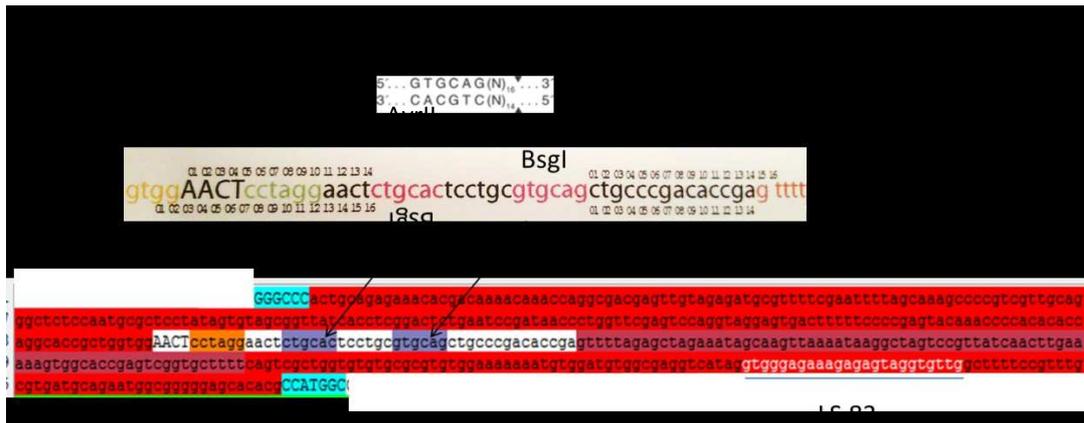
A.2. Cloning homology region 2 of your GOI:

- i- Digest pLS6-PFR2 by HpaI/MfeI to remove HR1-PFR2 1137pb
- ii- Purify the cut DNA on PCR clean up kit Macherey (optional)
- iii- Amplify HR2 by PCR
- iv- Do the InFusion reaction using the User Manual or [online tools](#).

B. Clone your seed in pLS6-GOI to obtain pLS7-GOI:

The seed is the 20nt which are a part of the sgRNA.

The pL6-X has a double BsgI cloning site to clone without scar (see figure, BsgI-recognition sites are highlighted in blue and its cut site is 14-16bp downstream) that can be replaced by the seed sequence you selected.



Steps for cloning of the seed to make functional the sgRNA:

B.1. You need 3 oligos:

=> 2 oligos for cloning the seed

=> 1 oligo to check colonies by PCR in combination with LS83

Again, you can do it either InFusion-based or ligase-based. Here is the design for InFusion-based cloning.

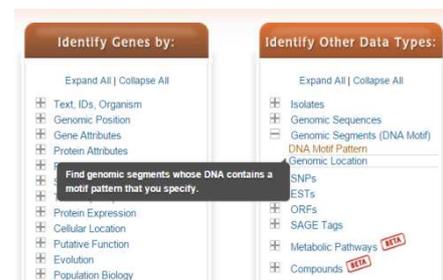
- Design of oligos for the cloning of the seed using InFusion system:

Oligo1: CAGGCACCGCTGGTGG -- (N)₂₀ ---GTTTTAGAGCTAGAAAT

CAGGCACCGCTGGTGG and GTTTTAGAGCTAGAAAT are necessary for InFusion-based cloning of the seed into the BsgI-cut plasmid.

- The (N)₂₀ correspond to your target site in respect with (N)₂₀ NGG

*We have observed that the efficiency of the editing is better if the seed is located near the homology region.



- The seed has to be unique. To avoid off-target you can check the specificity of your seed on TriTrypDB using [DNA Motif Pattern](#):

NNNNNNNNNNNNNNNNNNNNNNNGG: 20nt of your seed following by NGG or NAG

NNNNNNNNNNNNNNNAG: the last 12nt of your seed following by NGG or NAG

The last 12nt are the most important for the specificity of the CAS9. Be aware that CAS9 can also cut next to a NAG (less efficient).

Oligo2: reverse complement of Oligo1

Oligo3: In combination with LS83 (CAACACCTACTCTTTCTCCCAC), the 3rd oligo will be used for PCR/colonies to check cloning of the sgRNA, its sequence is the (N)₂₀, It needs to have a similar T_m, if necessary add some bases from the plasmid on the 5' to increase the T_m.

**The oligos for cloning are ordered category "cloning oligos" and oligo for the check is "unmodified oligo". (MWG)*

B.2. Annealing the seed oligos 1&2:

Way 1:

- Stock oligos are suspended to have 100µM
- Mix 10µl of each + 2.2µl NEB2 buffer
- use the annealing program (95 °C for 10 minutes; cooling from 95°C to 25°C using a -1.5°C/s ramp and 1min pause every 10°C and keeping at 4°C) and keep annealed oligos on ice all time.
- Dilute oligos to 0.5uM in cold 5mM/Tris pH8 (1/100)

Way 2: (best one)

- Stock oligos are suspended to have 100µM
- Mix 10 µl of each + 1X Taq buffer+ 0,5mM MgCl₂
- Incubate the tube 2 min in boil water. Then let it cooling at RT to 25°C in the water.
- Dilute oligos to 0.5uM in cold 5mM/Tris pH8 (1/100)

B.3. Digestion of the plasmid by BsgI (NEB):

- Ideally 10µg of pLS6-GOI is cut by BsgI restriction enzyme. 1µl enz/2µg/20µl/reaction + SAM, 2h then denature the enzyme according to the NEB recommendation.
- Purify (first purification) the cut DNA on PCR clean up kit Macherey
- Purify (second purification) from agarose gel (0.8%-0.5X-TBE) (optional, we don't do it anymore).

B.4. Seed InFusion cloning:

Prepare reaction on ice to avoid oligos des-annealing 10µl/reaction:

- 100ng of plasmid + 1.5ul from diluted annealed-oligos (0,5. uM)
- 2 µL of 5X infusion Buffer
- H₂O qsp 10µL

Reaction 15min/50°C.

Use 5µl or 10µl of reaction to transform 70µl competent cells.

=> Check colonies by PCR using Oligo3/LS83.

=> You can check minipreps by digestion using AvrII site (present in the BsgI-adapter).

Positive clones are PCR(+) and digestion(-)

Sequence the mini-preps of PCR-positives colonies using the gDNArev oligo LS83.

C. Transfection of parasites:

We had transfected in a strain which contains pT-CAS9 the pLS7-GOI linearized (paper) and circular, in both of the case the parasites were edited.

After several transfections the use of circular plasmids is more efficient.

Annexe II : Chromatin Immuno-precipitation protocol.

ChIP was performed as previously described in (Thomas *et al.*, 2009) with some modifications: For each assay, 3×10^9 *L. major* cells were incubated with 1% formaldehyde at 4°C for 5 min and stopped with 2.5 M glycine incubation for an additional 5 min. Chromatin purification and preparation of sheared DNA: spinned down cells were resuspended in 100uL of permeabilization buffer (100 mM KCl 10 mM Tris pH8.0 25 mM EDTA Protease inhibitors), followed by incubation with Digitonin 200uM final for 15 min RT. Next, we spinned down and resuspended the pellet in 100uL of cold NP-S Buffer (0.5mM Spermidine 0.075% NP-40 (IGEPAL CA-630) 50 mM NaCl 10 mM Tris-HCl pH 7.5 5mM MgCl₂ 1 mM CaCl₂). Pellet was washed 3 times with NP-S Buffer, resuspended in 40uL of NP-S Buffer and proceeded to MNase digestion for 5 min at 37°C. The reaction was stopped by adding 4ul of EDTA 0.5M pH8 (final 10mM) and centrifugated at 10000g by 10 min. Supernatant containing soluble mononucleosomes was collected in a new tube. Solubilized pellet with 100uL of NP-S, 0.2%SDS buffer was sonicated for 16 cycles of 30 sec on/30 sec off using a Bioruptor Pico sonication device (Diagenode). Sheared chromatine was centrifugated 10000g for 10 min and supernatant was added to saved soluble mononucleosomes. After checking chromatin shearing size we proceeded with immunoprecipitation assay following protocol from Lopez-Rubio *et al.*, with minor modifications (Lopez-Rubio *et al.*, 2013). Antibody anti-GFP ChIP grade from abcam (ab2090) was added to the chromatin sample in a 1/500 dilution and incubated overnight at 4°C with rotation. To collect immune complexes, we incubated the sample with 50ul per sample of protein G dyna-BEADS for 2 h at 4°C with rotation. Dyna-beads with chromatin complexes were washed for 5 min on a rotating platform with 1 ml of each of the buffers listed below: Low salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA (pH 8.0), 20 mM Tris-HCl (pH 8.1), and 150 mM NaCl) at 4°C. High salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA (pH 8.0), 20 mM Tris-HCl (pH 8.1), and 500 mM NaCl) at 4°C. LiCl immune complex wash buffer (0.25 M LiCl, 1% NP-40, 1% Deoxycholate, 1 mM EDTA (pH 8.0), and 10 mM Tris-HCl (pH 8.1)) at 4°C. TE buffer (10 mM Tris-HCl (pH 8.0) and 1 mM EDTA (pH 8.0)), two washes at RT. Immune complexes were eluted by adding 100 ul of fresh elution buffer (1% SDS and 0.1 M NaHCO₃) at RT for 15 min with rotation two times. ChIP and Input samples werw reverse cross-linked (65°C overnight), and the DNA was purified on a Qiagen QiaQuick[®] spin column and analyzed Bioanalyser (Agilent) to determine the average genomic fragment size achieved by sonication (300 bp).

7.2 Annexe III : Nascent Strands preparation

Use gloves, filter tips, Gibco water, Filter each solution and ice

I. Isolation Of Nascent Strands Of DNA

- ❖ Obtained 10^9 cells (Leish or Trypa) in exponential growth phase ($5 \cdot 10^6$ cells/mL)
- ❖ Wash 2 times in cold PBS (2500 rpm_10 min_4°C)
- ❖ gDNA extraction: phenol /chloroform
- ❖ Precipitate gDNA
- ❖ Sucrose gradient: (use Gibco water!!)

1. TEN20 buffer

10 mM Tris pH 7.5	→ 5 mL of 1M stock	} qsp 500 mL
1 mM EDTA pH 8	→ 1 mL of 0.5M stock	
20 mM NaCl	→ 10 mL of 1M stock	

2. 20% sucrose in TEN20 buffer

% sucrose final	FOR 1mL		FOR 4 mL	
	20% sucrose	TEN 20	20% sucrose	TEN 20
19%	950	50	3800	200
18%	900	100	3600	400
17%	850	150	3400	600
16%	800	200	3200	800
15%	750	250	3000	1000
14%	700	300	2800	1200
13%	650	350	2600	1400
12%	600	400	2400	1600
11%	550	450	2200	1800
10%	500	500	2000	2000
9%	450	550	1800	2200
8%	400	600	1600	2400
7%	350	650	1400	2600
6%	300	700	1200	2800
5%	250	750	1000	3000

- ❖ Start with 20%, Freeze between each layer of 300 μ L.
- ❖ Denaturated gDNA 10 min at 95°C, put back in ice
- ❖ Drop off the gDNA at the top of the gradient gently.
- ❖ Centrifugation 7h 45000 rpm swing55 4°C

- ❖ Collect 200µL fractions in new low binding tube, change tip for each fraction. Store at -20°C
- ❖ Save 20 µL of odd number fraction for migrated 10 µL in a 2% agarose Gel in TAE. Took a picture every 10 min. Use DNA ladder and keep in mind that 500 pb = 1000 bases ssDNA

II. Treatment of nascent strands (Repeat 2 times)

1. Pool Fraction

- ❖ Select fractions between 500 and 2000 bases
- ❖ 200 µl of pooled fractions of DNA

2. Précipitation:

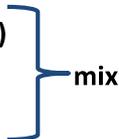
- ❖ 20 µl 3M Na acetate
- ❖ 1.2 ml cold EtOH (-20°C)
- ❖ Centrifugent 60 min /14000rpm /+4 °C
- ❖ Wash 70% cold EtOH (-20°C) cfg 10 min/14000rpm /+4 °C
- ❖ Dry samples
- ❖ **Add 35 µl of water**

3. Denaturation:

- ❖ Denaturation 95 °C/5 min
- ❖ Put on ice

4. T4 PNK treatment:

Add:

- ❖ 5 µl 10x T4 PNK buffer (buffer A)
 - ❖ 5 µl 1mM ATP
 - ❖ 2 µl RNase inhibitor
 - ❖ 3 µl T4 PNK (10u/µl)
 - ❖ Incubate at 37 °C for 1 hour
- 

5. 2nd precipitation

Add:

- ❖ 150 µl of water
- ❖ 20 µl 3M Na acetate
- ❖ 1.2 ml cold EtOH (-20°C)
- ❖ Cfg 60 min /14000rpm /+4 °C
- ❖ Wash 70% cold EtOH (-20°C) cfg 10 min/14000rpm /+4 °C
- ❖ Dry samples
- ❖ Add 40 µl of water

6. 2nd denaturation:

- ❖ Denaturation 95 °C/5 min
- ❖ Put on ice

7. λ-exo treatment:

Add:

- ❖ 2 µl RNase inhibitor
- Mix
- ❖ 5 µl 10x λ exonuclease buffer
- ❖ 3 µl 10x λ exonuclease (10u/µl)
- ❖ Incubate at 37°C/ON



Do not treat all the samples

8. RNAse treatment:

Inactivated RNAsine and λ exo by 95°C 5 min

Add:

- ❖ 1µl of RNAse
- ❖ 1H 37 °C

9. qPCR analysis

If necessary repeat a 3rd λ exo treatment.

8 Bibliographie

- Aladjem MI (2007) Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat Rev Genet* **8**: 588-600.
- Alonso G, Rastrojo A, Lopez-Perez S, Requena JM & Aguado B (2016) Resequencing and assembly of seven complex loci to improve the *Leishmania major* (Friedlin strain) reference genome. *Parasit Vectors* **9**: 74.
- Bailey H & Bishop WJ (1959) Leishman-Donovan bodies and donovaniasis; Sir William Boog Leishman, 1865-1926; Charles Donovan, 1863-1951. *Br J Vener Dis* **35**: 8-9.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA & Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.
- Bastien P, Blaineau C & Pages M (1992) *Leishmania*: sex, lies and karyotype. *Parasitol Today* **8**: 174-177.
- Bastien P, Blaineau C, Taminh M, Rioux JA, Roizes G & Pages M (1990) Interclonal variations in molecular karyotype in *Leishmania infantum* imply a 'mosaic' strain structure. *Mol Biochem Parasitol* **40**: 53-61.
- Bell SP & Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**: 128-134.
- Bell SP & Kaguni JM (2013) Helicase loading at chromosomal origins of replication. *Cold Spring Harb Perspect Biol* **5**.
- Bernards A, van Harten-Loosbroek N & Borst P (1984) Modification of telomeric DNA in *Trypanosoma brucei*; a role in antigenic variation? *Nucleic Acids Res* **12**: 4153-4170.
- Blaineau C, Bastien P, Rioux JA, Roizes G & Pages M (1991) Long-range restriction maps of size-variable homologous chromosomes in *Leishmania infantum*. *Mol Biochem Parasitol* **46**: 292-302.
- Borst P (2016) Maxi-circles, glycosomes, gene transposition, expression sites, transsplicing, transferrin receptors and base J. **205**: 39-52.
- Borst P & Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* **62**: 235-251.
- Britto C, Ravel C, Bastien P, Blaineau C, Pages M, Dedet JP & Wincker P (1998) Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene* **222**: 107-117.
- Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV & van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.
- Calvo-Alvarez E, Alvarez-Velilla R, Jimenez M, Molina R, Perez-Pertejo Y, Balana-Fouce R & Reguera RM (2014) First evidence of intraclonal genetic exchange in trypanosomatids using two *Leishmania infantum* fluorescent transgenic clones. *PLoS Negl Trop Dis* **8**: e3075.
- Cayrou C, Gregoire D, Coulombe P, Danis E & Mechali M (2012) Genome-scale identification of active DNA replication origins. *Methods* **57**: 158-164.
- Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J & Mechali M (2015) The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res* **25**: 1873-1885.
- Cayrou C, Coulombe P, Vigneron A, *et al.* (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res* **21**: 1438-1449.
- Chargui N, Amro A, Haouas N, *et al.* (2009) Population structure of Tunisian *Leishmania infantum* and evidence for the existence of hybrids and gene flow between genetically different populations. *Int J Parasitol* **39**: 801-811.

Chong JP, Hayashi MK, Simon MN, Xu RM & Stillman B (2000) A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proc Natl Acad Sci U S A* **97**: 1530-1535.

Cliffe LJ *et al.* (2016) Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription ... - PubMed - NCBI.

Cong L, Ran FA, Cox D, *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.

Crauwels P *et al.* (2015) Apoptotic-like Leishmania exploit the host's autophagy machinery to reduce T-cell-mediated parasite elimination. - PubMed - NCBI.

Cruz AK, Titus R & Beverley SM (1993) Plasticity in chromosome number and testing of essential genes in Leishmania by targeting. *Proc Natl Acad Sci U S A* **90**: 1599-1603.

Dame RT, Kalmykova OJ & Grainger DC (2011) Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in gram negative bacteria. *PLoS Genet* **7**: e1002123.

Dang HQ & Li Z (2011) The Cdc45.Mcm2-7.GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. *J Biol Chem* **286**: 32424-32435.

de Mendonca LZ, Resende LA, Lanna MF, *et al.* (2016) Multicomponent LBSap vaccine displays immunological and parasitological profiles similar to those of Leish-Tec(R) and Leishmune(R) vaccines against visceral leishmaniasis. *Parasit Vectors* **9**: 472.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J & Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607.

den Boer M, Argaw D, Jannin J & Alvar J (2011) Leishmaniasis impact and treatment access. *Clin Microbiol Infect* **17**: 1471-1477.

Devlin R, Marques CA, Paape D, Prohic M, Zurita-Leal AC, Campbell SJ, Lapsley C, Dickens N & McCulloch R (2016) Mapping replication dynamics in *Trypanosoma brucei* reveals a link with telomere transcription and antigenic variation. *Elife* **5**.

Dhar SK & Dutta A (2000) Identification and characterization of the human ORC6 homolog. *J Biol Chem* **275**: 34983-34988.

Diaz-Trivino S, del Mar Castellano M, de la Paz Sanchez M, Ramirez-Parra E, Desvoves B & Gutierrez C (2005) The genes encoding Arabidopsis ORC subunits are E2F targets and the two ORC1 genes are differently expressed in proliferating and endoreplicating cells. *Nucleic Acids Res* **33**: 5404-5414.

Dooijes D, Chaves I, Kieft R, Dirks-Mulder A, Martin W & Borst P (2000) Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. *Nucleic Acids Res* **28**: 3017-3021.

Downing T, Imamura H, Decuyper S, *et al.* (2011) Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res* **21**: 2143-2156.

Dubessay P *et al.* (2006) Cell cycle-dependent expression regulation by the proteasome pathway and characterization of the nuclear targeting signal of a *Leishmania major* Kin... - PubMed - NCBI.

Duncan SM, Myburgh E, Philipon C, Brown E, Meissner M, Brewer J & Mottram JC (2016) Conditional gene deletion with DiCre demonstrates an essential role for CRK3 in *Leishmania mexicana* cell cycle regulation. *Mol Microbiol* **100**: 931-944.

El-Sayed NM, Myler PJ, Bartholomeu DC, *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**: 409-415.

Eren RO, Reverte M, Rossi M, *et al.* (2016) Mammalian Innate Immune Response to a *Leishmania*-Resident RNA Virus Increases Macrophage Survival to Promote Parasite Persistence. *Cell Host Microbe* **20**: 318-328.

Erzberger JP & Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* **35**: 93-114.

Estevez AM & Simpson L (1999) Uridine insertion/deletion RNA editing in trypanosome mitochondria--a review. *Gene* **240**: 247-260.

Fiebig M, Kelly S & Gluenz E (2015) Comparative Life Cycle Transcriptomics Revises *Leishmania mexicana* Genome Annotation and Links a Chromosome Duplication with Parasitism of Vertebrates. *PLoS Pathog*, Vol. 11 p.^pp.

Fuller RS, Funnell BE & Kornberg A (1984) The *dnaA* protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* **38**: 889-900.

Gai D, Chang YP & Chen XS (2010) Origin DNA melting and unwinding in DNA replication. *Curr Opin Struct Biol* **20**: 756-762.

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH & Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Gasiunas G, Barrangou R, Horvath P & Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**: E2579-2586.

Genest PA, Baugh L, Taipale A, *et al.* (2015) Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Res* **43**: 2102-2115.

Geurts AM, Cost GJ, Freyvert Y, *et al.* (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**: 433.

Ghorbal M, Gorman M, Macpherson CR, Martins RM, Scherf A & Lopez-Rubio JJ (2014) Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat Biotechnol*.

Godoy PD, Nogueira-Junior LA, Paes LS, Cornejo A, Martins RM, Silber AM, Schenkman S & Elias MC (2009) Trypanosome prereplication machinery contains a single functional *orc1/cdc6* protein, which is typical of archaea. *Eukaryot Cell* **8**: 1592-1603.

Godoy PDdM, Nogueira-Junior LA, Paes LS, Cornejo A, Martins RM, Silber AM, Schenkman S & Elias MC (2009) Trypanosome Prereplication Machinery Contains a Single Functional *Orc1/Cdc6* Protein, Which Is Typical of Archaea.

Gommers-Ampt J, Lutgerink J & Borst P (1991) A novel DNA nucleotide in *Trypanosoma brucei* only present in the mammalian phase of the life-cycle. *Nucleic Acids Res* **19**: 1745-1751.

Gommers-Ampt JH, Teixeira AJ, van de Werken G, van Dijk WJ & Borst P (1993) The identification of hydroxymethyluracil in DNA of *Trypanosoma brucei*. *Nucleic Acids Res* **21**: 2039-2043.

Gossen M, Pak DT, Hansen SK, Acharya JK & Botchan MR (1995) A *Drosophila* homolog of the yeast origin recognition complex. *Science* **270**: 1674-1677.

Grabowski B & Kelman Z (2001) Autophosphorylation of archaeal *Cdc6* homologues is regulated by DNA. *J Bacteriol* **183**: 5459-5464.

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM & Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Handman E (2001) Leishmaniasis: current status of vaccine development. *Clin Microbiol Rev* **14**: 229-243.

Horn D (2014) Antigenic variation in African trypanosomes. *Mol Biochem Parasitol*, Vol. 195 p.^pp. 123-129.

Huberman JA, Spotila LD, Nawotka KA, el-Assouli SM & Davis LR (1987) The *in vivo* replication origin of the yeast 2 microns plasmid. *Cell* **51**: 473-481.

Iovannisci DM & Beverley SM (1989) Structural alterations of chromosome 2 in *Leishmania major* as evidence for diploidy, including spontaneous amplification of the mini-exon array. *Mol Biochem Parasitol* **34**: 177-188.

Ivens AC & Peacock CS & Wortley EA, *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**: 436-442.

Jacob F & Brenner S (1963) [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon]. *C R Hebd Seances Acad Sci* **256**: 298-300.

Jansen R, Embden JD, Gastra W & Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565-1575.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

Jullien N, Sampieri F, Enjalbert A & Herman JP (2003) Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res* **31**: e131.

Kaguni JM (2011) Replication initiation at the Escherichia coli chromosomal origin. *Curr Opin Chem Biol* **15**: 606-613.

Kamhawi S (2006) Phlebotomine sand flies and Leishmania parasites: friends or foes? *Trends Parasitol* **22**: 439-445.

Kamhawi S, Modi GB, Pimenta PF, Rowton E & Sacks DL (2000) The vectorial competence of Phlebotomus sergenti is specific for Leishmania tropica and is controlled by species-specific, lipophosphoglycan-mediated midgut attachment. *Parasitology* **121 (Pt 1)**: 25-33.

Kaye P & Scott P (2011) Leishmaniasis: complexity at the host-pathogen interface. *Nat Rev Microbiol* **9**: 604-615.

Kraeva N, Ishemgulova A, Lukes J & Yurchenko V (2014) Tetracycline-inducible gene expression system in Leishmania mexicana. *Mol Biochem Parasitol* **198**: 11-13.

Kumar D & Saha S (2012) A highly basic sequence at the N-terminal region is essential for targeting the DNA replication protein ORC1 to the nucleus in Leishmania donovani. *Microbiology* **158**: 1775-1782.

Kumar D, Mukherji A & Saha S (2008) Expression and subcellular localization of ORC1 in Leishmania major. *Biochem Biophys Res Commun* **375**: 74-79.

Lachaud L, Bourgeois N, Kuk N, Morelle C, Crobu L, Merlin G, Bastien P, Pages M & Sterkers Y (2013) Constitutive mosaic aneuploidy is a unique genetic feature widespread in the Leishmania genus. *Microbes Infect.*

Laffitte MC, Genoie MM, Mukherjee A, Legare D, Masson JY & Ouellette M (2014) Formation of linear amplicons with inverted duplications in Leishmania requires the MRE11 nuclease. *PLoS Genet* **10**: e1004805.

Lander ES (2016) The Heroes of CRISPR. *Cell* **164**: 18-28.

LeBowitz JH, Smith HQ, Rusche L & Beverley SM (1993) Coupling of poly(A) site selection and trans-splicing in Leishmania. *Genes Dev* **7**: 996-1007.

Leman AR & Noguchi E (2013) The Replication Fork: Understanding the Eukaryotic Replication Machinery and the Challenges to Genome Duplication. *Genes (Basel)*, Vol. 4 p. 1-32.

Leprohon P, Legare D, Raymond F, Madore E, Hardiman G, Corbeil J & Ouellette M (2009) Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant Leishmania infantum. *Nucleic Acids Res* **37**: 1387-1399.

Lohia A, Mukherjee C, Majumder S & Dastidar PG (2007) Genome re-duplication and irregular segregation occur during the cell cycle of Entamoeba histolytica. *Biosci Rep* **27**: 373-384.

Lombrana R, Alvarez A, Fernandez-Justel JM, Almeida R, Poza-Carrion C, Gomes F, Calzada A, Requena JM & Gomez M (2016) Transcriptionally Driven DNA Replication Program of the Human Parasite Leishmania major. *Cell Rep* **16**: 1774-1786.

Lopez-Rubio JJ, Siegel TN & Scherf A (2013) Genome-wide chromatin immunoprecipitation-sequencing in Plasmodium. *Methods Mol Biol* **923**: 321-333.

Lundgren M, Andersson A, Chen L, Nilsson P & Bernander R (2004) Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101**: 7046-7051.

Maine GT, Sinha P & Tye BK (1984) Mutants of S. cerevisiae defective in the maintenance of minichromosomes. *Genetics* **106**: 365-385.

Makarova KS, Wolf YI, Alkhnbashi OS, et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**: 722-736.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE & Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826.

Mancio-Silva L, Rojas-Meza AP, Vargas M, Scherf A & Hernandez-Rivas R (2008) Differential association of Orc1 and Sir2 proteins to telomeric domains in Plasmodium falciparum. *J Cell Sci* **121**: 2046-2053.

Marini JC, Miller KG & Englund PT (1980) Decatenation of kinetoplast DNA by topoisomerases. *J Biol Chem* **255**: 4976-4979.

Marques CA, Dickens NJ, Paape D, Campbell SJ & McCulloch R (2015) Genome-wide mapping reveals single-origin chromosome replication in Leishmania, a eukaryotic microbe. *Genome Biol* **16**: 230.

Marraffini LA & Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843-1845.

Martinez-Calvillo S, Stuart K & Myler PJ (2005) Ploidy changes associated with disruption of two adjacent genes on Leishmania major chromosome 1. *Int J Parasitol* **35**: 419-429.

Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K & Myler PJ (2003) Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* **11**: 1291-1299.

Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N & Oda M (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu Rev Biochem* **79**: 89-130.

Mauricio IL, Yeo M, Baghaei M, Doto D, Pratlong F, Zemanova E, Dedet JP, Lukes J & Miles MA (2006) Towards multilocus sequence typing of the Leishmania donovani complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). *Int J Parasitol* **36**: 757-769.

Maurício IL SJ, Miles MA. (2000) The strange case of Leishmania chagasi. - PubMed - NCBI.

Mechali M, Yoshida K, Coulombe P & Pasero P (2013) Genetic and epigenetic determinants of DNA replication origins, position and activation. *Curr Opin Genet Dev* **23**: 124-131.

Miller JC, Tan S, Qiao G, et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**: 143-148.

Minocha N, Kumar D, Rajanala K & Saha S (2011) Kinetoplast morphology and segregation pattern as a marker for cell cycle progression in Leishmania donovani. *J Eukaryot Microbiol* **58**: 249-253.

Minocha N KD, Rajanala K, Saha S (2011) Kinetoplast Morphology and Segregation Pattern as a Marker for Cell Cycle Progression in Leishmania donovani. *Journal of Eukaryotic Microbiology* **58**: 249-253.

Mojica FJ, Juez G & Rodriguez-Valera F (1993) Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol* **9**: 613-621.

Mojica FJ, Diez-Villasenor C, Soria E & Juez G (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol*, Vol. 36 p.^pp. 244-246. England.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J & Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174-182.

Muller CA, Hawkins M, Retkute R, et al. (2014) The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* **42**: e3.

Muller M, Padmanabhan PK & Papadopoulou B (2010) Selective inactivation of SIDER2 retroposon-mediated mRNA decay contributes to stage- and species-specific gene expression in Leishmania. *Mol Microbiol* **77**: 471-491.

Myler PJ, Venkataraman GM, Lodes MJ & Stuart KD (1994) A frequently amplified region in Leishmania contains a gene conserved in prokaryotes and eukaryotes. *Gene* **148**: 187-193.

North TW & Wyler DJ (1987) DNA synthesis in promastigotes of Leishmania major and L. donovani. *Mol Biochem Parasitol* **22**: 215-221.

O'Donnell M, Langston L & Stillman B (2013) Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya.

Pagès M, Bastien P, Veas F, Rossi V, Bellis M, Wincker P, Rioux JA & Roizès G (1989) Chromosome size and number polymorphisms in Leishmania infantum suggest amplification/deletion and possible genetic exchange. *Mol Biochem Parasitol* **36**: 161-168.

Pays E, Delauw MF, Laurent M & Steinert M (1984) Possible DNA modification in GC dinucleotides of *Trypanosoma brucei* telomeric sequences; relationship with antigen gene transcription. *Nucleic Acids Res* **12**: 5235-5247.

Peacock CS, Seeger K, Harris D, *et al.* (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet* **39**: 839-847.

Pelve EA, Lindas AC, Knoppel A, Mira A & Bernander R (2012) Four chromosome replication origins in the archaeon *Pyrobaculum calidifontis*. *Mol Microbiol* **85**: 986-995.

Peng D, Kurup SP, Yao PY, Minning TA & Tarleton RL (2015) CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. *MBio* **6**: e02097-02014.

Perez-Morga D & Englund PT (1993) The structure of replicating kinetoplast DNA networks. *J Cell Biol* **123**: 1069-1079.

Porteus MH & Carroll D (2005) Gene targeting using zinc finger nucleases. *Nat Biotechnol* **23**: 967-973.

Ravel C, Dubessay P, Bastien P, Blackwell JM & Ivens AC (1998) The complete chromosomal organization of the reference strain of the *Leishmania* Genome Project, *L. major*; Friedlin'. *Parasitol Today* **14**: 301-303.

Ravel C, Cortes S, Pratlong F, Morio F, Dedet JP & Campino L (2006) First report of genetic hybrids between two very divergent *Leishmania* species: *Leishmania infantum* and *Leishmania major*. *Int J Parasitol* **36**: 1383-1388.

Raymond F, Boisvert S, Roy G, *et al.* (2012) Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Res* **40**: 1131-1147.

Ready PD (2013) Biology of phlebotomine sand flies as vectors of disease agents. *Annu Rev Entomol* **58**: 227-250.

Reguera RM *ea* (2016) Current status on prevention and treatment of canine leishmaniasis. - PubMed - NCBI.

Reis-Cunha JL, Rodrigues-Luiz GF, Valdivia HO, *et al.* (2015) Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct *Trypanosoma cruzi* strains. *BMC Genomics* **16**: 499.

Reynolds DL, Hofmeister BT, Cliffe L, Siegel TN, Anderson BA, Beverley SM, Schmitz RJ & Sabatini R (2016) Base J represses genes at the end of polycistronic gene clusters in *Leishmania major* by promoting RNAP II termination. *Mol Microbiol*.

Rioux JA, Lanotte G, Serres E, Pratlong F, Bastien P & Perieres J (1990) Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. *Ann Parasitol Hum Comp* **65**: 111-125.

Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R & Bell SD (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116**: 25-38.

Rogers MB, Hilley JD, Dickens NJ, *et al.* (2011) Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res* **21**: 2129-2142.

Romano A, Inbar E, Debrabant A, *et al.* (2014) Cross-species genetic exchange between visceral and cutaneous strains of *Leishmania* in the sand fly vector. *Proc Natl Acad Sci U S A* **111**: 16808-16813.

Sacks D & Kamhawi S (2001) Molecular aspects of parasite-vector and vector-host interactions in leishmaniasis. *Annu Rev Microbiol* **55**: 453-483.

Santrich C, Moore L, Sherwin T, Bastin P, Brokaw C, Gull K & LeBowitz JH (1997) A motility function for the paraflagellar rod of *Leishmania* parasites revealed by PFR-2 gene knockouts. *Mol Biochem Parasitol* **90**: 95-109.

Scholler JK, Reed SG & Stuart K (1986) Molecular karyotype of species and subspecies of *Leishmania*. *Mol Biochem Parasitol* **20**: 279-293.

Shen B, Brown KM, Lee TD & Sibley LD (2014) Efficient gene disruption in diverse strains of *Toxoplasma gondii* using CRISPR/CAS9. *MBio* **5**: e01114-01114.

Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S & Cross GA (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* **23**: 1063-1076.

Simmons LA & Kaguni JM (2003) The DnaAcs allele of *Escherichia coli*: hyperactive initiation is caused by substitution of A184V and Y271H, resulting in defective ATP binding and aberrant DNA replication control. *Mol Microbiol* **47**: 755-765.

Simpson L (1987) The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. *Annu Rev Microbiol* **41**: 363-382.

Smith DF, Peacock CS & Cruz AK (2007) Comparative genomics: from genotype to disease phenotype in the leishmaniasis. *Int J Parasitol* **37**: 1173-1186.

Sollelis L, Ghorbal M, MacPherson CR, Martins RM, Kuk N, Crobu L, Bastien P, Scherf A, Lopez-Rubio JJ & Sterkers Y (2015) First efficient CRISPR-Cas9-mediated genome editing in *Leishmania* parasites. *Cell Microbiol* **17**: 1405-1412.

Sorek R, Kunin V & Hugenholtz P (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181-186.

Stanojic S, Sollelis L, Kuk N, Crobu L, Balard Y, Schwob E, Bastien P, Pages M & Sterkers Y (2016) Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes *Leishmania* and *Trypanosoma brucei* versus mammalian cells. *Sci Rep* **6**: 23142.

Sterkers Y, Lachaud L, Crobu L, Bastien P & Pagès M (2011) FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. *Cell Microbiol* **13**: 274-283.

Sterkers Y, Lachaud L, Crobu L, Bastien P & Pages M (2011) FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*. *Cell Microbiol* **13**: 274-283.

Sterkers Y, Lachaud L, Bourgeois N, Crobu L, Bastien P & Pages M (2012) Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. *Mol Microbiol* **86**: 15-23.

Stillman B (2005) Origin recognition and the chromosome cycle. *FEBS Lett* **579**: 877-884.

Struhl K, Stinchcomb DT, Scherer S & Davis RW (1979) High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proc Natl Acad Sci U S A* **76**: 1035-1039.

Sunkin SM, Kiser P, Myler PJ & Stuart K (2000) The size difference between *leishmania major* friedlin chromosome one homologues is localized to sub-telomeric repeats at one chromosomal end. *Mol Biochem Parasitol* **109**: 1-15.

Tait A (1980) Evidence for diploidy and mating in trypanosomes. *Nature* **287**: 536-538.

Tetaud E, Lefebvre M, M'Bang-Benet DE, Crobu L, Blancard C, Sterkers Y, Pages M, Bastien P & Merlin G (2016) TbFlabarin, a flagellar protein of *Trypanosoma brucei*, highlights differences between *Leishmania* and *Trypanosoma* flagellar-targeting signals. *Exp Parasitol* **166**: 97-107.

Thomas S, Green A, Sturm NR, Campbell DA & Myler PJ (2009) Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics* **10**: 152.

Tiengwe C, Marques CA & McCulloch R (2014) Nuclear DNA replication initiation in kinetoplastid parasites: new insights into an ancient process. *Trends Parasitol* **30**: 27-36.

Tiengwe C, Marcello L, Farr H, Gadelha C, Burchmore R, Barry JD, Bell SD & McCulloch R (2012) Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLoS One* **7**: e32674.

Tiengwe C, Marcello L, Farr H, *et al.* (2012) Genome-wide Analysis Reveals Extensive Functional Interaction between DNA Replication Initiation and Transcription in the Genome of *Trypanosoma brucei*.

Tiengwe C, Marcello L, Farr H, *et al.* (2012) Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell Rep* **2**: 185-197.

Tumova P, Uzlikova M, Jurczyk T & Nohynkova E (2016) Constitutive aneuploidy and genomic instability in the single-celled eukaryote *Giardia intestinalis*. *Microbiologyopen* **5**: 560-574.

Ubeda JM, Legare D, Raymond F, *et al.* (2008) Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome Biol* **9**: R115.

Ubeda JM, Raymond F, Mukherjee A, *et al.* (2014) Genome-Wide Stochastic Adaptive DNA Amplification at Direct and Inverted DNA Repeats in the Parasite *Leishmania*. *PLoS Biol*, Vol. 12 p.^pp.

van Luenen HG, Farris C, Jan S, *et al.* (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell* **150**: 909-921.

Vinayak S, Pawlowic MC, Sateriale A, Brooks CF, Studstill CJ, Bar-Peled Y, Cipriano MJ & Striepen B (2015) Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature* **523**: 477-480.

Wincker P, Ravel C, Blaineau C, Pages M, Jauffret Y, Dedet JP & Bastien P (1996) The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res* **24**: 1688-1694.

Woodward R & Gull K (1990) Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *J Cell Sci* **95 (Pt 1)**: 49-57.

Yan S, Myler PJ & Stuart K (2001) Tetracycline regulated gene expression in *Leishmania donovani*. *Mol Biochem Parasitol* **112**: 61-69.

Yan S, Martinez-Calvillo S, Schnauffer A, Sunkin S, Myler PJ & Stuart K (2002) A low-background inducible promoter system in *Leishmania donovani*. *Mol Biochem Parasitol* **119**: 217-223.

Yao C, Luo J, Hsiao CH, Donelson JE & Wilson ME (2007) *Leishmania chagasi*: a tetracycline-inducible cell line driven by T7 RNA polymerase. *Exp Parasitol* **116**: 205-213.

Yong Wang NE (2016) Advances and Applications of Single-Cell Sequencing Technologies: Molecular Cell. *molecular cell* **58**.

Zangger H, Ronet C, Desponds C, *et al.* (2013) Detection of *Leishmania* RNA virus in *Leishmania* parasites. *PLoS Negl Trop Dis* **7**: e2006.

Zhang WW & Matlashewski G (2015) CRISPR-Cas9-Mediated Genome Editing in *Leishmania donovani*. *MBio* **6**: e00861.

Zhang WW, Mendez S, Ghosh A, Myler P, Ivens A, Clos J, Sacks DL & Matlashewski G (2003) Comparison of the A2 gene locus in *Leishmania donovani* and *Leishmania major* and its control over cutaneous infection. *J Biol Chem* **278**: 35508-35515.

DNA replication dynamics and pre-replication complex in *Leishmania*. Implementation of the CRISPR/Cas9 system in this divergent eukaryote.

Leishmania, a protozoan parasite which causes a large range of clinical diseases worldwide, is characterized by a constitutive 'mosaic aneuploidy' *i.e.* each cell in a population possesses a unique combination of mono-, di- and trisomies for each of the 36 heterologous chromosomes which constitute its genome. Mosaic aneuploidy is generated and maintained via asymmetric chromosomal allotments during mitosis, which implies an unconventional regulation of the replication followed by a permissive segregation.

This study aimed to unravel DNA replication dynamic and to map the replication initiation sites in *Leishmania* using DNA combing and ChIP-seq analyses. Firstly we have characterized DNA replication fork parameters, one of the major findings was that *Leishmania* got the fastest replication speed and the largest interorigin distances among the eukaryotes tested so far. We have also estimated that *Leishmania major* genome possess 168 origins of replication.

To study the actors involved in DNA replication we realized that we first need to develop genetic tools. The CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR associated endonuclease 9) system is a new powerful technique for genome editing. Our second objective was then to adapt this system in *Leishmania*. We have chosen a two-plasmid strategy: one for the expression of single guide RNA and a second for the expression of endonuclease CAS9. The proof of concept has been based on the disruption of the paraflagellar rod-2 (PFR2) loci by the CRISPR-Cas9 system. In the second attempt, we have developed an inducible CRISPR-Cas9 system, both to obtain knock out and to perform endogenous tagging marker free. We used the system to investigate the function of the Origin Recognition Complex proteins. Despite some leak in the system, we obtained edited genome. We deleted *Orc1b* and *Orc1/Cdc6* respectively and monitored the cell cycle progression of the parasite. We found that the depletion of these nuclear proteins lead to a defect of growth and to the appearance of zoids (cells without nucleus). The endogenous tagging of *Orc1b* confirmed the localization that we obtained with an episomal expression vector, this will allow further investigation on the role of the protein.

In total, we found original replication dynamics and using CRISPR Cas9, we have demonstrated that *ORC1b* and *ORC1CDC6* are involved in the nuclear duplication of *Leishmania* suggesting their role in DNA replication.

Dynamique de La réplication de l'ADN et complexe pré-répliatif chez *Leishmania* sp.. Apport du système CRISPR-Cas9.

Leishmania est un parasite eucaryote divergeant responsable d'un large spectre de maladies à travers le monde. Ce parasite est caractérisé par une forme constitutive d'aneuploidie mosaïque, c'est dire qu'au sein d'une population chaque cellule comporte une combinaison de mono-, di- et trisomie de chacun de ses 36 chromosomes. L'aneuploidie mosaïque est générée et maintenue chez les générations suivantes grâce à un taux élevé de répartition asymétrique des chromosomes lors de la mitose. Ceci implique une régulation non-conventionnelle de la réplication suivie d'une ségrégation permissive des chromosomes.

L'objectif premier de cette étude était de comprendre la dynamique de la réplication de l'ADN ainsi que de cartographier les sites d'initiation de la réplication chez *Leishmania*, en utilisant la technique du peignage moléculaire d'une part et celle du ChIP-seq d'une autre part. Nous avons pu caractériser les différents paramètres de progression de la fourche de réplication, un des majeurs résultats qui ressort de cette étude est que *Leishmania* possède les plus grandes distances interorigines et réplique son génome avec une vitesse supérieure aux autres eucaryotes déjà étudiés. Nous avons également pu estimer que le génome de *Leishmania* possède environ 168 origines de réplication. Afin d'étudier les acteurs impliqués dans la réplication de l'ADN chez *Leishmania* nous avons développé l'outil génétique CRISPR/Cas9. Pour développer cet outil nous avons basé notre approche sur une stratégie à deux vecteurs : l'un permet l'expression du sgRNA et l'autre de l'endonucléase Cas9. La validation de cet outil génétique a été réalisée en détruisant le locus PFR2 codant une protéine flagellaire. Dans un second temps nous avons fait évoluer le CRISPR/Cas9 vers un système inductible pour réaliser les Knock Out et des tagging au locus endogène de protéines d'intérêt. Nous avons utilisé ce nouveau système pour étudier la fonction de deux protéines potentiellement impliquées dans le complexe de reconnaissance des origines. Malgré une fuite du système, nous avons pu faire le KO des gènes *Orc1b* et *Orc1/Cdc6* et suivre la progression du cycle cellulaire. Nous avons pu constater que la perte de ces gènes conduisait à un défaut de croissance ainsi qu'à l'apparition de cellule sans noyau. L'insertion d'un tag au locus endogène d'*Orc1b* nous a permis de confirmer la localisation que nous avions obtenu avec une construction épisomale et va nous permettre d'étudier plus précisément le rôle de cette protéine.

En conclusion, nous avons mis en évidence des paramètres de réplication originaux et démontré en utilisant le CRISPR/Cas9 que les protéines *Orc1B* et *Orc1/Cdc6* étaient impliquées dans la duplication du noyau de *Leishmania* suggérant un rôle dans la réplication de l'ADN.