



Relations génome, transcriptome et épigénome chez le champignon filamenteux *Trichoderma reesei*

Sarah Fajon

► To cite this version:

Sarah Fajon. Relations génome, transcriptome et épigénome chez le champignon filamenteux *Trichoderma reesei*. Biotechnologies. Université Paris-Saclay, 2023. Français. NNT : 2023UPASB020 . tel-04107594

HAL Id: tel-04107594

<https://pastel.hal.science/tel-04107594>

Submitted on 26 May 2023

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.

Relations génome, transcriptome et épigénome
chez le champignon filamenteux *Trichoderma reesei*

*Relationships between genome, transcriptome and epigenome in the
filamentous fungus Trichoderma reesei*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°581 : agriculture, alimentation, biologie, environnement, santé (ABIES),
Spécialité de doctorat : Biologie moléculaire et cellulaire
Graduate School : Biosphera. Référent : AgroParisTech

Thèse préparée à **IFP Energies nouvelles**, département Biotechnologie
Sous la direction de **Antoine MARGEOT**, Chercheur (HDR), le co-encadrement
de **Frédérique BIDARD-MICHELOT**, Ingénieure de Recherche et
Aurélié CHATAIGNON, Ingénieure de Recherche

Thèse soutenue à Paris-Saclay, le 27 mars 2023, par

Sarah FAJON

Composition du Jury

Membres du jury avec voix délibérative

Muriel VIAUD

Directrice de Recherche, INRAE (Université Paris-Saclay)

Stéphane LE CROM

Professeur, Sorbonne Université

Claire VENEULT-FOURREY

Directrice de Recherche, INRAE (Université de Lorraine)

Jean-Luc PARROU

Chargé de Recherche, CNRS Toulouse

Présidente

Rapporteur & Examineur

Rapporteur & Examinatrice

Examineur

Titre : Relations génome, transcriptome et épigénome chez le champignon filamenteux *Trichoderma reesei*

Mots clés : *Trichoderma reesei*, champignon filamenteux, épigénétique, ingénierie génétique

Résumé : Le champignon filamenteux *Trichoderma reesei* peut sécréter de grandes quantités d'enzymes cellulolytiques qui sont utilisées à l'échelle industrielle pour produire des sucres fermentescibles à partir de biomasse lignocellulosique. Chez les champignons, divers processus biologiques, tels que la réponse rapide aux changements environnementaux, la production de métabolites secondaires ou d'enzymes industrielles, impliquent des mécanismes de contrôle épigénétique via des changements d'état de la chromatine. L'acquisition de telles connaissances est cruciale pour la définition de stratégies visant à contrôler efficacement les productions fongiques. Cependant, la dynamique chromatinienne qui caractérise les changements entre divers processus de vie fongiques reste peu claire. Le présent projet propose de définir une signature épigénétique durant la production de cellulases chez *T. reesei* en démêlant la dynamique chromatinienne et en identifiant les éléments régulateurs clés.

Dans un premier temps, nous avons cherché à

caractériser trois modifications majeures des histones le long du génome de deux souches de *Trichoderma reesei* : le type sauvage et une souche hyper-productive utilisée dans l'industrie comme référence pour la production de cellulases.

La technique ChIP Seq a été utilisée dans les deux conditions d'induction et d'inhibition de la synthèse de cellulases pour générer la cartographie de l'euchromatine H3K4me3.

Dans la deuxième partie du projet, nous avons cherché à comprendre les rôles joués par les régulateurs épigénétiques majeurs lors de la production de cellulases. Au moyen de la technologie CRISPR-Cas9, nous avons éliminé les gènes codant pour les histones méthyltransférases KMT6 et DIM5, les ADN méthyltransférases DIM2 et RID1 et enfin la protéine de lecture HP1 dans les deux souches. Le phénotypage des mutants a révélé une instabilité phénotypique observée à la fois au stade du développement et de la production de cellulases sur les mutants $\Delta dim5$.

Title : Relationships between genome, transcriptome and epigenome in the filamentous fungus *Trichoderma reesei*

Keywords : *Trichoderma reesei*, filamentous fungus, epigenetics, genetic engineering

Abstract : The filamentous fungus *Trichoderma reesei* can secrete high amounts of cellulolytic enzymes which are used at an industrial scale to produce fermentable sugars from lignocellulosic biomass. In fungi, various biological processes, such as a rapid response to environmental changes, the production of secondary metabolites or industrial enzymes, involve epigenetic control mechanisms via chromatin status changes. The acquisition of such knowledge is crucial for the definition of strategies aiming at efficiently controlling fungal productions. However, the chromatin dynamics that characterize switches between various fungal life traits remain unclear. The present project proposes to define an epigenetic signature during cellulase production in *T. reesei* by unraveling the chromatin dynamics and identifying key regulatory elements.

Firstly, we aimed at characterizing three major

histone modifications along the genome of two *Trichoderma reesei* strains: the wild type and a hyper-productive strain used in industry as a cellulase producing reference. The ChIP Seq technique was used in both conditions of cellulase synthesis induction and inhibition to generate the mapping of euchromatin cue H3K4me3.

In the second part of the project, we strived to attain an understanding of the roles played by known epigenetics regulators during cellulase production. By means of the CRISPR-Cas9 technology, we knocked-out genes encoding the histone methyltransferases KMT6 and DIM5, the DNA methyltransferases DIM2 and RID1 and lastly the reader protein HP1 in both strains. Phenotyping the mutants revealed phenotypic instability observed during both development stage and cellulase production on the $\Delta dim5$ mutants.

TABLE DES MATIERES

TABLE DES MATIERES	1
LISTE DES FIGURES.....	4
LISTE DES TABLEAUX.....	9
INTRODUCTION GENERALE.....	11
INTRODUCTION BIBLIOGRAPHIQUE.....	14
I. INTRODUCTION A L'EPIGENETIQUE	14
I.1. Un peu d'histoire sur l'épigénétique	14
I.2. Les bases de l'épigénétique	15
I.2.1. L'unité nucléosomique	15
I.2.2. Les variants d'histones	17
I.2.3. Les modifications post-traductionnelles des histones	18
I.2.4. Les protéines responsables de l'établissement des modifications post-traductionnelles des histones	20
I.2.5. La méthylation de l'ADN	22
I.3. La chromatine et la régulation de l'expression des génomes : focus chez les champignons	24
I.3.1. Les différents stades de compaction de la chromatine	24
I.3.2. L'hétérochromatine constitutive	25
I.3.2.1. Modification de la structure chimique de l'ADN : la méthylation des cytosines	25
I.3.2.2. Modifications des protéines structurales : modifications post-traductionnelles des histones	27
I.3.3. L'hétérochromatine facultative	29
I.3.4. L'euchromatine	31
I.4. Modifications Epigénétiques et biotechnologie industrielle	31
I.4.1. Modifications et expression épigénétique ou l'art d'agir sur les régulateurs de la chromatine	32
I.4.2. Approche chimique de l'épigénétique	34
I.4.3. La technologie CRISPR-Cas9 : un outil au service de la modulation des promoteurs.....	34
I.4.4. Position chromosomique et expression	36
I.4.5. Variations épigénétiques induites par l'environnement	36
II. LE CHAMPIGNON TRICHODERMA REESEI.....	37
II.1. Origines et patchwork des champignons du genre <i>Trichoderma</i>	37
II.2. Qui est <i>Trichoderma reesei</i> ?	38
II.2.1. Cycle de vie, reproduction sexuée et conidiation de <i>Trichoderma reesei</i>	38
II.2.2. Reproduction sexuée	39
II.2.3. Reproduction asexuée.....	40
II.2.4. La production de cellulases chez <i>Trichoderma reesei</i>	41
II.3. Historique du séquençage de la souche QM6a de <i>T. reesei</i>	42
II.4. Les souches hyper-productrices de <i>Trichoderma reesei</i>	43
II.5. Régulation génétique de la production de cellulases chez <i>Trichoderma reesei</i>	47
II.6. Régulation épigénétique de la production de cellulases chez <i>Trichoderma reesei</i>	50
ENJEUX DE LA THESE.....	53
MATERIEL ET METHODES.....	55
I. MATERIEL BIOLOGIQUE.....	55
II. FERMENTATION	56
I.1. Phase de croissance en mode batch	56
II.1.1. Préparation des fioles de culture et ensemencement	56
II.1.2. Suivi de la phase de croissance	56
II.2. Phase de production en mode Fed-batch	57
II.2.1. Préparation des pompes Dasgip.....	57
II.2.2. Suivi de la phase de fed-batch.....	58
II.2.3. Arrêt de la culture	58
II.3. Détermination de la concentration de protéines extracellulaires	58
II.4. Activité enzymatique	59
III. EXTRACTION D'ARN	59
IV. ANALYSE TRANSCRIPTOMIQUE.....	60

V.	CHIP SEQ	61
V.1.	<i>Préparation des échantillons</i>	61
V.2.	<i>Immunoprécipitation de la chromatine</i>	61
VI.	SCHEMA EXPERIMENTAL DES FERMENTATIONS ET EXPERIENCES DE CHIP SEQ ET RNA SEQ QUI EN DECOULENT	65
VII.	DOT BLOT	65
VIII.	CARTOGRAPHIE ET ANALYSE BIOINFORMATIQUE DE LA MARQUE EPIGENETIQUE H3K4Me3	67
VIII.1.	<i>Contrôle qualité des données de séquençage</i>	68
VIII.2.	<i>Alignement des reads</i>	68
VIII.3.	<i>Détection et élimination des duplicats</i>	68
VIII.4.	<i>Peakcalling</i>	68
VIII.5.	<i>Normalisation et visualisation des données</i>	69
VIII.6.	<i>Analyse différentielle</i>	69
VIII.7.	<i>Génération des graphiques</i>	69
IX.	IDENTIFICATION DES GENES INVALIDES	69
X.	CASSETTES D'INVALIDATION.....	70
XI.	AMPLIFICATION DES PLASMIDES.....	70
XII.	AMPLIFICATION DES CASSETTES D'INVALIDATION.....	71
XIII.	TRANSFORMATIONS DE <i>T. REESEI</i> UTILISANT LA TECHNOLOGIE CRISPR-Cas9	71
XIV.	PREPARATION DES PROTOPLASTES.....	71
XV.	PREPARATION DU COMPLEXE CRISPR-Cas9.....	72
XVI.	TRANSFORMATION.....	73
XVII.	PURIFICATION DES TRANSFORMANTS	74
XVIII.	PCR DE VERIFICATION DES INVALIDATIONS	75
XIX.	SOUTHERN BLOT.....	76
XIX.1.	<i>Création des sondes</i>	76
XIX.2.	<i>Cultures, digestion, migration sur gel</i>	76
XIX.3.	<i>Transfert et révélation</i>	77
XX.	WESTERN BLOT	77
XXI.	COMPLEMENTATIONS.....	78
XXII.	PHENOTYPAGE SUR MILIEUX SOLIDES	79
XXII.1.	<i>Sur milieux minimum</i>	79
XXII.2.	<i>Sur milieux riches</i>	79
XXIII.	STABILITE SUR LE TEMPS.....	79
CHAPITRE 1.....		82
RESULTATS		85
I.	FERMENTATIONS.....	85
I.1.	<i>Fermentation de la souche QM6a</i>	86
I.2.	<i>Fermentation de la souche RutC30</i>	89
II.	RNA SEQ	91
II.1.	<i>Résultats du transcriptome de la souche QM6a</i>	92
II.2.	<i>Résultats du transcriptome de la souche RutC30</i>	93
III.	IMMUNOPRECIPITATION DE LA CHROMATINE : CHIP-SEQ	95
III.1.	<i>Vérification de la spécificité des anticorps</i>	97
III.2.	<i>Immunoprécipitations des échantillons QM6a</i>	99
III.3.	<i>Immunoprécipitation des échantillons RutC30</i>	100
IV.	CARTOGRAPHIE DE LA MARQUE EPIGENETIQUE H3K4Me3.....	102
IV.1.	<i>Contrôle Qualité</i>	103
IV.2.	<i>Mapping</i>	105
IV.3.	<i>Détection et élimination des duplicats</i>	106
IV.4.	<i>Peakcalling</i>	106
IV.5.	<i>Cartographie de la marque H3K4me3 chez la souche QM6a</i>	107
IV.6.	<i>Cartographie de la marque H3K4me3 chez la souche RutC30</i>	113
IV.7.	<i>Comparaison des deux souches et analyses des translocations</i>	118
V.	DIFFERENCE D'EXPRESSION ENTRE LES GENES TOUCHEES PAR H3K4Me3 ET CEUX QUI EN SONT DEPOURVUS	125
VI.	ANALYSES DIFFERENTIELLES INTRA SOUCHE DES DONNEES DE CHIP SEQ.....	127

VI.1.	Analyse différentielle de la marque H3K4me3 chez la souche QM6a	128
VI.2.	Analyse différentielle de la marque H3K4me3 chez la souche RutC30	137
VII.	COMPARAISON ENTRE LES SOUCHES DES RESULTATS DES ANALYSES DIFFERENTIELLES DES CHIP SEQ	148
DISCUSSION		152
CHAPITRE 2.....		156
RESULTATS		158
I.	CONSTRUCTIONS DES MUTANTS DANS LES SOUCHES QM6A ET RUTC30	158
I.1.	Identification des gènes à invalider	158
I.2.	Construction des mutants et validation PCR.....	160
II.	PRODUCTION ENZYMATIQUE DES MUTANTS VALIDES PAR PCR ET CHOIX DES CLONES ETUDIES.....	164
III.	VERIFICATION GENETIQUE DES INDIVIDUS $\Delta DIM5$ ET DISPARITION DE LA MARQUE H3K9ME3	171
III.1.	Vérification par Southern blot.....	171
III.2.	Western blot	173
III.3.	Complémentation	174
III.4.	Séquençage du génome du mutant R $\Delta dim5$ 12.....	176
IV.	CARACTERISATION DES INDIVIDUS $\Delta DIM5$ RETENUS	177
IV.1.	Production enzymatique des clones $\Delta dim5$	177
IV.2.	Phénotypage sur milieux solides.....	182
IV.3.	Stabilité sur le temps.....	188
V.	ANALYSES TRANSCRIPTOMIQUE.....	192
DISCUSSION		198
CONCLUSION GENERALE.....		203
BIBLIOGRAPHIE		206
ANNEXES		229

LISTE DES FIGURES

Figure 1 : Voies de production des biocarburants de 1ère et 2ème génération.	12
Figure 2 : Conrad Waddington.	14
Figure 3 : Assemblage de 147 pb d'ADN autour de l'octamère d'histones pour former le nucléosome.	16
Figure 4 : Structure linéaire des histones canoniques.	17
Figure 5 : Variants d'histones canoniques et leur structure protéique.....	18
Figure 6 : Les différentes modifications post-traductionnelles présentes sur la queue de l'histone H3.	19
Figure 7 : Code international des acides aminés et exemple de nomenclature pour la dénomination des modifications post-traductionnelles des histones.	19
Figure 8 : Effets des modifications des domaines N-terminaux des histones sur la stabilité du nucléosome.	22
Figure 9 : Modèle de la formation et de la propagation de l'hétérochromatine.....	28
Figure 10 : Complexe PRC2 chez deux ascomycètes et un basidiomycète.	30
Figure 11 : Modifications épigénétiques induites par intervention humaine.....	34
Figure 12 : CRISPR-Cas9, un outil d'ingénierie épigénétique.	35
Figure 13 : Classement phylogénique de <i>Trichoderma reesei</i>	38
Figure 14 : Reproduction sexuée chez un champignon ascomycète filamenteux hétérothallique et hermaphrodite.	39
Figure 15 : Patron morphologique de <i>Trichoderma atroviride</i> après une alternance entre lumière et obscurité 12h/12h.	40
Figure 16 : Schéma de dégradation de la cellulose par les cellulases principales.	41
Figure 17 : Généalogie de la souche hyper-productrice RutC30.....	45
Figure 18 : Réassemblage des chromosomes de <i>T. reesei</i> des souches QM6a et RutC30.....	46
Figure 19 : Réseau de régulation de la production d'enzymes cellulolytiques chez <i>T. reesei</i>	48
Figure 20 : Schéma expérimental des fermentations en fioles réalisées chez la RutC30 et la QM6a. .	65

Figure 21 : Représentation de la membrane de Dot blot.	66
Figure 22 : Plasmide contenant la cassette d'inactivation du gène rid.....	70
Figure 23 : Etapes de purification des transformants.	75
Figure 24 : Plasmide de complémentation pour les mutants Δ dim5.	78
Figure 25 : Plan expérimental de l'expérience de cinq semaines visant à établir un phénotype commun aux mutants R Δ dim5.	80
Figure 26: Schéma des étapes d'expérimentations nécessaires à la récolte des données de RNA seq et de ChIP seq.	84
Figure 27: Schématisation des tendances théoriques des concentrations en glucose, en cellulases, en biomasse et en sucre durant l'expérience de fermentation en fioles en fed-batch lactose.	85
Figure 28: Dessin expérimental des expériences de fermentation et ChIP Seq.	86
Figure 29: Suivi des fermentations en fiole de la souche Qm6a.	88
Figure 30: Suivi des fermentations en fiole de la souche RutC30.	90
Figure 31 : Nombre de gènes différentiellement exprimés chez la souche sauvage QM6a en condition lactose.	92
Figure 32 : Nombre de gènes différentiellement exprimés chez la souche hyper-productrice RutC30 en condition lactose.	94
Figure 33: Schématisation de la technique de ChIP-Seq.	96
Figure 34: Schématisation de l'expérience de dot blot.	97
Figure 35 : Résultat de dot blot démontrant la spécificité des anticorps utilisés pour l'expérience de Chip-Seq.....	98
Figure 36: Profils de digestion par nucléase micrococcale de la chromatine extraite des échantillons de mycélium de la souche QM6a.	99
Figure 37: Profil de digestion par nucléase micrococcale de la chromatine extraite des échantillons de mycélium de la souche RutC30.	101
Figure 38: Schématisation du déroulé de l'analyse des données de séquençage jusqu'à l'étape du peakcalling.....	103
Figure 39: Proportion de bases nucléiques le long des reads du fichier R1 de l'échantillon Glu1 de la souche QM6a.	104
Figure 40 : Schématisation de l'alignement des reads sur un génome de référence.	105

Figure 41: Heatmap de corrélation et analyse ACP des échantillons de la souche QM6a.	108
Figure 42: Représentation de la distribution de la marque H3K4me3 le long des chromosomes de la souche QM6a de l'échantillon Glu1.	109
Figure 43: Couverture totale du chromosome par la marque H3K4me3 en fonction du nombre de gènes présents sur celui-ci.....	111
Figure 44: Couverture de la marque H3K4me3 par rapport à la position de l'ATG des gènes.	112
Figure 45: Heatmap de corrélation et analyse ACP des échantillons de la souche RutC30.....	113
Figure 46: Représentation de la distribution de la marque H3K4me3 le long des chromosomes de la souche RutC30.....	115
Figure 47 : Couverture totale du chromosome par la marque H3K4me3 en fonction du nombre de gènes présents sur celui-ci.	116
Figure 48: Couverture de la marque H3K4me4 par rapport à la position ATG des gènes.	117
Figure 49 : Exemple de zones dépourvues de marques et entourées de zones couvertes.	118
Figure 50 : Diagramme de Venn représentant les ensembles de gènes associés aux ensembles de pics consensus des deux souches.....	119
Figure 51: Zone de translocation du chromosome I avec le chromosome VI.....	121
Figure 52: Comparaison de la distribution des marques H3K4me3 le long de l'insertion sur le chromosome VI d'un bout du chromosome I chez la souche RutC30 avec la zone correspondante chez la souche QM6a.....	122
Figure 53: Distribution de la marque H3K4me3 le long de l'insertion du bout de chromosome I sur le chromosome III chez la souche RutC30.....	123
Figure 54 : Visualisation IGV du gène cre1 et de la marque H3K4me3 qui lui est associée sur les réplicats Glu1.	124
Figure 55 : Représentation des groupes de gènes touchés ou non par H3K4me3 en fonction de leur expression chez les deux souches QM6a et RutC30 dans les deux conditions d'étude.	126
Figure 56 : Récapitulatif des étapes et résultats obtenus lors des analyses différentielles.	128
Figure 57 : Distribution de la longueur des pics.....	129
Figure 58: Volcano plot de l'analyse différentielle sur les données de la souche QM6a.....	131
Figure 59: Volcano plot de l'analyse différentielle sur les données de la souche RutC30.....	138

Figure 60 : Récapitulatif de la distribution des ensembles consensus de gènes associés à un pic H3K4me3 chez les deux souches.....	150
Figure 61 : Comparaison des gènes associés à des pics DA chez la RutC30 et chez la QM6a dans les deux conditions d'étude.	151
Figure 62: Domaines protéiques et arbres phylogénétiques des protéines des gènes dim5 et kmt6 et des gènes dim2 et rid.	159
Figure 63: Représentation simplifiée de la recombinaison homologue de la cassette d'inactivation au niveau des gènes ciblés lors de la transformation effectuée avec l'aide de la technologie CRISPR-Cas9.	160
Figure 64: Taux de mutants purifiés pour chacun des gènes invalidés chez les deux souches	161
Figure 65: Représentation schématique des régions amplifiées par les trois PCR de validation.	162
Figure 66: Gel de vérification de la PCR en duplex de l'inactivation du gène dim5 chez la souche RutC30 transformée.....	162
Figure 67: Amplifications PCR au locus du gène dim5 chez la souche RutC30.	163
Figure 68: Production enzymatique chez les différents clones issus de la souche QM6a.	166
Figure 69: Production enzymatique chez les différents clones issus de la souche RutC30.	169
Figure 70: Sites de restriction choisis pour les mutants Δ dim5.	171
Figure 71: Révélation de la membrane de Southern blot pour les mutants Δ dim5 de la souche QM6a et RutC30	172
Figure 72: Western blot réalisé sur les souches parentales RutC30 et QM6a ainsi que sur leurs mutants respectifs.	173
Figure 73: Complémentation du mutant Q Δ dim5 39.....	175
Figure 74: Complémentation des mutants R Δ dim5 12, R Δ dim5 28 et R Δ dim5 39.	175
Figure 75 : Changement de la base G en base C au niveau du codon GAG du gène X du mutant.....	177
Figure 76: Résultats des dosages de quantité d'enzymes produites et de leurs activités enzymatiques spécifiques chez la souche parentale QM6a et le mutant Q Δ dim5 39.	179
Figure 77: Résultats des dosages de quantité d'enzymes produites et de leurs activités enzymatiques spécifiques chez la souche parentale RutC30 et les mutants R Δ dim5 28 et R Δ dim5 39.	181
Figure 78 : Résultats des phénotypages de l'ensemble des mutants Δ dim5 sur milieux solides contenant diverses sources de carbone et sur milieux provoquant un stress oxydatif (ménadione) et un stress osmotique (NaCl).....	183

Figure 79 : Résultats des phénotypages sur milieu PDA avec ou sans exposition à la lumière.	185
Figure 80 : Résultats des phénotypages sur milieu PDA exposés à différentes températures.....	187
Figure 81 : Concentration en protéines extracellulaires dans les surnageants de culture de la souche QM6a et du mutant Q Δ dim5 39.....	189
Figure 82 : Concentration en protéines extracellulaires des surnageants de culture de la souche RutC30 et des mutants R Δ dim5 28 et R Δ dim5 39.	191
Figure 83: Concentration en protéines extracellulaires et biomasse sèche en condition de fed-batch lactose chez la souche RutC30 et ses deux mutants.....	193
Figure 84 : Gènes différentiellement exprimés chez les mutants R Δ dim5 28 et R Δ dim5 39 en comparaison avec la souche parentale RutC30.	194
Figure 85 : Diagrammes de Venn des ensembles des gènes différentiellement exprimés chez les mutants RutC30 en condition lactose.	195

LISTE DES TABLEAUX

Tableau 1 : Modifications post-traductionnelles des histones et enzymes responsables de l'apparition/disparition de ces marques.	21
Tableau 2 : Localisations et fonctions des méthylations post-traductionnelles principales.....	25
Tableau 3 : Protéines responsables de la mise en place de l'hétérochromatine constitutive ainsi que les enzymes responsables pour son confinement dans les zones dédiées.	29
Tableau 4 : Modificateurs épigénétiques pour lesquels l'implication dans la régulation du métabolisme secondaire a été montrée.	33
Tableau 5 : Récapitulatif des facteurs de transcription impliqués dans la synthèse enzymatique.	50
Tableau 6 : Souches de <i>T. reesei</i> utilisées lors de cette étude.....	55
Tableau 7: Anticorps utilisés lors des expériences de ChIP Seq.....	63
Tableau 8 : Dilution, température et temps d'incubation des anticorps utilisés lors des expériences de Dot blot.....	66
Tableau 9 : Préparation du complexe RNP.....	73
Tableau 10: Composition des tubes de transformation.....	74
Tableau 11 : Programme d'amplification utilisé pour les PCR.	75
Tableau 12 : Récapitulation des enzymes de restriction utilisées pour la digestion de l'ADN génomique lors des expériences de Southern blot ainsi que les profils de digestion attendus.	76
Tableau 13 : Schématisation du plan d'expérience en plaques DeepWell.	80
Tableau 14: Quantité d'ADN en nanogrammes obtenues pour les différentes expériences de ChIP réalisées sur la souche QM6a.....	100
Tableau 15: Quantité d'ADN en nanogrammes obtenues pour les différentes expériences de ChIP réalisées sur la souche RutC30.	101
Tableau 16: Nombre de pics H3K4me3 détectés dans chaque échantillon de la souche QM6a.....	107
Tableau 17: Tableau renseignant la taille totale de la librairie, le FRiP score et le nombre de reads dans les pics pour chaque échantillon.	107
Tableau 18: Couverture de H3K4me3 le long des chromosomes de la souche QM6a.	109
Tableau 19: Nombre de pics H3K4me3 détectés dans chaque échantillon de la souche RutC30.	113

Tableau 20: Tableau renseignant la taille totale de la librairie, le FRiP score et le nombre de reads dans les pics pour chaque échantillon.	114
Tableau 21: Couverture de H3K4me3 le long des chromosomes de la souche RutC30	115
Tableau 22 : Gènes associés aux pics différentiellement apposés dans la condition lactose chez la souche QM6a.	132
Tableau 23 : Gènes associés aux pics différentiellement apposés dans la condition glucose chez la souche QM6a.	135
Tableau 24: Gènes associés aux pics différentiellement apposés dans la condition lactose chez la souche RutC30.....	142
Tableau 25 : Gènes associés aux pics différentiellement apposés dans la condition glucose chez la souche RutC30.....	147
Tableau 26: Récapitulatif du nombre de transformants uninucléés validés par PCR et du taux de transformation final pour chaque souche.....	164

INTRODUCTION GENERALE

Depuis le début des années 2000, le développement des biocarburants a connu un essor exceptionnel. Plusieurs raisons à ce nouvel engouement pour les carburants verts expliquent cet intérêt croissant. Tout d'abord, en 1997, à la suite du protocole de Kyoto, la communauté internationale s'est engagée à lutter contre le réchauffement climatique en réduisant les émissions de gaz à effet de serre et plus particulièrement le dioxyde de carbone provenant des transports. D'autre part, les questions sur l'approvisionnement énergétique et l'épuisement des réserves d'énergies fossiles deviennent de plus en plus pressantes. A cela s'ajoute, en 2008, la crise au Moyen Orient et la flambée des cours du pétrole. Ainsi, l'utilisation de denrées végétales est dès lors apparue comme solution à cette crise et aux problèmes pressants du dérèglement climatique : la production de biodiesel s'appuyant principalement sur les plantes oléagineuses telles que le tournesol, le colza, le soja et les palmiers à huile et le bioéthanol produit à partir des plantes sucrières telles que la betterave ou la canne à sucre.

Il en découle une montée en puissance de la production mondiale de biocarburants. D'après l'association World Bioenergy, entre 2000 et 2012, la production annuelle d'éthanol passe de 17 milliards de litres à près de 86 milliards. La production de biodiesel, elle, est multipliée par vingt-sept, passant de 806 millions de litres à 21,5 milliards de litres.

Néanmoins, rapidement, de nombreux débats et inquiétudes s'élèvent à l'encontre de ces biocarburants dits de première génération ou classiques. En effet, en plus des coûts importants de production, l'usage des produits agricoles pour la production de biocarburants aurait des conséquences néfastes sur l'environnement (Fargione et al. 2008) et la sécurité alimentaire dans de nombreux pays (Bekunda et al. 2009). Elle impacterait également fortement les prix des matières premières, déstabilisant les marchés agricoles mondiaux déjà fragiles (Drabik, Ciaian et Pokrivčák 2016). Ceci participe donc fortement au développement de biocarburants de seconde génération.

De nombreux efforts sont donc mis en place pour le développement de ces nouveaux biocarburants. Ils sont obtenus à partir de plantes entières dédiées ou de fractions non comestibles de plantes pour l'alimentation. Il s'agit donc de biomasse lignocellulosique, extrêmement abondante mais dispersée, peu valorisée car considérée comme « déchet » de la production agricole et forestière (bois de découpe, pailles, résidus agricoles non utilisables).

Au sein de cette nouvelle génération de biocarburants, nous distinguons le biogazole et le biokérosène, obtenus par voie thermochimique et synthèse Fischer-Tropsch, du bioéthanol, obtenu par voie biochimique et fermentation. Dans ce

manuscrit, nous allons nous concentrer uniquement sur ce dernier.

La production de bioéthanol dit de seconde génération, nécessite un procédé de trois étapes (Figure 1). Premièrement, la biomasse lignocellulosique doit être prétraitée pour rendre accessible la cellulose à l'hydrolyse enzymatique. Puis, l'action de cocktails d'enzymes cellulolytiques produits par un champignon filamenteux, généralement *Trichoderma reesei*, libère les monosaccharides de glucose. Il s'en suit une étape de fermentation pour la production d'éthanol par la levure *Saccharomyces cerevisiae* qui va ensuite être purifié et concentré par distillation (Margeot et al. 2009). L'étape d'hydrolyse enzymatique représentait dans les années 2010 à elle seule au moins 30% du coût de production de bioéthanol. Il s'agit donc ici d'un point décisif à optimiser pour l'amélioration de la production et la rendre viable économiquement.

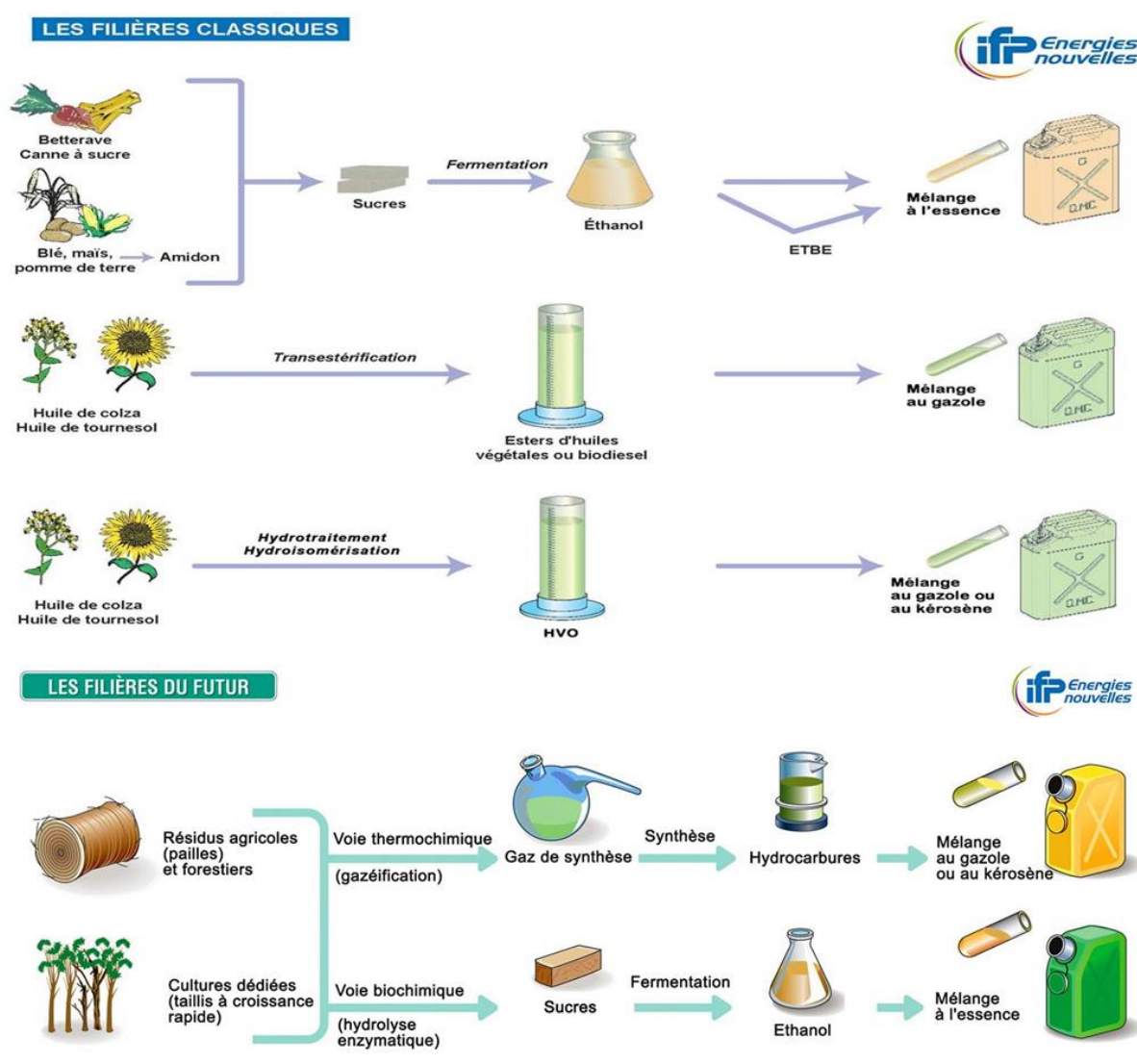


Figure 1 : Voies de production des biocarburants de 1ère et 2ème génération.
(IFPEN)

Ainsi, une des voies de réduction des coûts de l'étape d'hydrolyse enzymatique a

reposé sur l'amélioration des souches du champignon *Trichoderma reesei*. Cette optimisation a d'abord été basée sur la génération de mutants hyper-producteurs de la souche sauvage. Pour cela, des mutagénèses aléatoires ont été menées et ont permis la création de souches industrielles performantes (Robyn Peterson 2012). Avec l'avancée de la recherche et le développement des méthodes de biologie moléculaire il a été possible d'intervenir de façon ciblée sur le génome et générer des améliorations par ingénierie génétique (Druzhinina et Kubicek 2017; Aro et al. 2003; Ma et al. 2011). Par ailleurs, les travaux en biologie systémique ont permis de progresser dans la connaissance des mécanismes de production des cellulases par le champignon et aujourd'hui les verrous apposés sur le contrôle des gènes codant pour ces enzymes sont petit à petit levés (Bischof, Ramoni et Seiboth 2016).

De manière générale, la découverte de mécanismes de régulation épigénétique dans des organismes modèles tels que *Aspergillus nidulans* ou *Neurospora crassa* ont suscité de nouvelles investigations pour une compréhension plus pointue et une maîtrise complète des systèmes de production biologiques et pour le développement de nouveaux outils pour l'amélioration des souches industrielles. Chez *T. reesei* il s'est avéré que la délétion de gènes régulateurs globaux du métabolisme des cellulases avait également un impact sur l'organisation spatiale des chromosomes (Cao et al. 2019). Également, le séquençage du génome du champignon a permis de découvrir l'organisation regroupée des gènes de cellulases et leur localisation dans les régions subtélomériques hautement influencées par l'épigénétique (Li et al. 2017; Marie-Nelly et al. 2014; Martinez et al. 2008). Tout ceci, permet de formuler l'hypothèse d'une régulation épigénétique de la production d'enzymes chez *Trichoderma reesei*. Nous proposons donc dans cette étude d'établir un premier état des lieux de l'épigénome de ce champignon ainsi que l'implication de certains gènes majeurs de l'épigénétique, tout ceci, en condition de production de cellulases.

Ces travaux ont été réalisés au sein du département Biotechnologie d'IFP Energies nouvelles (IFPEN), sous la direction de M. Antoine Margeot, l'encadrement de Mme Frédérique Bidard-Michelot et Mme Aurélie Pirayre Chataignon ainsi qu'avec les conseils scientifiques et techniques de Mme Fabienne Malagnac.

INTRODUCTION BIBLIOGRAPHIQUE

I. INTRODUCTION A L'EPIGENETIQUE

I.1. Un peu d'histoire sur l'épigénétique

Le mot « épigénétique » a été inventé en 1942 par Conrad Waddington (Figure 2) biologiste du développement, pour décrire le lien entre l'hérédité dirigée par la génétique classique et la mise en place de la différenciation cellulaire lors de l'embryogénèse. L'épigénétique lui permettait alors de réconcilier généticiens et embryologistes en y classant tout mécanisme biologique inexplicable par la génétique.



Figure 2 : Conrad Waddington.

Après la découverte de la structure de l'ADN en 1953 et les mécanismes d'expressions génétiques, de nouveaux questionnements sont apparus concernant cette dimension supplémentaire de l'expression génique. Comment se faisait-il que des cellules contenant le même ADN puissent être si différentes ? Comment cette différenciation se maintient-elle ?

Aujourd'hui, la définition moderne de l'épigénétique découle de ces questionnements : l'épigénétique est tout mécanisme héréditaire de modulation d'expression des gènes, stable mais néanmoins réversible sans modification de la séquence de l'ADN. Il s'agit donc de modifications chimiques de l'ADN (Riggs 1975) ou de ses protéines structurales (Allfrey, Faulkner et Mirsky 1964) sans altération de la séquence génétique. Ces modifications permettent une expression différentielle de certains gènes cibles et sont, dans certains cas, transmis aux générations suivantes de cellules (Heard et Martienssen 2014). Néanmoins, dans des conditions environnementales particulières, ces marques épigénétiques peuvent être altérées et induire une toute autre expression des gènes et ainsi produire un phénotype nouveau.

L'épigénétique est aujourd'hui largement étudiée par les chercheurs du monde entier pour son implication majeure dans des mécanismes aussi variés que l'inactivation du chromosome X chez les mammifères (Sandrine Augui et Edith Heard 2008), le développement de cancers (Sharma, Kelly et Jones 2010) ou de maladies neuro-dégénératives (Berson et al. 2018), le polyphénisme chez *Apis mellifera* (Cameron, Duncan et Dearden 2013) ou encore la régulation de la floraison chez *Arabidopsis thaliana* (He, Michaels et Amasino 2003).

Aujourd'hui, nous savons que les champignons ne sont pas de reste lorsqu'il s'agit de régulation épigénétique. L'étude de leur épigénome est ainsi devenu un point central de l'amélioration des souches industrielles (Aghcheh et Kubicek 2015). Nous proposons ici de donner une vue d'ensemble sur les connaissances des mécanismes épigénétiques en lien avec l'ADN chez les champignons filamenteux puis nous nous concentrerons sur le producteur d'enzymes *Trichoderma reesei*.

I.2. Les bases de l'épigénétique

I.2.1. L'unité nucléosomique

Dans tout organisme vivant, la structure primaire de l'ADN reste la même. Il s'agit d'une chaîne d'acides désoxyribonucléiques bicaténaire antiparallèle formant une double hélice et maintenue par des liaisons hydrogènes entre bases complémentaires (Watson et al. 2014).

Dans les cellules eucaryotes comme dans les cellules procaryotes, l'ADN est compacté afin de pouvoir rentrer dans un espace cellulaire restreint : le noyau pour les eucaryotes, et le cytoplasme pour les procaryotes. En revanche, alors que l'ADN des procaryotes reste « dénudé » et ne s'associe pas à des protéines structurales, l'ADN des eucaryotes s'agrège en structures compactes complexes autour de protéines spécifiques (Oudet, Gross-Bellard et Chambon 1975). Cette compaction donne naissance à une structure appelée chromatine, importante pour la régulation de l'expression des gènes et de nombreux processus cellulaires. Plusieurs états de compaction de la chromatine ont été décrits au fil des années, permettant une meilleure compréhension de son rôle, au-delà de l'économie d'espace.

Le premier stade de compaction de la chromatine consiste en une fibre de 6 nm de diamètre composée de 147 pb d'ADN négativement chargé, enroulée 1,65 fois autour d'un octamère de protéines positivement chargées, appelées histones (Watson et al. 2014). Il existe quatre histones canoniques, composant une unité cœur centrale: histone H2A, H2B, H3 et H4 (Maeshima, Hihara et Eltsov 2010). Dans l'unité centrale, les histones sont organisées en dimère et couplées deux à deux chacune avec son homologue (H2A-H2A, H2B-H2B etc...). La longueur de l'ADN entre ces unités varie en fonction de l'espèce étudiée et de l'état chromatinien (entre 20 et 60 pb).

L'enroulement de la fibre d'ADN autour de l'octamère d'histones constitue l'unité de base de la chromatine et s'appelle un nucléosome (Figure 3) (Perche et al. 2003).

Une histone est composée d'un corps globulaire (« histone fold-domain » : HFD), d'un domaine C-terminal court ainsi qu'un domaine N-terminal beaucoup plus long (Figure 3) (Arents et Moudrianakis 1995). Il est fait référence à ce dernier domaine comme à la « queue » des histones en raison de sa longueur, sa structure d'acides aminés désordonnée et son accessibilité depuis l'extérieur du nucléosome. Le faible stade de compaction (6X) engendré par le nucléosome neutralise environ 60% des charges négatives totales de l'ADN grâce aux acides aminés (lysine et arginines notamment) situés sur les queues des histones (Strick et al. 2001). De plus, une histone additionnelle, H1, est située sur l'ADN « *linker* » (entre chaque matrice protéique) et joue ainsi le rôle de verrou entre l'ADN et le nucléosome. Elle permet ainsi une compaction plus importante de la chromatine par une élimination supplémentaire des charges négatives (Maeshima, Hihara et Eltsov 2010). Les histones canoniques sont en quantités similaires et très abondantes dans le noyau alors que l'histone H1 est nettement moins représentée.

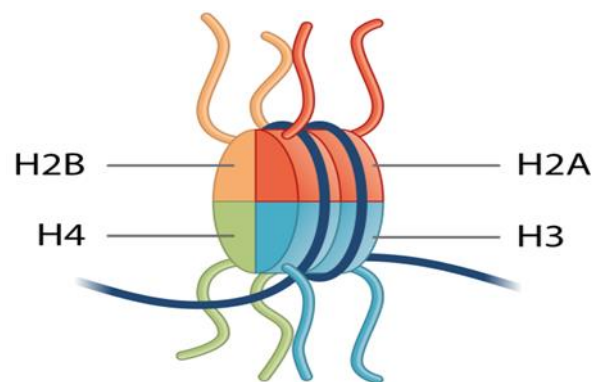


Figure 3 : Assemblage de 147 pb d'ADN autour de l'octamère d'histones pour former le
(Vincent Grandperret, IFPEN)

Il est également intéressant de noter que les quatre histones canoniques (Figure 4) font partie d'une famille de protéines extrêmement bien conservées dans le règne des eucaryotes (Luger, Dechassa et Tremethick 2012). Les gènes codant pour les histones canoniques ne possèdent pas d'intron et ne subissent donc pas d'épissage. Ils sont souvent organisés en cluster et s'expriment uniquement lors de la réplication de l'ADN afin de faire face au besoin de compaction des brins nouvellement synthétisés. Ainsi, les histones canoniques ne s'intègrent à l'ADN qu'à cette étape (Henikoff et Ahmad 2005).

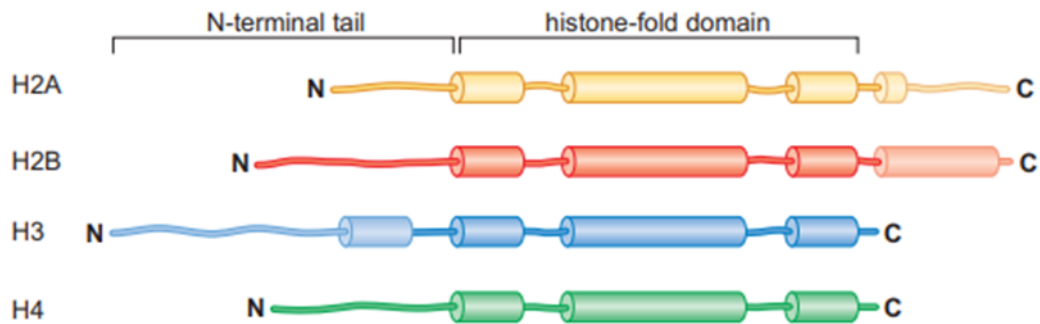


Figure 4 : Structure linéaire des histones canoniques.
(Watson et al. 2014)

1.2.2. Les variants d'histones

Pour une même histone, plusieurs isoformes existent. Il est possible de classer ces variants de deux manières. La première consiste en une classification structurale différenciant les variants selon leur degré d'homologie avec l'histone canonique (Figure 5). Une deuxième distingue les variants sur leurs caractéristiques fonctionnelles et leur implication dans des mécanismes cellulaires spécifiques (expression spécifique des tissus, réplication, réparation de l'ADN etc...). Contrairement aux histones canoniques, les gènes des variants contiennent des introns et sont donc sujets à l'épissage. Les variants sont exprimés tout au long du cycle cellulaire et sont insérés aux nucléosomes en remplacement des histones canoniques de façon indépendante à la réplication de l'ADN (Luger, Dechassa et Tremethick 2012). Ce remplacement est assuré par des facteurs spécifiques du remodelage de la chromatine ATP-dépendants ainsi que des protéines chaperonnes sur des localisations très particulières du génome (H2A.Z par exemple se situe au niveau des promoteurs des gènes alors que CenH3 est exclusivement au niveau des centromères).

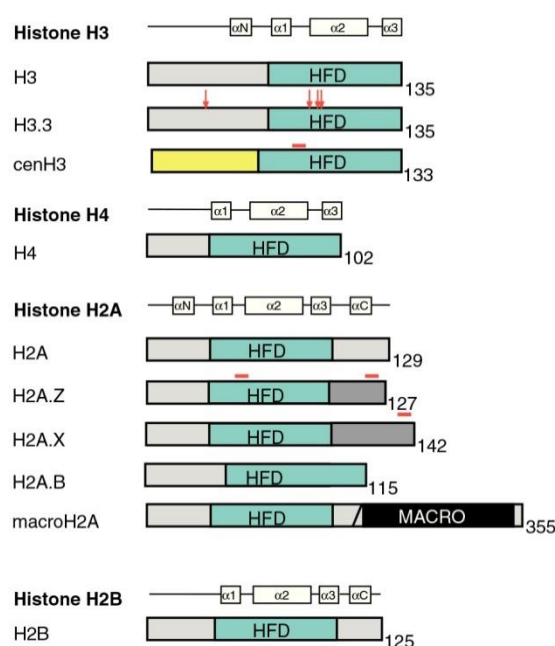


Figure 5 : Variants d'histones canoniques et leur structure protéique.

Le corps globulaire est mentionné comme HFD (Histone-Core Domain), les régions marquées de rouge localisent les variations dans la séquence d'acides aminés. Les flèches indiquent une modification ponctuelle alors que les modifications plus étendues et importantes sont signalées par les tirets (Allis et al. 2015)

1.2.3. Les modifications post-traductionnelles des histones

La formation des nucléosomes n'est que la première étape dans un processus complexe qui permet aux eucaryotes de condenser leur ADN en une fibre dense atteignant son maximum de compaction au cours de la mitose. Cette hypercondensation de la fibre nucléosomique est un processus dynamique permettant une accessibilité choisie de l'ADN en fonction des besoins et des contraintes dans lesquelles se trouvent la cellule. Ainsi, bien que pouvant également intervenir sur le domaine C-terminal, c'est majoritairement sur les résidus de la queue N-terminale des histones que vont se situer les modifications chimiques déterminant le remodelage de la chromatine et l'expression génétique (Figure 6). Le domaine N-terminal des histones étant d'une importance majeure pour la stabilité du complexe nucléosomique, toute modification chimique intervenant sur ses acides aminés crée un changement, *cis* ou *trans*, dans ses interactions avec l'ADN (Bannister et Kouzarides 2011).

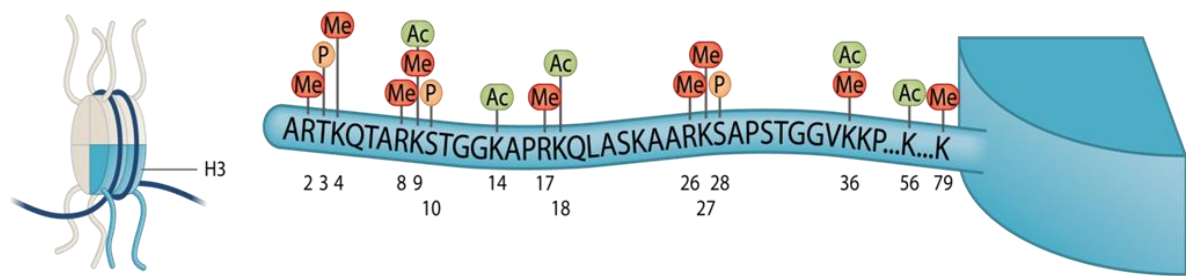


Figure 6 : Les différentes modifications post-traductionnelles présentes sur la queue de l'histone H3.

Me : méthylation ; P : phosphorylation ; Ac : acétylation
(Vincent Grandperret, IFPEN)

Ces modifications sont post-traductionnelles (PTM), c'est-à-dire qu'elles interviennent après la synthèse de l'histone et ne prennent donc pas part dans le déterminisme de sa structure. Elles consistent en l'addition d'un ou plusieurs groupes fonctionnels chimiques sur les acides aminés (aussi appelés « résidus ») de la queue de l'histone ciblée et sont aussi appelées « marques épigénétiques ». La Figure 7 ci-dessous renseigne sur la nomenclature utilisée pour y faire référence. Le nom de l'histone est cité en premier puis le résidu touché et enfin la modification qu'il a subie.

Code international des acides aminés

- G - Glycine (Gly)
- P - Proline (Pro)
- A - Alanine (Ala)
- V - Valine (Val)
- L - Leucine (Leu)
- I - Isoleucine (Ile)
- M - Methionine (Met)
- C - Cysteine (Cys)
- F - Phenylalanine (Phe)
- Y - Tyrosine (Tyr)
- W - Tryptophan (Trp)
- H - Histidine (His)
- K - Lysine (Lys)
- R - Arginine (Arg)
- Q - Glutamine (Gln)
- N - Asparagine (Asn)
- E - Glutamic Acid (Glu)
- D - Aspartic Acid (Asp)
- S - Serine (Ser)
- T - Threonine (Thr)

Histone ciblée	Résidu touché	Position sur la queue	Type de modification	Nomenclature
H3	Lysine	4	Triméthylation	H3K9me3
H4	Lysine	18	Acétylation	H3K18ac
H3	Arginine	17	Diméthylation	H3R17me2

Figure 7 : Code international des acides aminés et exemple de nomenclature pour la dénomination des modifications post-traductionnelles des histones.

http://svt.tice.ac-orleans-tours.fr/php5/publis/phylogene/nouv_donn_phylo/html/codeaa.htm

On recense aujourd'hui plus d'une centaine de modifications, localisées en globalité de la queue des quatre histones canoniques mais aussi sur leurs variants. Les plus connues et les plus étudiées sont les acétylations, les méthylations et les phosphorylations mais il en existe de nombreuses autres : ADP-ribosylation, ubiquitination, sumoylation, déimination, proline isomérisation, succinylation, hydroxylation, crotonylation, N-formylation etc (Bannister et Kouzarides 2011).

Ainsi, en fonction du type de modification mais également de l'acide aminé spécifique ciblé, ces PTM n'auront pas le même impact sur la chromatine. De fait, les acétylations sont connues pour déstabiliser directement les interactions ADN-histones par l'augmentation des charges négatives sur les queues et entraîner un état chromatinien favorable pour la transcription des gènes. De la même manière, la phosphorylation entraîne également le rejet entre charges négatives et donc une fragilisation du contact entre ADN et protéines structurales. Ainsi l'acétylation des résidus lysine 13 et 56 de l'histone H3 (H3K56ac, H4K16ac) sont notamment connues pour leur impact positif sur la transcription (Shogren-Knaak et al. 2006). Pendant que ces marques sont invariablement associées à une conformation de la chromatine favorable à l'activité transcriptionnelle, d'autres sont ambivalentes. Par exemple, en fonction de l'histone (H3 et/ou H4) et du résidu ciblé (arginine et/ou lysine), les méthylations vont donner un message favorable ou défavorable pour la transcription (Watson et al. 2014). Par exemple, les méthylations les plus étudiées pour leur impact positif sur la transcription des gènes sont H3K4me3, H3K36me3 alors que H3K27me3, H3K9me3, H4K20me3 sont connues pour leur répression transcriptionnelle (Freitag 2017; Freitag 2014). N'altérant pas les charges portées par les histones, les méthylations ne représentent pas un mécanisme direct de modulation de la conformation de la chromatine et n'interviennent pas dans la stabilité électrostatique des nucléosomes (Freitag 2017). En revanche, elles représentent un signal pour des complexes multi-protéiques modificateurs dont nous parlerons plus en détails plus loin dans ce manuscrit.

1.2.4. Les protéines responsables de l'établissement des modifications post-traductionnelles des histones

Les modifications post-traductionnelles sont catalysées par des protéines « éditrices ». Ces protéines sont capables de cibler de façon spécifique un résidu d'histone dans une région génomique bien déterminée. Elles interviennent de façon directe et chacune d'entre elles catalyse - ou « appose » - la modification d'un l'acide aminé particulier résultant en une marque épigénétique spécifique. Les acétylations des lysines des domaines N-terminaux des histones sont promues par des protéines histones acétyltransférases (HAT) contenant un site actif soit de type bromodomaine, soit chromodomaine (« *chromatine organization domain* »). Chez les champignons filamenteux, la méthylation du groupement amine des lysines est catalysée par des histones méthyltransférases (KMT) contenant un domaine SET (Freitag 2017). Il est

toutefois possible de supprimer ces PTM. Pour cela, des protéines « *eraser* » enlèvent, toujours de manière spécifique, ces modifications. Ainsi, les histones déacétylases (HDAC) sont chargées de retirer les acétylations des lysines des histones pendant que les lysines déméthylases (KDM) et arginines déméthylases (PRMT) s'occupent des méthylations (Tableau 1) (Freitag 2017).

Modification	Residue	Example	Enzyme (fungal homologs)
Acetylation	Lys (K)	H4K5/12ac	KAT1 (ScHat1, NcHAT-1)
		H3K9/14/18/36ac	KAT2 (ScGcn5, NcNGF-1)
		H3ac and H4ac	KAT4 (ScTaf1, NcHAT-3)
		H4K5/8/12/16ac	KAT5 (ScEsa1, NcHAT-4, An ESAA)
		H3K14ac	KAT6 (ScSas3, NcHAT-6)
		H4K16ac	KAT8 (ScSas2, NcHAT-5)
		H3ac	KAT9 (ScElp3, NcELP-3)
		H3K14ac	KAT10 (ScHap2)
		H3K56ac	KAT11 (ScRtt109)
		H3ac, H4ac	Rpd3 (ScRpd3, NcHDA-3, AnRPDA, FgRpd3)
Deacetylation	AcLys (acK)	H3ac?, H4ac?	Hda1 (ScHda1, NcHDA-1, AnHDAA, FgHdF2)
		H3ac, H4ac	Hos2 (ScHos2, NcHDA-2, AnHosA, FgHdF1)
		H3ac, H4ac	Hos3 (ScHos3, NcHDA-4, AnHosB, FgHdF3)
		H3K9/14/56ac, H4K16ac	SIRT1 (ScHst1, NcNST-1,)
		H3K9/14/56ac	SIRT2 (ScHst2, NcNST-2, An HSTA)
		H4ac?	SIRT1 (ScHst4, NcNST-3)
		?	SIRT3 (ScHst3, NcNST-4)
		H4ac?	SIRT4 (NcNST-5)
		?	SIRT5 (NcNST-6)
		Nucleolus	SIRT7 (NcNST-7)
Methylation	Lys (K)	H3K9me2, -3	KMT1 (SpClr4, NcDIM-5, AnCLRd, FgKMT1)
		H3K4me2, -3	KMT2 (SpSet1, NcSET-1, An, FgKMT2)
		H3K36me2, -3	KMT3 (ScSet2, NcSET-2, An, FgKMT3)
		H3K79me	KMT4 (ScDot1, NcDOT-1, FgKMT4)
		H4K20me	KMT5 (NcSET-, FgKMT5)
		H3K27me3	KMT6 (NcSET-7, FgKMT6)
	Arg (R)	H4R3me	PRMT1 (ScHmt1, NcPRM-1)
		Nonhistone proteins	PRMT3 (NcPRM-3)
	MeLys (meK)	H3R8me, H4R3me	PRMT5 (ScHsl7, SpSkb1, NcPP-2)
		H3K4me, H3K9me	KDM1 (SpLsd1, AnHDMA)
		H3K36me	KDM2 (ScJhd1)
		H3K9me	KDM3
Demethylation	MeLys (meK)	H3K9me, H3K36me	KDM4 (ScRph1)
		H3K4me	KDM5 (Jhd2, SpJmj2, SpLid1)
		H3K27me	KDM6
	Ser (S)	H3S10ph	Aurora B (ScIpl1, AnNIMA)
		H4S1ph	CKII (ScCkII)
	Thr (T)	H3T45ph	CDC7 (ScCdc7)

Tableau 1 : Modifications post-traductionnelles des histones et enzymes responsables de l'apparition/disparition de ces marques.

Les enzymes sont listées par leurs noms donnés chez les mammifères. Leurs homologues chez les champignons sont accompagnés par les initiales des organismes appropriés : Sc : *Saccharomyces cerevisiae*, Sp : *Schizosaccharomyces pombe*, Nc : *Neurospora crassa*, An : *Aspergillus nidulans*, Fg : *Fusarium graminearum* (Freitag M. 2014).

Outre ces aspects purement mécaniques, l'ensemble de ces marques épigénétiques « émettent » un signal global avec des informations parfois contradictoires. Il se créait alors des boucles de rétroaction interconnectées permettant

un « dialogue » entre les mécanismes modificateurs (Figure 8). Le paysage chromatinien dépend donc d'un contexte général épigénétique modulé par les besoins immédiats de la cellule (transcription, répression, recombinaison, réparation etc...) permettant un contrôle précis de la régulation des gènes (Paksa et Rajagopal 2017).

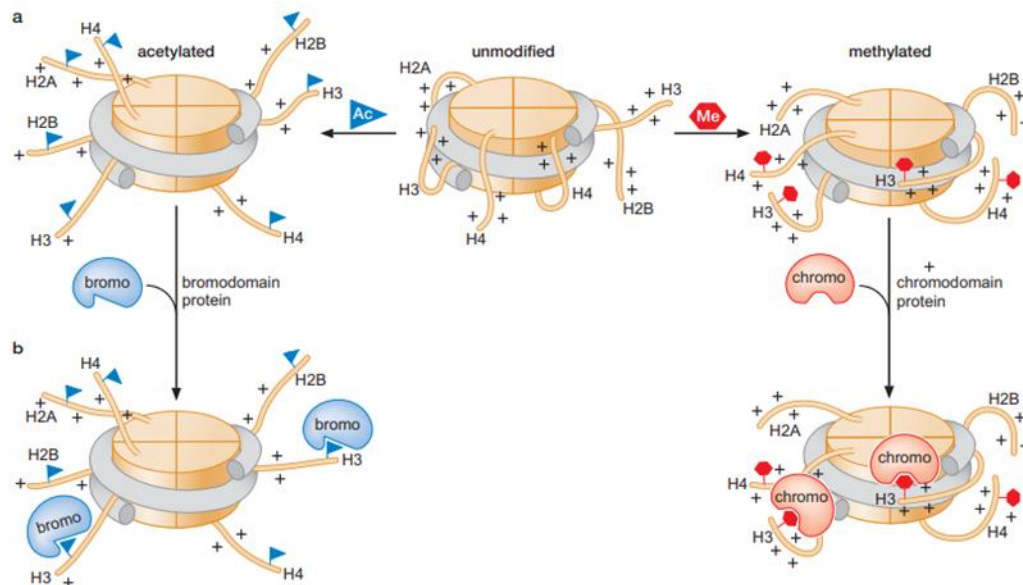


Figure 8 : Effets des modifications des domaines N-terminaux des histones sur la stabilité du nucléosome.

a) Les histones non modifiées ou ayant été méthylées s'associent de façon plus étroite à l'ADN nucléosomal que les histones acétylées. b) Les modifications post-traductionnelles peuvent créer des sites de reconnaissance pour des complexes modificateurs de la chromatine (Watson et al. 2014)

L'installation des modifications post-traductionnelles est donc un processus actif, transitoire et réversible impliquant la coordination de nombreux facteurs.

1.2.5. La méthylation de l'ADN

Un autre acteur majeur dans le langage épigénétique est la méthylation de l'ADN. Elle joue un rôle crucial dans les processus cellulaires de développement autant chez les plantes que chez les mammifères ou chez les champignons. Des études ont démontré que chez les plantes et chez les mammifères, un défaut dans la méthylation de l'ADN pouvait entraîner de graves anomalies de croissance et la mort embryonnaire. Au contraire, chez les champignons filamenteux, les défauts de la méthylation semblent avoir des impacts divers sur les phénotypes mais ne sont pas létaux (Jeon et al. 2015; Rountree et Selker 2010, 1997; Bewick et al. 2019; Grognet et al. 2019).

La méthylation de l'ADN est un mécanisme épigénétique contrôlant l'expression des gènes. Cette méthylation altère ainsi les affinités avec les protéines transcriptionnelles et permet de recruter de nouvelles protéines impliquées dans la répression de l'expression des gènes. La méthylation des acides nucléiques est connue pour son impact sur de nombreux processus biologiques tel que l'empreinte parentale des gènes, l'inactivation du chromosome X, la différenciation cellulaire ou bien encore l'extinction de gène (Reik et Lewis 2005; Sandrine Augui et Edith Heard 2008).

Les schémas de méthylations sont établis et maintenus à travers les réplifications de l'ADN grâce aux ADN méthyltransférases (DMT ou DNMT). Chez les mammifères, deux familles d'ADN méthyltransférases sont distinguées. La protéine DNMT1, responsable de la maintenance des patrons de méthylation chez les cellules somatiques après réplication, et les protéines du type DNMT3, responsables de la méthylation *de novo* de séquences jusqu'alors non méthylées chez les embryons (Huang et al. 2016). Chez les plantes, trois types de DNMTs sont recensées : MET1 pour la maintenance des méthylations, la chromométhylase 3 CMT3, spécifique des plantes à fleurs, et enfin les deux protéines DRM1/2 pour la méthylation *de novo* (Law et Jacobsen, 2010). Chez les champignons, quatre ADN méthyltransférases ont été identifiées : DIM2 (Defective In Methylation 2), DNMT1, DNMT5 et RID (RIP Defective). Pendant que RID est nécessaire pour la méthylation *de novo*, DIM2, DNMT1 et DNMT5 sont nécessaires à la maintenance des méthylations (Bewick et al. 2019; Huff et Zilberman 2014; Huang et al. 2016).

La répartition des méthylations le long de l'ADN n'est pas uniforme. De fait, nous retrouvons majoritairement des cytosines méthylées au niveau des sites CpG des parties centrales des chromosomes (centromères) et aux extrémités (télomères). Ces zones sont typiquement riches en séquences répétées et éléments transposables (jusqu'à 45% chez l'humain (Schulz, Steinhoff et Florl 2006)). De même, la méthylation parmi et le long des gènes n'est pas uniforme. En effet, la majorité des méthylations retrouvées sur les éléments transposables est localisée au niveau des promoteurs et des terminateurs, imposant une inhibition forcée de la transcription. D'un autre côté, la méthylation des promoteurs chez les gènes, en particulier domestiques, reste globalement faible (Gardiner-Garden et Frommer 1987). Rountree et Selker (1997) ont montré que l'association de complexes transcriptionnels fonctionnels avec des promoteurs méthylés est possible. En revanche, l'élongation de la transcription sera bloquée (Rountree et Selker 1997; Barry, Faugeron et Rossignol 1993). Toutefois, des expériences de RNA-seq ont démontré qu'il était possible de retrouver des méthylations dans le corps de gènes transcrits. En effet, Rountree et Selker (1997) ont également démontré par le marquage des ARN *in vivo* que si l'allèle d'un gène non méthylé est transcrit cinq fois plus rapidement que l'allèle méthylé de ce même gène, des ARN fonctionnels sont néanmoins retrouvés dans le noyau. Dans le même sens, Hellman et Chess (2007) ont démontré que chez les mammifères femelles, le chromosome X actif présente deux fois plus de méthylations dans le corps de ses gènes que le chromosome inactif mais présente en revanche une hypométhylation de leur

promoteur. La relation entre méthylation de l'ADN et transcription est donc plus fine qu'un simple jeu d'interrupteur entre inhibition et transcription comme pensé auparavant.

1.3. La chromatine et la régulation de l'expression des génomes : focus chez les champignons

1.3.1. *Les différents stades de compaction de la chromatine*

La modulation du paysage chromatinien est donc directement liée à la présence ou non de modifications épigénétiques post-traductionnelles sur les histones et/ou de méthylations sur les cytosines de l'ADN. L'activité transcriptionnelle et de nombreux mécanismes cellulaires dépendent de cette conformation. En fonction de son état de compaction, on distingue deux types de chromatine. Un premier état favorable à la transcription où la chromatine est appelée euchromatine et qui est associée à des régions génomiques riches en gènes. Un deuxième état, qui représente une forme chromatinienne d'inhibition transcriptionnelle appelé hétérochromatine. Cette forme hautement compacte de la chromatine peut être permanente et sera alors appelée hétérochromatine constitutive. Sinon, elle peut être un état transitoire de la chromatine dans un environnement donné et sera distinguée de son isoforme par le nom hétérochromatine facultative (Freitag M. 2014).

Alors que l'hétérochromatine constitutive est située dans des régions génomiques pauvres en gènes et riches en éléments répétés et /ou transposables, l'hétérochromatine facultative est quant à elle formée dans un environnement euchromatinien dans lequel des protéines de l'hétérochromatine sont venues apposer des marques épigénétiques (Dillon et al. 2005). L'hétérochromatine constitutive permet une ségrégation des régions génomiques et est typiquement localisée au niveau du centromère et des régions télomériques des chromosomes. Elle permet la répression de l'expression des transposons par l'inaccessibilité de l'ADN aux complexes protéiques. D'un autre côté, l'hétérochromatine facultative est réversible et est fondamentale pour une expression ciblée du génome.

Aujourd'hui, les principales marques associées à chacune des formes de la chromatine sont connues chez les eucaryotes. Ainsi, les méthylations des lysines K4, K36 et K79 ou l'acétylation de K9 de l'histone H3 sont généralement liées aux gènes transcriptionnellement actifs, alors que les méthylations des lysines K9 et K27 de H3 sont rattachées aux gènes inhibés (Tableau 2) (Lachner et Jenuwein 2002; Dillon et al. 2005).

Histone lysine	Function(s)
H1 K26	Transcriptional silencing
H3 K4	Transcriptional activation Transcriptional activation and elongation Transcriptional activation Transcriptional activation (in conjunction with ASH1-mediated methylation of H3 K9 and H4 K20)
H3 K9	Heterochromatic and euchromatic silencing; DNA methylation Euchromatic silencing; DNA methylation Euchromatic silencing Heterochromatic silencing; DNA methylation Heterochromatic silencing; DNA methylation Transcriptional activation (in conjunction with ASH1-mediated methylation of H3 K4 and H4 K20)
H3 K27	Euchromatic silencing Euchromatic silencing
H3 K36	Transcriptional elongation and silencing Transcriptional regulation
H3 K79	Demarcation of euchromatin
H4 K20	Cell cycle-dependent silencing, mitosis and cytokinesis [52,53] Heterochromatic silencing Transcriptional regulation Transcriptional activation (in conjunction with ASH1-mediated methylation of H3 K4 and H3 K9) Recruitment of checkpoint protein Crb2 to sites of DNA damage

Tableau 2 : Localisations et fonctions des méthylations post-traductionnelles principales.

Abréviations : *At*, *Arabidopsis thaliana*; *Dm*, *Drosophila melanogaster*; *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*; *Nc*, *Neurospora crassa*; *Sc*, *Saccharomyces cerevisiae*; *Sp*, *Schizosaccharomyces pombe* (Dillon et al. 2005).

1.3.2. L'hétérochromatine constitutive

Historiquement, l'hétérochromatine constitutive a été considérée comme inutile et peu importante car située dans des régions génomiques structurales ne contenant aucun gène. Pourtant, depuis peu, il a été avéré qu'elle est profondément impliquée dans la stabilité génomique et représente un élément majeur de défense génétique en évitant l'expression des éléments mobiles invasifs souvent responsables de maladies chez les humains (Sharma, Kelly et Jones 2010; Grewal et Jia 2007). De nombreux mécanismes complexes interdépendants sont impliqués dans la mise en place et le maintien de cette hétérochromatine. Les modifications post-traductionnelles des histones, en particulier la déacétylation globale des histones et la marque H3K9me3, mais également la méthylation des cytosines de l'ADN (5mC) sont les principaux facteurs mis en jeu pour sa formation chez les eucaryotes. Aussi, on retrouve une co-localisation marquée de ces acteurs au niveau de l'hétérochromatine constitutive.

1.3.2.1 Modification de la structure chimique de l'ADN : la méthylation des cytosines

Le degré de méthylation, sa distribution et sa fonction peuvent varier d'une espèce de champignon à l'autre (Zemach et al. 2010). Elle peut donc avoir divers effets

sur les processus biologiques fondamentaux des champignons tels que l'expression des gènes, la recombinaison génétique, la régulation de la structure chromatinienne ou encore l'inhibition de la transcription des éléments transposables (TE). Par exemple, le génome de *Neurospora crassa* contient 1,5% de méthylation alors que chez *Aspergillus flavus*, la quantité est négligeable (Foss et al. 1993 ; Liu et al., 2012). Toujours chez *N. crassa*, la méthylation de l'ADN sert principalement comme mécanisme de défense du génome pour éteindre les éléments transposables et les séquences répétées grâce au système RIP (« Repeat-Induced Point Mutation ») (Selker et al. 2003). *Magnaporthe oryzae* utilise également la méthylation de son ADN comme d'une régulation épigénétique au sein d'un système de défense génétique mais aussi comme régulation au cours de son développement. Ainsi, sa reproduction asexuelle dépend d'une répartition normale du méthylome le long du génome et les patrons de méthylation évoluent au cours du cycle pathogénique du champignon sans pour autant qu'une relation de régulation stricte entre les deux ait été démontrée (Jeon et al. 2015). En revanche, chez *Candida albicans*, la méthylation de l'ADN est très présente sur les gènes régulateurs de la transcription sans pour autant être présente sur les séquences répétées (Mishra, Baum et Carbon 2011).

Comme abordé plus haut dans ce manuscrit, trois ADN méthyltransférases de maintenance (DIM2, DNMT1 et DNMT5) et une ADN méthyltransférase *de novo* (RID) ont été identifiées chez les champignons. La combinaison des méthyltransférases 5mC est très diverse en fonction du champignon étudié. En effet, les basidiomycètes possèdent en majorité les gènes pour DNMT1 et DNMT5. Chez les ascomycètes en revanche, la combinaison la plus fréquente de gènes codant pour ces DMT est la présence des gènes pour DIM2, DNMT5 et RID ou seulement DIM2 et RID (Bewick et al. 2019). Jusqu'à présent, seule la présence des protéines DIM2 et RID a été démontrée chez *Trichoderma reesei* (Li et al. 2017).

DNMT1 et DNMT5 ciblent les séquences CG symétriques, avec une spécificité des cytosines internucléosomiques pour DNMT5, alors que DIM2 va se diriger vers les séquences répétées et leurs environs ainsi que des séquences marquées par H3K9me3 durant la croissance végétative (Bewick et al. 2019; Gladyshev 2017).

La protéine RID quant à elle est responsable des méthylations *de novo* de l'ADN et fait partie du système de défense génétique RIP cité plus haut et propre aux champignons (Hood, Katawczik et Giraud 2005; Freitag et al. 2002; Li, Chen et Wang 2018; Malagnac et al. 1999; Malagnac et al. 1997). Il s'agit d'un moyen de contrôle des éléments d'ADN invasifs (transposons, virus, séquences répétées etc...). Il a été observé pour la première fois chez *Neurospora crassa* entre la fertilisation et la karyogamie lors de son cycle de reproduction sexuée (Cambareri et al. 1989; Selker et al. 1987). Le mécanisme de défense opéré par RIP repère les séquences répétées et effectue une mutation sur les deux brins d'ADN, au nucléotide près (Hane et al. 2015). Un changement des cytosines en thiamines est opéré et environ 30% des C:G dupliqués peuvent être mutés en un seul cycle (Holliday et Pugh 1975). Ceci résulte en des régions

génomiques repérables car riches en A:T. Les méthylations se font donc sur les cytosines n'ayant pas été altérées par ces mutations (Du et al. 2015). Gladyshev et Kleckner (2014) ont démontré que la fréquence des mutations était influencée par la distance physique ainsi que l'accessibilité des séquences répétées. L'impact sur le développement et les mécanismes de ce système de défense sont très variables en fonction du champignon. En effet, si chez *Podospora anserina* (Grognet et al. 2019), *Trichoderma reesei* (Li, Chen et Wang 2018) ou *Aspergillus nidulans* (Lee et al. 2008) le RIP est indispensable pour la reproduction sexuée, ce n'est pas le cas chez *Neurospora crassa* (Gladyshev et Kleckner 2017).

Le RIP a suscité beaucoup d'intérêt car il a d'abord permis de formuler un modèle d'héritabilité et de maintenance des patrons symétriques de méthylations observés chez les mammifères (Holliday et Pugh 1975). Néanmoins, il a été démontré plus tard que chez *Neurospora crassa* la méthylation de l'ADN n'était pas limitée aux motifs symétriques tels que les sites CpG ou CpNpG et qu'une méthylation *de novo* des régions était possible (Singer, Marcotte et Selker 1995). Zemach et al. (2010) ont démontré que les schémas de méthylation peuvent être entièrement reprogrammés au cours du développement fongique dans les tissus végétatifs.

Chez de nombreux champignons, les méthylations des cytosines peuvent s'étendre en zones enrichies de longueur et de niveau de méthylation très variables. Chez *Phycomyces blakesleeanus*, ces clusters ne représentent que 0,04% du génome alors que chez *Agaricus bisporus*, 9,90% du génome est occupé par ces regroupements méthylés. De plus, si chez *A. bisporus* ces clusters peuvent s'étendre jusqu'à des longueurs de 142 kb, des clusters à peine plus grands que 230 pb ont été observés chez *Phanerochaete chrysosporium* (Bewick et al. 2019). Typiquement, ces clusters entourent des régions pauvres en gènes ou bien riches en éléments transposables avec une préférence marquée pour les sites CG ou CHH (H : A, T et C) (Jeon et al. 2015).

1.3.2.2 Modifications des protéines structurales : modifications post-traductionnelles des histones

En plus des méthylations des cytosines, certaines modifications des histones jouent un rôle clef dans la mise en place de l'hétérochromatine constitutive. Nous allons nous concentrer ici sur la marque H3K9me3 qui en est une ambassadrice reconnue.

Chez *Neurospora crassa*, la triméthylation de H3K9 est initiée par la méthyltransférase DIM5 (FgKMT1 chez *Fusarium graminearum*), faisant partie du complexe DCDC composé des protéines DIM-5, DIM-7, DIM-9, CUL4 et DDB1. Afin que le complexe protéique puisse apposer la marque, la déphosphorylation de la sérine 10 de l'histone H3 (H3S10) est nécessaire (Fischle et al. 2005) (Figure 9). Pour cela, la protéine phosphatase PP1 va permettre la désinhibition de l'action de DIM-5. Puis,

DIM-5 associé au reste du complexe DCDC, va reconnaître les séquences riches en AT résultant du RIP et trimétyler les histones H3 associées (Du et al. 2015). La protéine de l'hétérochromatine 1 (HP1), grâce à son chromodomaine, s'associe à la nouvelle marque. HP1 recrute alors l'ADN méthyltransférase DIM-2 grâce à son chromo-shadow domain. DIM-2 va alors apposer et étendre aux régions voisines les méthylations sur les cytosines de l'ADN (Figure 9). Il a été démontré que la suppression de *dim-2* chez *N. crassa* n'affecte pas l'activité de DIM-5 ni de HP1 et n'implique pas de redistribution de H3K9me3. Au contraire, chez le mutant *hpo* (« *heterochromatin protein one* ») une perte complète des méthylations de l'ADN dans les tissus végétatif est observée. De plus, HP1 est normalement co-localisée au niveau de H3K9me3, même au niveau de zones ponctuelles d'hétérochromatine. Or, chez les mutants *dim-5* ayant perdu l'intégralité des tri-méthylation de H3K9, HP1 est également absente et les méthylations des cytosines ne sont pas détectables. Il existe donc une route de régulation unidirectionnelle des histones vers l'ADN avec une action de DIM-5 et HP1 en amont de DIM-2 (Freitag 2017; Honda et Selker 2008; Lewis et al. 2009) avec HP1 créant un lien direct entre H3K9me3 apposée par DIM-5 et la méthylation de l'ADN faite par DIM-2 (Rountree et Selker 2010).

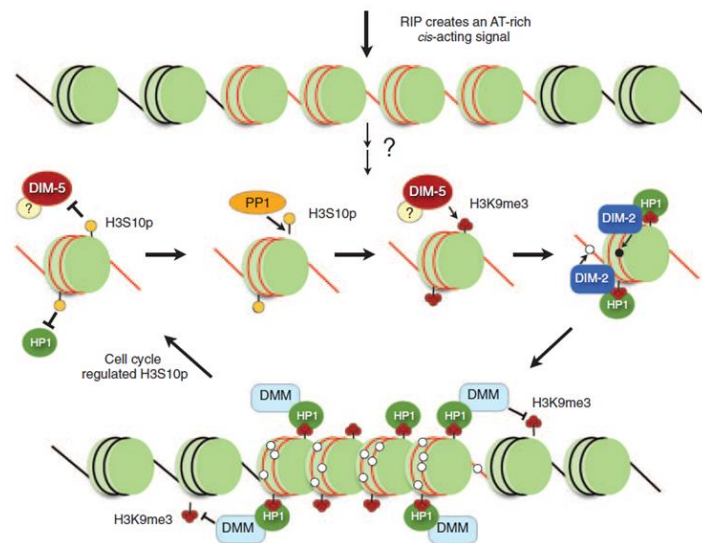


Figure 9 : Modèle de la formation et de la propagation de l'hétérochromatine.

Les mutations créées par le mécanisme RIP forment une zone riche en AT émettant un signal cis pour la formation de novo de l'hétérochromatine (Rountree et Selker 2010).

Néanmoins, le déploiement de l'hétérochromatine doit être maintenu dans des zones précises et déterminées (plus particulièrement les centromères et télomères) afin de ne pas atteindre des régions de l'euchromatine. Pour cela, chez *S. cerevisiae* et *S. pombe* par exemple, des éléments de frontières tels que des bouts redondants et inversés de gènes codant pour des ARN de transfert sont placés à l'intermédiaire entre

euchromatine et hétérochromatine. De façon surprenante, chez *N. crassa* la présence de HP1 et son association à la protéine DMM-1 (« *DNA Methylation Modulator-1* »), participe aux mécanismes permettant la ségrégation de l'hétérochromatine constitutive (Honda et al. 2010). De fait, la suppression de *dmm-1* entraîne d'importants défauts de croissance dus à l'étalement des méthylations de l'ADN sur les régions avoisinantes aux transposons « RIPés ». Ceci cause la mise sous silence de gènes critiques et la diffusion de l'hétérochromatine. Les effets néfastes de la disparition de *dmm-1* sont toutefois annulés lors de l'élimination de *dim-2* et un phénotype normal est rétabli (Tableau 3) (Rountree et Selker 2010).

<i>Protein</i>	<i>Domain(s)</i>	<i>Mutant phenotype</i>	<i>Principal activity</i>
RID	DMT	RIP defect	DMT and/or deaminase in ascogenous cells
DIM-5	HMT	No H3K9me3; no 5mC ^a	Trimethylation of H3K9
HP1	Chromo and Chromo-shadow	No 5mC	Adapter protein that binds H3K9me3 and recruits DIM-2 and DMM-1
DIM-2	DMT and 2 PXVXL-related	No 5mC	5mC in vegetative cells
DMM-1	JmjC; Cys-rich and 3 A:T-hooks	5mC patterns altered	?
DMM-2	Zn(II) ₂ Cys6	5mC patterns altered	?
PP1	Serine/threonine phosphatase	H3K9me3 and 5mC deficits	H3S10p phosphatase

Tableau 3 : Protéines responsables de la mise en place de l'hétérochromatine constitutive ainsi que les enzymes responsables pour son confinement dans les zones dédiées.

(Rountree et Selker 2010)

1.3.3. L'hétérochromatine facultative

Au sein de la division des ascomycètes, d'importantes différences dans la présence et la distribution de la marque H3K27me3, associée à l'hétérochromatine facultative, sont à noter. Ainsi, alors qu'elle est présente chez *Neurospora crassa*, *Magnaporthe oryzae*, *Cryptococcus neoformans* et *Fusarium graminearum*, aucune étude n'a encore pu démontrer sa présence chez *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* ou bien encore chez *Aspergillus nidulans* et *Ustilago maydis* (Connolly, Smith et Freitag 2013; Jamieson et al. 2016; Jamieson et al. 2013). Chez *N. crassa*, H3K27me3 recouvre environ 7% du génome, soit approximativement 800 gènes, répartis sur 200 domaines distincts, en majorité dans les régions télomériques (Jamieson et al. 2013). De même, chez *C. neoformans* on retrouve un enrichissement de la marque sur 5% du génome, concentré aux extrémités des chromosomes (Dumesic et al. 2015). En comparaison, H3K27me3 recouvre environ 30% des génomes de *F. graminearum* et *F. fujikuroi*, ce qui représente 3500 gènes, répartis sur des régions bien plus grandes et homogènes le long des chromosomes (Connolly, Smith et Freitag 2013; Studt et al. 2016). Cette répartition particulière des « blocs » de H3K27me3 coïncide avec les régions subtélomériques ainsi que d'anciennes régions subtélomériques intégrées au cœur du génome à la suite de fusions de chromosomes

et peu conservées au sein de l'espèce *Fusarium* (Connolly, Smith et Freitag 2013).

La méthylation de H3K27me3 est catalysée par l'histone méthyltransférase KMT6 (SET-7 chez *N. crassa*), formant partie du complexe protéique PRC2 (« *Polycomb Repressive Complexes 2* »). Trois protéines constituent ce complexe : KMT6, EED et SUZ12 (Shaver et al. 2010) (Figure 10). Une protéine supplémentaire, NcNPF, est nécessaire lors de l'action de ce complexe dans les régions subtélomériques (Jamieson et al. 2013). De plus, bien que ces protéines soient bien conservées chez les ascomycètes, la diversité du nombre de protéines finales formant PRC2 est importante. Ainsi, chez *F. graminearum* FgMsil, l'homologue de NcNPF, n'est pas requis. *C. neoformans* pour sa part, ne dispose pas de SUZ12 mais contient une protéine avec un chromodomaine se fixant sur H3K27me3 et étant impliqué dans des complexes HDAC (Dumesic et al. 2015). En outre, chez les champignons filamenteux, de nombreuses lignées ont perdu les gènes requis pour la synthèse des méthyltransférases du complexe PRC2 (Shaver et al. 2010).

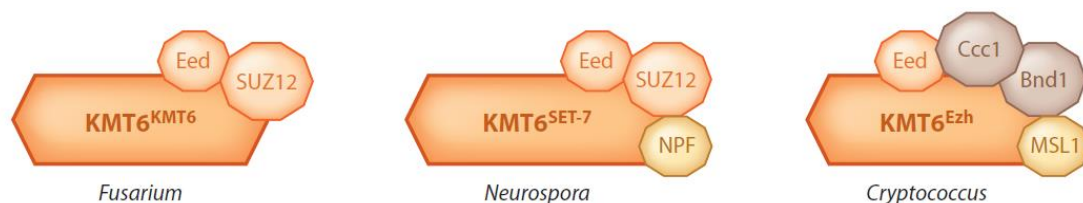


Figure 10 : Complexe PRC2 chez deux ascomycètes et un basidiomycète.

La conformation du complexe PRC2, responsable de la dépose de la marque épigénétique H3K27me3, est variable en fonction de l'espèce de champignon étudiée. (Freitag 2017)

Les protéines de PRC2 jouent un rôle majeur dans la régulation de l'expression génétique mais également du développement cellulaire. Ainsi, chez *Fusarium* la délétion de KMT6, EED ou SUZ12 entraîne une dérégulation du développement et de la différenciation avec des conséquences telles que l'infertilité, la perte de pathogénicité mais aussi la synthèse de métabolites secondaires. De manière globale, les mutants $\Delta kmt6$ affichent une levée de l'inhibition de la transcription des gènes (Connolly, Smith et Freitag 2013). Par exemple, chez *N. crassa*, la transcription d'une centaine de gènes, auparavant marqués ou non par H3K27me3, est activée. Pour autant, pas de phénotype hautement perturbé n'est observé chez ces mutants (Jamieson et al. 2016).

De façon tout à fait intéressante, Jamieson et al. (2016) ont également démontré que la perte des marques H3K9me3 de l'hétérochromatine constitutive cause une redistribution drastique des marques H3K27me2 et H3K27me3 de l'hétérochromatine facultative. Ceci permet le maintien des zones hyper condensées et une inhibition

conservée des gènes chez les mutants. Dans le même esprit, la délétion du gène codant pour HP1 entraîne une propagation des méthylations de H3K27 au niveau des zones de l'hétérochromatine constitutive provoquant la superposition avec H3K9me3. Ceci conforte l'hypothèse du rôle de HP1 pour la régulation de la diffusion de l'hétérochromatine chez les champignons filamenteux. En revanche, la désactivation des gènes pour une ou plusieurs protéines du complexe PRC2 n'a pas d'impact sur la redistribution des marques H3K9me3, ni 5mC, confirmant l'aspect facultatif de ce type d'hétérochromatine (Jamieson et al. 2016).

1.3.4. L'euchromatine

L'état ouvert de la chromatine se caractérise par un espacement irrégulier des nucléosomes le long des bras des chromosomes, une déplétion de l'histone « linker » H1, une forte présence des modifications post-traductionnelles H3K4me3, H3K36me3, H3K79me3 et H3K9ac ainsi que l'absence de la protéine HP1 (Dubey et Jeon 2017; van Vu, Pham et Nakayashiki 2013; Pham et al. 2015; Roth, Denu et Allis 2001). Chez les champignons et eucaryotes en général, le complexe COMPASS contenant l'histone méthyltransférase KMT2 est responsable des tri-méthylations de H3K4 et représente ainsi un facteur majeur de l'installation de l'euchromatine et de la transcription des gènes.

Cet état ouvert permet aux promoteurs et autres amplificateurs (« *enhancers* ») d'être accessibles aux facteurs de transcriptions, aux protéines régulatrices et l'accès des gènes à la machinerie transcriptionnelle. Il a été démontré que même si l'activité de la Polymérase II (Pol II) est initiée, la présence de nucléosomes, même incomplets, le long de l'ADN bloquera la progression de l'élongation (Chang et Luse 1997). Le complexe remodelleur de la chromatine SWI/SNF fait partie des acteurs intervenant au niveau des régions euchromatiniennes afin d'évincer les nucléosomes, de permettre leur glissement ou de les dérouler afin de faciliter l'ouverture de la fibre et l'accessibilité de l'ADN aux ARN polymérases pour l'expression des gènes. La structure du complexe SWI/SNF est hautement conservé au sein des eucaryotes bien que le nombre d'unités présentes varie en fonction de l'espèce étudiée (Savas et Skardasi 2018). Chez les levures, environ 6% des gènes sont régulés par ce complexe (Holstege et al. 1998).

1.4. Modifications Epigénétiques et biotechnologie industrielle

Les champignons sont utilisés par l'Homme pour leur capacité de biotransformation depuis des millénaires. Que ce soit pour la fabrication du pain, la production de vin, de miso ou de fromage bleu pour l'agroalimentaire ou bien d'antibiotiques, d'hormones ou d'enzymes pour les industries pharmaceutiques et chimiques. Les champignons sont une source majeure de molécules d'intérêt

représentant une source incommensurable d'améliorations dans notre vie quotidienne grâce à leur capacité à produire un éventail large de métabolites.

Dans ce chapitre nous nous intéresserons aux méthodes développées pour modifier les caractéristiques épigénétiques chez les champignons, soit pour la découverte de nouvelles molécules d'intérêts, soit pour améliorer les capacités de souches industrielles.

1.4.1. Modifications et expression épigénétique ou l'art d'agir sur les régulateurs de la chromatine

Des études récentes dans plusieurs espèces de champignons ont montré l'existence d'une régulation des clusters du métabolisme secondaire basée sur l'état de la chromatine. Cette régulation représente un niveau hiérarchique supérieur pour le contrôle coordonné de larges fragments chromosomiques. Ainsi, une des stratégies utilisées pour identifier de nouveaux métabolites d'intérêt est d'activer l'expression de clusters d'intérêts ou inconnus en modifiant les gènes agissant directement sur l'état de la chromatine (Strauss et Reyes-Dominguez 2011). Une découverte décisive dans le domaine a été la découverte de l'implication du gène *laeA*, codant pour une protéine impliquée dans la conservation des paysages chromatinien, chez l'espèce *Aspergillus*. L'invalidation puis la complémentation du gène de cette méthyltransférase putative a permis de découvrir son rôle comme régulateur de la synthèse de stérigmatocystine (Bok et Keller 2004; Strauss et Reyes-Dominguez 2011). Perrin et al. (2007) ont ensuite démontré que 13 clusters de métabolites secondaires sur 22 et qu'entre 20 et 40% des gènes impliqués dans leur biosynthèse étaient réprimés chez les mutants $\Delta laeA$ en comparaison avec la souche sauvage. Il a été proposé qu'un mécanisme modulé par LaeA supprime les marques d'hétérochromatine permettant le positionnement de marques d'activation. Le rôle de cette protéine dans la régulation globale de métabolites secondaires a été confirmé dans *T. reesei* (Seiboth et al. 2012).

D'autres régulateurs de l'épigénome ont été utilisés pour identifier ou déréguler des clusters de métabolites secondaires (Tableau 4). En effet, il a été démontré pour la première fois chez *A. nidulans* que la suppression des HDAC menait à l'activation de deux clusters produisant de la stérigmatocystine et de la pénicilline (Albright et al. 2015; Shwab et al. 2007). Chez *Calcarisporium arbuscula*, un autre champignon filamenteux, la suppression du gène *hdaA* a mené à l'altération générale du métabolisme secondaire et à la production de nombreuses molécules jusqu'alors inhibée (Mao et al. 2015). D'un autre côté, la disparition de l'histone méthyltransférase KMT6 entraîne la disparition des marque H3K27me3, un déficit de croissance important et la stérilité mais amène en contrepartie, l'expression constitutive de mycotoxines, pigments et autres métabolites chez *Fusarium graminearum* (Connolly, Smith et Freitag 2013). Chez *F. fujikuroi*, la disparition du gène *kmt6* entraîne également la transcription de quatre clusters de métabolites secondaires jusqu'à présent inhibés et l'accumulation de

nouveaux composés (Studt et al. 2016).

Classe de modifications	Action	Complexe/Protéine	Modification	Organisme	
Ecriture	Méthyltransférase	Bre2/Ash2/CclA (COMPASS)	H3K4me2/3	A. nidulans, A. oryzae, F. fujikuroi, F. graminearum, P. fici	
		Set1 (COMPASS)	H3K4me1/2/3	F. graminearum, F. verticillioïdes, M. oryzae	
		Clr4/Dim-5	H3K9me3	A. nidulans, E. festucae, F. verticillioïdes	
		Kmt6 (PRC2)	H3K27me3	E. festucae, F. graminearum, F. fujikuroi	
		Set2	H3K36me3	F. fujikuroi, F. verticillioïdes	
		Ash1	H3K36me3	F. fujikuroi	
		Dot1	H3K79me3	A. flavus	
		Hmt1	H4R3me2	A. flavus, F. graminearum	
	Acétyltransferase	Gcn5 (SAGA/Ada)	H3K9ac H3K14ac H3K18ac H3K27ac	A. nidulans, F. fujikuroi, F. graminearum, T. atroviride, U. maydis	
		Esa1	H4K12ac	A. nidulans	
Sas3		H3ac	F. graminearum, M. robertsii, P. microspora		
Hat1		H3/H4ac	P. microspora		
Effaceur	Deméthylase	Rph1	H3K36me3	A. nidulans F. fujikuroi	
		Jhd2	H3K4me3	A. flavus, A. nidulans	
	Deacétylase	Rpd3	H3/H4ac	A. nidulans	
		Hos2	H3/H4ac	A. nidulans, F. fujikuroi	
		Hda1	H3/H4ac	A. nidulans, C. arbuscular, F. fujikuroi, F. asiaticum, M. oryzae, P. chrysogenum, P. fici, P. microspora	
		Hos3	H3/H4ac	F. fujikuroi	
		Hst1	H4K16ac	A. nidulans	
		Hst4	H3K9K18K56ac	A. nidulans, A. oryzae	
		Lecteur	Ssp1 (COMPASS)	H3K4me3	A. oryzae
			HP1	Heterochromatin	A. nidulans, F. graminearum
Snt2	H3K4me3 H3ac		A. nidulans, A. flavus		

Tableau 4 : Modificateurs épigénétiques pour lesquels l'implication dans la régulation du métabolisme secondaire a été montrée.

(Pfannenstiel et Keller 2019)

La modification de la chromatine peut aussi agir sur l'expression des enzymes industrielles. Ainsi, chez *Penicillium oxalicum*, la délétion du gène *hepA* (homologue du gène codant pour HP1 chez *N. crassa*) entraîne chez le mutant une surexpression de 95% des gènes différentiellement exprimés en comparaison avec la souche sauvage. En revanche, dans les 5% sous exprimés nous retrouvons les gènes codant pour les cellulases CBH1, CBH2, EGI, EGII et BGL1 et la LPMO cel61A (Zhang, Qu et Qin 2016). Alors que la présence de HepA est normalement associée à un état hautement compact de l'ADN et une inhibition de l'expression génique, il semblerait que *hepA* joue ici un rôle de régulateur positif de l'expression de ces gènes.

Modifier les gènes qui contrôlent l'état de la chromatine semble donc être une piste intéressante pour identifier de nouveaux métabolites, ou mieux, contrôler leur

production. Cette stratégie pourrait donc aussi être utilisée pour la production d'enzymes chez *Trichoderma reesei*. Toutefois, la manipulation des régulateurs globaux ne fonctionne pas de manière unique chez l'ensemble des champignons et une étude cas par cas est nécessaire afin de saisir toute la complexité des mécanismes régulateurs.

1.4.2. Approche chimique de l'épigénétique

L'utilisation de petites molécules modificatrices de l'épigénétique (Figure 11) telles que la 5-azacytidine et l'acide hydroxamique suberoylanilide, un inhibiteur d'ADN méthyltransférase et un inhibiteur d'histone déacétylase respectivement, engendre des restructurations de grande ampleur de l'épigénome et du métabolome secondaire chez *Cladosporium cladosporioides* (Williams et al. 2008). Il est intéressant de noter également, que la 5-aza-2'-deoxycytidine entraîne chez *Humicola grisea* une accumulation de transcrits des gènes de cellobiohydrolases et du gène de xylanase *xyn2* avec une augmentation de l'activité de ce dernier (Manfrão-Netto et al. 2018). De même, l'inhibition des déacétylases par l'application de trichostatine A chez *Alternaria alternata*, *Penicillium expansum* et *Aspergillus oryzae* entraîne la production de plusieurs nouveaux composés (Shwab et al. 2007). Toutefois, l'utilisation de ces molécules reste essentiellement à but fondamental et exploratoire (Springer et Schmitz 2017) et est difficilement applicable dans un contexte industriel du fait de leurs effets secondaires importants sur la dérégulation du métabolisme général des organismes mais aussi du coût de leur utilisation à grande échelle.

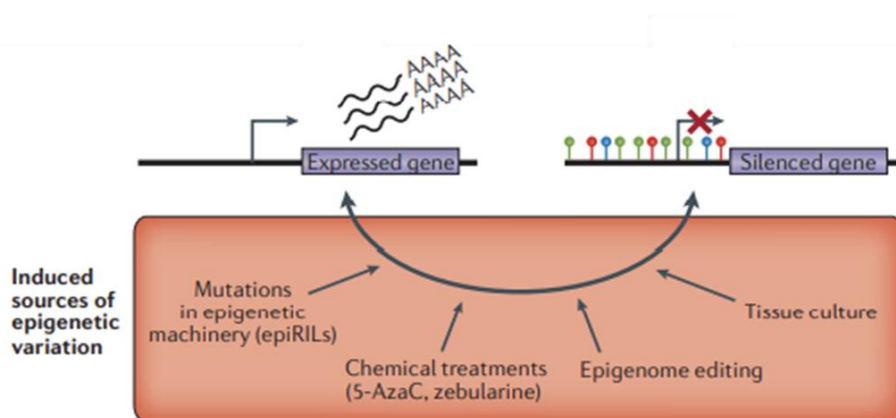


Figure 11 : Modifications épigénétiques induites par intervention humaine.
(Springer et Schmitz 2017)

1.4.3. La technologie CRISPR-Cas9 : un outil au service de la modulation des promoteurs

La technologie CRISPR-Cas9 offre également de nouvelles perspectives sur les modifications génétiques possibles pour l'amélioration des souches de champignons industriels ou de plantes à intérêts agronomiques. Ainsi, la délétion multiplexe de

gènes en utilisant plusieurs ARN guides dans un même système de transformation est de plus en plus courante. Ceci permet la suppression multiple et simultanée de voies métaboliques intermédiaires et éviter la co-production de protéines non voulues.

Des applications à l'épigénétique ont été récemment développées grâce à ce complexe CRISPR-Cas9 contenant une nucléase « *dead* » cas9 (dCas9), dont l'activité catalytique a été neutralisée par deux mutations (Figure 12). Ainsi, l'unité CRISPR de reconnaissance des séquences ciblées d'ADN fonctionnant toujours, le complexe va pouvoir se diriger au site choisi sans avoir la capacité de le modifier. Ce système CRISPR-dCas9 peut alors être fusionné avec une ADN méthyltransférase, déméthyltransférase ou autres protéines modificatrices et être utilisé comme convoyeur de complexes protéiques. Ainsi allié à sa précision, il va permettre une dérégulation épigénétique au gène près (Bikard et al. 2013; Springer et Schmitz 2017) (Figure 12).

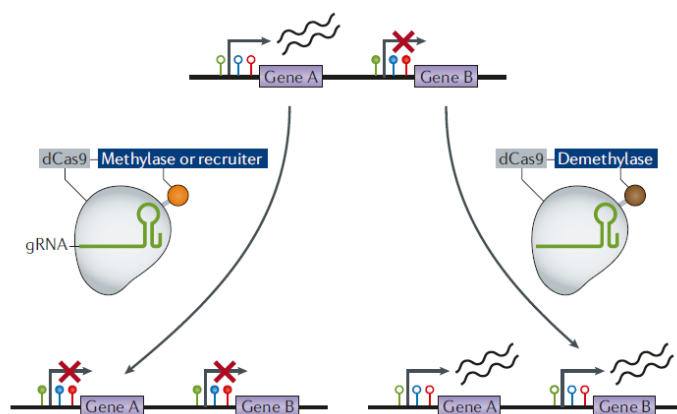


Figure 12 : CRISPR-Cas9, un outil d'ingénierie épigénétique.
(Springer et Schmitz 2017)

Ceci a également déjà été démontré chez les mammifères (Liu, X. Shawn et al. 2016; Amabile et al. 2016; Park, Keung et Khalil 2016) et ouvre des perspectives intéressantes pour la génération « d'épimutants » aux phénotypes intéressants. De plus, la répression de l'expression des transgènes est une clef de voute dans l'ingénierie des génomes des plantes à intérêts agronomiques. Le développement de systèmes permettant de modifier le profil épigénétique des nouveaux gènes seraient donc d'une valeur appréciable (Springer et Schmitz 2017). Si cette technique a été utilisée chez *S. cerevisiae* pour le transport et l'utilisation ciblée d'une histone acétyltransférase (Yamamoto et al. 2022), elle ne l'a pas encore été chez les champignons filamenteux et pourrait devenir une piste intéressante pour mieux réguler l'état chromatinien d'un gène sans toucher à la totalité de l'épigénome. Toutefois, cette méthodologie nécessite de connaître au préalable le paysage chromatinien du gène ciblé.

1.4.4. Position chromosomique et expression

Les techniques d'expression hétérologue ou homologue de gènes ou clusters de gènes sont déjà très répandues chez les champignons filamenteux. Chez *Trichoderma reesei* l'utilisation d'un système d'expression synthétique pour une production renforcée de lipase B de *Candida antarctica* a été mise en place avec succès en 2019 (A. Rantasalo et al. 2019). Un autre exemple d'expression hétérologue est le clonage du cluster de la pleuromutiline, antibiotique sécrété par le basidiomycète *Clitopilus passeckerianus*, chez *A. oryzae* (Bailey et al. 2016). L'expression hétérologue de gènes uniques ou de promoteurs choisis est source d'améliorations majeure chez les champignons producteurs d'enzymes. Ainsi, des gènes normalement sujets à la répression catabolique peuvent en être libérés par la délétion et la substitution de leur promoteur par un promoteur constitutivement exprimé (Ma et al. 2011).

Tout ceci pourrait être appliqué dans un contexte épigénétique par l'insertion de ces clusters ou de ces gènes uniques dans des régions chromatiniennes ouvertes. Cela fut expérimenté pour la première fois lors de la découverte des centromères et télomères. En effet, Henikoff et al. (1990) et Gottschling (1990) ont, de façon indépendante, démontré une expression différentielle de gènes hétérologues lorsque insérés dans des régions d'hétérochromatine ou d'euchromatine chez *Drosophila* et *Saccharomyces*, respectivement (Gottschling et al. 1990; Henikoff 1990). Toutefois, la connaissance des paysages chromatiniens des champignons filamenteux est indispensable pour cibler l'insertion de ces gènes dans des régions d'euchromatine.

1.4.5. Variations épigénétiques induites par l'environnement

Chez les plantes, la méthylation de l'ADN exhibe une variation naturelle qui peut influencer leurs caractéristiques phénotypiques. Des variants épigénétiques (ou épimutants) avec des intérêts agronomiques ont été identifiés dans plusieurs espèces et notamment dans le riz (Chen et Zhou 2013). Il semble cependant que la méthylation de l'ADN varie peu en fonction des conditions environnementales et cette stabilité pourrait être un facteur crucial pour l'utilisation d'approche épigénétique dans l'amélioration de l'agriculture. Ainsi, si un épimutant présente une reprogrammation globale de la méthylation adaptée à des conditions environnementales stressantes, ce variant pourrait être sélectionné et ses caractéristiques seraient stables dans le temps (Weigel et Colot 2012).

Il y a peu d'informations sur l'existence de variants épigénétiques chez les champignons filamenteux. Les cas décrits représentent plutôt un changement phénotypique impactant négativement les capacités du microorganisme (Silar 2019). Toutefois, la domestication de souches sauvages de *Penicillium* sur des fromages a montré une perte rapide de certains traits et un gain de capacité à produire des arômes d'intérêts. Si des mutations ont pu être détectées dans le génome des souches adaptées, aucune ne semble pouvoir expliquer les nouveaux phénotypes. Cette

évolution pourrait donc s'expliquer non par une évolution génomique mais par un héritage épigénétique transgénérationnel (Bodinaku et al. 2019). Une voie d'amélioration des champignons industriels pourrait donc être la sélection de variants épigénétiques et/ou une adaptation épigénétique des microorganismes à des conditions environnementales d'intérêts.

II. LE CHAMPIGNON *TRICHODERMA REESEI*

II.1. Origines et patchwork des champignons du genre *Trichoderma*

Les champignons du genre *Trichoderma* font partie de la division des ascomycètes, c'est-à-dire l'embranchement des champignons formant des ascospores contenues dans des asques. Ils sont dans la sous-division des pezizomycètes et sont classés dans les sordariomycètes. Associés pendant longtemps à leurs téléomorphes *Hypocrea*, il aura fallu attendre l'arrivée des techniques de PCR « *fingerprinting* » et les méthodes d'analyses des ADN ribosomaux (gènes ITS1, ITS2, 5.8S) pour en avoir la confirmation (Kuhls et al. 1996).

Si la première description du genre *Trichoderma* remonte à 1794 (Persoon 1794), aujourd'hui des centaines d'espèces ont été décrites et le genre *Trichoderma* fait désormais partie des champignons saprophytiques et biotrophiques les plus répandus au monde (Atanasova 2014). Depuis peu, un outil de caractérisation moléculaire basé sur un « code barre oligonuléotidique » (TrichOKEY), les biopuces phénotypiques (« *phenotype microarrays* ») (Bochner, Gadzinski et Panomitros 2001) ainsi que des ressources en ligne comme TrichoBLAST permet une identification sûre des nouvelles espèces identifiées (Druzhinina et Kubicek 2005). L'index Fungorum database (<http://www.indexfungorum.org/Names/Names.asp>) liste aujourd'hui 512 espèces pour *Hypocrea* et 433 pour *Trichoderma*. Par ailleurs, la Sous-commission internationale de taxonomie des *Trichoderma* et *Hypocrea* (ISTH : <http://www.isth.info/biodiversity/index.php>) répertorie 104 espèces caractérisées au niveau moléculaire et les classes en cinq sections.

S'il est si compliqué de distinguer les différentes espèces de *Trichoderma*, c'est que l'on retrouve les champignons de ce genre dans une très grande variété d'habitats mais avec des variations morphologiques mineures. On le retrouve fréquemment dans les sols naturels et agricoles mais aussi sur le bois et les plantes en décomposition (*T. hamatum*, *T. viride*, *T. piluliferum*, *T. reesei*, *T. asperellum* etc...). *Trichoderma* spp se développe également comme champignons endophytes sur les plantes vivantes (*T. caribbaeum*, *T. virens*, *T. brevicompactum* etc...) ou bien comme parasite de champignons macroscopiques (*T. aggressivum*, *T. longibrachiatum*, *T. harzianum* etc...). Assez fréquemment, on les retrouve aussi dans les environnements aquatiques

(*T. harzianum*, *T. matsushimae*, *T. aeroaquaticum* etc...). Depuis 2003, *T. longibrachiatum* a été identifié comme pathogène pour les mammifères, dont les humains en immunodépression (Kredics et al. 2003; Atanasova 2014). Le genre *Trichoderma* produit un large éventail de pigments pouvant aller du vert clair-jaune au rouge en passant par l'incolore. De plus, la pigmentation des conidies est également très variée avec différentes nuances de vert, gris et marron (Castle et al. 1998).

II.2. Qui est *Trichoderma reesei* ?

Trichoderma reesei est un champignon filamenteux isolé durant le Seconde Guerre Mondiale dans les Iles Salomon alors qu'il dégradait les équipements en coton des soldats américains. Une équipe de scientifiques a été envoyée sur place afin d'en déterminer l'élément responsable (Mandels et Reese 1957). La souche sauvage identifiée fut prénommée QM6a en référence au grade militaire Quarter Master et rattachée à l'espèce *Trichoderma viride* (Figure 13). Ce n'est que vingt ans plus tard que l'espèce fut reconnue comme à part entière et nommée d'après son découvreur Elwyn T. Reese.

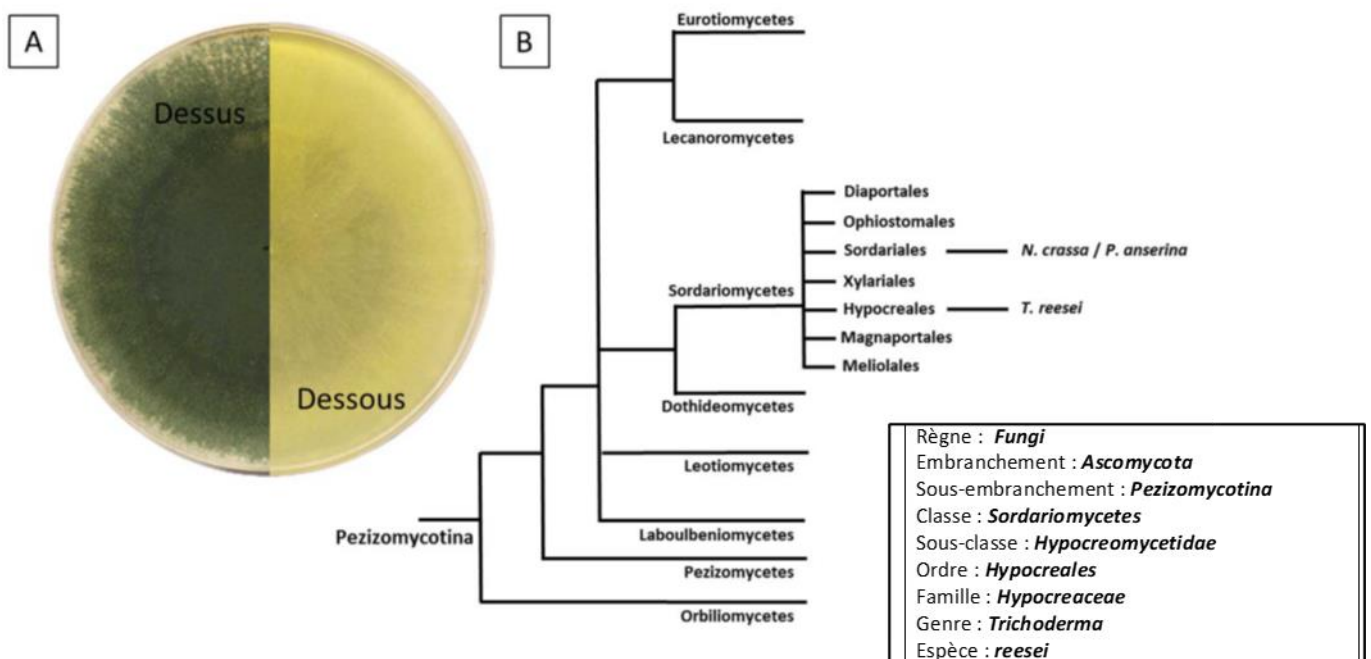


Figure 13 : Classement phylogénique de *Trichoderma reesei*.
 (Laetitia CHAN HO TONG 2017)

II.2.1. Cycle de vie, reproduction sexuée et conidiation de *Trichoderma reesei*

Comme décrit par Kuhls et al. en 1996, les champignons des genres *Trichoderma* et *Hypocrea* sont anamorphes/téléomorphes. Plus précisément, la souche sauvage de

T. reesei est l'isolat anamorphe de l'espèce *Hypocrea jecorina* (Kuhls et al. 1996). Aujourd'hui, même si certains articles font encore référence à *Hypocrea* ces deux genres sont regroupés sous le nom de *Trichoderma* après décision de la communauté scientifique (<http://www.isth.info/>).

De manière générale, il existe deux types de reproduction chez les champignons et que nous retrouvons chez *T. reesei*. La reproduction sexuée impliquant la fusion de deux noyaux haploïdes, et la reproduction asexuée par voie végétative.

II.2.2. Reproduction sexuée

Les ascomycètes filamenteux ou *Pezizomycotina* sont hermaphrodites, c'est-à-dire qu'ils différencient à la fois les structures sexuées mâles et femelles et peuvent féconder ou être fécondés (Leslie et Klein 1996). Chez *T. reesei* la reproduction sexuée est de type hétérothallique bipolaire. Il existe donc deux types sexuels différents, déterminés par deux idiomorphes au locus *MAT* du chromosome III : *MAT1-1* et *MAT1-2*. La reproduction sexuée s'effectue entre individus de types sexuels compatibles : les structures femelles (ascogones) d'une souche *MAT1-1* ne peuvent être fécondées que par les gamètes mâles *MAT1-2* et vice versa (Figure 14).

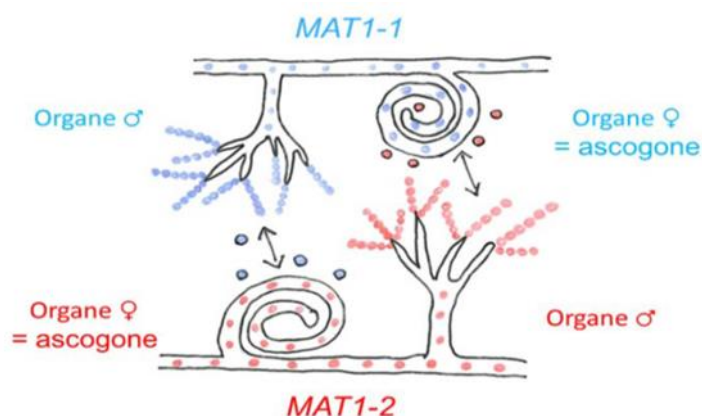


Figure 14 : Reproduction sexuée chez un champignon ascomycète filamenteux hétérothallique et hermaphrodite.

(Laetitia CHAN HO TONG 2017)

Parmi les isolats naturels, seuls les thalles des ascospores de la souche hétérocaryotique CBS999.97 sont capables de former des fructifications en confrontation avec une souche de type sexuel opposé. La souche QM6a de *T. reesei* est de type sexuel *MAT1-2* (Seidl et al. 2009) et peut donc être croisée avec une ascospore CBS999.97 de type sexuel *MAT1-1*. La stérilité de QM6a femelle provient de la présence d'une version non fonctionnelle du gène *IDC1*. La fertilité de la souche QM6a peut être restaurée en introduisant une version fonctionnelle du gène *IDC1* provenant de la souche CBS999.97 (Linke et al. 2015).

II.2.3. Reproduction asexuée

Outre les intérêts académiques pour la reproduction asexuée (ou conidiation) des champignons, la production de spores a été extensivement étudiée à des fins commerciales. En effet, les *Trichoderma* spp utilisés comme agents de biocontrôle contre les maladies des plantes sont vendus sous forme de spores (Chernin et Chet 2002) afin de faciliter leur conservation. Ainsi, comprendre les facteurs contrôlant leur production est d'un intérêt majeur pour leur production optimale et leur commercialisation.

La reproduction asexuée est majoritaire chez les champignons filamenteux et implique la production de spores après un temps de croissance végétative. Les spores contribuent à la survie de l'organisme dans des conditions environnementales défavorables à la croissance et permettent la dispersion du matériel génétique d'un site vers de multiples autres, plus favorables. La transition d'un mycélium vers la spore est dépendant de plusieurs facteurs combinés (Steyaert et al. 2010). Le pH du milieu, la lumière, la présence de calcium extracellulaire, la présence de blessures physiques du mycélium, de composés organiques volatiles ou bien le ratio carbone/azote (C : N) du milieu sont autant de paramètres importants pour le déclenchement de la conidiation chez *Trichoderma* et la plupart des autres *Pezizomycotina*. Une alternance des cycles lumière/obscurité a été démontrée comme optimale pour l'induction de la sporulation (Betina et Zajacová 1978; Gutter 1957; Gressel et Hartmann 1968; Schmoll, Esquivel-Naranjo et Herrera-Estrella 2010). Lors de ces expériences, il a été découvert que seules les cellules les plus jeunes, situées en périphérie du mycélium répondent à la stimulation lumineuse et sporulent. Ceci résulte en la formation de patrons circulaires de photoconidiation correspondant aux périodes d'obscurité précédant une exposition à la lumière.

Le mycélium ou hyphe végétative produit un appareil reproducteur asexuel vertical qui va ensuite se segmenter en plusieurs cellules indépendantes. Le développement de cellules spécialisées pour la production de conidies, appelées phialides, prend place au niveau de ces embranchements. Les conidies vont ensuite se pigmenter puis mûrir jusqu'à leur libération (Figure 15) (Betina et Zajacová 1978). Il est important de noter que chez *T. reesei*, les conidies sont toujours mononuclées.

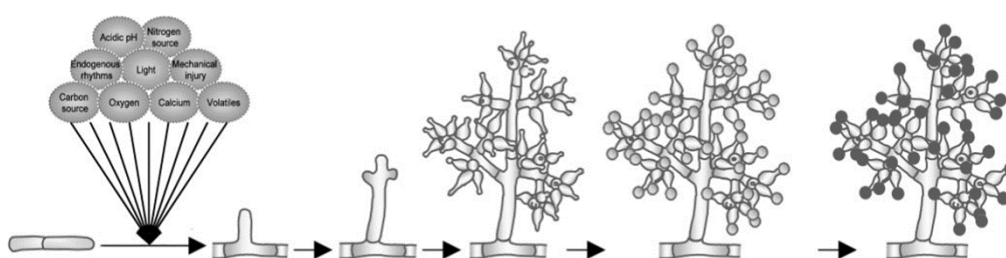


Figure 15 : Patron morphologique de *Trichoderma atroviride* après une alternance entre lumière et obscurité 12h/12h.

(Steyaert et al. 2010)

II.2.4. La production de cellulases chez *Trichoderma reesei*

En tant qu'organisme saprophyte, *T. reesei* possède un large éventail d'enzymes extracellulaires capables de dégrader les parois végétales et autres biomatériels (lignocellulose et polysaccharides dérivés etc...) afin de les utiliser comme sources d'énergie et de carbone. Ces enzymes communément appelées « cellulases » rassemblent en fait une variété d'enzymes différentes.

Ainsi, nous retrouvons les cellulases à proprement parlé, qui regroupent cinq endoglucanases (EGI, EGII, EGIII, EGVI, EGV), deux exoglucanases (CBHI et CBHII) et deux β -glucosidases (BGLI/Cel3a, BGLII/Cel1a) qui permettent l'hydrolyse de la cellulose (Derntl, Mach-Aigner et Mach 2014) (Figure 16), composant principal de la mixture restante après pré-traitement de la biomasse durant le processus de fabrication de bioéthanol de seconde génération. Les enzymes CBHI, CBHII, EGI et EGII dans une proportion approximative de 60 : 20 : 10 : 10, peuvent représenter à elles seules jusqu'à 90 % de la quantité totale de protéines sécrétées alors que les β -glucosidases représentent moins de 1 % (Herpoël-Gimbert et al. 2008; Markov et al. 2005; Margeot et al. 2009; Gusakov 2011).

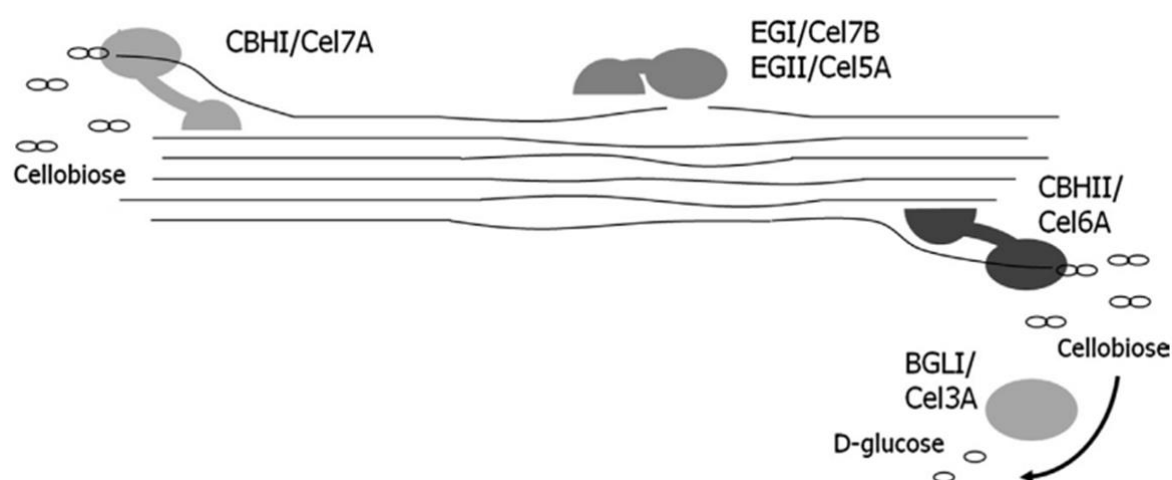


Figure 16 : Schéma de dégradation de la cellulose par les cellulases principales.
(Derntl, Mach-Aigner et Mach 2014)

En plus des cellulases, nous retrouvons également dans le cocktail enzymatique de *T. reesei*, des hémicellulases qui hydrolysent l'hémicellulose (hétéropolymère constitué de pentoses) et qui sont représentées majoritairement par les xylanases XYNI, XYNII, XYNIII et XYNIV, la mannanase MANI et la xyloglucanase Cel74A.

Enfin, les enzymes auxiliaires LPMO (« *Lytic Polysaccharide MonoOxygenases* ») aident à la déconstruction des substrats cellulolytiques par mécanismes oxydatifs et catalysent ainsi la réaction de dégradation.

Ces enzymes cellulolytiques sont aussi appelées « CAZymes » de l'anglais « *Carbohydrate-Active enZymes* » et sont donc caractérisées par leur action sur les carbohydrates. Cette famille compte cinq classes d'enzymes : les glycosides hydrolases (GH), qui représentent la majorité des CAZymes, les GlycosylTransferases (GT), les Polysaccharides Lyases (PL), les Carbohydrate Esterases (CE) et enfin les enzymes Auxiliaires. Il faut toutefois noter que cette appellation particulière ne compte pas seulement des enzymes participant au métabolisme des sucres complexes mais aussi à leur synthèse, et à leur modification. Nous retrouvons par exemple de nombreuses CAZymes dans la synthèse des parois cellulaires. Récemment les LPMO ont été rajoutées dans la liste des CAZymes du fait de leurs actions synergiques avec ces dernières.

D'un point de vue de la production enzymatique, la souche QM6a issue de l'environnement n'est pas suffisamment performante pour son exploitation. Ainsi, de nombreuses stratégies d'amélioration ont été mises en place au cours du temps et ont permis une amélioration du cocktail enzymatique et de la productivité de synthèse des enzymes extracellulaires.

II.3. Historique du séquençage de la souche QM6a de *T. reesei*

Plusieurs campagnes de séquençage de l'isolat naturel ont abouti à l'assemblage complet de son génome. Ainsi, en 2008, le Joint Genome Institute (JGI) séquence en grande partie le génome de 33,4 Mb et l'assemble en 89 scaffolds avec 9129 gènes prédits (Martinez et al. 2008). Puis, Marie-Nelly et al. (2014) suivi par Jourdiier et al. (2017) utilisent la technique de capture de conformation 3D de la chromatine (Hi-C) et l'outil bioinformatique GRAAL (« *Genome Re-Assembly Assessing Likelihood* ») pour créer une matrice de contact des séquences d'ADN et organiser les scaffolds en 7 chromosomes entiers, en accord avec le caryotype obtenu par (Carter et al. 1992). Ce nouvel assemblage a permis l'identification partielle des scaffolds correspondant aux centromères (zones contenant un pourcentage en AT supérieure à 67%) mais les intervalles entre les différents scaffolds ne sont pas connus.

Ce n'est qu'avec la combinaison de la technologie Pacific Biosciences (PacBio) et celle du séquençage Illumina que Li et al. (2017) réussissent à combler les zones entre scaffolds, compléter les régions télomériques et améliorer la qualité du premier séquençage. Finalement, le génome se révèle plus grand que précédemment (34,9 Mb) et compte cette fois 10 877 gènes soit 1630 nouveaux gènes. La plupart des séquences nouvellement séquencées correspondent aux centromères, aux 14 télomères et à 2329 zones riches en AT. Ces zones sont associées au chargement de la cohésine sur l'ADN

de *Trichoderma reesei* (permettant aux deux chromatides sœurs de garder leur cohésion lors de la méiose et mitose) et mesurent généralement plus de 3 kb. Il existe chez la QM6a 167 de ces blocs riches en AT, séparés en moyenne par une distance de 13 kb. Toutefois, de nombreux blocs riches en AT plus petits (de 500 pb à 3 kb) sont distribués le long des bras des chromosomes (Li et al. 2017).

En 2008, un premier inventaire des enzymes actives sur les carbohydrates (CAZymes) de *T. reesei* a recensé seulement 200 glycosides hydrolases (GH) contre 211 en moyenne chez les autres sordariomycètes étudiés (Martinez et al. 2008). Les auteurs observent aussi que les CAZymes ne sont pas réparties aléatoirement mais regroupées dans certaines parties du génome. La comparaison de la synténie avec d'autres génomes montre que ces concentrations d'enzymes ont pour origine une relocalisation des gènes et non une duplication (Martinez et al., 2008). Un tiers d'entre eux sont regroupés et contiennent des gènes impliqués dans le métabolisme secondaire du champignon. Plus récemment, une analyse de la position des gènes réalisée à partir du réassemblage du génome a montré que 13 des 35 groupes de CAZymes étaient localisées dans les régions subtélomériques (Marie Nelly et al, 2014). Pour finir, une analyse de fonction des gènes situés dans ces régions a été effectuée à partir du réassemblage du génome et démontre d'un enrichissement de gènes codant pour des protéases extracellulaires, des polyketide synthases (PKS)–non-ribosomal peptide synthase (NRPS) et des protéines mixtes (PKS–NRPS) (Druzhinina et al. 2016).

En 2019, le département Biotechnologie à IFP Energies nouvelles a entrepris une réannotation structurale experte du génome de la QM6a en utilisant les annotations de Martinez et al. (2008) et de Li et al. (2017) et des données de transcriptomes. Cette nouvelle annotation a permis de mettre en évidence de nombreuses erreurs d'annotations mais aussi des phénomènes d'épissage alternatif et la présence d'ARN non codants (article en cours de rédaction). Cette annotation contient 10160 gènes et sera utilisée pour les analyses des résultats épigénomiques et des transcriptomes du présent travail.

II.4. Les souches hyper-productrices de *Trichoderma reesei*

L'ensemble des souches actuellement utilisées par les industriels pour la production d'enzymes ont été générées à partir de la souche sauvage femelle stérile QM6a (Figure 17). En effet, bien que capable de synthétiser un cocktail enzymatique suffisamment puissant pour endommager les équipements de l'armée américaine, la souche QM6a est loin de pouvoir être exploitée à un haut niveau de production. Ainsi, de lourds efforts de la part des industriels ont été faits afin d'améliorer génétiquement *Trichoderma reesei* et augmenter sa production de cellulases. Les premières campagnes d'amélioration génétique ont été lancées en 1970 et étaient essentiellement basées sur la mutagénèse aléatoire (Viikari, Vehmaanperä et Koivula 2012). Elles ont permis de

créer des souches à forte capacité de sécrétion de cellulases avec des protocoles de production adaptés (Cherry et Fidantsef 2003). Au-delà de la productivité, la recherche s'intéresse également à l'amélioration des activités enzymatiques spécifiques et la capacité à produire des cellulases à partir d'une palette de sucres inducteurs plus variée (Seiboth, Ivanova et Seidl-Seiboth 2011) mais aussi à la modification de la morphologie du champignon pour faciliter le procédé de fermentation. Actuellement, l'ingénierie de souches peut s'effectuer à l'aide de méthodes de modification plus efficaces comme CRISPR-Cas9 (Liu et al. 2015; Harkki et al. 1991).

Aujourd'hui, une des souches hyper-productrices de *T. reesei* de référence dans le domaine public est la souche RutC30 (Eveleigh et Montenecourt 1979). Elle a été obtenue grâce à 3 étapes de mutagénèse dont le but final était de s'affranchir de la répression catabolique et d'augmenter la production de cellulases (Figure 17). En effet, l'expression de gènes codant pour les cellulases est sujette à deux niveaux de régulation : la répression catabolique et l'induction. Dans le premier cas, tant qu'il y a du glucose dans le milieu, les gènes permettant d'utiliser d'autres sources de carbones sont réprimés. A l'inverse, afin de déclencher l'induction de la production de cellulases l'organisme doit être placé en présence de cellulose ou de ses dérivés comme le sophorose ou le lactose (Kubicek, C., & Penttilä, M. 1998). Les souches industrielles utilisées à IFPEN sont toutes issues de la souche RutC30.

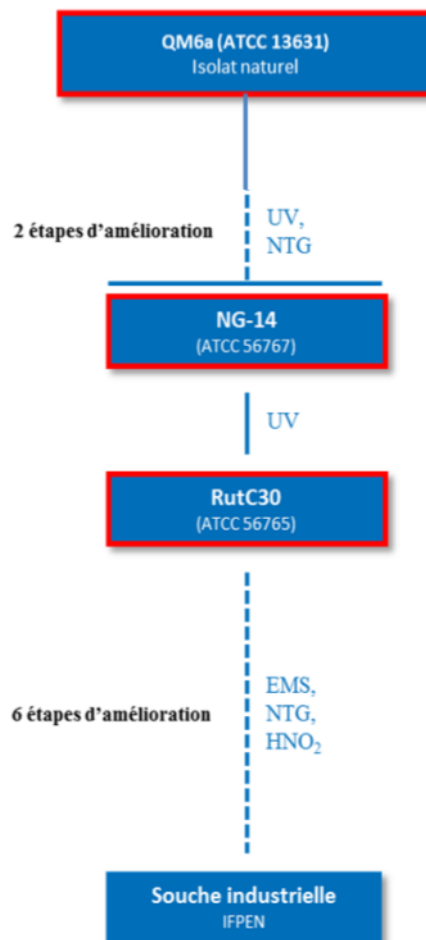


Figure 17 : Généalogie de la souche hyper-productrice RutC30.

Les agents mutagènes utilisés et le nombre de mutagénèses réalisées par rapport à la QM6a sont mentionnés. Les traits pleins correspondent à une seule étape de sélection tandis que les traits en pointillé correspondent à plusieurs étapes. Les souches entourées de rouge sont celles pour lesquelles un séquençage est disponible. UV : rayonnement ultra-violet ; NTG : nitrosoguanidine ; EMS : éthylméthylsulfonate ; HNO₂ : acide nitreux. (Thèse de Laetitia CHAN HO TONG, 2017)

Ainsi, à partir de la souche QM6a, la souche intermédiaire M7, plus disponible à ce jour, a été obtenue. Après traitement chimique sur cette dernière, une nouvelle souche hyper-productrice NG14 a été sélectionnée (Montenecourt et Eveleigh 1977). La souche hyper-productrice RutC30 décent directement de la souche NG14 après une nouvelle mutagénèse par UV et criblage pour une forte activité cellulolytique et une résistance au 2-deoxyglucose, un analogue du glucose non métabolisable (Montenecourt et Eveleigh 1977; Montenecourt et Eveleigh 1979).

Le séquençage du génome des souches a révélé de nombreux changements génétiques dans RutC30 par rapport à QM6a dont 179 SNV (« Single Nucleotide Variant »), 5 InDel (Insertion ou Délétion) et 8 variations structurales (4 translocations,

3 délétions et une inversion) (Le Crom et al. 2009; Vitikainen et al. 2010; Koike et al. 2013). Le caryotype de RutC30 établi par Jourdier et al. (2017) identifie les réarrangements chromosomiques avec deux fragments d'ADN du chromosome I transloqués vers les chromosomes III et VI et un fragment du chromosome VI transloqué sur le chromosome I (Figure 18). En plus de cela, une troncation de 2,5 kb du gène *cre1* responsable de la répression catabolique ainsi qu'une délétion de 85kb sur la partie télomérique du chromosome VI sont à dénombrer. Il faut noter que le caryotype de la souche NG14 est identique à celui de la souche RutC30 (Jourdier et al. 2017).

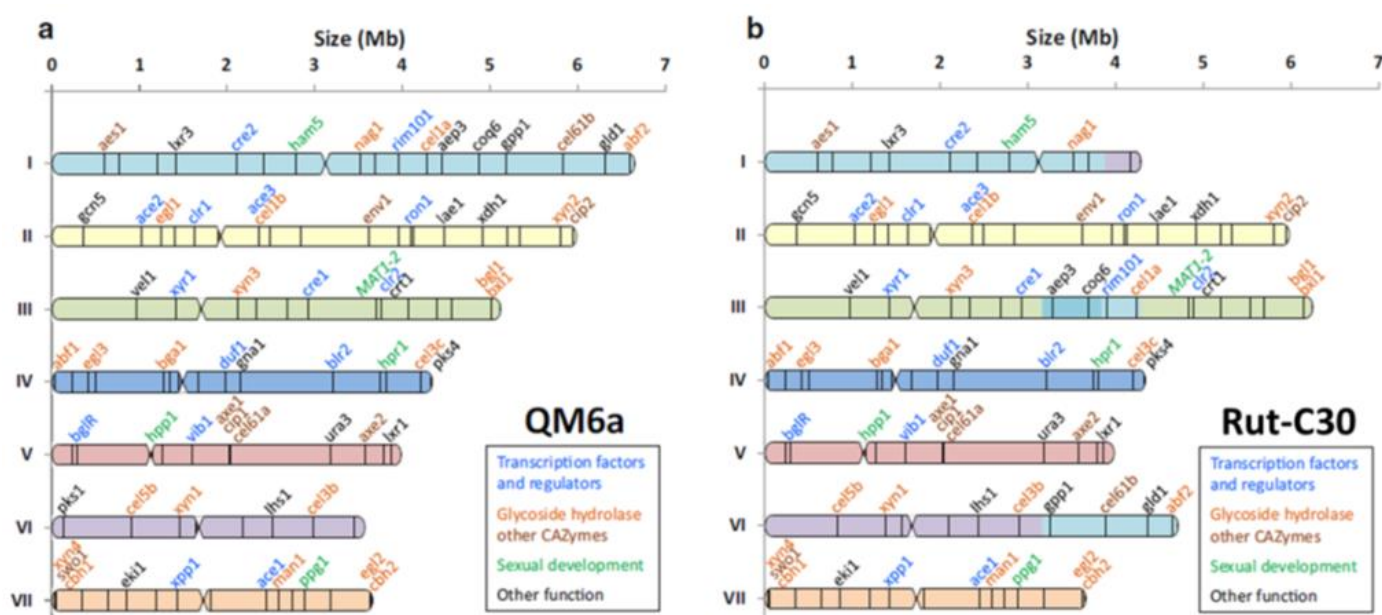


Figure 18 : Réassemblage des chromosomes de *T. reesei* des souches QM6a et RutC30.
(Jourdier et al. 2017)

S'il est difficile d'établir un lien formel entre mutation et phénotype de par le nombre important de modifications génétiques (Le Crom et al. 2009), un lien direct entre amélioration de la production de cellulases et génotype a été établi pour la troncation du gène *cre1* ainsi que pour la mutation du gène *gls2* codant pour la sous-unité alpha de la glucosidase II, responsable de la glycosylation des protéines dans le réticulum endoplasmique (Ilmén, Thrane et Penttilä 1996; Geysens et al. 2005). Il a également été établi que le facteur de transcription BglR contient un SNP responsable d'une production accrue de cellulase en présence de cellobiose (Nitta et al. 2012).

Une autre question concerne le rôle des translocations dans le phénotype hyper-producteur des souches RutC30 et NG14. La souche QM9978, obtenu après deux tours de mutagenèse de QM6a, ne produit plus de cellulases. Une étude génomique et transcriptomique dans cette souche a montré qu'une translocation entre deux

chromosomes était responsable. Or, le site de translocation est localisé à plus de 1000 pb du début du gène en dehors de la région promotrice. Les auteurs ont formulé entre autres hypothèses que le changement d'expression de *vib1* pourrait provenir d'un contexte épigénétique différent (Ivanova et al. 2017). Il est donc possible que les importants réarrangements chromosomiques observées dans les souches RutC30 et NG14 aient un rôle dans le phénotype hyper-producteur de ces souches (Nikitin et al. 2008).

II.5. Régulation génétique de la production de cellulases chez *Trichoderma reesei*

La régulation génétique de la production d'enzymes cellulolytiques chez *T. reesei* a été et est encore un sujet de recherche très étudié. Aujourd'hui, nous savons que de nombreux gènes sont impliqués dans ces processus régulateurs et les mécanismes d'activation ou de répression de la production sont loin d'être simples (Harman et Kubicek 1998; Shida, Furukawa et Ogasawara 2016).

Afin de déclencher la production enzymatique, le champignon doit être placé dans un environnement contenant au moins un inducteur de la transcription des gènes de cellulases. Plusieurs existent : la cellulose, le lactose, le xylose, le cellobiose etc... et de manière générale les polysaccharide de plantes (Harman et Kubicek 1998).

En présence de ces inducteurs, une régulation au niveau des promoteurs des gènes de synthèse des cellulases doit être réalisée par des facteurs de transcriptions dédiés. Nous allons ici nous concentrer sur les principaux (Figure 19).

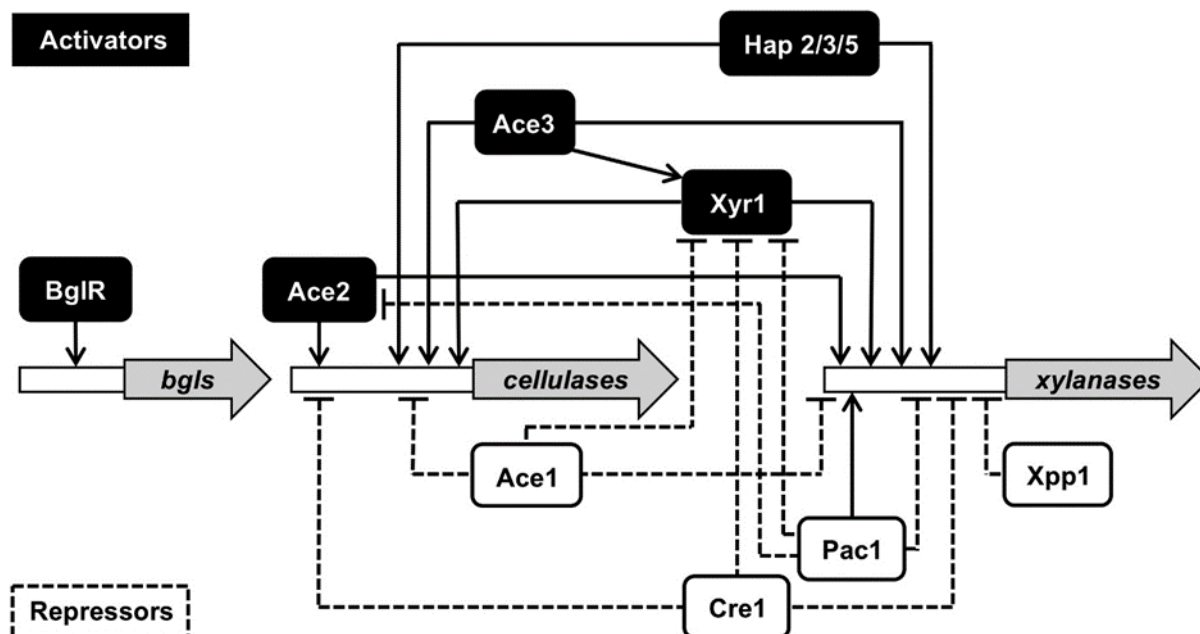


Figure 19 : Réseau de régulation de la production d'enzymes cellulolytiques chez *T. reesei*.
(Shida, Furukawa et Ogasawara 2016)

Le gène *xyr1* (« *Xylanase Regulator 1* ») code pour le facteur de transcription principal Xyr1 de l'activation des gènes de cellulases et d'hémicellulases. Ainsi, Stricker et al. (2006) ont démontré qu'en son absence l'expression des gènes de cellulases *cbh1*, *cbh2* et *egl1* était nulle lorsque le champignon est cultivé en présence de cellulose ou de sophorose. De même, ils ont démontré qu'il était nécessaire pour l'expression des hémicellulases *xyn1* et *xyn2* ainsi que la β -glucosidase *bgl1* dans des milieux inducteurs dédiés.

Également, le gène codant pour la protéine ACE2 (« *Activator of Cellulases 2* ») joue un rôle important dans la régulation positive de la synthèse de cellulases puisqu'il a été démontré que la délétion du gène *ace2* provoque une réduction pouvant aller de 30 et 70% des transcrits des principaux gènes de cellulases lorsque cultivé sur cellulose (Aro et al. 2001).

Le facteur de transcription BglR (« *Beta-Glucosidase Regulator* ») régule positivement l'expression des gènes spécifiques des β -glucosidases intracellulaires lorsqu'en présence de cellobiose, à l'exception de *bgl1*, qui est sous l'influence de XYR1.

Il existe également plusieurs répresseurs de l'expression des gènes impliqués dans la synthèse d'enzymes cellulolytiques. Un des plus important est le facteur de transcription CRE1 impliqué dans la répression catabolique par le glucose. Cette répression catabolique permet au champignon de privilégier l'absorption des sucres simples au détriment des sucres complexes grâce à l'inhibition de la synthèse d'enzymes associées à l'absorption de sources de carbone plus complexes (Kubicek et al. 2009). L'élimination de *cre1* avait déjà été obtenue par mutagenèse classique avec

une version tronquée du gène chez RutC30 et résultait en une synthèse accrue de cellulases à la fois en conditions d'induction et de non induction (Mello-de-Sousa et al. 2014; Montenecourt et Eveleigh 1979).

Malgré son nom, la protéine ACE1 (« *Activator of Cellulases 1* ») est en fait un répresseur de l'expression des gènes de cellulases *cbh1*, *cbh2*, *egl1* et *egl2* et d'hémicellulases *xyn1* et *xyn2* lorsque le champignon est cultivé en présence de sophorose ou de cellulose (Aro et al. 2003). De plus, il semblerait que les niveaux de transcription de *xyr1* soient supérieurs dans une souche délétée de *ace1* que chez une souche sauvage. Cela impliquerait que la répression de ACE1 sur la production des cellulases se fasse par l'intermédiaire d'une répression faite sur *xyr1* (Mach-Aigner et al. 2008)

D'autres facteurs de transcription activateurs ou répresseurs de cellulases existent, le Tableau 5 recense la majorité d'entre eux.

Nom	Fonction	Référence
ACE1	Répresseur de l'expression des gènes de cellulases <i>cbh1</i> , <i>cbh2</i> , <i>egl1</i> et <i>egl2</i> et d'hémicellulases <i>xyn1</i> et <i>xyn2</i>	(Aro et al. 2003)
ACE3	Essentiel pour l'expression des gènes de cellulases and xylanases	(Häkkinen et al. 2014)
ACE2	Active les CBH et EG	(Aro, Pakula et Penttilä 2005)
BGLR	Régule à la hausse les gènes de β -glucosidases mais son action peut déclencher le CCR	(Nakari-Setälä et Penttilä 1995)
Complexe HAP2/3/5	Participe à l'ouverture de la chromatine et active la transcription des gènes de cellulases.	(Zeilinger et al. 2001)
PAC1	Répresseur de la production de cellulases et de <i>Xyr1</i> et <i>Ace2</i>	(He et al. 2014)
PMH20	Activateur de la production de cellulases et d'hémicellulases	(Häkkinen et al. 2014)
PMH25	Activateur de la production de cellulases et d'hémicellulases	(Häkkinen et al. 2014)
PMH29	Activateur de la production de cellulases et d'hémicellulases	(Häkkinen et al. 2014)
XPP1	Régule la production des xylanases et active la production des hémicellulases	(Derntl et al. 2017)
CRE1	Represseur catabolique	(Kubicek et al. 2009)
VE1	Activateur de la production de cellulases	(Liu, Kuimei et al. 2016)
VEL2	Activateur de la production de cellulases	(Liu, Kuimei et al. 2016)
VIB1	Régule indirectement l'expression des gènes de cellulases par la régulation indirecte de CLR2	(Ivanova et al. 2017)
RXE1	Activateur de la production de cellulases. Modulateur de <i>Xyr1</i>	(Wang et al. 2019)

Nom	Fonction	Référence
Mat1-2-1	Activateur lié à <i>Xyr1</i>	(Zheng et al. 2017)
RCE1	Répresseur de la production de cellulases par antagonisme avec <i>Xyr1</i>	(Cao et al. 2017)
ARA1	Répresseur de la production de cellulases	(Benocci et al. 2019)
BLR1/BLR2	Augmente la transcription des gènes de cellulases mais peut altérer leur sécrétion	(Schmoll 2018)
AZF1	Activateur des cellulases et xylanases	(Antonieto et al. 2019)
XYL1	Nécessaire pour l'induction des β -galactosidases et de <i>cbh1</i> en présence de lactose	(Seiboth et al. 2007)
RCE2	Répresseur de la production de cellulases par compétition avec ACE3	(Chen, Yumeng et al. 2021)
CLR2	Activateur de la production de cellulases et hémicellulases	(Häkkinen et al. 2014)
XYR1	Activateur principal des gènes de cellulases et d'hémicellulases	(Kubicek et al. 2009)

Tableau 5 : Récapitulatif des facteurs de transcription impliqués dans la synthèse enzymatique.

II.6. Régulation épigénétique de la production de cellulases chez *Trichoderma reesei*

L'amélioration épigénétique des souches de *Trichoderma reesei* est aujourd'hui un domaine de recherche émergent. Les régulations épigénétiques doivent être décrites et comprises afin de pouvoir utiliser ce nouveau levier d'amélioration chez le champignon.

Ainsi, comme discuté plus haut, la suppression du gène régulateur *xyr1* chez *Trichoderma reesei* élimine toute induction de xylanases et cellulases chez l'organisme. Au contraire, sa surexpression entraîne une augmentation dans leur expression (Bischof, Ramoni et Seiboth 2016). Il a donc été de premier intérêt de manipuler l'expression de *xyr1* afin d'augmenter la production d'enzymes ou de rendre leur expression constitutive. De plus amples études à son sujet ont démontré son rôle majeur dans la modulation de la structure de la chromatine. De fait, Xyr1 interagit de manière directe avec le complexe modulateur de la chromatine SWI/SNF et permet l'éviction des nucléosomes au niveau des promoteurs des gènes *cbh1* et *cbh2* lors de l'expression des cellulases (Cao et al. 2019). La délétion de *xyr1* provoque une compaction dense de la chromatine résultant en l'inactivation des gènes. Ceci est dû à une modification d'accessibilité à l'ADN, déclenchant *in vivo* une hypersensibilité à la méthylation des sites de Xyr1 et une protection accrue sur les sites du répresseur

catabolique CRE1 (Mello-de-Sousa et al. 2015).

En effet, il a été démontré que CRE1 s'associe aux sites de reconnaissance situés dans les séquences proches des promoteurs de *cbh1*. En conditions de non-induction, des nucléosomes occupent les promoteurs alors qu'au moment de la mise en conditions d'induction de synthèse des cellulases, ces nucléosomes sont enlevés et repositionnés sur les sites de reconnaissance de CRE1. Ceci empêche alors l'association de la protéine avec les régions promotrices et active la transcription (Ries et al. 2014). Les mutants $\Delta cre1$ affichent une perte de positionnement des nucléosomes dans les promoteurs de *cbh1* et *cbh2* même en présence de glucose. Ceci indique l'importance de CRE1 pour un positionnement et une distribution correcte des nucléosomes dans le gène *cbh1* (Ries et al. 2014).

De plus, il est important de noter que la version tronquée de *cre1* chez la souche RutC30 n'a pas le même impact que si le gène avait été délété entièrement. Au contraire, la version *Cre1-96* de la RutC30 influence positivement la régulation des gènes grâce à son interaction avec Xyr1 pour garder un état chromatinien ouvert constitutivement (Mello-de-Sousa et al. 2014).

Dans le même esprit, Zeilinger et al. (2003) ont montré que le gène *cbh2* est également régulé au niveau de sa région promotrice par deux motifs séquentiels interdépendants fonctionnant comme un unique facteur épigénétique nommé CAE. Celui-ci est impliqué dans le positionnement des nucléosomes autour et sur les motifs pour la transcription comme la TATA box. Une mutation simultanée sur les deux motifs du CAE entraîne un déplacement des nucléosomes normalement situés sur la TATA box et autorisant ainsi la transcription. En fait, ces séquences spécifiques servent de sites de reconnaissance pour une protéine ou un complexe protéique dépendant de CRE1 et du complexe Hap2/3/5, intimement liés au positionnement des nucléosomes et à l'état chromatinien (Zeilinger et al. 2003). Une compréhension plus poussée de ces mécanismes épigénétiques au niveau même des promoteurs des gènes pourrait être clef pour permettre une dérégulation choisie et ciblée de l'épigénome local.

Pour finir, la méthyltransférase putative LAE1 pourrait être responsable de la modulation des marques de l'hétérochromatine H3K9me3 et de l'euchromatine H3K4me3 chez *Trichoderma reesei* selon Karimi-Aghcheh et al. (2013). D'un autre côté, Seiboth et al. (2012) ont démontré que l'action de LAE1 sur l'expression des gènes des CAZymes n'était pas due à une modulation des modifications post-traductionnelles des histones. En revanche, il semblerait que XYR1 et LAE1 soient interdépendants dans leur fonctionnalité. De plus amples recherches doivent donc être faites à ce sujet afin de démontrer si en effet, LAE1 et le complexe Velvet dont il fait partie jouent un rôle dans la régulation épigénétique de l'expression des gènes codant pour les cellulases chez *Trichoderma reesei*.

Ainsi, s'il a été démontré qu'il existe un lien entre épigénétique, production de cellulases et gènes régulateurs grâce à l'implication de ces derniers dans la plasticité

chromatinienne locale, l'enjeu est désormais de comprendre les mécanismes interdépendants entre facteurs de transcription, complexes remodeleurs, marques épigénétiques et expression génique tout en mesurant l'importance quantitative de ces interactions sur la production enzymatique. Si pour le moment aucune application directe de l'épigénétique pour l'amélioration des souches industrielles n'a à notre connaissance été mise en place, il semble que cette voie de recherche soit prometteuse pour le futur.

ENJEUX DE LA THESE

Dans le contexte de la production de bioéthanol de seconde génération, mon sujet de thèse propose de poursuivre et approfondir les travaux de biologie systémique initiés au département Biotechnologie d'IFPEN (Ivanova et al. 2017; Pirayre et al. 2020; Poggi-Parodi et al. 2014; Lichius et al. 2015; Portnoy, Margeot, Linke et al. 2011; Le Crom et al. 2009). Ces travaux basés sur des données transcriptomiques et génomiques ont guidé les stratégies actuelles d'amélioration du microorganisme reposant sur l'ingénierie génétique. Ainsi, grâce aux résultats obtenus, des nouvelles souches ont été développées conduisant à une production optimisée d'enzymes cellulolytiques (Brevet IFPEN FR2006495-N/Réf: 8096/10).

Ces dernières années, les connaissances obtenues en biologie systémique à IFP Energies nouvelles et provenant de la littérature suggèrent l'implication de modifications épigénétiques de la chromatine lors de la production de cellulases (Portnoy, Margeot, Seidl-Seiboth et al. 2011; Portnoy, Margeot, Linke et al. 2011; Poggi-Parodi et al. 2014; Jourdier et al. 2017; Zhang, Qu et Qin 2016; Karimi-Aghchah et al. 2013; Strauss et Reyes-Dominguez 2011; Sukumaran et al. 2021). Toutefois, les dynamiques chromatinienne prenant place lors de cette production n'ont pas été explorées. L'acquisition de telles connaissances est donc un levier important pour la définition de nouvelles stratégies visant à contrôler plus efficacement la production enzymatique de *Trichoderma reesei*.

Le projet de thèse vise donc à une meilleure compréhension des mécanismes épigénétiques intervenant dans le comportement de *T. reesei*, notamment lors de la production enzymatique. Cette étude a également pour but de mettre en évidence l'incidence sur les dynamiques épigénétiques qu'ont eu les mutagenèses résultant en l'hyper-productivité de la RutC30.

Dans un premier temps, des cartographies de la distribution des marques H3K4me3, H3K9me3 et H3K27me3 chez deux souches, l'isolat sauvage QM6a de *T. reesei* et la souche hyper-productrice de référence RutC30, dans un état d'induction (production de cellulases) et de non-induction (pas de production de cellulases) ont été réalisées partiellement (seule la cartographie de H3K4me3 a été menée jusqu'au bout) avec deux objectifs : (1) comprendre de façon individuelle chez les deux souches l'implication des mécanismes épigénétiques lors de la synthèse enzymatique et (2) si les mutations de la RutC30 ont été moteurs de changements dans l'épigénétique de cette dernière. Ces premières expériences ont permis d'évaluer les différences entre la souche sauvage et la souche hyper-productrice et de rendre compte des modifications épigénétiques qui surviennent lors de la production de cellulases pour chacune des souches. Si l'acquisition, l'analyse et l'intégration de données de séquençage haut débit de type génomique et transcriptomique sont devenus relativement courantes à IFPEN, les méthodes permettant d'obtenir des données épigénomiques (ChIP-Seq) mais aussi

leur exploitation et leur intégration dans un système multi-omique ne sont pas maîtrisées. Les résultats de ces expériences ont été complétés par des données de transcriptome.

Enfin dans un second temps, nous avons choisi de nous intéresser à des gènes majeurs dans la mise en place de l'hétérochromatine afin de déterminer si l'un d'entre eux jouait un rôle régulateur dans la production enzymatique de *Trichoderma reesei*. Les gènes *hp1* (protéine lectrice de H3K9me3), *dim2* (responsable de la marque 5mC), *dim5* (responsable de la marque H3K9me3), *kmt6* (responsable de la marque H3K27me3) et *rid* (responsable de la marque 5mC *de novo*) ont été invalidés dans la souche hyper-productrice RutC30 et la souche sauvage QM6a. Dans un contexte inducteur, la levée de compaction chromatinienne au niveau global du génome par la suppression tour à tour de ces gènes, a pour but un éclairage sur l'importance de ces derniers dans deux fonds génétiques différents dont l'un a été optimisé pour la production de cellulases. Les mutants aux phénotypes de production enzymatique les plus distincts de leur souche parentale ont ensuite été phénotypés afin d'évaluer de manière plus ample l'impact qu'a eu la disparition de la marque épigénétique visée sur le comportement du champignon.

MATERIEL ET METHODES

I. MATERIEL BIOLOGIQUE

Les souches de *Trichoderma reesei* utilisées dans cette thèse sont répertoriées dans le Tableau 6.

Nom	Origine	Référence
QM6a (ATCC 13631)	Sauvage	(Mandels et Reese 1957)
RutC30 (ATCC 56765)	Mutagénèses aléatoires à partir de la QM6a	(Eveleigh et Montenecourt 1979)
QΔdim5 39	QM6a <i>dim5::hph</i>	
RΔdim5 12	RutC30 <i>dim5::hph</i>	
RΔdim5 28	RutC30 <i>dim5::hph</i>	
RΔdim5 39	RutC30 <i>dim5::hph</i>	

Tableau 6 : Souches de *T. reesei* utilisées lors de cette étude.

Les solutions utilisées dans cette section sont référencées en Annexe I et en Annexe II.

Une banque de spores pour chacune des souches a été créée afin d'assurer la qualité (provenance de la souche) et le contrôle (quantité de spores introduites) de l'ensemencement des milieux de cultures. Pour les souches QM6a et RutC30, un cryotube de spores de *T. reesei* du stock commun du laboratoire a servi à ensemercer 150 mL de milieu PDA liquide tiède (Potato Dextrose Agar). Pour les mutants, la banque de spores s'est faite à partir d'un transformant génétiquement vérifié et dont les spores de la génération P3 ont été récoltés puis coulés dans du milieu PDA dans les mêmes conditions que pour les souches parentales. Ensuite, 50 mL de milieu encore liquide ont été distribués dans 3 boîtes de pétri de 140 mm de diamètre. Une fois solidifiées, celles-ci ont été placées quatre jours en chambre de culture avec une alternance de 18 heures de lumière et 6h d'obscurité afin de permettre la sporulation. Une fois l'apparition du duvet de spores vert à la surface des boîtes, 5 mL de CTS50 ont été utilisés pour légèrement racler la surface de la gélose à l'aide d'un râteau. Les spores ainsi collectées ont été lavées deux fois au CTS50 stérile puis comptées sur une cellule de Malassez. Elles ont ensuite été conservées à une concentration de 5×10^7 spores.mL⁻¹ dans du CTS50 stérile à -80°C.

II. FERMENTATION

Les fermentations sont réalisées en deux étapes permettant une exploitation maximale du potentiel du champignon. La première phase de la fermentation, dite en mode batch, est consacrée à la croissance de la biomasse. Cette étape se déroule en présence de glucose jusqu'à consommation complète du sucre. La seconde phase de fermentation est la phase de production enzymatique. Au contraire de la première phase, celle-ci se déroule en Fed-batch, c'est-à-dire que le sucre est introduit à débit constant dans le milieu de culture. Le sucre choisi peut être inducteur de la synthèse enzymatique, du lactose dans notre cas, ou bien non inducteur de la synthèse, du glucose.

Les solutions utilisées dans cette section sont référencées en Annexe III.

I.1 Phase de croissance en mode batch

II.1.1. Préparation des fioles de culture et ensemencement

Les fermentations de *T. reesei* ont été effectuées dans des fioles Erlenmeyer en verre de 250 mL, préalablement stérilisées à l'autoclave (30 min à 120°C). Un bouchon en mousse recouvert de papier aluminium assure la stérilité après autoclavage. Pour le développement optimal du mycélium lors de la phase de croissance de la biomasse, 2.5×10^7 spores de la souche sauvage Qm6a de *T. reesei* ont été introduits dans 57 mL de milieu salin complet contenant une concentration de glucose finale de 12.5 g.L⁻¹.

Une pesée initiale de chaque fiole ainsi qu'une mesure du pH initial du milieu complet avant début de fermentation sont effectuées pour le suivi expérimental. Les fioles sont ensuite placées dans un Infors (de la marque Suisse Infors HT et de type Multitron) à 30°C avec une agitation à 150 rpm.

II.1.2. Suivi de la phase de croissance

Après 24 h, 1,5 mL de culture est prélevé stérilement. Une mesure de pH est effectuée puis une centrifugation de cinq minutes à 1300 rpm permet de récupérer le surnageant. Il est ensuite conservé à 4°C pour une quantification enzymatique ultérieure.

La détermination de la concentration en biomasse est réalisée par la mesure du poids sec. Pour cela des coupelles contenant un filtre de cellulose (filtres en fibres de verre GF/C, rétention de 1,2 mm de marque Whatman) sont placées à l'étuve pour dessiccation 24 h avant le prélèvement. A 48 h après le début de l'expérience, 6 mL de culture de chaque fiole sont prélevés. Les coupelles contenant les filtres sont pesées à vide après 24 h de dessiccation. Le mycélium est ensuite récupéré par passage sous

pression de la culture à travers le filtre de cellulose. Les coupelles, contenant cette fois les filtres et le mycélium, sont à nouveau placées à l'étuve à 105°C pendant 48 h puis pesées. La masse sèche est alors déterminée par différence entre la masse avant et après filtration de la suspension cellulaire sur la membrane.

Les surnageants de culture sont récupérés et 1 mL est conservé pour une quantification enzymatique ultérieure. Un test de mesure de pH est effectué sur les tubes de surnageants. Dans les conditions normales d'expérimentations, il est attendu une valeur proche de 3,5. A l'aide de soude concentrée à 12%, un réajustement de la culture à pH 4,3 doit être fait. Après homogénéisation, 1,5 mL de culture en plus est prélevé pour une seconde vérification du pH et la concentration en glucose est vérifiée au Glucostat Beckman Glucose Analyser 2. Le principe de cet appareil repose sur la détection de l'eau oxygénée libérée lors de la transformation du glucose en acide gluconique par une glucose oxydase immobilisée sur une membrane. L'oxydation de H_2O_2 par une électrode de platine libère un flux d'électrons ; le courant ainsi créé est proportionnel à la concentration en glucose. Dix microlitres de solution à tester sont injectés dans l'analyseur préalablement étalonné par une solution standard de glucose. Afin de pouvoir démarrer la phase de fed-batch, la quantité de glucose dans le milieu doit être inférieure à 0,3 g.L⁻¹.

II.2. Phase de production en mode Fed-batch

II.2.1. Préparation des pompes Dasgip

En préparation de la phase de fed-batch, la solution d'urée concentrée à 250 g.L⁻¹ est ajoutée stérilement dans les solutions de glucose et de lactose utilisées et concentrées à 53,3 g.L⁻¹, pour une concentration finale en urée de 5,6 g.L⁻¹. L'injection en continu des solutions est assurée par un module multipompes Dasgip MP8 contrôlées avec le logiciel DASGIP Control 4.0 afin de maîtriser le débit de sucres introduit dans les milieux.

Il est nécessaire de préalablement rincer les pompes Dasgip avec les solutions de sucre contenant l'urée. Pour cela, les lignes d'alimentation des pompes sont branchées sur les flacons Schott des solutions et de nouvelles aiguilles sont placées à la sortie de lignes. Un programme de rinçage avec un pulse de 6 mL à 20 mL.h⁻¹ est effectué afin d'assurer la propreté du système de distribution.

Une fois le rinçage terminé, les aiguilles sont plantées dans les bouchons en mousse des fioles afin que l'aiguille ressorte dans le « ciel » des Erlenmeyer. Les fioles sont branchées soit sur le Schott contenant la solution de glucose soit sur le Schott contenant la solution de lactose. Le débit de sucre est de 0,3 mL.h⁻¹.

II.2.2. Suivi de la phase de fed-batch

A 24h après le début de la nouvelle phase, 6 mL de culture sont prélevés à l'aide d'une pipette graduée à usage unique, en glissant délicatement l'outil entre le bouchon mousse et le col de l'Erlenmeyer. Ce prélèvement est utilisé pour mesurer le poids sec et le pH comme décrit précédemment. S'il n'est pas compris dans l'intervalle 3,8 – 5,0, le pH doit être réajusté à 4,3 avec de la soude 12% ou de l'acide sulfurique 15%. Un millilitre de surnageant est conservé pour les analyses enzymatiques ultérieures.

II.2.3. Arrêt de la culture

Le fed-batch est stoppé après 48 h de culture. Avant le dernier prélèvement, une pesée finale des fioles est nécessaire pour le suivi de perte hydrologique des cultures. Les mêmes mesures que pour le suivi de la phase fed-batch sont à faire avant arrêt des cultures : mesure de poids sec sur 6mL d'échantillon, mesure pH et conservation des surnageants.

Le matériel biologie des cultures est ensuite utilisée pour les expériences ChIP (Chromatine ImmunoPrecipitation) et pour les extractions d'ADN et d'ARN.

Le programme de distribution des sucres est stoppé, les aiguilles des bouchons des fioles de culture puis les lignes d'alimentation des bouchons des Schott sont débranchées. Il est important d'effectuer ces actions dans cet ordre précis afin qu'aucun sucre supplémentaire ne tombe par gravité dans les Erlenmeyers. Les lignes d'alimentation doivent être ensuite nettoyées à l'eau tiède pour éliminer les résidus de sucre grâce au même programme de rinçage qu'utilisé auparavant. Elles sont ensuite remplies d'éthanol 20% jusqu'à la prochaine expérimentation.

II.3. Détermination de la concentration de protéines extracellulaires

La concentration protéique des surnageants récoltés tout au long de l'expérience de fermentation permettent de déterminer s'il y a en effet une production différentielle entre les cultures ayant été exposées à du glucose (milieu non inducteur de production de cellulases) et celles ayant été exposées à du lactose (milieu inducteur de production de cellulases) lors de la phase de Fed-batch.

Les dosages sont réalisés avec le robot Freedom Evo de TECAN. La méthode de Bradford est utilisée à l'aide du kit Quick Start Bradford protein assay kit (Bio-Rad). Ce dosage est très sensible et permet de quantifier des concentrations de protéines très faibles. Le robot effectue deux dosages à la suite l'un de l'autre pour une meilleure appréciation de la précision de la mesure. Nous nous servons de ces deux mesures

pour représenter les erreurs standards sur les graphiques dédiés.

II.4. Activité enzymatique

Les activités enzymatiques sont mesurées sur des substrats chromogéniques pNPL (activité cellobiohydrolase CBH1) ou pNPG (activité β -glucosidase Cel3a) qui libère du *para*-nitrophénol (pNP) à la suite de l'hydrolyse enzymatique des molécules de *para*-nitrophényl- β -lactoside ou *para*-nitrophényl- β -D-glucopuranoside, respectivement. Le pNP relâché est révélé par l'addition de 100 μ L de Na₂CO₃ à 2% et l'absorbance est mesurée à 410 nm. Une gamme de concentration connue de *para*-nitrophénol (allant de 12,5 à 200 μ M) permet de déterminer ensuite le nombre de moles de pNP libérées dans le milieu par les enzymes et d'en déduire leur activité. Les essais sont effectués dans des plaques 48 puits après 30 minutes d'incubation à 50°C avec un robot Freedom Evo de TECAN.

L'activité enzymatique est exprimée comme activité spécifique en IU Nkat.mgP⁻¹ (nmol esp-nitrophenol.s⁻¹).mg enzymes⁻¹.

III. EXTRACTION D'ARN

L'extraction d'ARN est réalisée à partir de 8 mL d'échantillon prélevés en fin de fermentation, filtrés sur un filtre de cellulose (filtres en fibres de verre GF/C, rétention de 1,2 mm de marque Whatman) et congelés dans l'azote liquide avant d'être conservé à -80 °C.

Le culot obtenu est inséré dans un tube de Lysis matrix C du Fast RNA Pro Red Kit (MP Bio), contenant 700 μ L de tampon RLT (Qiagen) additionné de 7 μ L de solution de β -mercaptoéthanol pur (SIGMA). Les cellules sont ensuite broyées au Fastprep (MP Bio) 40 secondes force 6 et centrifugées cinq minutes à 14 000 rpm et 4 °C.

La phase liquide est transférée sur une colonne QIAshredder (Qiagen) pour éliminer les débris cellulaires. L'éluât est collecté après centrifugation (2 min, 10 000 rpm, 4 °C) et additionné de 0,5 volume d'éthanol absolu. L'homogénéisation a lieu par va-et-vient successifs avec la pipette. Le mélange est ensuite déposé sur colonne RNeasy (Qiagen) et l'éluât est éliminé après centrifugation (1 min à 10 000 rpm). La colonne est rincée avec 350 μ L de tampon RW1 (Qiagen) pour fixer l'ARN, et l'éluât éliminé après centrifugation (1 min, 12 000 rpm, 4°C).

Pour se débarrasser d'éventuelles contaminations d'ADN, les échantillons sont traités à la DNase (RNase-Free DNase Set, Qiagen) : 80 μ L du mélange DNase - Tampon RDD (10 μ L de DNase pour 70 μ L de tampon RDD) sont déposés au centre de la colonne RNeasy, suivis d'une incubation de 35 min à température ambiante. Le rinçage

s'effectue en trois étapes : tout d'abord par l'ajout et l'élimination par centrifugation de 350 μ L de tampon RW1 (1 min, 12 000 rpm, 4 °C) suivis de deux rinçages similaires avec 500 μ L de tampon RPE (deux minutes de pose sur la colonne sont nécessaires avant centrifugation).

Après élimination du filtrat la colonne est essorée par un cycle supplémentaire de centrifugation (1 min, 12 000 rpm, 4 °C). La colonne est ensuite transférée dans un tube Eppendorf 1,5 mL pour être éluée avec 100 μ L d'eau DEPC (UltraPure DEPC-Treated Water, Invitrogen).

La quantité d'ARN obtenue est évaluée par mesure au Qubit 2.0. La qualité des ARN extraits est, quant à elle, évaluée au Bioanalyseur (Agilent).

IV. ANALYSE TRANSCRIPTOMIQUE

Les RNA Seq (construction des banques inclus) de la souche RutC30 ont été réalisés sur la plateforme Génomique de l'Institut de Biologie de l'École normale supérieure (IBENS) par séquençage Illumina. Les librairies sont préparées avec le kit RNA-Seq library preparation TruSeq Stranded mRNA kit (Illumina). Elles sont ensuite multiplexées sur des « *flowcells* » pour des séquençages de 50 pb sur une machine HiSeq 1500 (Illumina). Les *reads* passant le contrôle qualité Illumina ont ensuite été raccourcis (« *trimmed* »), alignés sur le génome de référence (Tableau 6) puis filtrés et enfin quantifiés. La normalisation des données et analyse différentielle sont réalisées grâce à DESeq, de la « *toolbox* » Bioconductor (version 1.8.3). La méthode utilisée par DESeq est basée sur l'hypothèse que seul un nombre restreint de gènes sont différentiellement exprimés et utilise donc une méthode de normalisation par la médiane. L'analyse différentielle est faite sur les données en condition « lactose vs glucose » générées lors des expériences en fed-batch comme décrites dans la section II. Elle repose sur un test statistique pour lequel les données possèdent une distribution négative binomiale et dont la variance et la moyenne relative sont déterminées par une régression locale. Une correction est apportée aux p-values grâce à la méthode de Benjamini et Hochberg (Benjamini et Hochberg 1995). Deux filtres sur l'expression sont ensuite appliqués et un gène est alors considéré comme différentiellement exprimé si son adjusted p-value est inférieure à 0,001 et que la valeur absolue de son $\log_2(\text{FC})$ est supérieure à 2 (sauf indication contraire dans le manuscrit). Le FC (« *Fold Change* ») fait ici référence au ratio du nombre de *reads* dans la condition testée (fed-batch lactose) contre le nombre de *reads* dans la condition de référence (fed-batch glucose).

Les RNA Seq de la souche QM6a ont été réalisés par la plateforme du JGI dans le même cadre que les expériences de ChIP Seq et ont subi les mêmes procédures de séquençage et d'analyses que les échantillons de la RutC30.

V. CHIP SEQ

L'expérience de ChIP Seq consiste à fixer la liaison existante entre protéines et ADN, de digérer l'ADN non lié puis de sélectionner grâce à des anticorps les portions d'ADN liés aux protéines d'intérêt. Le lien entre protéines et ADN est alors cassé et l'ADN récupéré pour séquençage ultérieur.

Les solutions utilisées ici sont référencées en Annexes IV.

V.1. Préparation des échantillons

Afin d'effectuer l'expérience de ChIP Seq (« *Chromatine ImmunoPrecipitation followed by Sequencing* »), il est nécessaire de faire subir aux cultures un prétraitement de « *crosslink* ». Ce prétraitement permet de chimiquement coller les protéines présentes dans le noyau et en interaction avec l'ADN, et ce, de manière définitive. Ainsi, le but est de préserver toutes les interactions ADN-histones présentes.

Cette fixation est réalisée à l'aide d'une solution de formaldéhyde obtenue par le mélange du contenu d'une ampoule de 10 mL de formaldéhyde 16% mélangé avec 150 mL de PBS pour obtenir une concentration finale en formaldéhyde de 1%.

Après arrêt des fermentations, chaque mycélium sélectionné pour les expériences de ChIP seq est filtré sur gazes stériles puis rincé abondamment avec du buffer PBS 1X jusqu'à élimination du colorant jaune qui peut, sinon, diminuer l'efficacité de la fixation. Le matériel biologique est ensuite transféré dans une solution de formaldéhyde 1% et incubé à 30°C sous agitation 150 rpm pendant 15 min. Il est important que l'ensemble du mycélium soit en contact avec le formaldéhyde. La fixation est arrêtée par l'addition de 4 mL d'une solution de glycine à 2,5 M suivie d'une incubation de cinq minutes à 30 °C et sous agitation 150 rpm. La fiole est ensuite conservée sur glace jusqu'à la prochaine étape.

En chambre froide, les échantillons fixés au formaldéhyde sont filtrés sous vide et rincés abondamment à l'aide de PBS préalablement refroidi à 4°C. Pour finir, le mycélium est transféré dans un tube Falcon de 50mL et congelé dans de l'azote liquide. Les échantillons sont stockés à -80 °C.

V.2. Immunoprécipitation de la chromatine

Ce protocole se déroule sur trois jours et il est recommandé de ne pas interrompre l'expérience une fois lancée. Le premier jour est dédié à la fragmentation de la chromatine en morceaux d'ADN attachés à un, deux et/ou trois nucléosomes. Durant le second jour, l'extraction spécifique de l'ADN lié aux nucléosomes portant les marques épigénétiques d'intérêts est effectuée puis, le troisième jour, deux traitements

permettent d'obtenir un ADN libre de protéines et d'ARN.

Jour 1 : Fragmentation de la chromatine

Un broyage du mycélium au mortier et pilon dans l'azote liquide est effectué afin d'extraire l'ADN. Pour obtenir une quantité suffisante d'ADN extrait à la fin de l'expérience, un tube pour la marque épigénétique H3K4me3, deux tubes pour la marque H3K9me3 et trois tubes pour la marque H3K27me3 sont utilisés. Environ 250 mg de mycélium broyé sont utilisés pour chaque tube.

Pour chaque échantillon, un volume de tampon de lyse contenant du CaCl_2 et des inhibiteurs de protéases (Roche) est introduit pour obtenir une concentration finale de mycélium à 250 mg.mL^{-1} . Après resuspension de la poudre de mycélium, une incubation de 10 min dans la glace est faite, avec une resuspension périodique du précipité toutes les deux minutes.

Un thermo-mixer à 37°C doit être préchauffé durant ce moment d'attente.

Une fois les dix minutes écoulées, un volume adapté d'enzyme nucléase micrococcale à raison de $5 \text{ }\mu\text{L/mL}$ de tampon de lyse, est ajouté. Les tubes sont ensuite incubés pendant 30 minutes sous agitation à 1400 rpm dans le thermomixer préalablement préchauffé.

Durant l'incubation, la température d'une centrifugeuse doit être descendue à 4°C .

La digestion est arrêtée en ajoutant un excès d'EGTA à raison de $50 \text{ }\mu\text{L/mL}$ de tampon de lyse. Deux centrifugations consécutives à 4°C pendant 15 minutes à vitesse maximale sont nécessaires pour la séparation du surnageant contenant la chromatine digérée des débris de cellules. Un dosage au Qubit 2.0 (Invitrogen) à partir de $20 \text{ }\mu\text{L}$ de l'échantillon total de chromatine digérée permet de déterminer la concentration en ADN extrait.

Pour chaque marque épigénétique à cibler, un volume correspondant à $5 \text{ }\mu\text{g}$ d'ADN est transféré dans un tube Eppendorf LoBind®. Du tampon de lyse sans CaCl_2 ni inhibiteur de protéases sert à compléter le volume jusqu'à $1,1 \text{ mL}$. Puis, $30 \text{ }\mu\text{L}$ de billes magnétiques (Dynabeads® Protein A de Life Technologies) sont ajoutés à chaque tube avant 4h d'incubation en chambre froide (4°C) sur agitateur orbital à 150 rpm.

A la fin de l'incubation, les tubes sont passés sur portoir magnétiques trois minutes à 4°C afin de récupérer les surnageants. Pour chaque marque épigénétique étudiée, 1 mL est prélevé et transféré dans un tube Eppendorf LoBind®.

Les anticorps sont ajoutés à leur tube désigné et sont laissés pour incubation durant la nuit à 4°C et avec agitation orbitale à 150 rpm (Tableau 7).

Résidu ciblé	Numéro catalogue	Marque	Type	Volume ajouté
H3K4me3	07-473	Millipore	Polyclonal	2 µL
H3K9me3	39161	Active Motif	Polyclonal	5 µL
H3K27me3	17-622	Millipore	Polyclonal	4 µL

Tableau 7: Anticorps utilisés lors des expériences de CHIP Seq.

Enfin, 100 µL du surnageant de départ sont congelés à -80°C pour une analyse d'échantillon Input et 300 µL du surnageant sont gardés afin de vérifier la bonne digestion de la chromatine et ainsi la visualisation de l'échelle nucléosomique.

Jour 2 : Immunoprécipitation de l'ADN

Au matin, 30 µL de billes magnétiques par tube sont ajoutés et incubés à nouveau 4h à 4°C avec agitation orbitale à 150 rpm. Après 4h d'incubation, une série de lavage est nécessaire afin de purifier l'ADN récolté avec les billes magnétiques. Ils s'opèrent dans cet ordre précis :

- Lavage au tampon de lyse (x2)
- Lavage au tampon de lyse + 0,5 M NaCl
- Lavage au tampon de LiCl (x2)
- Lavage au Tris EDTA pH 8 (TE)

Pour chaque lavage, une rapide (2s) centrifugation des tubes dans une petite centrifugeuse de paillasse à vitesse minimale permet de précipiter toutes les billes magnétiques. Puis, les tubes sont placés trois minutes sur portoir magnétique afin de garder les billes et de se débarrasser du surnageant. Un millilitre de la solution de lavage est nécessaire à chaque étape. Les incubations entre chaque lavage sont de 10 min à 4°C sous agitation orbitale à 150 rpm.

Lors des derniers lavages, le préchauffage d'un thermo-mixer et d'une solution de TES à 65°C est nécessaire ainsi que la décongélation d'un tube d'échantillon Input.

Après le dernier lavage au TE, les billes magnétiques sont resuspendues dans 60 µL du TES préchauffé à 65°C puis les tubes sont placés dans le thermo-mixer chaud

sous agitation maximale (1400 rpm) pendant 15 minutes. Les surnageants sont récupérés dans de nouveaux tubes Eppendorf LoBind® et les billes sont à nouveau lavées de la même manière. Ensuite, 100 µL de TES et 20 µL de SDS 10% sont ajoutés à l'échantillon Input décongelé.

Les réplicats techniques des marques sont finalement mélangés et l'ensemble des tubes sont incubés à 65°C pendant 8h puis 20°C.

Jour 3 : Purification de l'ADN immunoprécipité

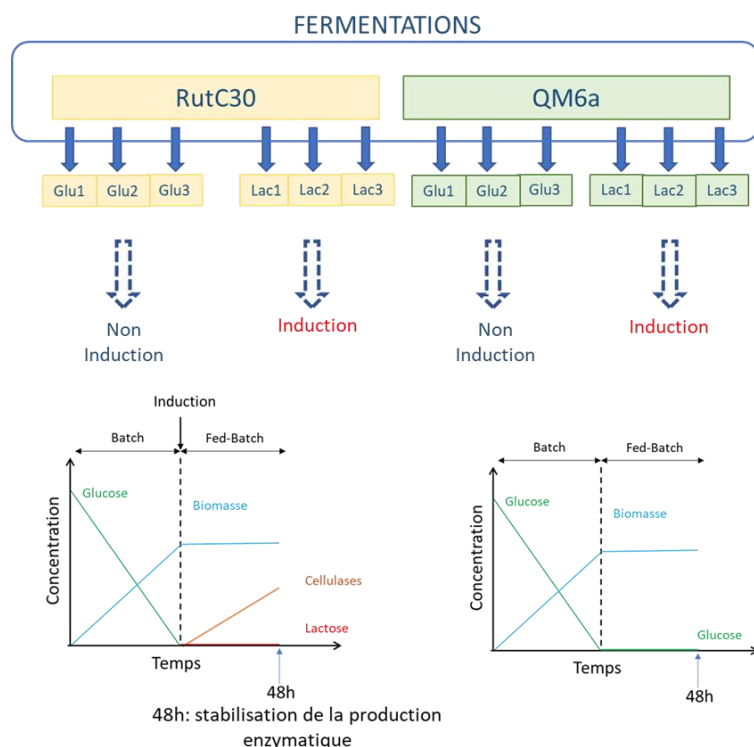
Au matin, 1,2 µL et 2,2µL de RNase A sont ajoutés (avec une enzyme concentrée à 20 mg.mL⁻¹) dans les tubes des marques épigénétiques et dans le tube d'échantillon Input, respectivement, puis incubés 2h à 37°C avec agitation à 600 rpm. Puis, 1,2 µL et 2,2 µL de Protéinase K sont ensuite ajoutés aux échantillons des marques épigénétiques et à l'échantillon d'Input, suivi d'une incubation de 2h à 55°C, toujours avec une agitation à 600 rpm.

Pour une purification optimale des échantillons d'ADN, le kit de purification pour produit PCR de Qiagen est utilisé. La quantité d'ADN récupéré est dosée au Qubit à l'aide du kit Qubit dsDNA HS Assay Kit grâce à 5 µL de solution éluée.

A partir du surnageant récupéré après la digestion du mycélium, un traitement de 1h à la RNase A à raison de 1 µL par 100 µL de solution est fait à 37°C avec un agitation de 600 rpm.

Ceci est suivi d'un traitement à la Protéinase K et SDS 1% à 55°C pendant 8h. L'ADN est ensuite purifié sur colonne Qiagen du kit QIAquick PCR Purification Kit. L'échelle nucléosomique est visualisable sur gel d'agarose 1,5% avec une révélation au BET.

VI. SCHEMA EXPERIMENTAL DES FERMENTATIONS ET EXPERIENCES DE CHIP SEQ ET RNA SEQ QUI EN DECOULENT



Séquençages

ChIP Seq

Six échantillons ont été utilisés pour chacune des marques chez les deux souches
Un contrôle pour chaque condition a été réalisé chez les deux souches

Total: 36 ChIP Seq + deux contrôles ont été séquencés

RNA Seq

Six échantillons ont été utilisés chez les deux souches

Total : 12 échantillons totaux ont été séquencés

Figure 20 : Schéma expérimental des fermentations en fioles réalisées chez la RutC30 et la QM6a.

Six fioles de fermentation ont été utilisées pour la génération des échantillons de ChIP Seq et de RNA Seq. Chez les fioles en fed-batch lactose, la quantité d'enzymes extracellulaires doit augmenter tandis qu'aucune synthèse enzymatique ne doit avoir lieu à l'intérieure des fioles en fed-batch glucose. Au total pour les deux souches, 36 échantillons de ChIP Seq ont été séquencés ainsi que deux contrôles input. Pour les RNA Seq, 12 échantillons totaux ont été séquencés (Figure 20).

VII. DOT BLOT

Trois peptides commerciaux provenant de chez Active motif ont été utilisés : H3K4me3 (81042) H3K9me3 (81047) et H3K27me3 (81052). Les anticorps primaires utilisés pour les expériences de Dot blot sont les mêmes que ceux utilisés pour l'étape d'immunoprécipitation du ChIP Seq (Tableau 8). L'anticorps secondaire est un anticorps chèvre anti-lapin couplé HRP (Péroxydase de raifort) pour la détection (172-1019 Biorad). Tous les anticorps sont dilués dans du TBS-T 1X +1% lait (Annexe V) et incubés selon les données renseignées dans le Tableau 8.

Anticorps primaire	Dilution anticorps primaire	Temps d'incubation et température	Dilution anticorps secondaire	Temps incubation anticorps secondaire
H3K4me3	1:5000	1h, temp. ambiante	1:3000	1h
H3K9me3	1:2000	Sur la nuit, 4°C	1:3000	2h30
H3K27me3	1:2000	Sur la nuit, 4°C	1:3000	2h30

Tableau 8 : Dilution, température et temps d'incubation des anticorps utilisés lors des expériences de Dot blot.

Un rectangle d'environ 6cmx3cm de membrane PVDF haute affinité avec 0,2 µm de diamètre de pores (1620177 Biorad) a été utilisé pour les tests d'affinité de chaque anticorps (Figure 21).

Peptide/Quantité	330 ng	100 ng	50 ng	10 ng	5ng
K4me3	o	o	o	o	o
K9me3	o	o	o	o	o
K27me3	o	o	o	o	o

Figure 21 : Représentation de la membrane de Dot blot.

Toutes les étapes du mode opératoire se déroulent dans des boîtes en plastique sous agitation légère, peptides vers le haut, à température ambiante (sauf contre-indication pour incubation des anticorps primaires Tableau 8).

L'activation de la membrane PVDF est faite par immersion dans du méthanol 100% pendant 15 s suivie d'une équilibration dans de l'eau distillée pendant 2 min puis d'une étape dans du tampon TBST-T 1X pendant 5 min. Un volume de 2 µL des peptides commerciaux sont ensuite déposés selon le schéma ci-dessus pendant que la membrane est encore humide. Le séchage de la membrane se fait à température ambiante pendant 1h30. Un volume de 50 mL de TBS-T 1X + 5% de lait sert à bloquer la membrane pendant 1h. Puis, l'anticorps primaire dilué dans 15mL de TSB-T 1X + 1% de lait selon le Tableau 8 est incubé avec la membrane. Elle est ensuite rincée rapidement 2 fois dans du TBS-T 1X puis lavée 3x5min dans du TBS-T 1X.

L'anticorps secondaire est ensuite incubé selon les conditions décrites dans le Tableau 8. Deux rinçages rapides sont ensuite suivis de 2x10 min puis 2x5 min de lavage

dans du TBS-T 1X.

La visualisation du signal émis par la HRP se fait à l'œil nu grâce au kit colorimétrique 1-Step Ultra TMB-Blotting Solution de Thermo Scientific après 5 min d'incubation dans le noir.

VIII. CARTOGRAPHIE ET ANALYSE BIOINFORMATIQUE DE LA MARQUE EPIGENETIQUE H3K4me3

Les données étudiées ont été générées selon le protocole de ChIP Seq décrit dans la section ChIP Seq de ce Matériel et Méthodes à partir de cellules de mycélium de deux souches distinctes de *Trichoderma reesei*. La souche sauvage QM6a et la souche hyper-productrice modèle RutC30. Pour chacune de ces souches, deux conditions de culture du champignon sont étudiées ici : une condition de non-induction de production de cellulases (condition de culture sur glucose) et une condition d'induction de production de cellulases (condition de culture sur lactose). Trois marques épigénétiques sont étudiées : la triméthylation de la lysine 4 de l'histone H3 (H3K4me3), la triméthylation de la lysine 9 de l'histone H3 (H3K9me3) et la triméthylation de la lysine 27 de l'histone H3 (H3K27me3). Trois réplicats biologiques ont été faits pour chaque marque de chaque condition. Un total de 18 échantillons a donc été généré pour les immunoprécipitations. Deux échantillons *input* servant de contrôle ont également été produits pour chaque condition.

Le séquençage des données issues des ChIP Seq de la souche RutC30 a été réalisé grâce à la technologie Illumina NextSeq 500/550 (NextSeq High output, *Paired end* 2x50 nt) au laboratoire The Next Generation Sequencing (NGS) Core Facility - Plateforme de Séquençage Haut Débit I2BC - UMR9198.

Le séquençage des données issues des ChIP-Seq de la souche QM6a a également été réalisé grâce à la technologie Illumina NextSeq 500/550 (NextSeq High output, *Paired end* 2x150 nt) mais sur la plateforme du Joint Genome Institute (JGI).

L'analyse bioinformatique des données de ChIP Seq a été réalisée dans un premier temps sur le cluster de bioinformatique cluster IFB-Core sur un système Linux. Le code généré pour l'analyse est écrit dans le langage bash.

Dans un second temps, le logiciel open source R (version 4.0.3) (R Core Team 2022) a été utilisé, avec l'interface RStudio (version 1.4.1103). La toolbox Bioconductor 3.13 contenant divers package d'analyses bioinformatiques a servi à procéder à l'analyse différentielle des données. Le langage statistique R a servi pour l'écriture du script.

VIII.1. Contrôle qualité des données de séquençage

Les renseignements sur le niveau de duplication des reads ainsi que l'éventuelle présence des adaptateurs et des informations sur la qualité générale des données de séquençage brutes sous format FASTQ sont contrôlés grâce à la fonction fastqc (version 0.11.9) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) sur le cluster de IFB-Core.

VIII.2. Alignement des reads

L'alignement des reads sur le génome de référence est faite avec Bowtie (version 1.2.3) (Langmead et al. 2009). Les fichiers de sortie sont générés sous format SAM.

Le mode d'alignement « -v » est utilisé afin de permettre une détection des ratés d'alignement des paires de base sur l'ensemble de la longueur des reads par rapport au génome de référence. Tout read possédant plus de 2 erreurs d'alignement sera éliminé de l'étude. Le paramètre « -m » a été réglé sur 1, ne permettant pas ainsi le multimapping des reads.

Pour les données issues du séquençage de la souche QM6a, un trimming de 3 paires de bases en 3' a été effectué sur les reads avec l'option « -3 ».

VIII.3. Détection et élimination des duplicats

Samtools (version 1.9) (Danecek et al. 2021) a été utilisé pour la détection et l'élimination des duplicats ainsi que la génération des fichiers BAM. L'option « -n » de samtools sort a servi à ordonner les reads par nom. Les fichiers générés ont ensuite été utilisés par la fonction samtools fixmate avec les options -r et -m pour, respectivement, supprimer les reads non alignés et assigner un score ms (mate score) à tous les reads pairés.

La fonction samtools sort a ensuite été utilisée afin de réordonner les reads par coordonnées. La fonction samtools markdup avec l'option -r a servi à supprimer les reads dupliqués et à générer les fichiers BAM finaux. Enfin, la fonction samtools index a permis la création des fichiers indexés BAI.

VIII.4. Peakcalling

L'outil macs2 (version 2.2.7.1) (Zhang et al. 2008) a été utilisé pour le peak calling. Les deux fichiers d'input en lactose ont été utilisés comme contrôles de détection des pics. La longueur totale du génome de 35 Mb a été utilisée comme région de détection. Une q-value de 0,05 a servi de limite de détection de pics faux positifs. L'option broad permet la détection de domaines larges caractéristiques des marques des histones.

VIII.5. Normalisation et visualisation des données

Dans un but de visualisation, une normalisation par la méthode RPKM (« *Reads per kilo base per million mapped reads* ») des données a été faite grâce à l'outil deeptools bamCoverage (version 3.5.0) avec une taille de génome de 35 Mb.

La visualisation se fait sur l'outil IGV (version 2.9.2) (Robinson et al. 2011).

VIII.6. Analyse différentielle

L'analyse différentielle a été réalisée avec l'outil DiffBind (version 3.2.5) (Rory Starkrory.stark@cruk.cam.ac.uk 2017) de la toolbox Bioconductor 3.13.

La séquence d'analyse a été suivie de manière linéaire comme indiquée dans le manuel « *DiffBind: Differential binding analysis of ChIPSeq peak data* ». L'ensemble des fichiers .bam finaux générés dans la première partie de l'analyse ainsi que les fichiers .xls du peakcalling sont intégrés dans la matrice de données. La totalité des reads sont comptés puis la taille de la librairie est normalisée.

La condition de culture en glucose est choisie comme condition contrôle de l'analyse différentielle et le consensus peakset est paramétré pour prendre en compte tous les échantillons d'une même marque, les deux conditions confondues. Seuls les pics étant présents dans au moins 3 échantillons du consensus peakset sont considérés pour l'analyse grâce à un minoverlap de 3.

VIII.7. Génération des graphiques

Les heatmaps, les analyses ACP et les Volcano plots ont été réalisés à partir de l'outil DiffBind (version 3.2.5). L'outil deepTools (version 3.5.0) a été utilisé pour la génération des graphiques de couverture de la marque le long des chromosomes.

IX. IDENTIFICATION DES GENES INVALIDES

Les orthologues des gènes *dim5* (protéine XP_957479.2), *dim2* (protéine XP_959891.1), *hp1* (protéine XP_957632.1), *kmt6* (protéine XP_965043.2) et *rid* (protéine AAM27408.1) (Annexe VI) chez *N. crassa* ont été identifiés chez *T. reesei* sur fungidb (<https://fungidb.org/fungidb/app>). Les annotations ont été révisées grâce à l'utilisation de données transcriptomiques générées par Pirayre et al. 2020 et l'alignement avec les orthologues respectifs de *N. crassa*. Les domaines protéiques principaux ont été déterminés sur Geneious Prime (version 2022.2.1) (Kearse et al., 2012) avec l'outil InterProScan (version 91.0) (Blum et al. 2021) et schématisés avec le programme DOG 2.0 pour la visualisation des structures de domaines protéiques.

L'alignement des protéines a été fait avec l'outil MUSCLE version 3.8.31) (Edgar 2004) en générant un fichier phylip interleaved utilisé ensuite par Geneious Prime pour la construction de l'arbre phylogénétique avec la technique du Neighbor-joining.

X. CASSETTES D'INVALIDATION

Pour l'inactivation des gènes, des plasmides contenant une cassette d'inactivation (Annexe VII) ciblant spécifiquement un locus ont été construites par l'entreprise de biotechnologie Twist Bioscience. Pour cela, la cassette d'inactivation comprend le gène de résistance à l'hygromycine *hph* bordé par des fragments de 200 pb homologues aux régions de part et d'autre du locus du gène cible (Figure 22).

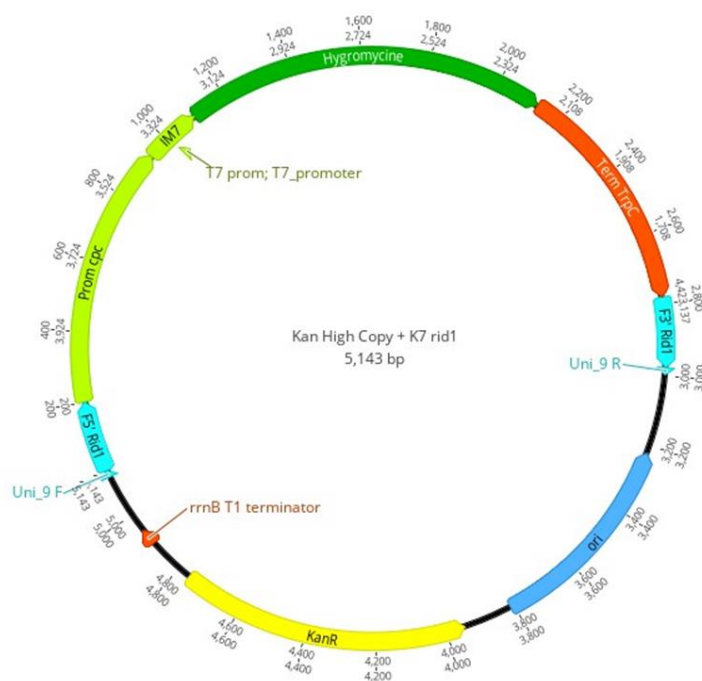


Figure 22 : Plasmide contenant la cassette d'inactivation du gène *rid*.

XI. AMPLIFICATION DES PLASMIDES

Les cellules compétentes (DH5α, Top10, 10-beta) utilisées ont été achetées auprès de New England Biolabs (NEB). 50 µL de cellules sont mises en contact avec 1 à 5 µL de plasmide (1 pg à 100 ng d'ADN). Le tube réactionnel est placé dans la glace pendant 30 minutes. La transformation se fait par un choc thermique à 42°C pendant 30 secondes. Les cellules sont ensuite remises dans la glace pour cinq minutes. Une phase de régénération permet ensuite au gène de résistance de s'exprimer. Pour cela les cellules sont reprises dans 950 µL de milieu SOC fourni et incubées 1h à 37°C avec

une agitation de 225 rpm. Ensuite, 100 µL de culture sont étalés sur milieu gélosé LB contenant l'antibiotique nécessaire à la sélection des clones ayant intégré le plasmide. Après une incubation à 37°C sur la nuit, les colonies sont repiquées. Les plasmides sont ensuite extraits avant d'être utilisés comme matrice PCR pour l'amplification des cassettes de substitution (Plasmid Plus Midi kit, Qiagen).

XII.AMPLIFICATION DES CASSETTES D'INVALIDATION

Les cassettes d'invalidation ont été amplifiées par PCR avec les amorces dessinées à l'aide du logiciel de bio-informatique Geneious Prime (Annexe VIII). Les produits PCR obtenus sont purifiés à l'aide du QIAquick PCR Purification Kit (Qiagen) et utilisés pour transformer les protoplastes

XIII.TRANSFORMATIONS DE *T. REESEI* UTILISANT LA TECHNOLOGIE CRISPR-Cas9

La transformation de *T.reesei* est basée sur la production de protoplastes obtenus après la suppression de la paroi cellulaire à l'aide d'enzymes lytiques. Les protoplastes purifiés sont ensuite capables d'intégrer de l'ADN en présence de CaCl₂ et de polyéthylène-glycol (PEG). Pour faciliter une intégration à un site précis du génome, la méthode CRISPR-Cas9 a été employée en utilisant les conditions optimisées sur *T. reesei* au département Biotechnologie, IFPEN (Dallongeville et Mathis, article en préparation).

Ainsi, une production *in vitro* des crRNA et des tracrRNA sous forme d'ARN guide en un seul fragment est utilisée. L'ARN guide doit d'ailleurs conserver la structure tige boucle très importante pour la reconnaissance par la nucléase Cas9. De la même manière, la nucléase est également produite *in vitro*, et l'association de l'ARN guide se fait donc avec une Cas9 sous forme protéique.

XIV.PREPARATION DES PROTOPLASTES

Des protoplastes des souches ont été produit afin de procéder à des transformations. Dans ce protocole, tout le matériel utilisé a été au préalable stérilisé à l'autoclave et toute manipulation du mycélium est faite sous hotte stérile.

Pour procéder, 5 à 10 fioles de Roux contenant 200mL de milieu PD (Potato Dextrose) ont été stérilisées à 110°C pendant 30 minutes puis ont étéensemencées avec 300 µL d'un cryotube de spores de la souche désirée. Elles sont ensuite placées

horizontalement dans la chambre de culture à 30°C pendant quatre à cinq jours, sans agitation. Lorsqu'un biofilm jaunâtre (il est important qu'il n'y ait pas eu sporulation) suffisamment épais est formé à la surface du milieu de culture, il est temps de procéder à la formation des protoplastes.

La totalité des fioles est alors filtrée sur une épaisseur de quatre gazes stériles ouvertes au-dessus d'un grand Erlenmeyer. Le mycélium est récupéré à l'aide d'une spatule et transvasé dans une bouteille contenant 200 mL de tampon KPAm (Annexe II) puis incubé 30 minutes minimum à 37°C afin de rincer l'ensemble des tissus fongiques. Le mycélium est alors filtré à nouveau sur des gazes puis transféré dans un tampon contenant 3 g de Glucanex (Novozymes VinoTaste® Pro) dissous dans 100 mL de KPAm préalablement stérilisé par filtration. Ce mélange est incubé 2 à 3h à 37°C sous agitation à 150 rpm. La formation des protoplastes est suivi par observation au microscope : il faut que la totalité du mycélium ait été digéré et que la solution présente une couleur irisée trouble, sans amas de cellules.

Une fois la digestion terminée, deux gazes ouvertes sont placées sur un verre fritté n°1 et la solution est filtrée. Seuls les protoplastes passent au travers du filtre. Délicatement, 23 mL de protoplastes sont transférés dans 4 tubes Corex puis centrifugés 10 minutes à 4000 g à 4°C.

Les solutions de CTS10 et CTS50 doivent être refroidies à 4°C et les étapes de centrifugation et resuspension suivantes sont effectuées :

- Resuspension dans 4x20 mL de CTS10 à 4°C + centrifugation 10 min 4000 g à 4°C
- Resuspension dans 2x20 mL de CTS10 à 4°C + centrifugation 5 min 4000 g à 4°C
- Resuspension dans 1x20 mL de CTS10 à 4°C + centrifugation 5 min 4000 g à 4°C
- Resuspension dans 5 mL de CTS50 à 4°C

Après la dernière étape, les protoplastes sont comptés sur cellules de Malassez puis centrifugés à nouveau cinq minutes à 4°C pour ajuster la concentration finale à 2.10^8 cellules.mL⁻¹ dans du CTS50. Les protoplastes peuvent être utilisés directement pour une transformation ou conservés dans des cryotubes à -80°C.

XV.PREPARATION DU COMPLEXE CRISPR-Cas9

Une réaction de couplage entre l'ARN guide et la protéine Cas9 pour former le complexe ribonucléoprotéique (RNP) est nécessaire avant la transformation des protoplastes de *Trichoderma reesei*. La protéine Cas9 utilisée est produite par le groupe NEB (EnGen® Spy Cas9 NLS) et est fournie avec son tampon NEBuffer™ 3.1. Les ARN

guides ont été dessinés à l'aide du logiciel de bio-informatique Geneious Prime grâce à la méthode de (Doench et al. 2016) (Annexe VIII) et commandés chez le groupe Twist Bioscience.

Le protocole pour les réactions de couplage est décrit dans le Tableau 9.

Solution	1 x Réaction	3 x Réactions
Cas9 [20µM]	1µL	3µL
ARN guide [30µM]	6µl	18µl
H₂O DEPC	2µL	6µL
Tampon Cas9 10X [NEBuffer™ 3.1]	1µL	3µL
Mélanger délicatement et incuber à température ambiante pendant 10 minutes		
Conserver à 4°C		

Tableau 9 : Préparation du complexe RNP.

XVI. TRANSFORMATION

Pour une transformation, les boîtes de pétri contiennent du PDA et du saccharose afin d'éviter l'explosion par choc osmotique des protoplastes. Elles sont constituées de deux couches. Une première couche faite au préalable et contenant du PDA additionné de 0,2 M de saccharose et 50 µg.mL⁻¹ d'hygromycine (ou de généticine pour les complémentations). Une deuxième couche appelée surcouche et contenant du PDA additionné 0,8 M de saccharose et 50 µg.mL⁻¹ d'hygromycine (ou de généticine pour les complémentations) et dans laquelle les protoplastes seront coulés avant étalement. Les boîtes pour le témoin de régénération ne contiennent pas d'antibiotiques.

Trois témoins sont réalisées lors d'une transformation : témoin de régénération, témoin de transformation et témoin de sensibilité. La composition des transformations est décrite dans le Tableau 10.

	Complexe RNP	ADN exogène
Témoin Régénération	NON	NON
Témoin Transformation	NON	5 µg du plasmide pUT1140
Témoin sensibilité	NON	NON
Transformation	10µL de RNP	5 µg de cassette d'inactivation

Tableau 10: Composition des tubes de transformation.

Dans des tubes Falcon de 50 mL, 50 µL de protoplastes concentrés à 2.10^8 protoplastes.mL⁻¹ de solution sont mélangés avec le CTS50, les ADN exogènes (plasmide du gène de résistance à l'antibiotique pour insertion ectopique ou cassette d'inactivation) et le complexe RNP si nécessaires puis incubés dix minutes à température ambiante. Une fois l'incubation finie, 1 mL de PEG à 60% est ajouté délicatement dans chacun des tubes. Le mélange est agité doucement à la main puis incubé 20 minutes à température ambiante. Enfin, 900 µL de CTS50 sont ajoutés.

Pour finir, 30 mL de milieu PDA additionné 0,8 M de saccharose en surfusion sont ajoutés. Le mélange s'effectue par plusieurs aller-retours avec la pipette puis est coulé dans cinq boîtes contenant 8 mL de sous-couche préparés au préalable. Les milieux pour le témoin de régénération ne contiennent pas d'hygromycine contrairement aux autres. Après gélification des milieux, les boîtes sont incubées à 30°C pendant deux à trois jours, jusqu'à apparition des premiers mycéliums.

XVII. PURIFICATION DES TRANSFORMANTS

Les protoplastes pouvant être plurinucléés, il est nécessaire d'effectuer plusieurs étapes de purifications pour obtenir un transformant issu d'un seul noyau. Les transformants (G1) sont repiquées une première fois sur un milieu sélectif contenant de l'hygromycine ou de la généticine. Cette étape est réalisée manuellement à l'aide de cure-dents stériles. Environ 48 transformants sont récupérés dans des microplaques de 48 puits et incubés à 30°C jusqu'à conidiation (G2). A partir de là, un deuxième repiquage est effectué à l'aide du robot colonies-piqueur Tecan. Ce deuxième repiquage a pour objectif de prélever des spores et de les isoler sur une boîte de milieu sélectif (G3) afin d'obtenir des mutants homocaryotiques. Les thalles issus de spores bien isolés sont alors repiqués au centre d'une boîte de Pétri de milieu sélectif et incubés à 30°C durant une semaine afin de permettre une sporulation optimale (G4) pour récolte et conservation. La Figure 23 décrit les étapes de purification des

transformants.

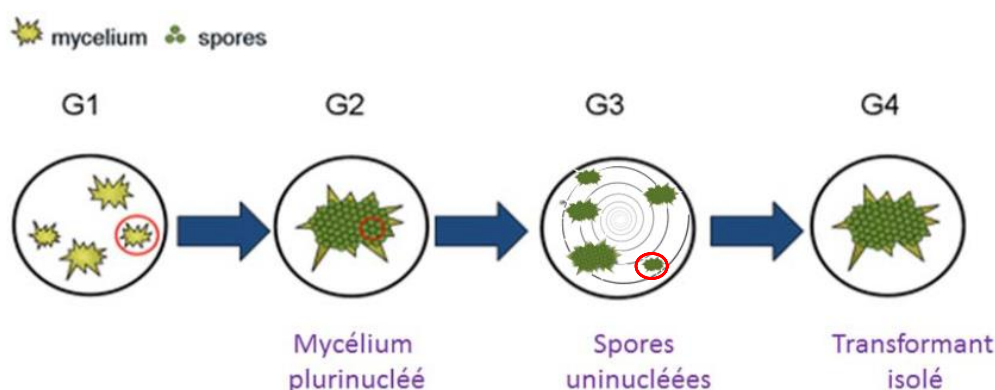


Figure 23 : Etapes de purification des transformants.

G1 : génération 1 : il s'agit de la boîte sur laquelle les protoplastes ont été étalés après la transformation

G2 : génération 2 : les transformants sont repiqués sur milieu sélectif dans des plaques 48 puits

G3 : génération 3 : les conidies sont isolées par étalement circulaire au robot

G4 : génération 4 : un mycélium isolé est repiqué pour donner un transformant purifié
(Laetitia CHAN HO TONG 2017)

XVIII. PCR DE VERIFICATION DES INVALIDATIONS

Pour les PCR de vérification de l'insertion, le mélange réactionnel pour la PCR contient 5 µL du tampon de réaction 5X Herculanse II, 0,5 µL de dNTP, 1 µL de chaque amorce (Annexe VIII) à 10 µM, 0,3 µL de la Herculanse Fusion DNA Polymerase (Agilent) et la matrice ADN. Les cycles PCR récapitulés dans le Tableau 11 ont été utilisés :

Nombre de cycle	Température	Temps
1	95°C	2 min
30	95°C	20 s
	Tm—5°C	20s
	72°C	30s/kb
1	72°C	5 min
1	12°C	∞

Tableau 11 : Programme d'amplification utilisé pour les PCR.

Les amorces ont été créées avec le programme Geneious Prime (version 2020.0) créé par Biomatters.

XIX. SOUTHERN BLOT

XIX.1. Création des sondes

Les sondes permettant la révélation par chimioluminescence ont été générées par PCR grâce au kit PCR DIG Probe Synthesis Kit de Roche (Cat. No. 11 636 090 910) et sont disponibles en Annexe VIII.

XIX.2. Cultures, digestion, migration sur gel

Afin d'obtenir suffisamment de biomasse pour l'extraction d'ADN génomique, les mutants ainsi que les souches mères ont été cultivés sur plaque DeepWell 24 puits dans le milieu EJ (Annexe) à l'aide du système de culture « Duetz-system » dans une chambre de culture Infors durant 7 jours à 30°C et sous agitation à 240 rpm.

Après récolte du mycélium, l'extraction d'ADN se fait selon la méthode du kit d'extraction Macherey-Nagel® Nucleospin – Genomic DNA from soil (Annexe IX).

Une fois extrait, l'ADN génomique est digéré dans un volume total de 150 µL à 37°C durant 12h, puis l'enzyme est inactivée à 60°C durant 20 min. Le produit de digestion est ensuite concentré au SpeedVac™ (Thermofisher) pour atteindre un volume final d'environ 10 µL. Le Tableau 12 ci-dessous récapitule les enzymes de restrictions utilisées pour chacun des mutants et leurs souches mères ainsi que les profils de digestion attendus.

Gène	Enzymes choisies	Profils attendus
<i>dim2</i>	BglII	WT : 5 kb et 7 kb Mutant : 11,5 kb
<i>dim5</i>	BamHI + EcoRI	WT : 3,6 kb Mutant : 2,1 kb et 3 kb
<i>hp1</i>	EcoRI	WT : 11,2 kb Mutant : 5,3 kb et 7,6 kb
<i>rid</i>	BglII + BamHI	WT : 2,5 kb et 2,7 kb Mutant : 5,3 kb

Tableau 12 : Récapitulation des enzymes de restriction utilisées pour la digestion de l'ADN génomique lors des expériences de Southern blot ainsi que les profils de digestion attendus.

La migration de l'ADN est effectuée sur gel d'agarose 1% au Tris Acétate EDTA (TAE) pendant 2h30 à 90V. Enfin, la qualité de la digestion est vérifiée grâce au bromure

d'éthidium (BET) et révélation sous UV.

Le gel est ensuite plongé pendant 10 minutes dans 250 mM de HCl afin de dépuriner l'ADN et faciliter le transfert sur membrane. Puis, il est immergé dans une solution de dénaturation (Annexe) 2x 15 minutes avant d'être abondamment rincé à l'eau distillée. Il est ensuite traité avec une solution de neutralisation (Annexe) également durant 2x 15 minutes. Enfin, le gel est équilibré dans du SSC 20X (Saline Sodium Citrate) pendant au moins 10 minutes.

XIX.3. Transfert et révélation

L'ADN du gel est enfin transféré sur une membrane de nylon (Merck® HNWP04700), par transfert capillaire humide avec du SSC 20X durant toute la nuit. Il est ensuite fixé à la membrane grâce à une exposition aux UV pendant 3 min à 120 mJ.

Le reste des étapes de préhybridation et d'hybridation se font grâce aux tampons et au protocole du kit DIG-probe (Roche, 2008). L'ensemble des incubations des membranes sont faites dans des tubes hermétiques placés dans un four à 67°C sous rotation permanente. La révélation par chimioluminescence a été réalisée avec l'appareil ChemiDoc (BioRad®).

XX. WESTERN BLOT

Afin de procéder à la vérification par Western blot de la disparition de la marque H3K9me3 chez les mutants *Δdim5*, une extraction des histones à partir des souches sauvages et des mutants vérifiés par Southern blot a été réalisée. L'ensemble des solutions nécessaires sont listées en Annexe X.

Les extraits nucléaires totaux ont été culotés à partir d'une solution contenant 5×10^7 protoplastes resuspendus dans 3 mL de solution New Cell Lysis Buffer et d'inhibiteur de protéases après centrifugation à 300 G durant 5 min. Le culot est récupéré dans 300 µL de solution Nuclei Lysis Buffer puis soumis à sonication (Diagenode bioruptor plus sonicateur) durant 5 min en alternances de 5 secondes afin d'extraire les protéines nucléaires.

La concentration en protéines extraites est mesurée grâce au Qubit Protein assay kit (Invitrogen).

Une électrophorèse sur gel SDS-polyacrilamide 10% de 35 µg de protéines nucléaires totales, préalablement diluées dans un tampon Laemmli a ensuite été réalisée. Le transfert des protéines sur membrane PVDF auparavant activée comme

décrit dans la section VII. Dot blot de ce Matériel et Méthodes a été réalisé par transfert semi-sec. Le traitement de la membrane a ensuite suivi le même protocole que décrit dans la section VI. Dot blot de ce manuscrit. L'anticorps primaire anti-H3K9me3 (39161 Active Motif) utilisé pour le ChIP Seq décrit au Chapitre 1 de ce manuscrit a été utilisé pour la reconnaissance des histones marquées par H3K9me3 ainsi que l'anticorps secondaire chèvre anti-lapin couplé HRP (Péroxydase de raifort) pour la détection (172-1019 Biorad).

XXI. COMPLEMENTATIONS

Pour construire le plasmide de complémentation (Annexe XI), le locus du gène *dim5* ainsi que le gène de résistance à la généticine (G418 sur la Figure 24) ont été assemblés dans le plasmide pUC19, préalablement digéré par les enzymes EcoRI et BamHI, grâce au Gibson Assembly Kit (New England Biolabs) selon les recommandations du fabricant. Le locus *dim5* a été amplifié grâce aux amorces Dim5_Fwd et Dim5_Rev (Annexe VIII).



Figure 24 : Plasmide de complémentation pour les mutants $\Delta dim5$.

La partie bleue G418 est le gène de résistance à la généticine, la partie jaune est le gène *dim5* à réintroduire et la partie rouge pUC19 est le squelette plasmidique.

Deux plasmides de complémentation sont séquencés pour s'assurer qu'ils ne contiennent pas de mutations. Ils sont ensuite transformés dans des protoplastes des mutants *Q $\Delta dim5$ 39*, *R $\Delta dim5$ 12*, *R $\Delta dim5$ 28* et *R $\Delta dim5$ 39* qui sont sélectionnés sur milieu PDA contenant 50 $\mu\text{g}/\mu\text{L}$ de généticine. Les transformants sont soumis à un crible de purification comme décrit dans la section XVII de ce Manuscrit et Méthodes.

Afin de s'assurer que la complémentation ait bien fonctionné, les transformants

sont inoculés sur milieu minimum (MM) solide (Annexe I) contenant 2% de sucre glucose, lactose ou sorbitol. Les boîtes sont placées durant sept jours à 30°C dans l'obscurité. La complémentation est considérée comme réussie lorsque les phénotypes sauvages ont été rétablis sur les trois milieux.

XXII. PHENOTYPAGE SUR MILIEUX SOLIDES

Afin de caractériser le phénotype des mutants $\Delta dim5$ ceux-ci ont été placés dans différentes conditions environnementales comprenant des variations dans la source de carbone disponible, un stress oxydatif, un stress osmotique, différentes températures ou bien encore l'exposition à la lumière.

Deux réplicats techniques ont été effectués pour chaque mutant. Toutes les expériences ont duré sept jours.

XXII.1. Sur milieux minimum

Les mutants $\Delta dim5$ ont été inoculés sur trois milieux solides minimum contenant différents sucres à 2% (glucose, lactose ou sorbitol) ainsi que sur deux milieux minimums contenant du glucose à 2% et une molécule induisant un stress (ménadione à 30uM ou NaCl 0,5M). Les boîtes ont ensuite été maintenues dans le noir à 30°C avant observation du phénotype.

XXII.2. Sur milieux riches

Les mutants $\Delta dim5$ ont également été introduits sur milieux PDA et placés dans des environnements divers. Une condition de lumière sous lumière permanente et une condition d'obscurité totale ont été créées afin d'observer les différences dans le développement du champignon à 30°C en fonction de l'exposition lumineuse. Pour finir, dans l'obscurité, trois températures différentes (30°C, 34°C et 37°C) ont été testées afin d'observer les phénotypes mutés.

XXIII. STABILITE SUR LE TEMPS

L'étude de l'instabilité phénotypique des mutants $R\Delta dim5$ après cinq générations avait pour but de déterminer si l'instabilité observée sur milieux liquides et solides tend vers un phénotype commun. Pour ce faire, quatre milieux de culture liquide EJ (Annexe I) avec les sources de carbone suivante ont été testés :

- Lactose + Cellulose
- Lactose
- Glucose + Cellulose

- Glucose

Deux boîtes DeepWell stérilisées ont été inoculées avec 100 μ L de spores selon le plan d'expérience du Tableau 13 suivant :

Lactose + Cellulose	RutC30	<i>RΔdim5 28</i>	<i>RΔdim5 39</i>	QM6a	<i>QΔdim5 39</i>
Lactose	RutC30	<i>RΔdim5 28</i>	<i>RΔdim5 39</i>	QM6a	<i>QΔdim5 39</i>
Glucose + Cellulose	RutC30	<i>RΔdim5 28</i>	<i>RΔdim5 39</i>	QM6a	<i>QΔdim5 39</i>
Glucose	RutC30	<i>RΔdim5 28</i>	<i>RΔdim5 39</i>	QM6a	<i>QΔdim5 39</i>

Tableau 13 : Schématisation du plan d'expérience en plaques DeepWell.

La répartition des souches se fait par colonnes tandis que les quatre milieux se répartissent par lignes.

Puis elles ont été placées dans un incubateur Infors à 30°C et 240 rpm grâce système de culture «Duetz-system». Le plan d'expérience suivant a été mis en place (Figure 25) :

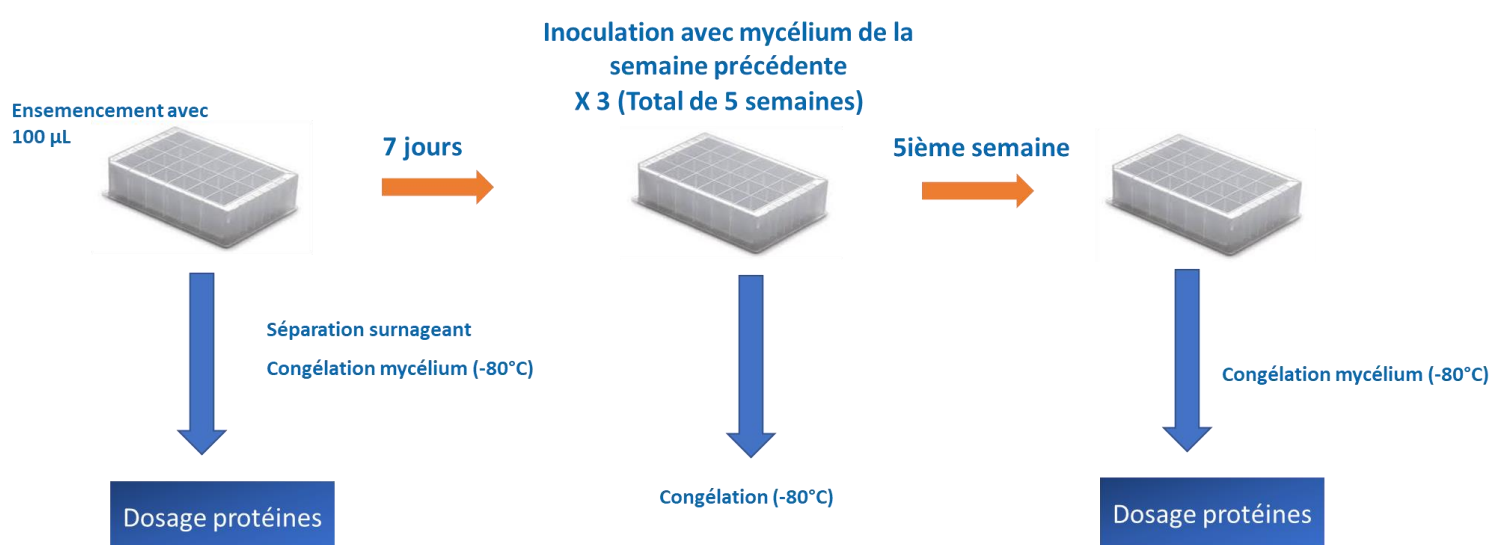


Figure 25 : Plan expérimental de l'expérience de cinq semaines visant à établir un phénotype commun aux mutants *R Δ dim5*.

L'inoculation d'une semaine à l'autre sur milieu frais est faite grâce 200 µL de surnageant contenant du mycélium. Puis, la séparation du surnageant et du mycélium est faite par centrifugation à 7000 g pendant 5 min sur la centrifugeuse Sorvall Lynx 6000 (Thermo Scientific). Un millilitre de surnageant est récupéré afin de procéder au dosage de la concentration enzymatique. La congélation du mycélium est faite directement en plaçant les plaques au congélateur à -80°C.

Le dosage de protéines a été effectué avec la méthode de Bradford telle que décrite en section II.3 de ce Matériel et Méthodes.

CHAPITRE 1

La première partie du travail de cette thèse consiste en l'élaboration d'une cartographie de marques épigénétiques sur deux souches de référence de *Trichoderma reesei* : la souche hyper-productrice RutC30 et la souche sauvage QM6a. Deux conditions ont été étudiées : une condition d'induction de production de cellulases et une condition de non-induction.

La génération des données épigénétiques pour les trois marques H3K4me3, H3K9me3 et H3K27me3 est présentée ici mais seule l'analyse et l'interprétation des résultats pour la première de ces marques seront abordées.

L'étude, chez ces deux souches, de la distribution de la marque H3K4me3, caractéristique de l'euchromatine, est aujourd'hui nécessaire à une meilleure compréhension du système de production biotechnologique qu'est *Trichoderma reesei*. En effet, si jusqu'à présent peu d'informations sont disponibles à ce sujet, il semble néanmoins établi qu'une régulation épigénétique entre en jeu lors de la production d'enzymes chez cet organisme (Chapitre II.2.2.3.3 Régulation épigénétique de la production de cellulases chez *Trichoderma reesei*). Néanmoins, nous ne pouvons pas affirmer savoir à quelle échelle (systémique ou à l'échelle du gène) et sous quelle forme (zones génomiques très enrichies en marque euchromatinienne, présence sur des groupements de gènes particuliers, patrons de distribution de la marque etc...) le contrôle de la régulation génétique par la présence d'euchromatine se manifeste. En outre, nous ignorons s'il existe une redistribution de H3K4me3 lorsque le champignon est placé en conditions inductrices. L'identification de mécanismes épigénétiques lors de la production enzymatique serait donc un premier pas vers une représentation plus complète des dynamiques mises en place au sein de chaque souche. Enfin, nous méconnaissons si la conformation sauvage du paysage épigénétique et ses potentiels mécanismes régulateurs ont été bouleversés par les mutagénèses menant à la production accrue de cellulases. Si ceci était le cas, cela impliquerait la présence d'un système « muté » de régulation de la production enzymatique et donc de potentiels leviers pour l'amélioration de celle-ci.

Ainsi, la constitution d'une cartographie de référence chez la QM6a et chez la RutC30 dans les deux conditions citées plus haut permettra dans un premier temps de répondre aux interrogations concernant la distribution générale de la marque sur les deux génomes. Un éclairage plus précis sera également apporté individuellement au sein de chacune des souches sur les mécanismes d'ouverture chromatinienne lors d'une mise en condition d'induction de production en contraste avec la condition basale de non-induction. Ensuite, une comparaison de ces paysages chromatiniens pour une même condition, d'induction ou de non-induction, mais entre souches nous instruira sur les changements épigénétiques ayant lieu entre le génome de la souche sauvage et celui de la souche hyper-productrice. De plus, les paysages chromatiniens seront mis en regard de l'expression des gènes pour chacune des souches à l'aide d'un transcriptome pour chacune des conditions, dans le but de comprendre les liens entre dynamique épigénétique et expression génétique.

Pour cela, un protocole fed-batch en fioles mimant les conditions de production industrielles d'induction de cellulases ou de non-induction a été réalisé grâce à l'utilisation de la technique de fermentation en fioles alimentées décrite dans (Jourdi et al. 2012). Dans un deuxième temps, à partir de prélèvements de mycéliums effectués lorsque l'induction est stabilisée, une immunoprécipitation de la chromatine suivie d'un séquençage haut débit (ChIP-Seq : « *Chromatin Immunoprecipitation followed by Sequencing* ») a été réalisée pour les modifications post-traductionnelles H3K4me3 (euchromatine), H3K9me3 (hétérochromatine constitutive) et H3K27me3 (hétérochromatine facultative) le long du génome des deux souches (Figure 26). La préparation des librairies et les séquençages de la souche QM6a ont été réalisés par le Joint Genome Institute (JGI) dans la cadre d'un partenariat avec un réseau de laboratoires impliqué dans l'étude épigénétique des champignons filamenteux (projet au *JGI Community Science Program: Epigenomics of carbon recycling fungi*). Également, une extraction d'ARN sur ces mêmes mycéliums a servi au séquençage et à l'analyse par bioinformatique des transcriptomes. Les séquençages des ChIP-Seq et des transcriptomes de la souche RutC30 ont été réalisés au laboratoire The Next Generation Sequencing (NGS) Core Facility - Plateforme de Séquençage Haut Débit I2BC - UMR9198. Pour finir, une analyse approfondie des données a été réalisée pour la marque H3K4me3 avec un travail sur la mise au point du pipeline bioinformatique dans le but d'obtenir une cartographie de l'euchromatine chez *Trichoderma reesei* (Figure 26).

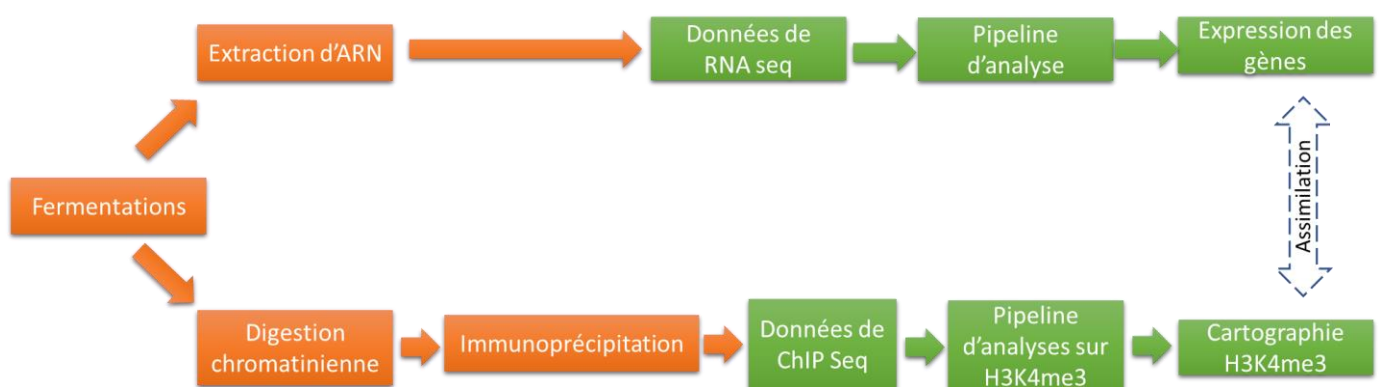


Figure 26: Schéma des étapes d'expérimentations nécessaires à la récolte des données de RNA seq et de ChIP seq.

Étapes d'expérimentations : orange

*Analyse des informations relatives au transcriptome et l'épigénome de *Trichoderma reesei* : vert*

RESULTATS

I. FERMENTATIONS

Les fermentations sur la souche RutC30 ont été réalisées avant le commencement de la thèse en 2018 par un post-doctorant du laboratoire, Vincent Grandperret, tandis que les fermentations sur la souche QM6a ont été effectuées durant la première partie de cette thèse. Les protocoles utilisés ont été dessinés pour des souches de types hyper-productrices donc quelques ajustements ont été nécessaires notamment au moment de l'inoculation de la souche QM6a dans le milieu.

Les fermentations se déroulent en deux temps (Figure 27). Dans un premier temps, une phase de batch a lieu pour tous les réplicats pendant laquelle le champignon va consommer la totalité du glucose introduit au début de l'expérience et produire de la biomasse. Une fois le glucose consommé entièrement (au bout de 48h environ), la phase de fed-batch peut commencer. C'est à cette étape que se différencient les conditions d'expérimentation. Certaines fioles vont se voir placées en fed-batch avec introduction de glucose, ce qui ne déclenche pas de production enzymatique, tandis que le reste des fioles est placé en fed-batch avec introduction de lactose. Ces dernières fioles sont donc celles chez lesquelles nous attendons une induction de production de cellulases.

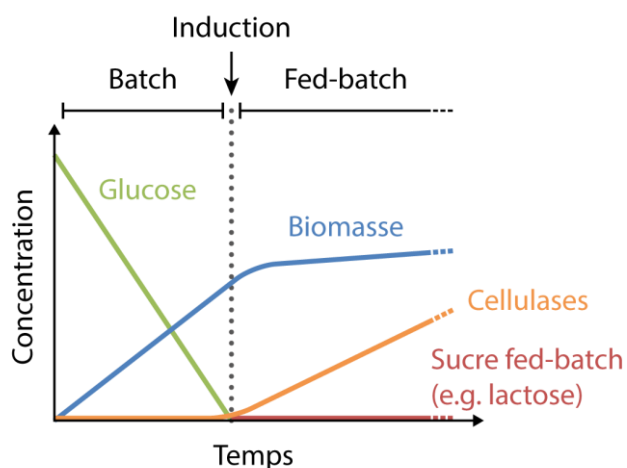


Figure 27: Schématisation des tendances théoriques des concentrations en glucose, en cellulases, en biomasse et en sucre durant l'expérience de fermentation en fioles en fed-batch lactose.

La concentration théorique en cellulases des fioles en fed-batch glucose n'est pas censée augmenter.

Tous les jours, un prélèvement permettant un dosage de la production de

protéines extracellulaires (à 90% composés de cellulases) et d'activité enzymatique est effectué. D'autres vérifications telles que des mesures de concentration en glucose (durant la phase de batch) et de biomasse produite permettent de surveiller le bon déroulement des fermentations et de décider quelles fioles seront utilisées pour le reste des expériences. Au total, 96h sont nécessaires à la réalisation de ce protocole.

Pour chaque souche et chaque condition, trois réplicats biologiques ont été produits (Figure 28).

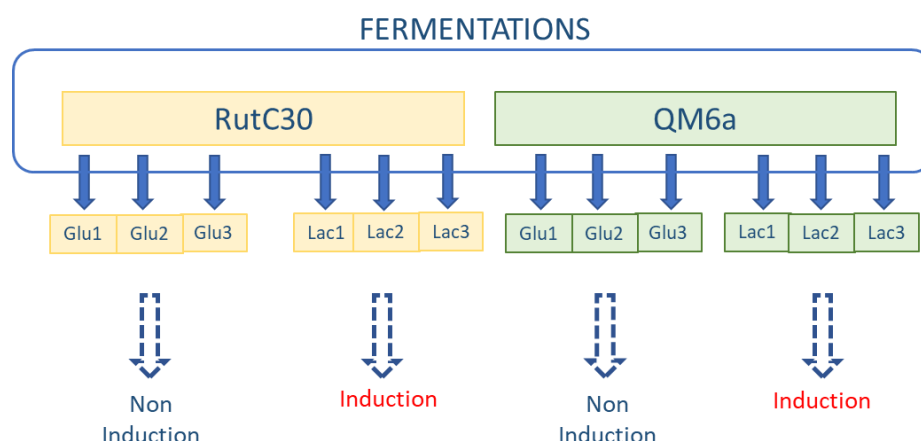


Figure 28: Dessin expérimental des expériences de fermentation et ChIP Seq.

Les fermentations sont réalisées dans deux conditions : condition d'induction de production de cellulases (lactose) et condition de non-induction de production de cellulases (glucose). Pour chaque souche et chaque condition, trois réplicats ont été produits.

I.1. Fermentation de la souche QM6a

Nous allons présenter ici les résultats des expériences en fioles effectuées avec la souche QM6a qui ont été utilisées pour les échantillons de ChIP Seq et de RNA Seq. Nous avons trois réplicats biologiques en glucose (Glu1, Glu2 et Glu3) et trois réplicats biologiques en lactose (Lac1, Lac2 et Lac3).

Un travail d'optimisation pour l'expérience avec la QM6a a été nécessaire pour la mise en place du protocole de fermentation. En effet, le taux d'ensemencement a dû être calculé avec précision car un problème d'agglomération des cultures avec la formation de pellets est apparu dès le début des manipulations. Une quantité finale de $2,5 \cdot 10^5$ spores au total dans les 57 mL de milieu a été déterminée comme optimale pour minimiser l'agrégation des cultures.

La quantité d'enzymes produite par la souche sauvage après 48h de fed-batch étant extrêmement faible même lors de l'induction en lactose, seule l'activité enzymatique démontre d'une différence significative entre les deux conditions. Une distinction claire entre les fioles en condition glucose et les fioles en condition lactose peut se faire grâce à l'activité spécifique des cellobiohydrolases (pNPL) (Figure 29 A).

Après le début de la phase de fed-batch, la croissance de la biomasse est attendue comme stationnaire. Mais contrairement à cela, une augmentation nette de la quantité de biomasse a eu lieu dans certaines des fioles au cours des 48h de fed-batch (Figure 29 B). Le protocole utilisé ayant été dessiné et optimisé pour les souches hyper-productrices telles que la RutC30, la croissance observée chez la QM6a lors de la phase de fed-batch résulte probablement de la disponibilité d'une quantité suffisamment importante du sucre pour que le champignon puisse produire de la biomasse.

Concernant le pH des cultures, au bout de 48h de batch, le milieu s'acidifie. Le pH est alors réajusté à une valeur aux alentours de 4,3 par ajout de soude 12%. Après 24h de fed-batch, le pH est à nouveau vérifié. Il doit être compris entre 3,8 et 5 afin de permettre le bon déroulement de l'expérience. C'est pour cela que pour le Lac2 ce paramètre a dû être modifié avec de l'acide sulfurique 15% car après une journée en fed-batch le pH était déjà au-dessus de 5 (Figure 29 C). Nous avons néanmoins décidé de conserver cette fiole car sa production de biomasse n'est pas aberrante par rapport aux autres (Figure 29 B) et son activité enzymatique pNPL détectée reste satisfaisante malgré une différence avec les deux autres réplicats (Figure 29 C).

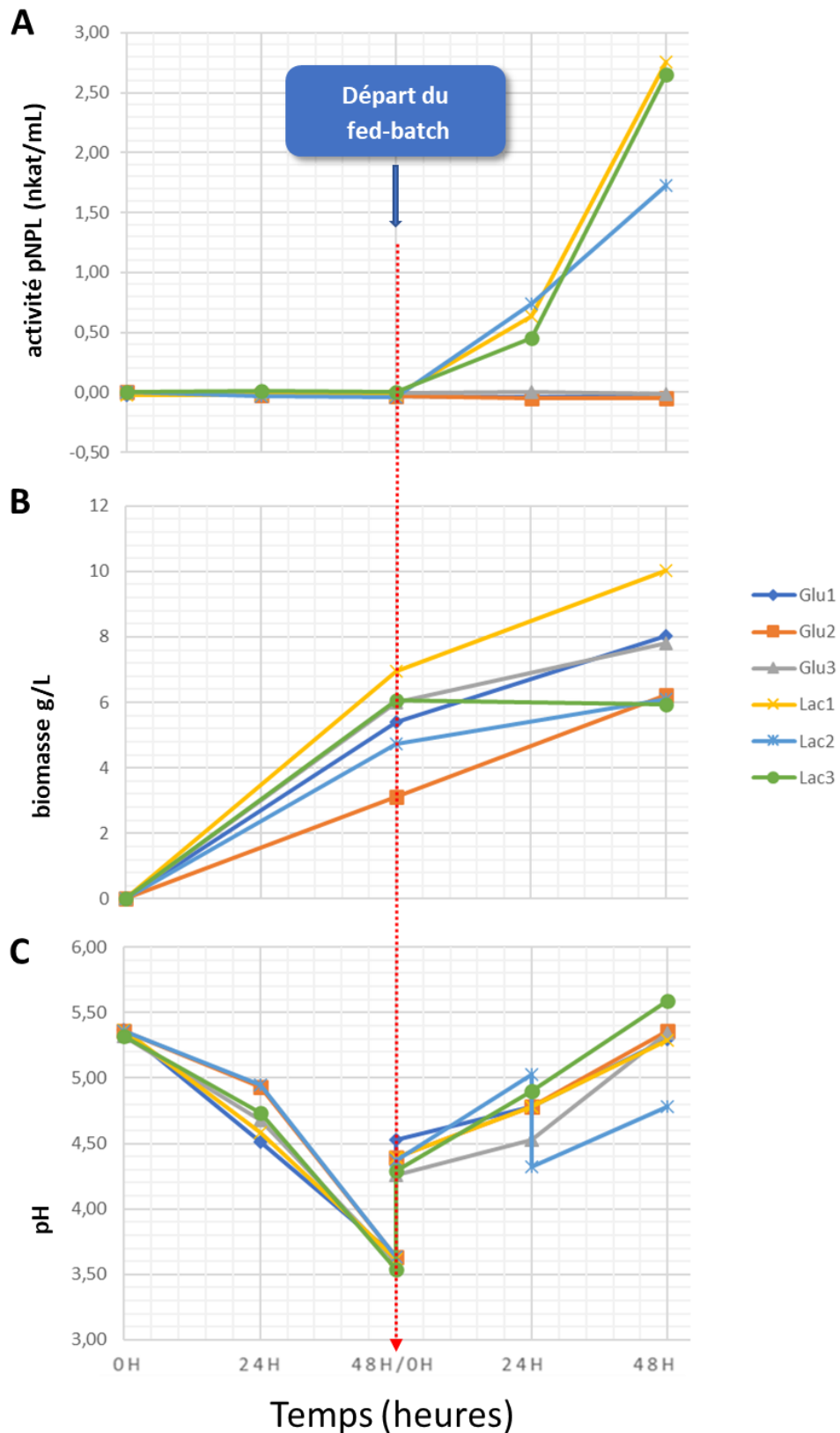


Figure 29: Suivi des fermentations en fiole de la souche Qm6a.

(A) Activité spécifique enzymatique pNPL des fermentations en fiole. (B) Concentration de la biomasse dans les cultures en fioles après 48h de batch et 48h de fed-batch. (C) Profils du pH des cultures en fioles. La fiole du réplikat Lac2 affichait une valeur légèrement au-dessus de 5 après 24h de fed-batch. Un réajustement a été opéré afin de redescendre le pH à 4,3.

I.2. Fermentation de la souche RutC30

La Figure 30 ci-dessous, renseigne sur les paramètres de suivi des expériences de fermentation de la souche RutC30. La quantité de protéines sécrétée dans le milieu au cours du temps est, contrairement à la QM6a, suffisamment importante pour distinguer les fioles ayant été en fed-batch glucose de celles en fed-batch lactose. Nous n'avons donc pas jugé nécessaire de faire une mesure de l'activité enzymatique pour vérifier le bon déroulement des fermentations. De plus, chez *Trichoderma reesei* les protéines extracellulaires produites sont majoritairement des cellulases (Herpoël-Gimbert et al. 2008) et une simple mesure par dosage de Lowry sur les surnageants de cultures permet de quantifier les protéines cellulolytiques sécrétées. Ainsi, une nette différenciation des échantillons lactose est faite grâce à une concentration finale en cellulases à 48h de fed-batch de $3,5 \text{ g.L}^{-1}$, tandis que les échantillons glucose restent aux alentours de $0,4 \text{ g.L}^{-1}$ (Figure 30 A). Les valeurs de l'échantillon Lac2 étant extrêmement proches de celles de Lac3, il n'est pas possible de les distinguer sur le graphique.

Comme attendue, en fed-batch en lactose la croissance de biomasse reste stationnaire puisque désormais le sucre n'est plus utilisé pour la croissance mais pour la production enzymatique. En revanche, une augmentation nette de la quantité du mycélium a lieu dans les fioles en condition glucose comme observé chez la QM6a (Figure 30 B).

Concernant le pH, toutes les fioles ont suivi la tendance attendue. Il est intéressant de remarquer le regroupement des échantillons par conditions avec les fioles en glucose affichant un pH final entre 4,6 et 4,8 tandis que les fioles en lactose ont gardé un pH plus faible entre 4,1 et 4,3. Il est possible que ces tendances soient dues à la croissance du mycélium plus importante dans les fioles en glucose, entraînant une augmentation du pH. D'un autre côté, l'arrêt de la croissance dans les fioles lactose aurait permis un maintien stationnaire du pH après le départ du fed-batch (Figure 30 C).

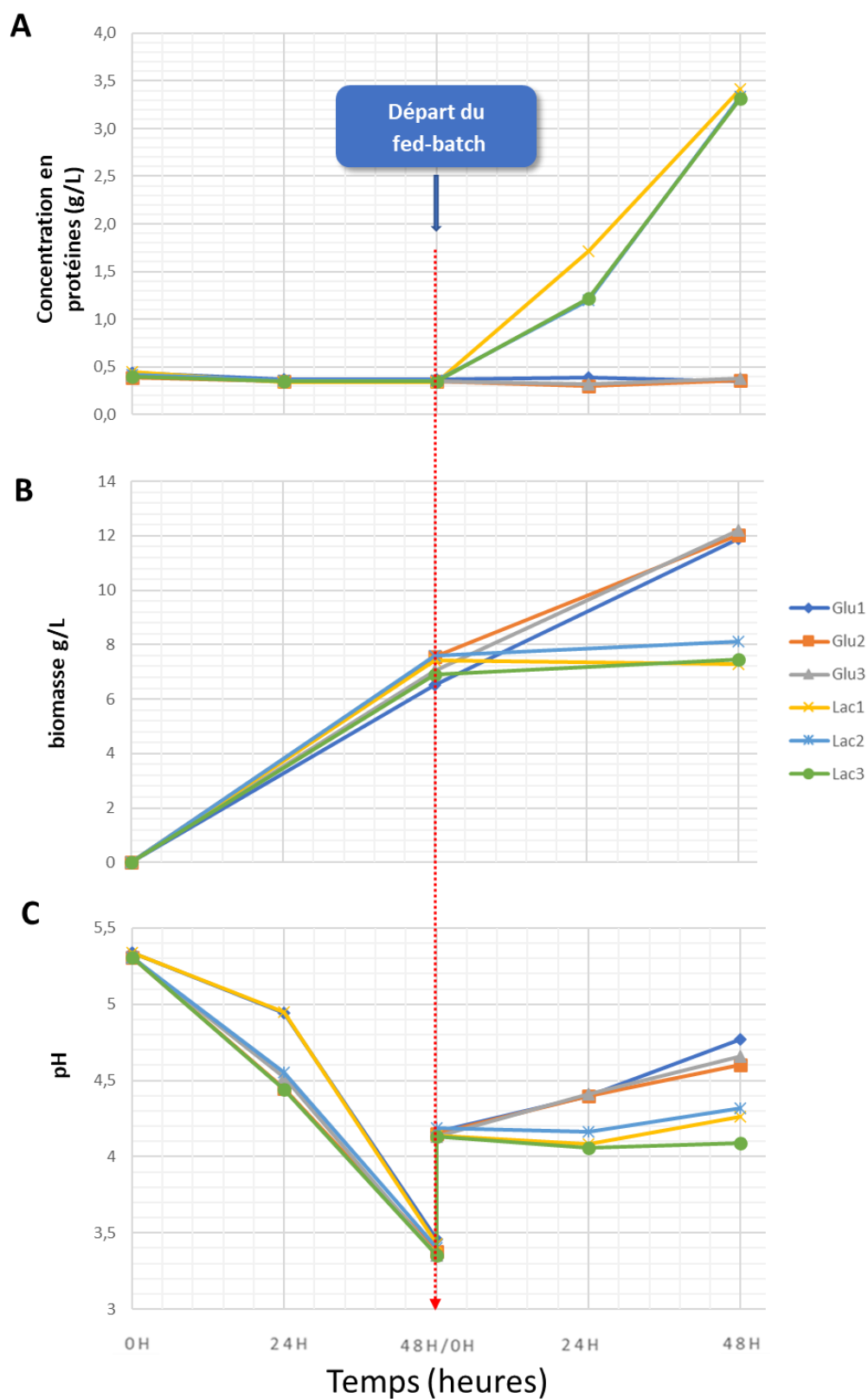


Figure 30: Suivi des fermentations en fiole de la souche RutC30.

(A) Suivi des quantités de protéines sécrétées dans le milieu au cours des fermentations. Un dosage de Lowry a été effectué pour chaque fioles chaque jour. Les valeurs pour les fioles de Lac2 et Lac3 sont très proches et ne permettent pas une distinction claire des courbes de suivi. (B) Concentration de la biomasse dans les cultures en fioles après 48h de batch et 48h de fed-batch. (C) Profils du pH des cultures en fioles.

Une fois les expériences en fioles finies, le prélèvement des échantillons de mycélium de chacune d'entre elles a été réalisé pour les RNA Seq et les ChIP Seq.

II. RNA SEQ

Afin de réaliser une étude du transcriptome, 8 mL de chaque échantillon ont été prélevés, congelés et ont servi pour l'extraction d'ARN totaux. Leur qualité a été vérifiée par analyse de leur profil électrophorétique. Les résultats sont disponibles en Annexe XII et Annexe XIII. Les ARN ont ensuite été envoyés à la plateforme IBENS et au JGI où les librairies, séquençages Illumina et les analyses différentielles ont été réalisées. Les résultats traités nous ont ensuite été délivrés.

Nous ne souhaitons pas ici nous étendre en détails sur les résultats de ces transcriptomes en tant que tels car ils ont été réalisés dans un but de support additionnel aux résultats de ChIP Seq. Également, ils ont aussi servi de vérification supplémentaire au bon déroulé des expériences en fioles puisqu'ils permettent la visualisation de l'activation de l'expression des gènes codant pour les enzymes cellulolytiques et autres facteurs majeurs de l'induction enzymatique en condition lactose. Dans le reste de ce manuscrit, nous ferons référence aux gènes différenciellement exprimés comme DE.

Nous discuterons donc plus en profondeur ces résultats au regard des informations épigénétiques et ne traiterons ici que des informations générales extraites.

L'expression des gènes en glucose a servi d'expression contrôle pour l'analyse différentielle et deux filtres ont été utilisés sur les données. Nous avons ici considéré comme différenciellement exprimés en lactose, les gènes possédant une « *adjusted p-value* » inférieure à 0,001 et une valeur absolue du log2FC supérieure à 2. Également, nous considérons comme exprimés les gènes possédant un nombre RPKM (normalisé par la longueur du gène) de « *reads* » supérieur ou égal à 10.

II.1. Résultats du transcriptome de la souche QM6a

Chez la souche sauvage QM6a, un total de 279 gènes sont différentiellement exprimés en lactose, avec 206 gènes surexprimés contre 73 sous-exprimés (Figure 31).

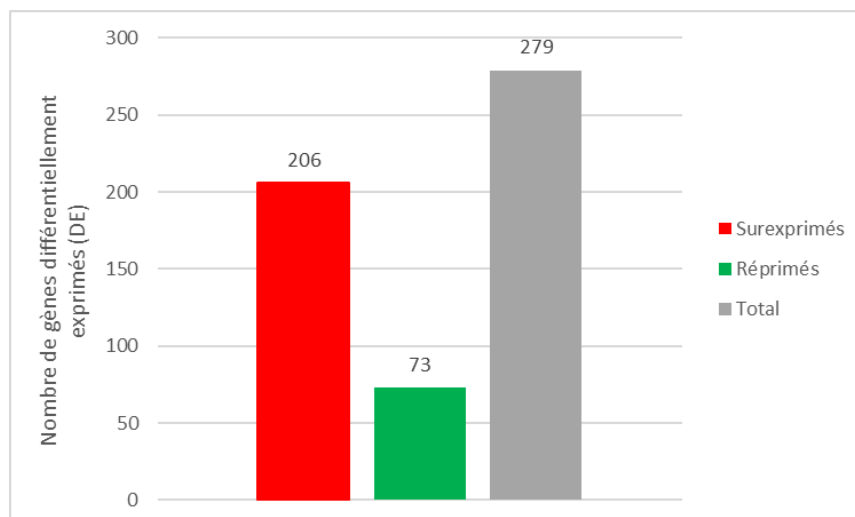


Figure 31 : Nombre de gènes différentiellement exprimés chez la souche sauvage QM6a en condition lactose.

Parmi les 206 gènes surexprimés en lactose nous en comptons 56 codant pour des cellulases et autres facteurs impliqués dans la réponse cellulolytique du champignon. Nous retrouvons notamment les gènes majeurs impliqués dans l'induction des cellulases tels que *xyr1*, *ace3*, *clr2* et *rxr1* (facteurs de transcription), et les gènes codant pour les cellulases et hemicellulases *bgl1* et *bgl2* (β -glucosidase), *bxl1* (β -xylosidase), *cbh1* et *cbh2* (cellobiohydrolases), *xyn2*, *xyn3*, *xyn4* et *xyn6* (endoxylanase), *egl1*, *egl2* et *egl3* (endoglucanases), *cel61a* (LPMO) etc... Nous retrouvons également cinq gènes de transporteurs MFS (« *Major Facilitator Superfamily* ») impliqués dans le transport de lactose et autres sucres, dont le gène *crt1*, essentiel pour l'induction de la production de cellulases (Zhang et al. 2013), ainsi que le transporteur de xylose, cellobiose et mannose Tr69957.

Parmi les gènes surexprimés en lactose ne participant pas à la production de cellulases, nous retrouvons aussi les orthologues de gènes impliqués dans le cycle de reproduction asexué du champignon ainsi que de la fonctionnalité de la paroi cellulaire : CRO1, PhiA, Rbt5p, HOC1, etc... (Berteaux-Lecellier et al. 1998; Melin, Schnürer et Wagner 2003; Mao, Zhang et Wong 2003; Neiman et al. 1997). Également, une importante variété de glycosides hydrolases, mis à part les cellulases, sont présentes puisque pas moins de 29 gènes codant pour 20 familles différentes de ces CAZymes sont surexprimés en lactose. Parmi ces 29 gènes, 22 ont déjà été clonés et caractérisés dans *Trichoderma reesei*.

En comparaison, seuls les transporteurs MFS TrSTR3 du xylose et STP1 du glucose/cellobiose et 13 gènes ou orthologues de gènes codant pour des glycosides hydrolases, une glycosyltransferase et une carbohydate esterase font partie des 73 gènes sous-exprimés en lactose. La grande majorité d'entre eux sont des gènes codant pour des protéines aux fonctions variées, dont souvent seul le domaine actif a été déterminé à l'instar de la peptidase putative contenant des domaines S8/S53, famille d'enzymes incluant des endopeptidases et des exopeptidases.

Ainsi, nous observons bien une induction de la transcription des gènes cellulolytiques chez la QM6a en condition d'induction en lactose par rapport à la condition contrôle en glucose. Ceci confirme la capacité de réponse et l'activation de la synthèse enzymatique de la souche sauvage à un environnement propice à la production de cellulases. De ce fait, bien que la quantité mesurée de protéines cellulolytiques extracellulaires reste faible en condition lactose (résultats présentés en I.1), ces résultats d'expression génique couplés à l'augmentation de l'activité enzymatique confirment à nouveau le bon déroulement de nos expériences.

II.2. Résultats du transcriptome de la souche RutC30

L'étude du transcriptome de la souche RutC30 dans les conditions de fermentation réalisées dans cette thèse a été effectué plusieurs fois auparavant au laboratoire et des résultats précédemment obtenus ont déjà été publiés. Nous avons donc décidé de nous baser sur les résultats de (Pirayre et al. 2020) pour valider les nôtres.

La Figure 32 ci-dessous récapitule le nombre de gènes chez la RutC30 surexprimés ou sous-exprimés en condition lactose.

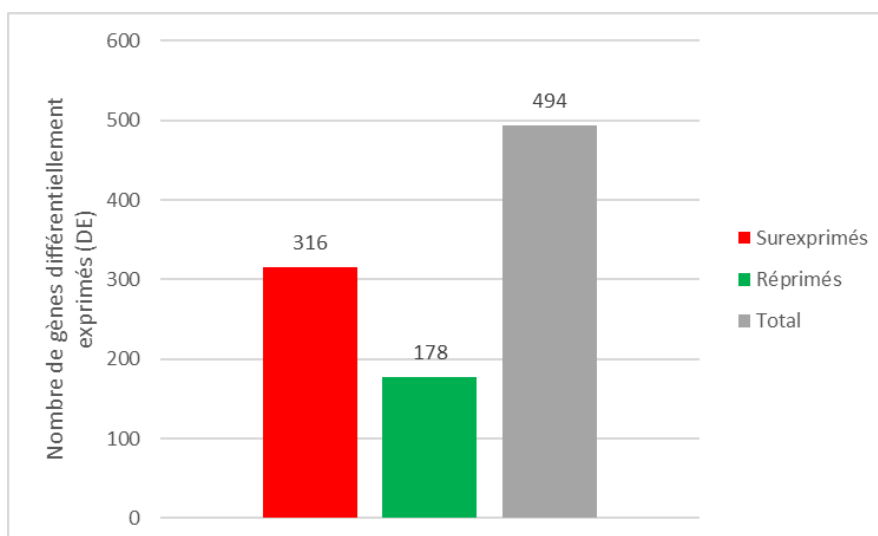


Figure 32 : Nombre de gènes différentiellement exprimés chez la souche hyper-productrice RutC30 en condition lactose.

Ainsi, nous retrouvons parmi les 316 gènes surexprimés en lactose 89 dont la fonction est liée à la production d'enzymes cellulolytiques. Nous retrouvons les principales cellulases impliquées dans la dégradation de sucres complexes (Foreman et al. 2003) : CBH1 et CBH2 (cellobiohydrolases), EGI, EGII et EGIII (endoglucanases), BGL1, BGL2, CEL3C et CEL3D (β -glucosidases), BXL1 (β -xylosidase), XYL1 (D-xylose reductase), XYN1, XYN3 et XYN4 (xylanases), CIP2/GE1, AXE1, AES1 (hémicellulases) (Li et al. 2008), CEL61A et CEL61B (LPMO). De plus, les facteurs de transcriptions majeurs de l'induction sont bien présents : XYR1, ACE3, CLR2.

En plus de cela, 23 gènes putatifs de transporteurs MFS sont surexprimés en lactose dont dix d'entre eux participeraient au transport de sucres connus. Nous comptons parmi eux les gènes codant pour CRT1, Tr69957 (TrB0016W) comme pour la QM6a mais aussi d'autres transporteur MFS putatifs facilitant l'assimilation du lactose, du fucose, du maltose ou encore des cellodextrines.

Également, nous notons une forte représentation des glycosides hydrolases, dont les cellulases, puisque 42 gènes regroupant 28 familles différentes sont notés comme surexprimés en lactose. Pour rappel, (Li et al. 2017), reportaient 228 séquences de gènes, potentiels et vérifiés, codant pour des CAZymes dans le génome de *Trichoderma reesei*. Nous retrouvons donc presque un quart de ces gènes surexprimés dans les conditions mises en place au laboratoire avec du lactose comme source de sucre inducteur et dans des conditions de fed-batch.

Concernant les gènes sous-exprimés en lactose, nous retrouvons ceux codant pour le transporteur MFS STR3 de xylose et pour le transporteur MFS STP1 de glucose/cellobiose. L'orthologue putatif du gène codant pour le transporteur de glucose Hxt chez *F. oxysporum* est aussi sous-exprimés dans cette condition.

Parmi les gènes non impliqués dans la production de cellulases, douze autres transporteurs sont également présents dont celui de la N-acetylglucosamine (GlcNAc) NGT1 ainsi que dix glycosides hydrolases, une putative glycosyltransférase et une putative carbohydate estérase.

Tous ces résultats sont bien en accord avec les résultats décrits dans Pirayre et al, (2020) dans les mêmes conditions d'expérimentations. Au vu des résultats similaires obtenus pour la souche sauvage, nous avons pu progresser avec confiance sur la suite des expériences quant au bon déroulement des fermentations en fioles et la qualité du matériel biologique récolté.

III. IMMUNOPRECIPITATION DE LA CHROMATINE : CHIP-SEQ

Une fois le volume pour le RNA Seq récolté, le reste du mycélium a été utilisé pour les expériences de ChIP Seq. Cette technique sert à étudier les interactions protéines/ADN et permet de générer une cartographie des zones où se situent ces liens. Cette technologie repose sur une immunoprécipitation (IP) dirigée vers les protéines cibles et le séquençage de l'ADN qui leur est lié.

Pour ce faire, une digestion complète de la chromatine est nécessaire afin de libérer un maximum de fragments d'ADN courts, contenant seulement un à quatre nucléosomes (Figure 33 (1)). Cette découpe minutieuse de l'ADN permet une immunoprécipitation (IP) fine par les anticorps et donc une cartographie précise. Notre but ici étant l'ADN lié aux histones H3 portant sur leur queue N-terminale les triméthylations des lysines 4, 9 et 27 (Figure 33 (2)). Pour rappel, H3K4me3, H3K9me3 et H3K27me3, sont respectivement spécifiques des zones d'euchromatine, d'hétérochromatine constitutive et d'hétérochromatine facultative. Il est important que les anticorps utilisés possèdent une haute spécificité pour leur épitope afin d'assurer une extraction uniquement des nucléosomes portant la marque d'intérêt. Un échantillon utilisé par la suite comme contrôle appelé « *input* » est également réalisé en parallèle. L'*input* correspond à l'ADN total digéré mais non immunoprécipité. Il est extrait avant ciblage par anticorps de l'ADN utilisé pour le reste des échantillons et correspond donc à l'ADN total disponible (Figure 33 1bis). Après dissolution des complexes protéiques (Figure 33 (3)), un séquençage de l'ADN récolté est effectué et une cartographie des marques est faite le long des génomes de référence (Figure 33 (4)).

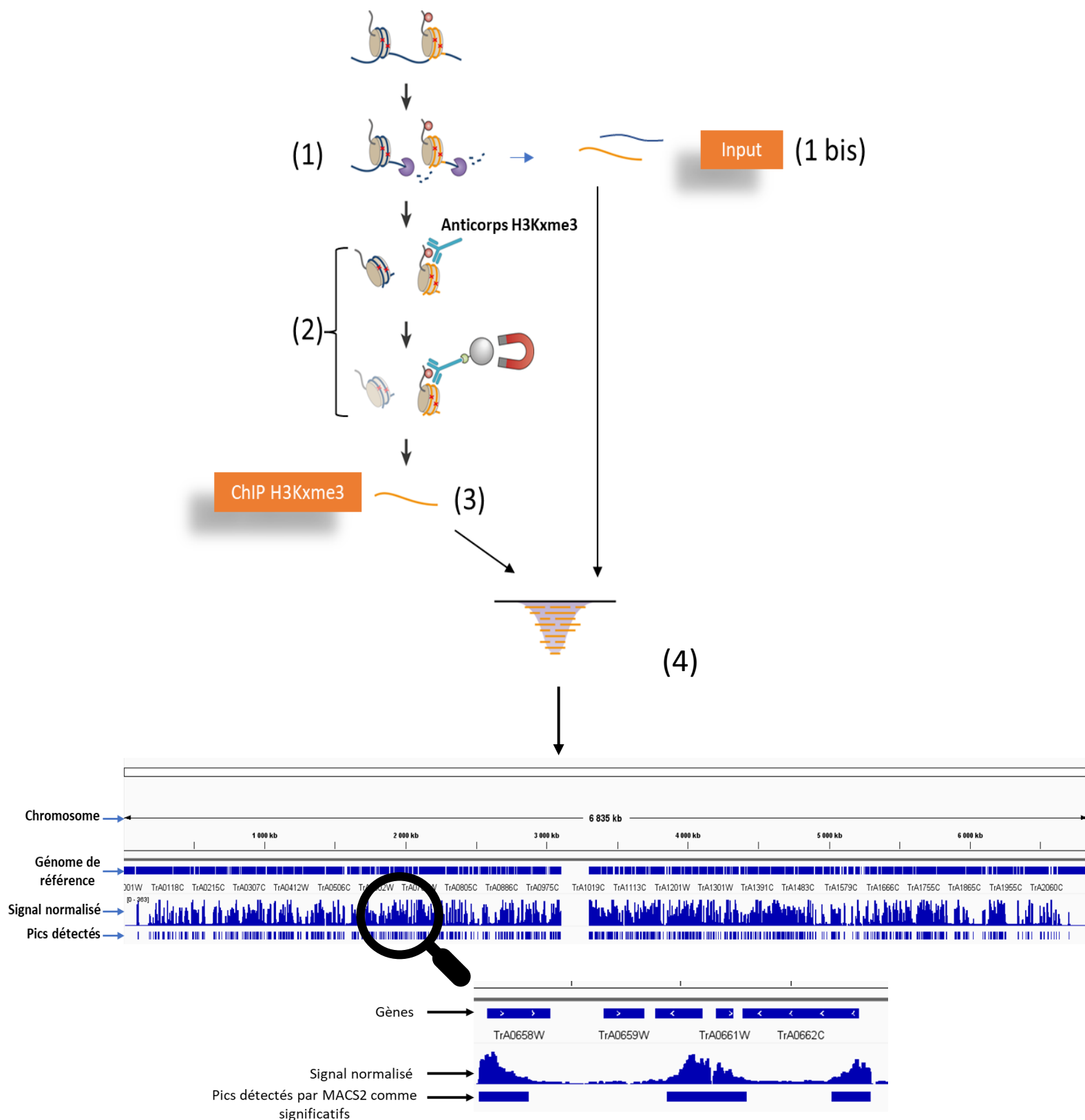


Figure 33: Schématisation de la technique de ChIP-Seq.

Une fragmentation de la chromatine est nécessaire à la séparation des nucléosomes (1). Une fois libérés de la chromatine, l'immunoprécipitation a lieu et les nucléosomes ciblés sont extraits du reste grâce à des billes magnétiques (2). Le séquençage de l'ADN récolté est fait (1 bis) (3) et permet la cartographie des zones touchées par les marques épigénétiques voulues (4).

III.1. Vérification de la spécificité des anticorps

Un des points importants à vérifier pour que l'expérience d'immunoprécipitation soit réussie est la spécificité des anticorps utilisés. En effet, il est nécessaire que l'anticorps soit dirigé uniquement vers le résidu d'histone ciblé (lysine 4, 9 ou 27 triméthylée de H3) afin d'éviter toute réactivité croisée et de s'assurer que le signal observé après analyse provienne seulement de la marque épigénétique choisie.

Ainsi, nous avons effectué des expériences de « dot blot » à partir de peptides commerciaux purs afin de vérifier qu'il n'y ait pas de réactivités croisées chez nos anticorps. Plusieurs concentrations de peptides commerciaux pour les trois marques visées ont été testés (330ng, 100ng, 50ng, 10ng et 5ng) afin de vérifier que même à forte concentration de protéines, les anticorps ne reconnaissent que leur résidu de prédilection. Les trois anticorps utilisés pour les immunoprécipitations ont été testés lors de trois expériences indépendantes (il n'y a pas eu réutilisation de membrane) (Figure 34).

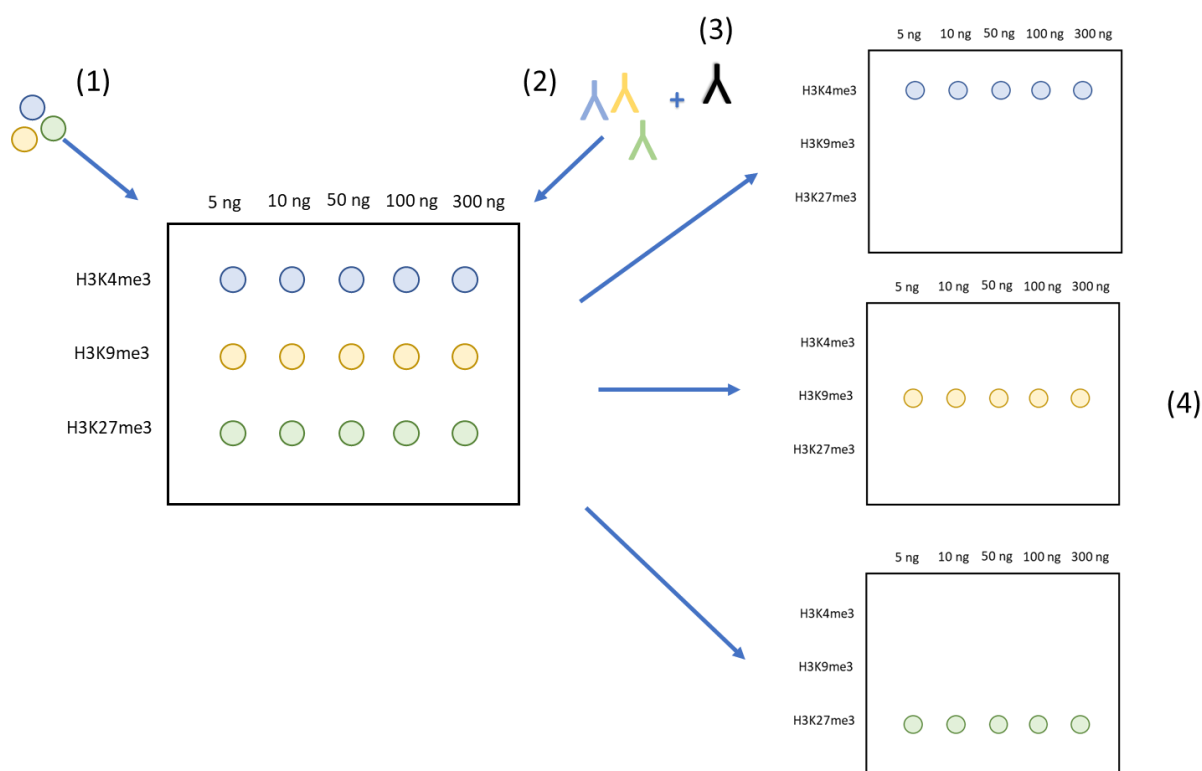


Figure 34: Schématisation de l'expérience de dot blot.

(1) Les peptides commerciaux sont déposés sur la membrane PVDF activée selon les concentrations et les dispositions indiquées. (2) Chaque anticorps primaire est testé individuellement (trois expériences différentes) sur la membrane afin de vérifier leur spécificité. (3) L'anticorps secondaire est ensuite ajouté. (4) Une révélation colorimétrique a été utilisée pour visualiser le résultat.

Les résultats présentés en Figure 35 suggèrent que la spécificité individuelle de chaque anticorps est impeccable et ne souffre d'aucune réactivité croisée entre les résidus choisis. En effet, dans le Figure 35A nous observons un signal détectant la marque H3K4me3, même pour la plus petite concentration, tandis qu'aucune réaction croisée n'est observée pour les autres peptides. De la même manière, un signal est visible pour la membrane correspondant au test de l'anticorps spécifique de H3K9me3 et celle de H3K27me3 (Figure 35B et Figure 35C).

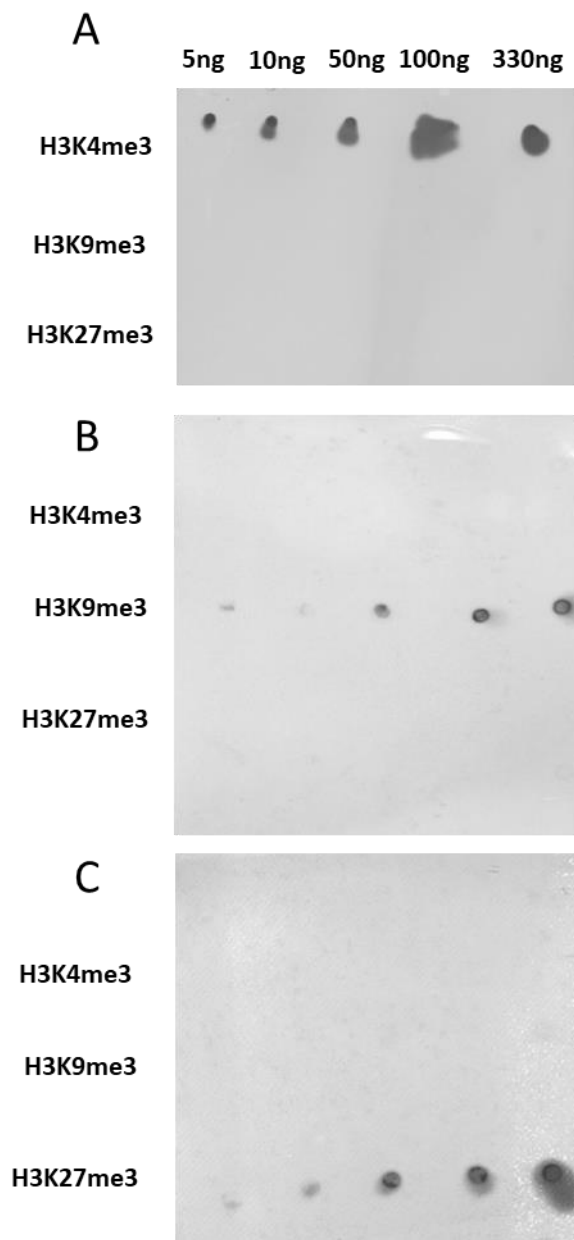


Figure 35 : Résultat de dot blot démontrant la spécificité des anticorps utilisés pour l'expérience de Chip-Seq.

III.2. Immunoprécipitations des échantillons QM6a

Pour chacune des conditions, trois réplicats biologiques ont été générés.

Un des premiers critères de sélection des échantillons choisis pour le séquençage est la qualité de la digestion de la chromatine. Une nucléase micrococcale a été utilisée pour obtenir des fragments d'ADN contenant entre 1 et 4 nucléosomes, permettant ainsi une immunoprécipitation précise des morceaux d'ADN liés aux histones modifiées. Les gels suivants (Figure 36) sont les gels d'électrophorèses permettant la visualisation de cette fragmentation chez les échantillons sélectionnés.

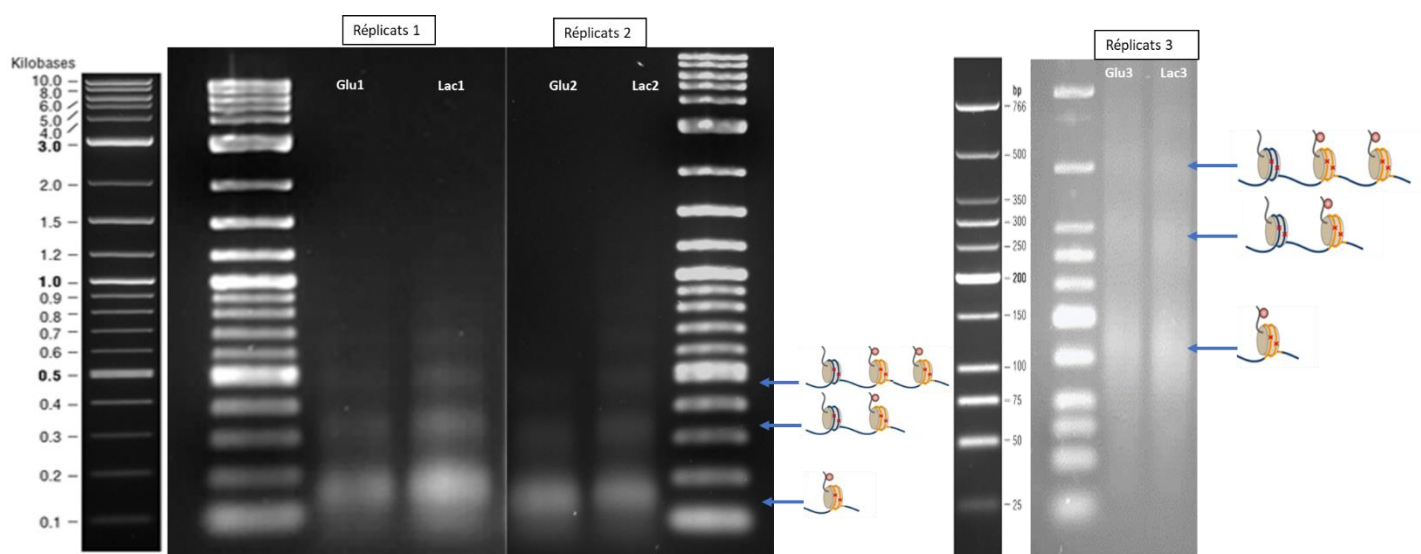


Figure 36: Profils de digestion par nucléase micrococcale de la chromatine extraite des échantillons de mycélium de la souche QM6a.

Pour les réplicats 1 et 2 les gels originaux ont été rassemblés sur la même figure pour une meilleure visibilité puisque la même échelle moléculaire a été utilisée. Le réplicat 3, ne possédant ni la même qualité ni la même échelle, est exposé à part.

Finalement, en se basant sur la qualité des digestions de la chromatine et la quantité d'ADN finale récoltée, les échantillons (six échantillons immunoprécipités pour chaque marque et deux *input* finaux) ont été sélectionnés. Pour chacun d'entre eux les quantités d'ADN obtenus sont répertoriées dans le Tableau 14.

Sucre utilisé lors de l'étape de fed-batch	Réplicat	H3K4me3	H3K9me3	H3K27me3	Input
Glucose	1	98 ng	17 ng	12,6 ng	190,4 ng
	2	93,6 ng	10,7 ng	41,6 ng	306 ng
	3	82,8 ng	18,4 ng	82,8 ng	X
Lactose	1	73,3 ng	23,6 ng	17,8 ng	433,1 ng
	2	107,5 ng	22,5 ng	22,5 ng	531 ng
	3	44,8 ng	48,2 ng	28,4 ng	X

Tableau 14: Quantité d'ADN en nanogrammes obtenues pour les différentes expériences de ChIP réalisées sur la souche QM6a

Entre autres, une digestion satisfaisante de la chromatine a été un point de difficulté majeure qu'il a fallu surmonter en testant différents ratios « quantité de mycélium/concentration en nucléase ».

Également, la récolte d'une quantité d'ADN suffisante après la précipitation de la chromatine a été un autre obstacle à surmonté puisque les quantités finales d'ADN sont à peine de l'ordre de quelques nanogrammes, voire indétectable dans certains cas. Or, le JGI demande une quantité minimale de 12 ng d'ADN précipité pour chacune des trois marques et le contrôle. La marque H3K4me3 étant abondante dans le génome en comparaison avec les marques de l'hétérochromatine il n'a pas été difficile d'atteindre le minima requis pour H3K4me3. En revanche, il a fallu le rassemblement de deux et trois réplicats techniques pour les marques H3K9me3 et H3K27me3, respectivement, pour atteindre ce critère. L'*input* étant un séquençage de l'ADN fragmenté initial sans immunoprécipitation, une grande quantité a pu être récoltée facilement.

Les échantillons ont ensuite été envoyés tels quels à la plateforme de séquençage, sans génération de la banque de séquençage au laboratoire.

III.3. Immunoprécipitation des échantillons RutC30

Les résultats d'immunoprécipitations sur la souche RutC30 présentés dans ce chapitre ont été générés par Vincent Grandperret lors de son passage en post-doctorat au laboratoire. Avant envoi à la plateforme de l'I2BC, il a également préparé la librairie NGS (« *Next Generation Sequencing* ») permettant le séquençage.

Les mêmes critères de sélection ont été appliqués pour ces échantillons que pour ceux de la QM6a : la qualité de la digestion de la chromatine (Figure 37) ainsi que la quantité d'ADN final obtenu (Tableau 15).

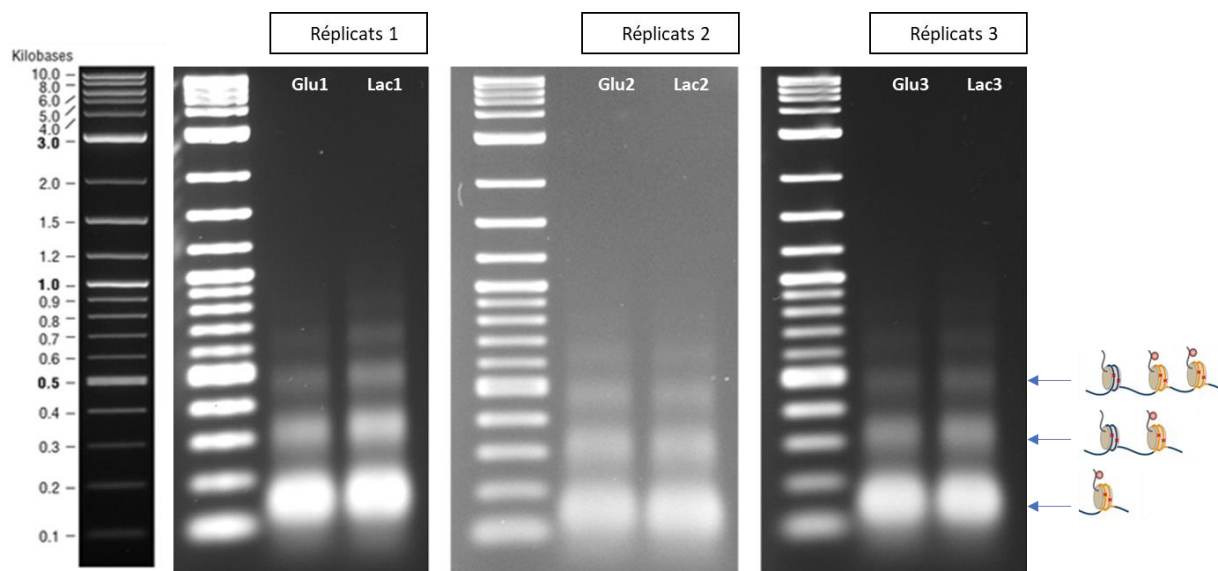


Figure 37: Profil de digestion par nucléase micrococcale de la chromatine extraite des échantillons de mycélium de la souche RutC30.

Sucre utilisé lors de l'étape de fed-batch	Réplikat	H3K4me3	H3K9me3	H3K27me3	Input
Glucose	1	75,90 ng	18,42 ng	ND	X
	2	29,61 ng	10,14 ng	ND	49,20 ng
	3	99,90 ng	9,81 ng	ND	258,00 ng
Lactose	1	38,4 ng	33,3 ng	ND	X
	2	9,75 ng	12,12 ng	1,5 ng	253,8 ng
	3	26,94 ng	13,17 ng	1,77 ng	318,00 ng

Tableau 15: Quantité d'ADN en nanogrammes obtenues pour les différentes expériences de ChIP réalisées sur la souche RutC30.

ND fait référence à Non Détectable.

Dans le cas de la souche RutC30, aucun réplikat technique n'a été rassemblé pour les marques H3K9me3 et H3K27me3.

La préparation de la librairie NGS a ensuite été nécessaire avant envoi des échantillons. Pour cela, des cycles PCR ont été appliqués avec des amorces spécifiques permettant le séquençage en paire (« *paired-end* ») des fragments d'ADN. Pour pallier les quantités non détectables récoltées pour la marque H3K27me3, un nombre un peu plus important de cycles a été effectué (13 ou 14 en fonction du réplikat contre 11 pour

le reste des échantillons des autres marques).

Pour finir, une vérification sur gel d'électrophorèse de la qualité de la librairie après les cycles PCR a été réalisé afin de s'assurer que les cycles d'amplification n'aient pas détérioré l'ADN (Annexe XIV).

IV. CARTOGRAPHIE DE LA MARQUE EPIGENETIQUE H3K4me3

A présent, nous allons nous concentrer sur la cartographie et l'analyse de la marque de l'euchromatine, H3K4me3. La qualité des données tirées pour cette marque a permis d'effectuer des analyses plus simples que celles nécessaires pour H3K9me3 ou H3K27me3. Du travail est encore requis à l'exploitation des informations sur ces dernières. De ce fait, nous avons choisi de ne présenter ici que les résultats sur la marque de l'euchromatine.

Le séquençage des données a été fait en « *paired-end* ». Cela implique qu'un séquençage aux deux extrémités des fragments d'ADN a été effectué, permettant ainsi la génération de données de haute qualité. Les plateformes de séquençage se sont chargées de l'élimination des adaptateurs de la librairie ainsi que de l'élimination des séquences obtenues (« *reads* ») n'étant pas en paire. Les fichiers de données nous ont ensuite été envoyés sous format fastq. La Figure 38 présentée ci-dessous schématise les étapes principales de l'analyse des données de ChIP Seq jusqu'à génération des fichiers contenant la cartographie des pics significatifs détectés par « *peakcalling* ». Dans un premier temps nous allons nous intéresser aux résultats intermédiaires générés par cette première phase d'analyse chez les deux souches. Elles ont permis de rassembler un nombre important d'informations, notamment la qualité du séquençage, le contenu des fichiers et une visualisation de la distribution des *reads* sur le génome de référence. Puis, nous présenterons les résultats du *peakcalling* de la souche sauvage et de la souche hyper-productrice. Cette étape a permis la détection des pics significatifs des échantillons immunoprécipités (IP) au regard du contrôle qu'est l'*input*, dont nous avons discuté auparavant. Nous nous sommes intéressés de manière globale à la distribution de la marque H3K4me3 le long des deux génomes de références et

avons également porté attention à la comparaison de cette distribution entre les souches.

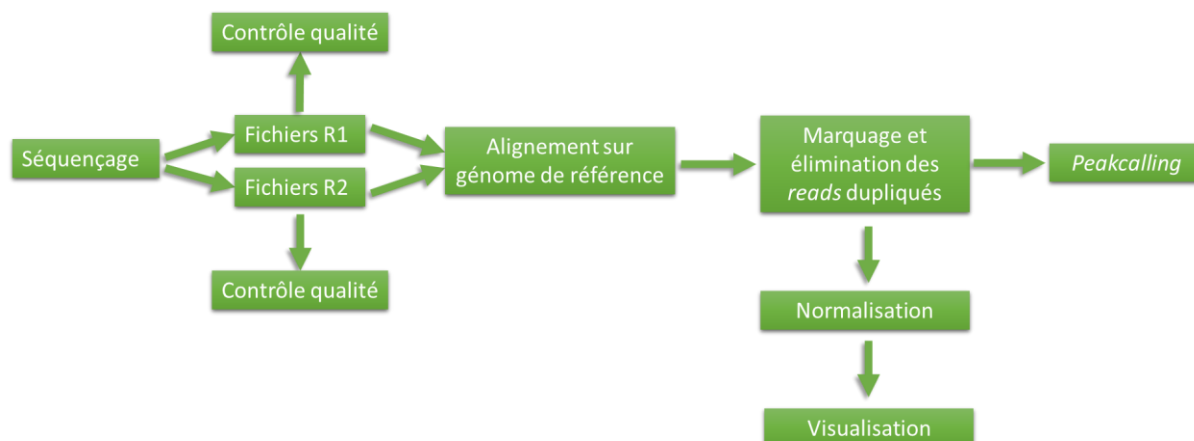


Figure 38: Schématisation du déroulé de l'analyse des données de séquençage jusqu'à l'étape du peakcalling.

Pour la souche RutC30 deux fichiers pour chaque échantillon nous ont été fournis : un fichier pour les séquences R1 (*reads* 1 correspondant à l'adaptateur « *forward* » du séquençage) et un fichier pour les séquences R2 (*reads* 2 correspondant à l'adaptateur « *reverse* » du séquençage). Pour la QM6a, bien qu'également séquencés en *paired-end*, le JGI nous a fourni des fichiers fastq uniques avec les séquences R1 et R2 fusionnées. Nous avons néanmoins préféré régénérer des fichiers séparés pour des raisons que nous aborderons plus tard dans ce manuscrit.

IV.1. Contrôle Qualité

Un contrôle qualité des données de séquençage de chaque échantillon a été réalisé à partir des fichiers fastq envoyées par les plateformes grâce à la fonction fastqc (version 0.11.9) (Babraham Bioinformatics). En plus de s'assurer de la qualité du séquençage, cette première étape permet également de prendre une première série de décisions concernant le traitement des données récoltées.

Dans les six échantillons respectifs des deux souches, le paramètre donnant la qualité de séquençage par base (« *Per base sequence quality* ») est validé permettant ainsi de progresser en sachant que le séquençage est de bonne qualité.

Néanmoins, le paramètre « *Per base sequence content* » ressort avec une alerte rouge pour tous les échantillons. Ce paramètre donne le pourcentage de bases nucléiques à chaque position de la longueur des *reads* et ressort en alerte si une

différence trop importante existe dans la proportion des bases complémentaires (10% pour une alerte orange et 20% pour une alerte rouge). Par exemple, dans le fichier R1 de l'échantillon Glu1 de la QM6a, à la base deux de l'ensemble des *reads*, nous retrouvons plus de 50% de G. En revanche, moins de 30% de C sont détectés (Figure 39). La proportion entre ces bases complémentaires n'est pas respectée à au moins une position le long des *reads*. La différence en proportion est supérieure à 20%, donc le contrôle qualité n'est pas passé.

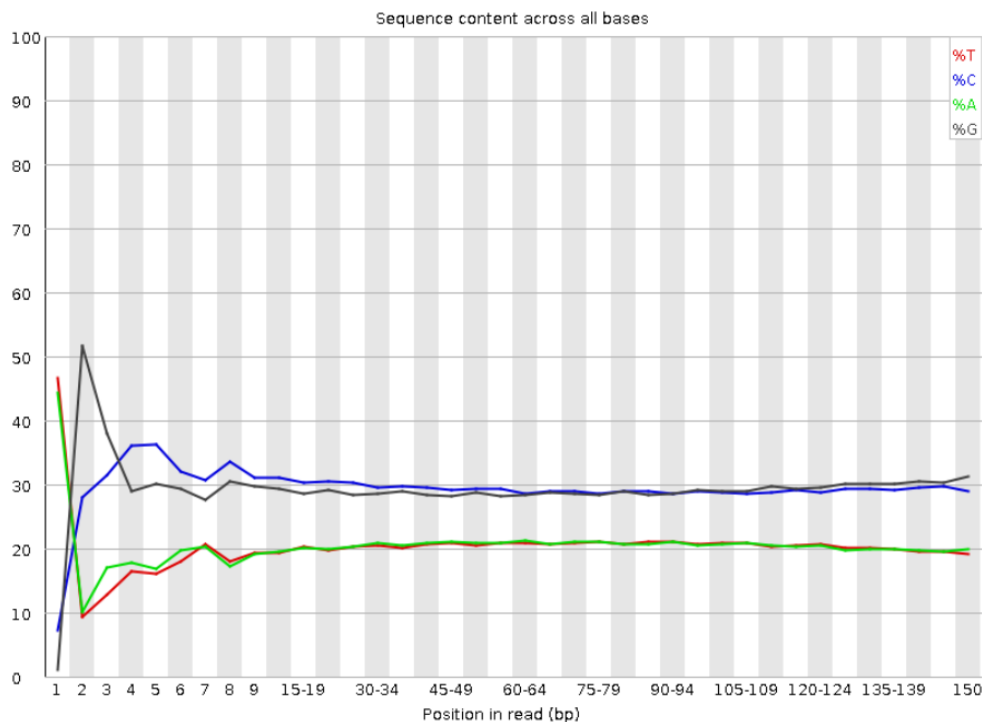


Figure 39: Proportion de bases nucléiques le long des reads du fichier R1 de l'échantillon Glu1 de la souche QM6a.

Un autre paramètre important est le niveau de duplicats dans l'échantillon (« *Sequence Duplication levels* »). Celui-ci ressort chez les deux souches en alerte orange pour tous les fichiers à l'exception du fichier R2 de l'échantillon Glu1 de la RutC30. Les duplicats correspondent à un empilement de plusieurs *reads* à une même position du génome et peuvent démontrer d'une complexité de librairie faible avec peu d'information (Figure 40). Ils proviennent d'une suramplification du même fragment d'ADN et peuvent indiquer, si en trop grande quantité, d'un échec de l'immunoprécipitation ou d'un biais PCR important. Pour ressortir en alerte orange, le

taux de duplicats dans nos échantillons est supérieur ou égal à 20% du nombre total de *reads*.



Figure 40 : Schématisation de l'alignement des reads sur un génome de référence.

(A) Empilement de reads à une même localisation génomique résultant en l'apparition de duplicats, d'une complexité de librairie faible et d'une perte d'information. (B) Distribution normale souhaitée des reads le long du génome de référence.

IV.2. Mapping

Afin d'effectuer l'alignement de nos *reads* sur les génomes de références de la QM6a et de la RutC30, nous avons choisi l'outil Bowtie (version 1.2.3, (Langmead et al. 2009)), qui est un outil adapté à la longueur de nos séquences courtes.

Nous avons d'abord tenté un alignement sans toucher à la qualité des *reads*, tout en ne permettant pas l'alignement multiple (ou « *multimapping* »), ni plus de deux erreurs d'alignement sur la totalité de la longueur des séquences. Au vu des premiers résultats médiocres d'alignements (en dessous de 50% de *reads* alignés, l'alignement est considéré comme mauvais) des échantillons QM6a sur leur génome de référence ainsi qu'à l'alarme d'échec concernant le contenu par base de nos séquences, nous avons décidé d'effectuer un rabotage de 3 paires de bases du côté de mauvaise qualité des *reads* sur tous nos échantillons de cette souche. Après cette intervention, l'alignement correct de nos *reads* sur le génome de la QM6a a significativement augmenté et est en moyenne sur les six échantillons à 86,7% (avec un minimum de 79,44%). Il est globalement considéré qu'un taux de 70% d'alignement est une bonne valeur pour cette étape (Annexe XV) (<https://learn.gencore.bio.nyu.edu/>). Concernant la souche RutC30, nous avons procédé de la même façon mais il s'est avéré que même sans rabotage, la proportion d'alignement sur le génome de référence de la souche était acceptable puisqu'en moyenne 83,47% des *reads* se sont alignés (avec un minimum à 82,09%) (Annexe XV).

IV.3. Détection et élimination des duplicats

Une fois l'alignement des *reads* effectué, une élimination des duplicats a été nécessaire puisque le contrôle qualité nous a informé d'une quantité non négligeable de ces derniers dans nos échantillons.

Afin de procéder à cette élimination nous avons choisi d'utiliser l'outil samtools (version 1.9) (Li et al. 2009) afin de ne garder que les meilleurs *reads* pour l'analyse. Pour cela, un « *mate score* » est associé à chacun d'entre eux et lorsqu'un amas de duplicats est trouvé, seul celui avec le meilleur score sera gardé. Sur les six échantillons, une moyenne de 85% de *reads* mappés ont été conservés pour la suite de l'analyse (Annexe XVI).

Nous avons donc ensuite continué l'analyse avec les fichiers « propres » ne contenant maintenant que des *reads* correctement alignés en paires sur le génome de référence, sans *multimapping* et sans duplicats.

IV.4. Peakcalling

Une fois les données nettoyées nous avons procédé à l'étape du *peakcalling* qui permet de distinguer les zones du génome possédant un enrichissement de la marque épigénétique H3K4me3. Pour ce faire, l'utilisation des deux échantillons *input* permet de faire une comparaison entre le signal du séquençage de l'ADN total et celui du séquençage de l'ADN qui devrait avoir été enrichi sur certaines zones bien précises par l'immunoprécipitation. Les *input* définissent le seuil de bruit basal du signal et permettent une normalisation des données des six échantillons sur ce contrôle. Sachant que nous utilisons des données séquencées en *paired-end*, la fonction MACS2 (version 2.2.7.1) (Zhang et al. 2008) utilise la taille réelle du fragment d'ADN (d). Une fois la longueur réelle moyenne des fragments déterminés, le sommet du pic est estimé se situer à $d/2$. Ensuite, un test statistique est utilisé pour discriminer du bruit de fond les pics significatifs.

IV.5. Cartographie de la marque H3K4me3 chez la souche QM6a

Pour chacun des échantillons de la QM6a, le nombre de pic détectés par MACS2 est renseigné dans le Tableau 16 ci-dessous. Il est notable que le nombre de pics est assez similaire dans l'ensemble des échantillons. C'est un premier résultat encourageant pour la suite de l'analyse puisque cela indique une première homogénéité du signal.

Condition	Rep1	Rep2	Rep3
Glucose	3327	3497	3292
Lactose	3356	3379	3265

Tableau 16: Nombre de pics H3K4me3 détectés dans chaque échantillon de la souche QM6a.

Avec l'aide de l'outil DiffBind, un jeu de pics consensus à l'ensemble des 6 échantillons a été généré. Pour ce faire, chaque pic est analysé et doit être présent dans au minimum 3 échantillons sur 6 afin d'être intégré dans l'ensemble de pics consensus. Un total de 3185 pics a ainsi été conservé sur l'ensemble des fichiers. Un comptage des *reads* dans les échantillons est fait et le score FRiP (« *Fraction of Reads in Peaks* ») est calculé. Pour chaque réplicat, ce dernier donne la proportion de *reads* présents dans les pics conservés par rapport au nombre total de *reads* dans l'échantillon (Tableau 17). Il fournit une bonne information quant à l'enrichissement des pics en *reads* et la qualité de l'immunoprécipitation. D'après les lignes directrices données par le consortium ENCODE (Landt et al. 2012), les échantillons possédant un FRiP supérieur à 1% sont considérés comme bons. Il semble donc que nos immunoprécipitations ont très bien fonctionné et qu'une grande majorité des *reads* présents dans nos échantillons sont contenus dans des pics, témoignant de la spécificité de nos anticorps.

Echantillon	Taille librairie (Nbre de paires)	FRiP	<i>Reads</i> dans pics (nbre de paires)
Glu1_K4	6 878 990	74%	5 090 453
Glu2_K4	5 937 409	78%	4 631 179
Glu3_K4	5 135 885	67%	3 441 043
Lac1_K4	6 877 409	70%	4 814 186
Lac2_K4	7 085 484	69%	4 888 984
Lac3_K4	7 137 836	58%	4 139 945

Tableau 17: Tableau renseignant la taille totale de la librairie, le FRiP score et le nombre de *reads* dans les pics pour chaque échantillon.

Ensuite, une normalisation est opérée par Diffbind sur la taille moyenne de l'ensemble des librairies par application d'un facteur de normalisation afin de les rendre comparables entre elles.

Basé sur le comptage normalisé des *reads*, une « *heatmap* » de corrélation et un graphique d'analyse ACP des échantillons ont été générées (Figure 41). La corrélation des échantillons de même condition est très claire avec un score minimal de 0,8 entre deux échantillons du même groupe et un regroupement marqué par une corrélation plus basse entre lactose vs glucose. Également, une nette démarcation des deux conditions est faite sur le graphique de l'ACP par la première composante principale expliquant 59% de la variabilité de nos données. Les réplicats biologiques de chaque condition sont séparés majoritairement par la seconde composante principale expliquant 20% de la variabilité.

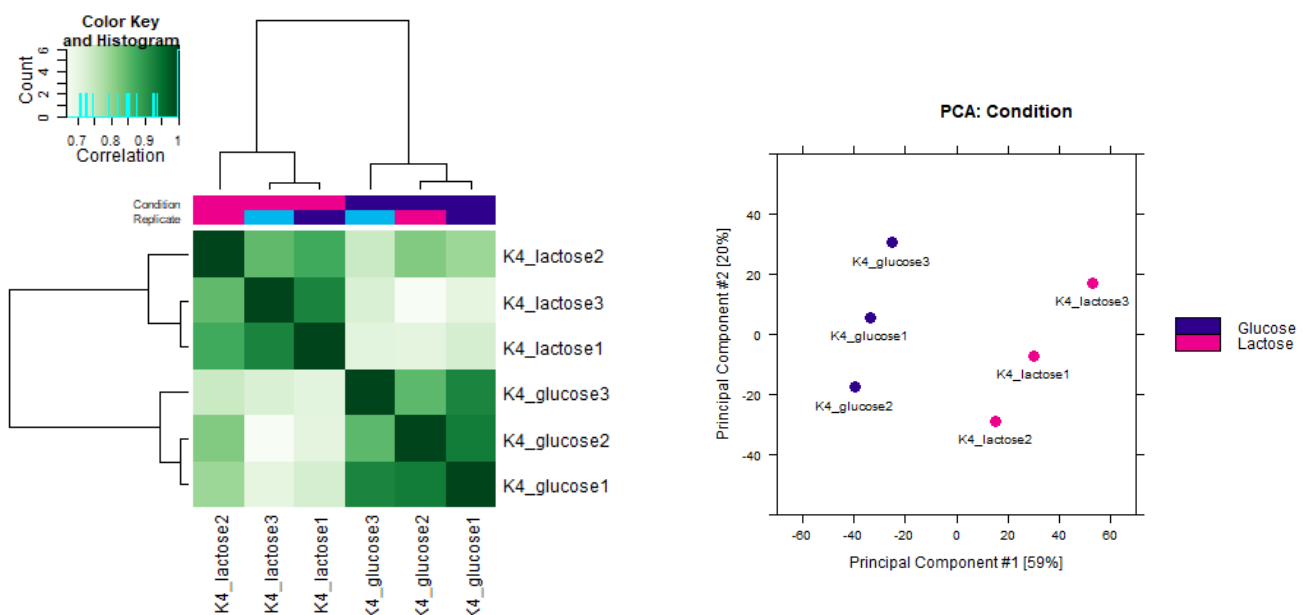


Figure 41: Heatmap de corrélation et analyse ACP des échantillons de la souche QM6a.

La séparation par condition des échantillons selon le premier axe est attendue et est signe d'une bonne qualité des données. La séparation des réplicats par le second axe est signe de certaines différences qui ne sont pas en lien avec les conditions mais qui sont peut-être liées aux manipulations puisque nous retrouvons le même schéma de distribution des échantillons le long de cet axe (glu3-lac3, gluc1-lac1, glu2-lac2).

En moyenne, les pics de H3K4me3 mesurent, conditions confondues, 1844 pb. Ils recouvrent, en moyenne pour les trois échantillons de la condition glucose, 17,43% du génome et 17,93% pour les échantillons en lactose.

Pour chaque chromosome, le pourcentage moyen de couverture de la marque a

été calculé et normalisé en fonction de leur longueur respective (Tableau 18).

Chromosome	Glucose	Lactose
ChrI	19,24%	19,71%
ChrII	18,17%	18,65%
ChrIII	17,53%	17,98%
ChrIV	16,90%	17,48%
ChrV	19,09%	19,73%
ChrVI	13,13%	13,50%
ChrVII	16,19%	16,80%

Tableau 18: Couverture de H3K4me3 le long des chromosomes de la souche QM6a.

Ces informations s'additionnent avec la Figure 42 ci-dessous, représentant la couverture de la marque H3K4me3 le long des chromosomes de la souche QM6a. Ceci a été représenté grâce à la moyenne de l'accumulation du nombre de *reads* sur des fenêtres de 10 pb à partir de fichiers BigWig normalisés en RPKM (« *Reads Per Kilobase per Million mapped reads* »). L'échantillon représenté ici est celui du réplicat Glu1. Grâce à la bonne reproductibilité des résultats dans l'ensemble des échantillons, nous avons choisi de ne représenter ici que celui-ci mais les résultats exprimés ci-dessous sont valables conditions confondues. Les centromères des chromosomes sont représentés par un trait rouge. Les deux flèches rouges (sur le chromosome VI) indiquent les îlots de forte présence de H3K4me3 après des zones de déplétion tandis que les étoiles renseignent sur certaines zones précises carencées en marque.

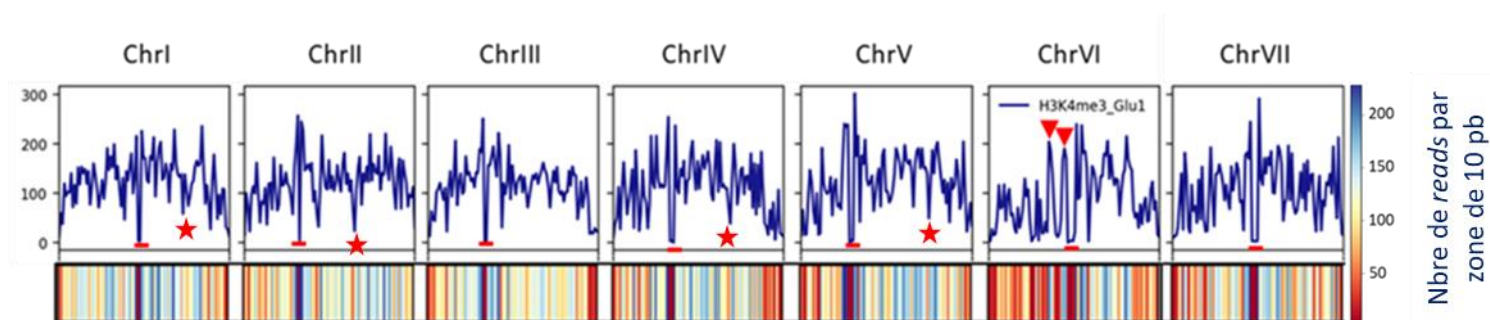


Figure 42: Représentation de la distribution de la marque H3K4me3 le long des chromosomes de la souche QM6a de l'échantillon Glu1.

Les centromères des chromosomes sont représentés par un trait rouge. Les deux flèches rouges indiquent les îlots de forte présence de H3K4me3 et les étoiles renseignent sur certaines zones précises carencées en marque.

A l'échelle macroscopique, la distribution de la marque le long du génome n'est pas uniforme. Des déplétions systématiques sont observées au niveau des télomères

et des centromères, ainsi qu'au niveau de zones ponctuelles carencées bien définies (zones renseignées par les étoiles rouges par exemple). Si les centromères de chaque chromosome sont facilement repérables grâce leur absence de marque, un enrichissement marqué en H3K4me3 dans les régions péri-centromériques est également présent.

Concernant le chromosome VI, nous notons que la partie à gauche du centromère est globalement appauvrie par rapport au reste du génome. Deux zones complètement déplétées entourant deux îlots fortement enrichis (renseignées par les triangles rouges) sont aussi présentes et participent au profil atypique de ce chromosome. La partie à droite du centromère est également marquée par une déplétion importante au niveau du télomère qui est probablement en partie la cause de son pourcentage de couverture bas (Tableau 18). De plus, ce chromosome serait à priori celui possédant un centromère le plus long participant donc à l'absence de H3K4me3.

Une visualisation des données de tous les échantillons sur l'outil IGV (« *Integrative Genome Viewer* ») corrobore cette représentation de la distribution macroscopique de la marque H3K4me3. Les motifs de regroupement par îlots ainsi que les déplétions sont également clairement visibles.

La couverture globale de H3K4me3 est quasiment identique dans les deux conditions, avec une légère augmentation en lactose (Tableau 18). La déplétion du chromosome VI nous a fait penser à une possible corrélation entre le nombre de gènes par chromosome et leur couverture en H3K4me3. La Figure 43 représente ainsi la couverture de la marque sur chaque chromosome en fonction du nombre de gènes présents sur celui-ci, dans les deux conditions. Le coefficient de corrélation de Pearson a été calculé dans les deux cas et est de 0,63 et de 0,60, en glucose et lactose respectivement. La p-value du test du χ^2 de Pearson est à 0,13 et 0,16, respectivement en glucose et lactose, et n'est donc pas significative. A priori, il est donc difficile de tirer des conclusions fortes à ce sujet mais il semble qu'il n'existe pas de corrélation entre couverture du chromosome par H3K4me3 et nombre de gènes.

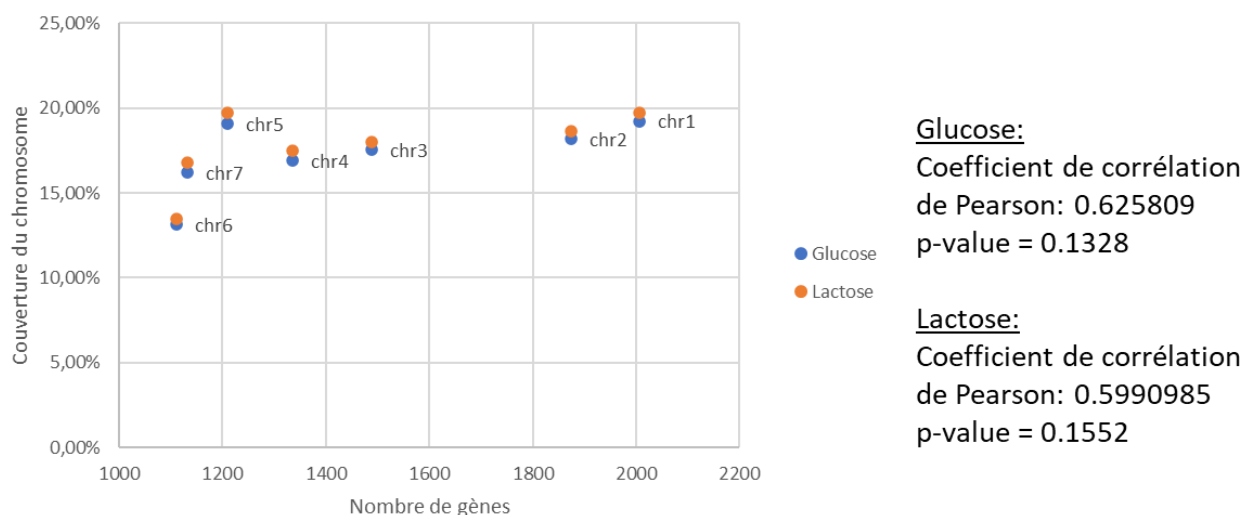


Figure 43: Couverture totale du chromosome par la marque H3K4me3 en fonction du nombre de gènes présents sur celui-ci.

Nous allons maintenant nous intéresser aux gènes liés à H3K4me3. Pour les 3185 pics identifiés, 4398 gènes leurs sont « associés ». Nous avons considéré comme associés aux gènes les pics se situant soit dans le corps du gène, soit à une distance de 1000 pb en amont ou 500 pb en aval de la séquence codante (CDS). Il est fréquent que par sa longueur, un pic recouvre plusieurs CDS de gènes. Également, dans le cas où un pic se situe en dehors d'une séquence codante, nous avons choisi de ne prendre en compte que la séquence la plus proche. En réalité, sur ces 4398 gènes, 4383 (99,6%) se situent au niveau d'un CDS, seuls 10 pics se situent au niveau du promoteur et 6 au niveau du terminateur d'un gène. Sachant que le génome de la souche QM6a compte 10160 gènes, nous comptons donc un peu plus de 43% d'entre eux touchés par H3K4me3. Parmi ces gènes du consensus, nous en comptons 4313 possédant une expression (RPKM > 10).

La Figure 44 ci-dessous représente le recouvrement des gènes par la marque H3K4me3 calculé grâce au nombre de *reads* présent en amont et en aval de l'ATG. Le maximum de couverture se fait au niveau de l'ATG avec un léger débordement à la fin du promoteur. Nous notons également une déplétion marquée un peu avant le codon d'initiation. Cette zone correspond certainement au site d'initiation de la transcription TSS (« *Transcription Start Site* ») où les nucléosomes sont évincés afin de permettre le positionnement de la polymérase.

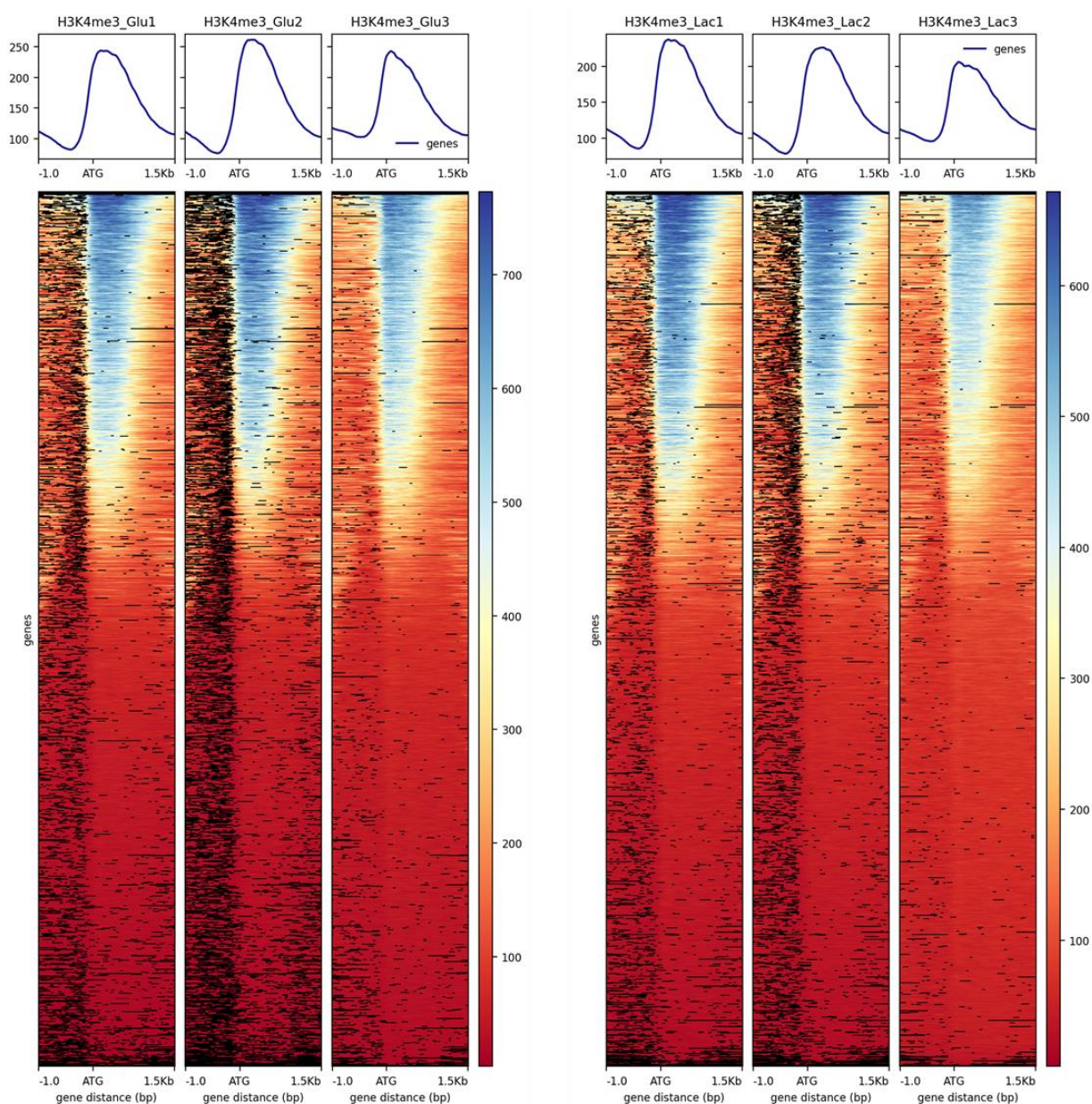


Figure 44: Couverture de la marque H3K4me3 par rapport à la position de l'ATG des gènes.

Le calcul de l'accumulation des reads sur une longueur de 1kb en amont et 1,5kb en aval de l'ATG est faite sur l'ensemble des gènes du génome grâce à une fenêtre de 10 bp et une normalisation du nombre de reads en RPKM. Une heatmap est générée et une courbe représentant le score attribué à la région correspondante est tracée.

IV.6. Cartographie de la marque H3K4me3 chez la souche RutC30

Pour chaque échantillon RutC30, le nombre de pics détectés par MACS2 est renseigné dans le Tableau 19 ci-dessous. Le nombre de pics est très proche dans l'ensemble des réplicats et témoigne de l'homogénéité du signal.

Condition	Rep1	Rep2	Rep3
Glucose	3375	3406	3291
Lactose	3393	3551	3390

Tableau 19: Nombre de pics H3K4me3 détectés dans chaque échantillon de la souche RutC30.

Avec l'aide de l'outil DiffBind, un jeu de pics consensus à l'ensemble des 6 échantillons a été construit de la même façon que pour la souche QM6a. Un total de 3300 pics a été conservé sur l'ensemble des fichiers. Après comptage des *reads*, le FRiP score est déterminé (Tableau 20) et une normalisation sur la taille moyenne de la librairie est faite. Une matrice de corrélation ainsi qu'une analyse ACP ont été effectuées (Figure 45).

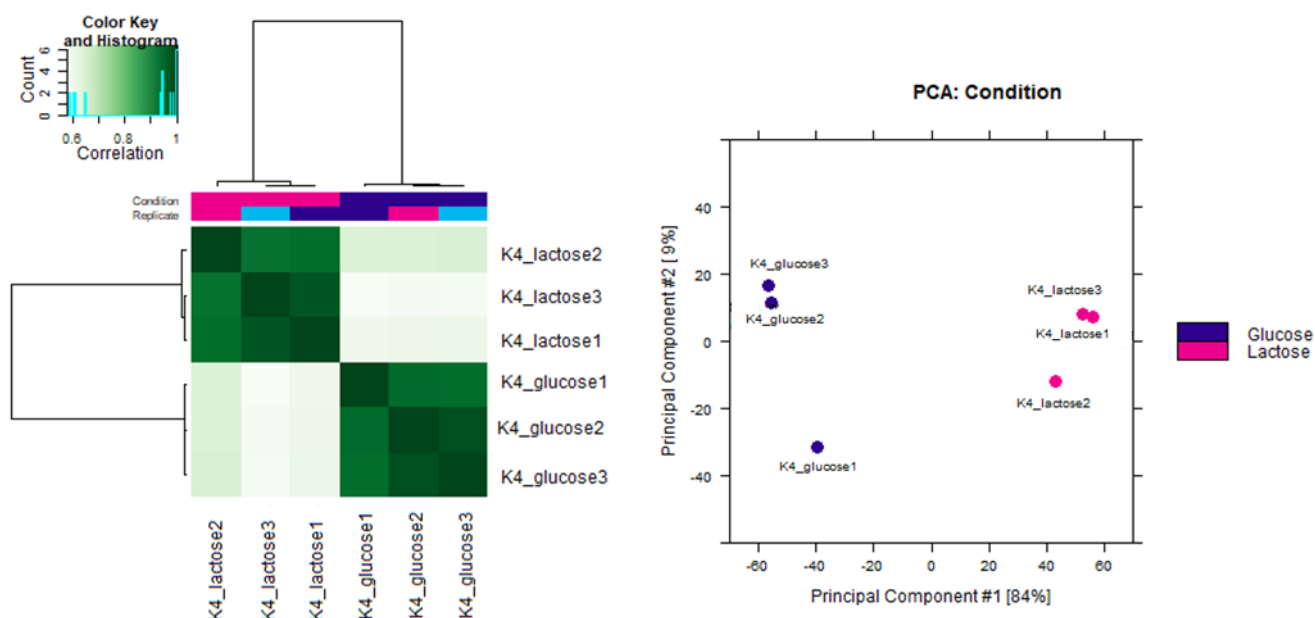


Figure 45: Heatmap de corrélation et analyse ACP des échantillons de la souche RutC30.

Echantillon	Taille librairie (nombre de paires)	FRiP	Reads dans pics (nombre de paires)
Glu1_K4	14 817 800	73%	10 816 994
Glu2_K4	8 466 792	74%	6 265 426
Glu3_K4	9 498 413	73%	6 933 841
Lac1_K4	8 042 984	78%	6 273 528
Lac2_K4	8 128 711	71%	5 771 385
Lac3_K4	9 112 569	75%	6 834 427

Tableau 20: Tableau renseignant la taille totale de la librairie, le FRiP score et le nombre de reads dans les pics pour chaque échantillon.

Les scores FRiP de nos échantillons sont très bons puisqu'ils sont largement supérieurs à 1% du nombre total de reads. Ceci est preuve d'un enrichissement important en *reads* au niveau des pics. L'immunoprécipitation a donc à priori bien fonctionné.

La distinction entre échantillons glucose et lactose est très bien marquée sur la *heatmap* avec des scores de corrélation supérieurs à 0,9 à l'intérieur des groupes de réplicats de même condition. En revanche, nous pouvons voir que la corrélation entre les échantillons de conditions différentes est bien moins élevée avec des scores entre 0,55 et 0,65. Ceci est complété par le graphique de l'analyse ACP où une séparation franche des deux groupes est faite selon l'axe de la première composante principale comptant pour 84% de la variabilité des échantillons. La seconde composante principale représente 9% de la variabilité et sépare principalement les échantillons Glu1_K4 et Lac2_K4 de leur groupe respectif.

Nous voyons ici que chez cette souche le paramètre « condition » a une influence plus forte sur la variabilité des données que chez la QM6a. Ceci est aussi en adéquation avec les résultats de fermentations où nous observons que les répétas des fermentations RutC30 sont meilleurs que ceux de la souche QM6a, probablement à cause du système expérimental mieux adapté. Aussi, ses mutations l'ont rendu plus sensible à l'induction ce qui rend donc les différences entre conditions plus importantes.

En moyenne, les pics de H3K4me3 mesurent, conditions confondues, 1986 pb. Ils recouvrent, en moyenne pour les trois échantillons en glucose, 19,41% du génome et 18,89% en lactose. Pour chaque chromosome, le pourcentage moyen de couverture de la marque a été calculé en fonction de leur longueur respective (Tableau 21).

Chromosome	Glucose	Lactose
ChrI	20,85%	20,07%
ChrII	20,39%	20,24%
ChrIII	20,82%	20,08%
ChrIV	19,02%	18,51%
ChrV	21,49%	20,83%
ChrVI	15,38%	14,91%
ChrVII	17,80%	17,38%

Tableau 21: Couverture de H3K4me3 le long des chromosomes de la souche RutC30

Une déplétion de H3K4me3 sur le chromosome VI par rapport aux autres est à remarquer comme pour la QM6a. Ce chromosome a subi chez la RutC30 une translocation résultant d'un échange entre l'extrémité d'un des bras du chromosome VI (402 kb) et l'extrémité d'un des bras du chromosome I (1,63 Mb) (Figure 18).

La Figure 46 ci-dessous représente la distribution de la marque H3K4me3 le long des sept chromosomes de la souche Rut30 de l'échantillon Lac1. Grâce à la bonne reproductibilité des résultats dans l'ensemble des échantillons, nous avons choisi de ne représenter ici que cet échantillon-là mais les résultats exprimés ci-dessous sont valables toutes conditions confondues. Grâce à cette figure, il est possible de visualiser l'emplacement des enrichissements et des déplétions des régions chromosomiques. Les centromères sont renseignés par un trait rouge en bas de la courbe.

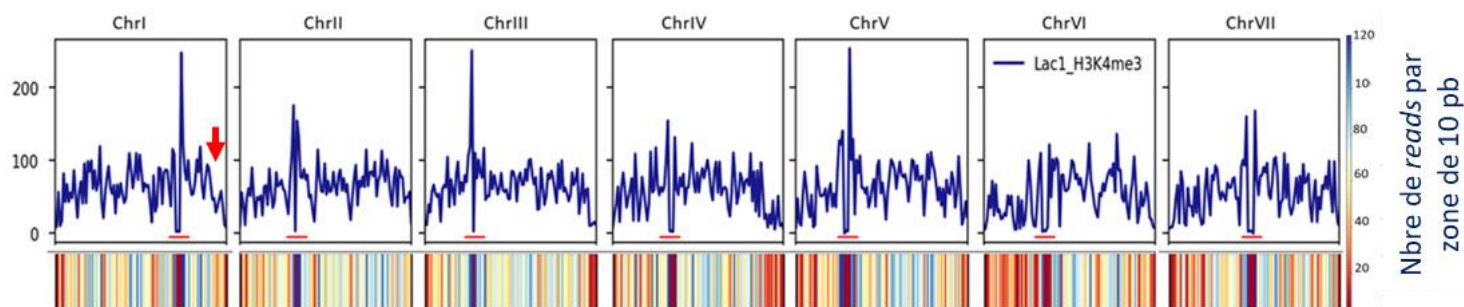


Figure 46: Représentation de la distribution de la marque H3K4me3 le long des chromosomes de la souche RutC30.

La flèche rouge indique la zone de translocation du chromosome. Les traits rouges marquent les centromères.

Ainsi, il apparaît de façon saillante sur les chromosomes I, II, III, V et VII que si le centromère est une région caractérisée par l'absence de marque (renseigné par un trait rouge sur la figure), ses régions directement adjacentes en sont fortement enrichies comme chez la QM6a. Aussi, une chute de la présence de la marque sur l'extrémité droite du chromosome I est notable (symbolisée par la flèche rouge), correspondant à

une région transloquée avec ajout de l'extrémité du chromosome VI. Également, nous observons une carence en marque au niveau de la région terminale du chromosome IV, visible de la même façon chez la QM6a sur ce chromosome.

La Figure 47 confirme que, chez la RutC30, la couverture en marque des chromosomes ne semble pas être corrélée au nombre de gènes présents sur ceux-ci puisque les coefficients de corrélation entre ces deux paramètres sont de seulement 0,20 et 0,23, en glucose et lactose respectivement. De plus, les p-value des deux tests du χ^2 de Pearson sont toutes les deux non significatives.

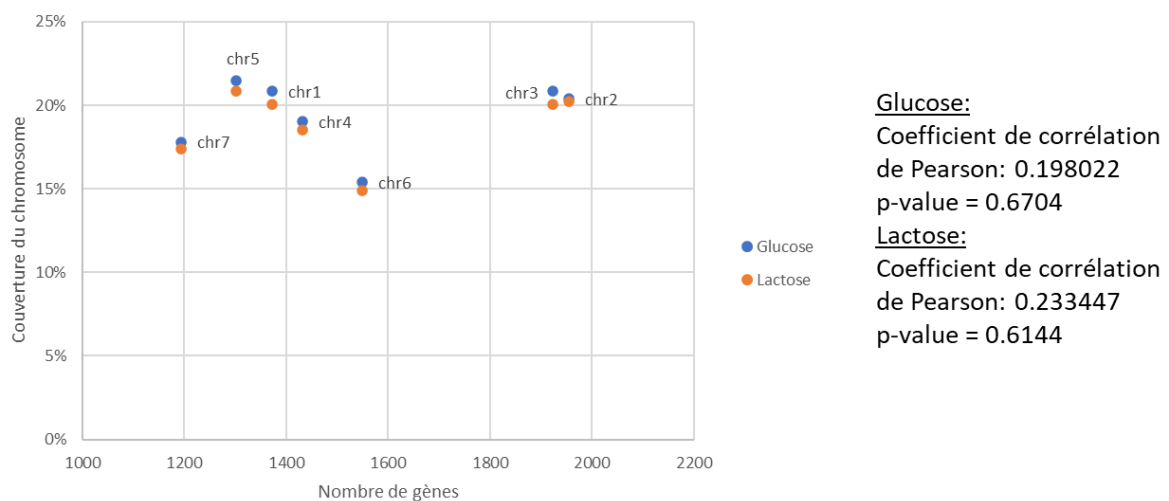


Figure 47 : Couverture totale du chromosome par la marque H3K4me3 en fonction du nombre de gènes présents sur celui-ci.

Nous avons ensuite procédé à l'étude des gènes touchés par H3K4me3. Pour les 3300 pics, 4628 gènes leurs sont « associés » sous les mêmes critères que pour la souche QM6a. Sur ces 4628 gènes, 4606 (99,5%) sont associés à un pic présent à l'intérieur de leur séquence codante. Seulement 17 gènes sont associés à un pic présent au niveau de leur promoteur et 6 gènes sont associés à un pic situé au niveau de leur terminateur. De même que pour la QM6a, nous retrouvons un recouvrement entre gènes et marque avec un positionnement « à cheval » entre début de la séquence codante et fin du promoteur. La déplétion de la marque avant l'ATG est également présente chez la souche RutC30 (Figure 48).

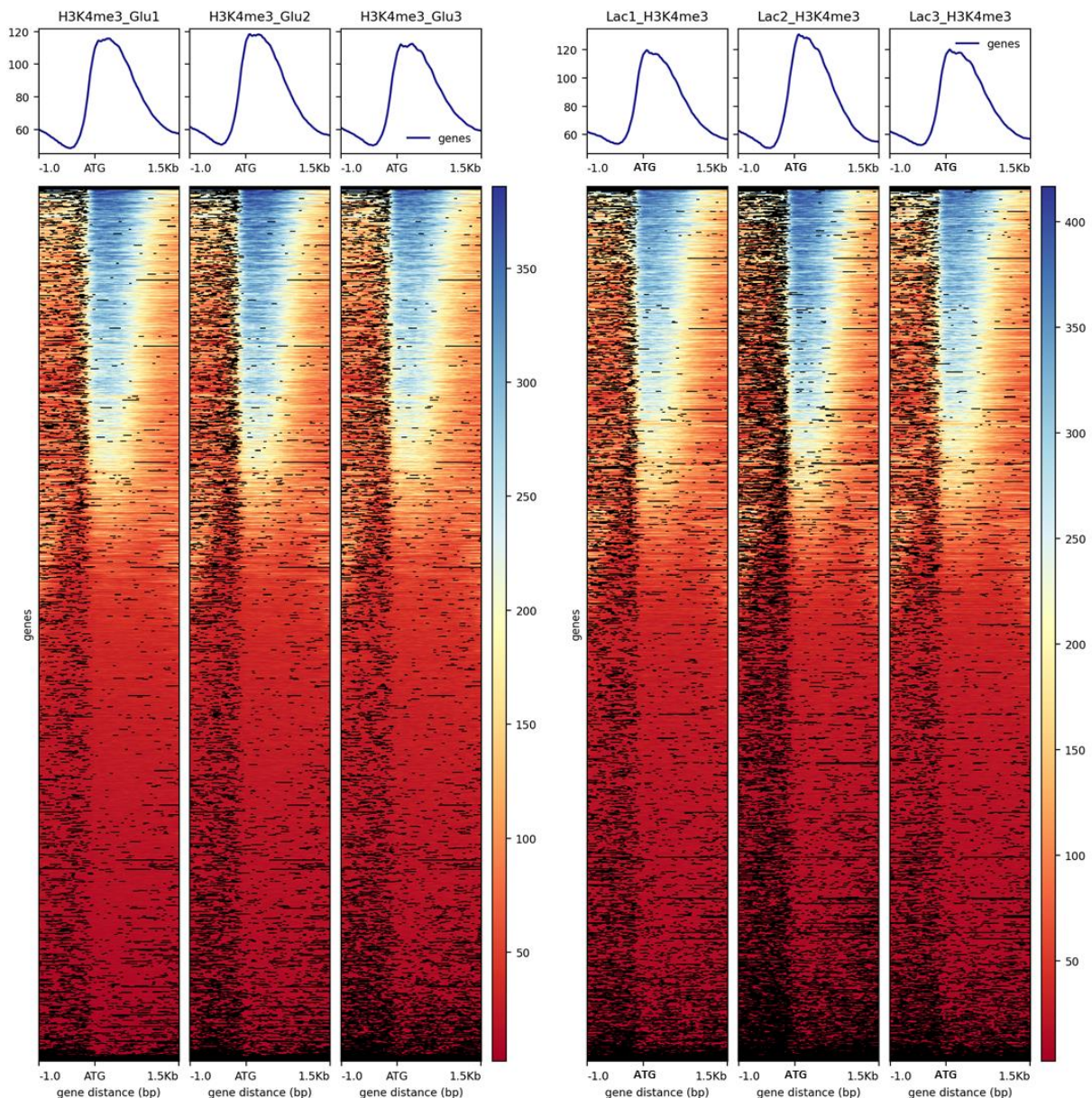


Figure 48: Couverture de la marque H3K4me4 par rapport à la position ATG des gènes.

IV.7. Comparaison des deux souches et analyses des translocations

Glucose et lactose confondu, nous observons que dans les conditions dans lesquelles nous avons réalisé les expériences, la souche QM6a possède moins de marques H3K4me3 le long de son génome que la RutC30 et que la distribution de cette marque reste globalement identique entre les deux souches avec une déplétion centromérique et télomérique et avec un enrichissement péracentromérique. Enfin, le nombre de pics dans chacun des réplicats est également semblable entre souche avec en moyenne 3357 pics pour RutC30 et 3372 pics pour QM6a en glucose et 3444 pics pour RutC30 et 3333 pics pour QM6a en lactose. Nous remarquons également que 55,2% et 54,5% des gènes de la QM6a et de la RutC30, respectivement, ne sont pas touchés par H3K4me3. Ceci représente dans les deux cas plus de la moitié des gènes des souches et après visualisation sur IGV il s'est avéré que ces gènes sont situés comme attendus dans les régions déplétées subtélomériques mais aussi tout le long des bras des chromosomes. Nous notons que certains gènes sont également dépourvus de marque alors que les régions adjacentes sont fortement couvertes (Figure 49).

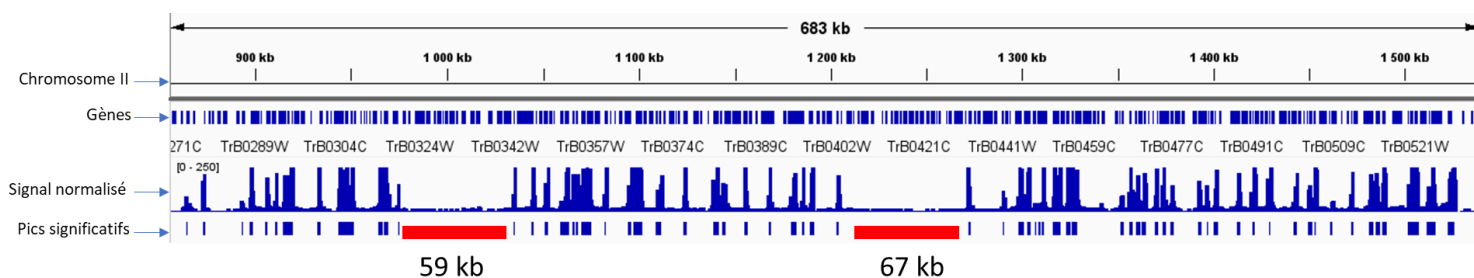


Figure 49 : Exemple de zones dépourvues de marques et entourées de zones couvertes.

Les zones dépourvues de marques sont représentées ici en rouge. Leur taille en kilobase est également renseignée.

Il est intéressant de voir que ces zones enrichies sont identiques entre échantillons toutes conditions confondues mais aussi entre souches. Pour finir, après analyse manuelle chez les deux souches des gènes surexprimés, il s'avère que la grande majorité d'entre eux (196 sur 206 pour la QM6a et 243 gènes sur 316 pour la RutC30) se trouve à l'intérieur de ces zones déplétées. Parmi eux, nous retrouvons notamment de nombreux gènes de cellulases et d'acteurs de leur production : CBH1, CBH2, BGL1, EGL3, CEL3C etc... .

La distribution bien caractéristique des pics au niveau de l'ATG (début de la séquence codante) des gènes est retrouvée autant chez l'une que chez l'autre confirmant donc la répétabilité des expériences et la robustesse de l'analyse menée jusqu'à présent.

Ensuite, nous avons comparé les ensembles de gènes associés aux consensus de pics des deux souches et avons généré un diagramme de Venn afin de visualiser les gènes communs entre la QM6a et la RutC30 et les gènes qui leur sont spécifiques

(Figure 50). Il apparaît ici clairement que la quasi-totalité des gènes associés aux pics sont partagés entre les deux souches, confirmant une nouvelle fois que la distribution de la marque le long des chromosomes des deux souches est très similaire.

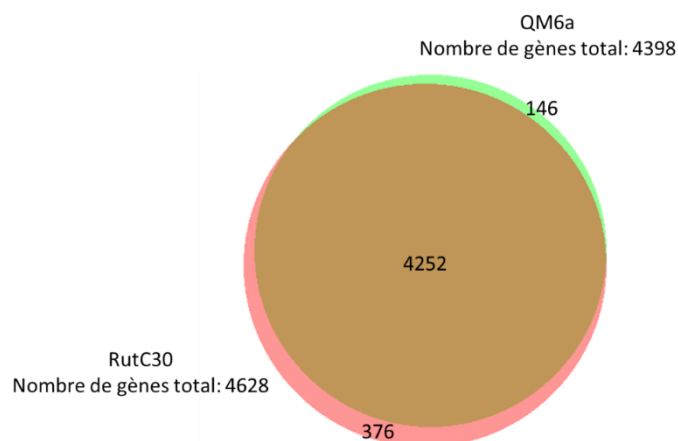


Figure 50 : Diagramme de Venn représentant les ensembles de gènes associés aux ensembles de pics consensus des deux souches.

La RutC30 est représentée en rouge et la QM6a en vert. La quasi-totalité des gènes liés aux ensembles de pics sont partagés entre les deux souches.

La liste des gènes spécifiques aux deux souches est disponible en Annexe XVII et Annexe XVIII. Parmi les 146 gènes spécifiques à la QM6a, peu de spécialisation fonctionnelle est visible. En revanche, 13 gènes associés à la production de cellulases sont retrouvés chez RutC30 ainsi que cinq gènes associés à des mécanismes épigénétiques.

Il semble donc qu'il n'y ait pas de changement majeur dans la distribution de la marque épigénétique chez les deux souches puisque 89% du nombre total de gènes des deux consensus sont partagés par la QM6a et la RutC30. Néanmoins, il est intéressant de noter qu'il existe un petit groupe de gènes touchant à la production de cellulases qui sont uniques à la souche hyper-productrice et qui possède donc un profil épigénétique différent de celui de la souche sauvage.

Nous nous sommes par ailleurs penchés sur la question des translocations chromosomiques et de l'impact que celles-ci auraient pu avoir sur la distribution de la marque tout leur long et au niveau des régions de jonction. Afin d'illustrer la translocation du chromosome I (Figure 18), une extraction de la zone à partir de l'outil IGV est présentée en Figure 51. Elle montre la présence de la marque le long de la translocation ainsi que les pics significatifs détectés par MACS2. La flèche rouge indique l'endroit de jonction entre le chromosome I et l'extrémité transloquée du chromosome VI.

Il est clairement visible à cet endroit qu'aucune marque H3K4me3 n'est présente ni chez la RutC30, ni chez la QM6a. Après analyse, il se trouve que cette région soit enrichie en AT (taux de AT > 30% sur plus de 500 bp), ce qui pourrait expliquer cette absence de marque. La distribution de la marque ensuite observée sur le reste du chromosome correspond exactement à celle trouvée sur le chromosome VI de la QM6a. Cette translocation n'a donc eu aucun impact sur la distribution de H3K4me3, autant sur la zone transloquée que sur le reste du chromosome de la RutC30.

Enfin, la translocation du chromosome I vers le chromosome VI (Figure 52) et la translocation sur vers chromosome III (Figure 53) ne présentent pas non plus de changement sur la disposition des marques. Ces résultats suggèrent que le remodelage intense du caryotype de la souche RutC30 n'a pas d'impact sur la distribution des marques le long du génome.

RutC30 (CHRI-CHRVI)

QM6a (CHRVI)



Figure 51: Zone de translocation du chromosome I avec le chromosome VI.

À gauche de la flèche est le chromosome I, à droite est l'extrémité transloquée du chromosome VI. Comparaison de l'extrémité transloquée du chr. VI de la souche RutC30 avec la zone correspondante du même chromosome de la souche QM6a. La distribution de la marque est sensiblement la même, aucun remodelage du paysage épigénétique n'est observé

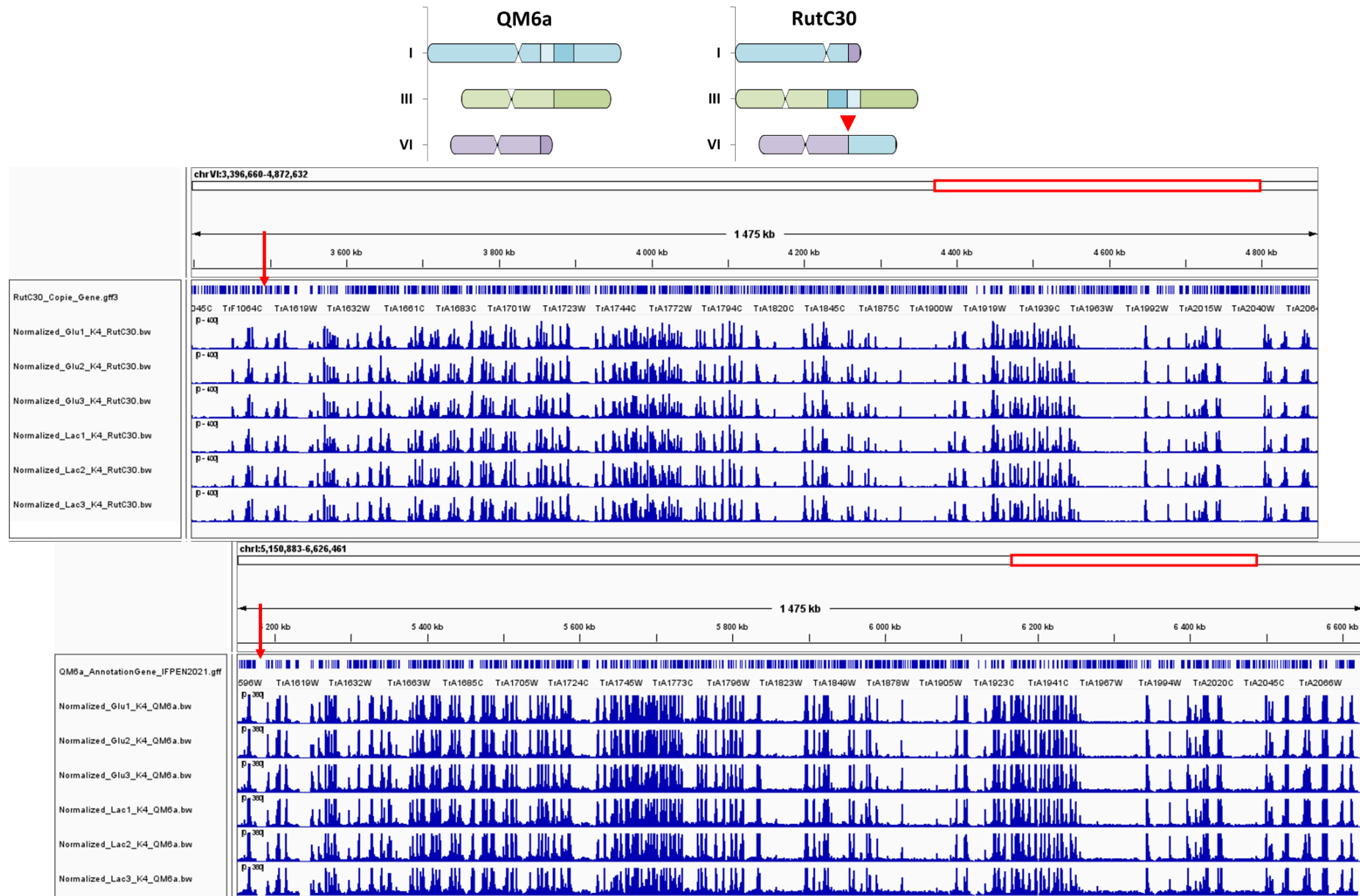


Figure 52: Comparaison de la distribution des marques H3K4me3 le long de l'insertion sur le chromosome VI d'un bout du chromosome I chez la souche RutC30 avec la zone correspondante chez la souche QM6a.



Figure 53: Distribution de la marque H3K4me3 le long de l'insertion du bout de chromosome I sur le chromosome III chez la souche RutC30.
L'insertion inversée du bout du rDNA possède également la même distribution que chez la souche sauvage (données non renseignées).

Nous nous sommes également intéressés à certains évènements spécifiques comme la troncature du gène *cre1* chez la RutC30. Nous avons aussi vérifié la délétion de 85Kb intervenant sur la région télomérique du chromosome VI.

Concernant le gène *cre1*, il se trouve dans une zone d'une quarantaine de kilobases entièrement déplétée de marque H3K4me3 à l'exception du pic (présent à la fois en lactose et en glucose) qui lui est associé (Figure 54). Le gène *cre1* étant tronqué chez la RutC30, il ne reste quasiment rien de la marque H3K4me3 qui lui est associée chez la QM6a.

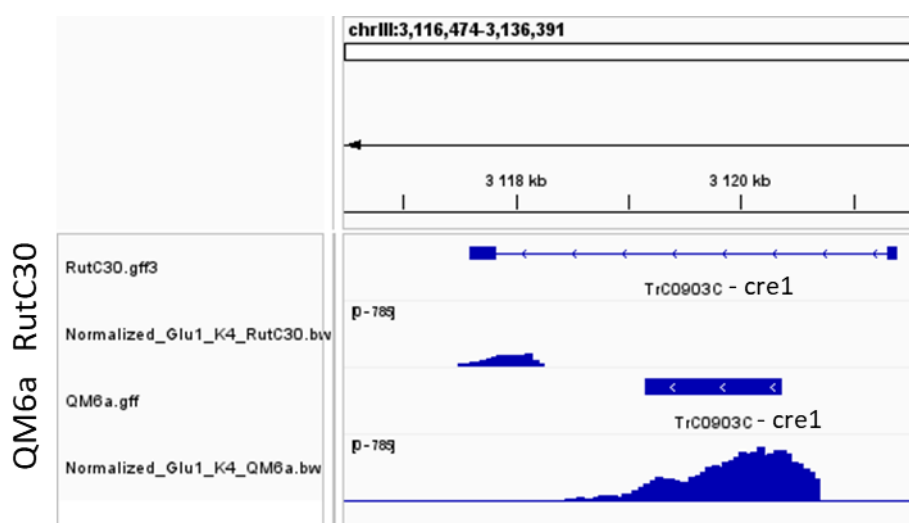


Figure 54 : Visualisation IGV du gène *cre1* et de la marque H3K4me3 qui lui est associée sur les réplicats Glu1.

Sur la souche RutC30 le gène a été tronqué et par conséquent très peu de marque lui est associée. Seul la partie 3' du gène est encore touchée. En revanche, sur la souche QM6a le gène est intact et H3K4me3 est très présente sur l'ensemble du gène.

Chez la RutC30, les pourtours de la délétion de 85kb sont également vides de marque et il semble que chez la QM6a la partie correspondant à celle ayant été supprimée n'est pas non plus touchée par H3K4me3. S'il est difficile de tirer des conclusions à ce sujet du fait de la position télomérique et subtelomérique de l'évènement il semble toutefois qu'il n'existe pas de différence entre les souches au niveau de ce remodelage chromosomique.

V. DIFFERENCE D'EXPRESSION ENTRE LES GENES TOUCHES PAR H3K4ME3 ET CEUX QUI EN SONT DEPOURVUS

A la suite de la cartographie de H3K4me3 le long du génome des deux souches, nous nous sommes interrogés sur l'impact que pouvait avoir la marque sur la transcription des gènes. En effet, H3K4me3 étant associée à une conformation ouverte de la chromatine, nous nous attendons à observer une transcription plus importante chez les gènes touchés par cette marque que ceux qui ne le sont pas.

Afin de vérifier cette hypothèse, nous avons mené une analyse statistique de type t-test sur les données de transcriptomique générées auparavant et avons comparé les deux groupes de gènes dans les deux conditions, de façon indépendante. La Figure 55 ci-dessous représente l'expression moyenne (normalisée par RPKM) des gènes des deux groupes et le nombre total de gènes par intervalle d'expression.

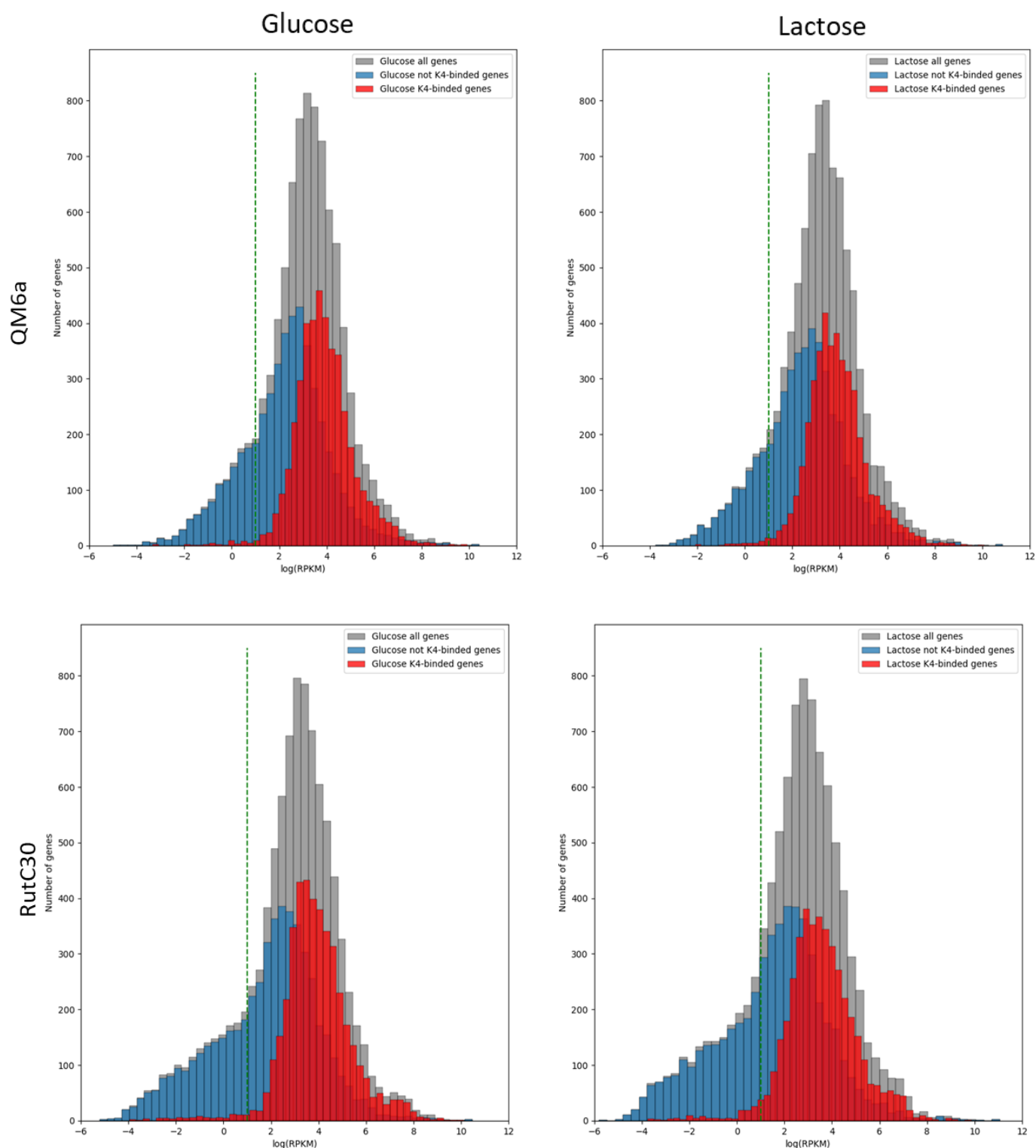


Figure 55 : Représentation des groupes de gènes touchés ou non par H3K4me3 en fonction de leur expression chez les deux souches QM6a et RutC30 dans les deux conditions d'étude.

La ligne verte discontinue marque le seuil minimal du log(RPKM) pour lequel nous considérons que les gènes sont exprimés.

Les gènes non touchés par H3K4me3 (en bleu) sont significativement (p -value $< 0,01$) moins exprimés que ceux qui sont associés à la marque (en rouge) et ce, chez les deux souches et dans les deux conditions. De plus, nous avons fixé comme seuil d'expression d'un gène $\log(\text{RPKM})=1$ (ou $\text{RPKM}=10$) et nous remarquons que la quasi-totalité des gènes touchés par H3K4me3 sont considérés comme exprimés contrairement à ceux qui ne sont pas associés à la marque.

Il existe donc une corrélation positive entre apposition de la marque H3K4me3 sur les gènes et expression active de ceux-ci bien que cela ne soit pas une condition nécessaire et suffisante à la transcription puisque l'on voit ici que des gènes non marqués possèdent également un niveau d'expression élevée.

L'analyse différentielle qui suit va maintenant permettre de distinguer les gènes qui sont associés à des pics H3K4me3 différentiellement apposés sur le génome entre les deux conditions d'étude pour chacune des souches.

VI. ANALYSES DIFFÉRENTIELLES INTRA SOUCHE DES DONNÉES DE CHIP SEQ

L'ensemble des analyses différentielles effectuées ont été faites avec l'outil DiffBind et ont été déroulées sur les données intra-souches inter-conditions. Diffbind se base sur la méthode et les calculs créés par l'outil DeSeq2. Nous avons choisi de prendre comme condition de référence la condition en fed-batch glucose. Un récapitulatif des étapes mises en place lors des analyses différentielles chez les deux souches est visible en Figure 56 ci-dessous. Le consensus de pics, le score FRiP dans les pics ainsi que les tailles de bibliothèques normalisées sont les résultats présentés dans les sections IV.5 et IV.6 de ce chapitre.

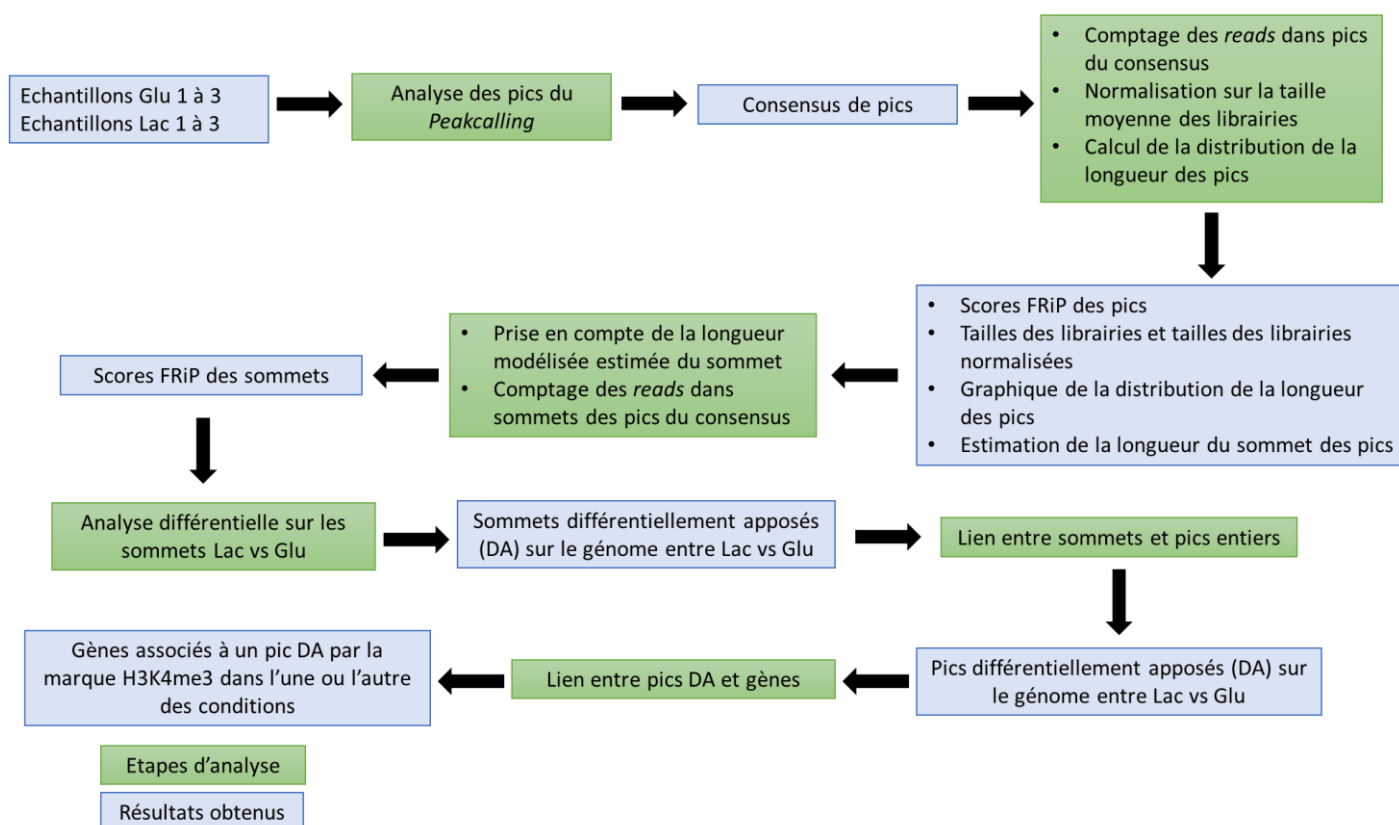


Figure 56 : Récapitulatif des étapes et résultats obtenus lors des analyses différentielles.

VI.1. Analyse différentielle de la marque H3K4me3 chez la souche QM6a

Pour cette analyse, le jeu de pics consensus généré pour les six échantillons de la souche QM6a présenté plus haut a été utilisé. Ainsi, 3185 pics ont été pris en compte. Une normalisation sur la taille de la librairie moyenne est effectuée pour chacun des échantillons grâce à l'application d'un facteur de normalisation propre à chacun comme mentionné dans la section de la cartographie de la marque chez la QM6a.

Ensuite, une estimation de la longueur moyenne en paire de base du sommet des pics sur l'ensemble des échantillons est calculée. Nous avons choisi de la modéliser par une longueur des sommets correspondant à la longueur moyenne entre la valeur minimale et la valeur du premier quartile (Figure 57).

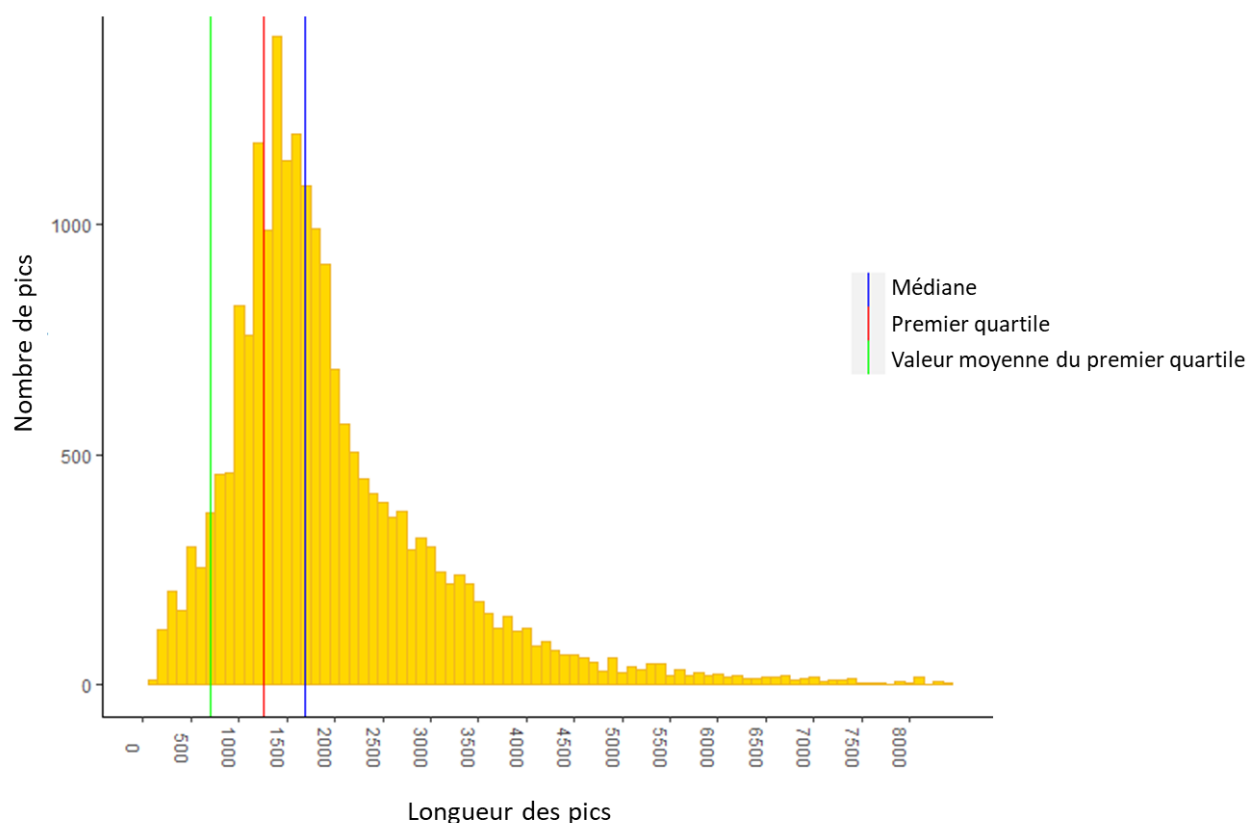


Figure 57 : Distribution de la longueur des pics.

La longueur du sommet des pics (vert) a été choisie comme étant la valeur moyenne entre la valeur de longueur minimale des pics et la valeur de longueur du premier quartile (rouge).

Pour la souche QM6a cette longueur est de 850 pb (par rapport à une longueur moyenne de pic de 1844 pb). Puis un comptage des *reads* dans le sommet des pics est effectué de la même façon que lorsque l'entièreté des pics était prise en compte. De cette façon les scores FRiP sur les sommets sont calculés. La prise en compte des sommets permet d'effectuer l'analyse différentielle uniquement sur la zone la plus enrichie des pics et donc de comparer des données plus significatives. Ceci est applicable dans le cas de H3K4me3 car le pic est considéré comme suffisamment fin pour ce genre d'analyse. Dans le cas d'une marque dont le domaine est plus large (comme H3K9me3 par exemple) l'assimilation d'un pic à son sommet ne serait pas une méthode pertinente.

A partir de là, l'analyse différentielle est exécutée selon la méthode de DeSeq2

(Love, Huber et Anders 2014). Les « *log2FoldChange* » (nous y feront aussi référence comme logFC) sont calculés et déterminent la différence du nombre de *reads* pour chaque sommet entre conditions. Un test statistique est appliqué pour chaque sommet et établit si cette différence est significative. Une p-value corrigée selon la procédure de Benjamini-Hochberg est appliquée afin de s'assurer que les sommets ressortant comme différentiellement apposés (nous y ferons aussi référence sous l'abréviation DA dans le reste du manuscrit) ne sont pas des faux positifs. Cette correction de la p-value entre dans la catégorie de type FDR, pour « *False Discovery Rate* », et sera mentionnée comme telle plus loin dans ce manuscrit. On y fera également référence comme q-value.

Le but étant d'obtenir une liste de pics avec une abondance de *reads* significativement différentes entre les deux conditions nous avons choisi d'appliquer deux filtres sur l'analyse différentielle afin d'obtenir les résultats les plus robustes. Ainsi, après une première visualisation de la distribution en Volcanoplot des sommets des pics, nous avons choisi de filtrer et éliminer ceux dont la q-value était supérieure à 0,01 et dont la valeur absolue du logFC était inférieure 2.

De cette façon, seuls 18 sommets correspondant à 18 pics différentiellement apposés en glucose et 20 sommets correspondant à 20 pics DA en lactose ont été conservés sur les 3185 pics de l'ensemble consensus utilisé pour l'analyse. La Figure 58 représente la répartition en Volcanoplot des sommets après analyse différentielle et renseigne les seuils appliqués pour filtrer les sommets significativement apposés.

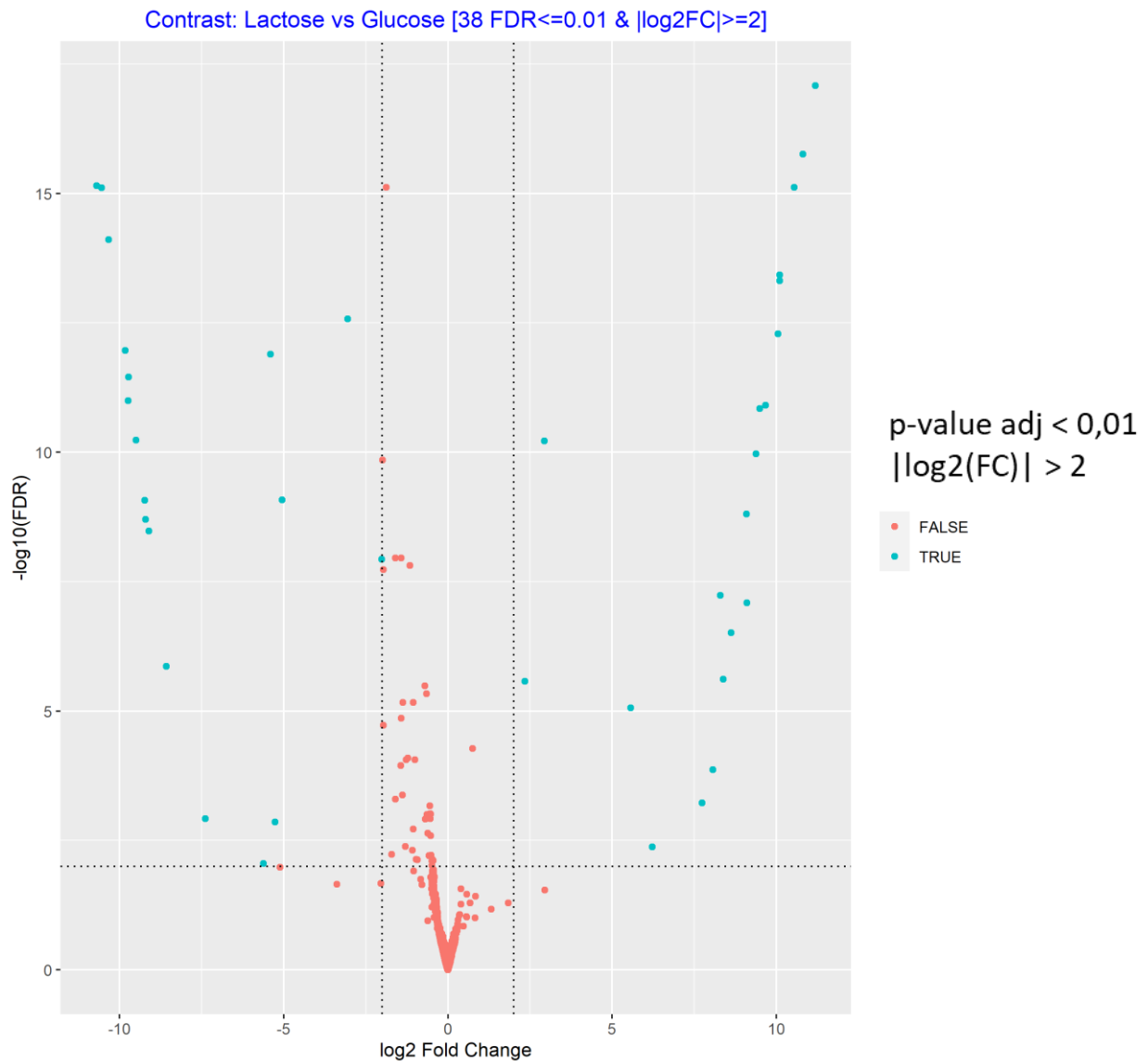


Figure 58: Volcano plot de l'analyse différentielle sur les données de la souche QM6a.

Les sommets de pics en bleu-vert sont ceux ayant passé avec succès les filtres sur la q -value et le logFC. Ceux en orange ont été écartés pour le reste de l'analyse.

Chaque sommet DA est « restitué » à son pic afin de pouvoir associer les gènes différentiellement touchés par la marque. A nouveau, nous considérons un gène touché par la marque si celle-ci se trouve à moins de 1000 pb de l'ATG et 500 pb de la fin du gène. Ainsi, 19 gènes sont associés aux pics DA en lactose et 19 aux pics DA en glucose.

Le Tableau 22 ci-dessous rassemble la liste des 19 gènes associés aux pics différentiellement apposés dans la condition lactose ainsi que les informations relatives à l'analyse et à leur expression.

ID-TFW	Protéine	Commentaire	Log2FC ChIP Seq	Adj p-value ChIP Seq	Log2FC Expression	RPKM Expression Lactose
TrA1372C	BGL2 / CEL1A	Intracellular β -glucosidase	10,81	1,76E-16	8,46	969,06
TrB0812W	ACE3	Fungal transcription factor activator of cellulase	8,30	5,79E-08	2,79	238,49
TrC0438W	Xyr1	Transcription factor xylanase regulator 1	10,09	4,90E-14	2,92	573,78
TrC1166C	CRT1	MFS Lactose transporter	10,05	5,16E-13	8,20	6584,92
TrB0855C	STP1	MFS glucose/cellobiose transporter	2,93	6,03E-11	-2,48	1836,44
TrD1300W	cfa14D (clone ID)	Putative cellulase synthesis associated MFS transporter	10,10	3,77E-14	7,36	793,70
TrB0908W	DIM2 (<i>N. crassa</i>)	Putative DNA methyltransferase	9,37	1,08E-10	-0,41	4,21
TrA2055W	ILV3 (<i>S. cerevisiae</i>)	Putative Dihydroxy-acid dehydratase	8,38	2,44E-06	0,51	58,60
TrF0801C	ARI9 (<i>T. arundinaceum</i>)	Putative E3 ubiquitin-protein ligase	9,49	1,42E-11	0,28	338,43
TrE0685C	atfB (<i>A. oryzae</i>)/ HAC1 (<i>C. albicans</i>)	Putative transcription factor	9,08	1,55E-09	-0,34	98,29
TrF0804W	Hsp70 (<i>Homo sapiens</i>)	Putative Heat shock protein family	7,73	5,96E-04	0,34	275,46
TrE0204W	HOC1 (<i>S. cerevisiae</i>)	Putative α -1,6-mannosyltransferase	9,67	1,24E-11	2,65	322,69
TrB0922W	ATP9 (<i>S. cerevisiae</i>)	Putative Mitochondrial ATP synthase protein 9	2,34	2,65E-06	-0,65	43,15
TrD0622W	Putative SWI-SNF protein	Involved in chromatin remodeling	6,22	4,20E-03	1,47	360,71
TrG0353W	Putative SWI-SNF protein	Involved in chromatin remodeling	8,62	3,09E-07	1,65	1214,07
TrC1167C	Putative NAD(P)-binding Rossmann-fold domain protein	NA	11,18	8,25E-18	6,99	NA
TrB0880C	Unknown function	NA	9,09	8,07E-08	2,03	180,92
TrB1333W	Unknown function	NA	8,07	1,36E-04	1,45	126,49
TrD0899W	Unknown function	NA	5,56	8,71E-06	1,80	560,74

Tableau 22 : Gènes associés aux pics différentiellement apposés dans la condition lactose chez la souche QM6a.

Les logFC et les q-values de l'analyse différentielle du ChIP Seq sont donnés ainsi que les logFC de l'analyse différentielle du transcriptome et le nombre de reads par gènes (normalisé en RPKM). Les gènes en gras sont ceux touchés par un évènement d'apparition de la marque en lactose. Les logFC en rouge indiquent les gènes surexprimés dans cette condition tandis que les logFC en vert indique les gènes sous exprimés dans cette condition.

Sur les 19 gènes associés aux pics DA en lactose, 16 sont touchés par une marque H3K4me3 qui n'est pas présente en condition glucose. Ces résultats montrent un fort clivage entre les deux conditions en ce qui concerne l'apposition de la marque et témoigne d'une spécificité bien particulière de l'apparition de H3K4me3 chez la QM6a en condition lactose.

Nous notons que dans cette condition, tous les gènes sauf celui codant pour l'orthologue de DIM2 chez *N. crassa* sont considérés comme exprimés (RPKM>10). Également, huit gènes associés à un pic DA sont aussi surexprimés dans cette condition (gènes codant pour les protéines CEL1A, ACE3, Xyr1, CRT1 et pour les orthologues des protéines putatives cfa14D, HOC1, TrC1167C et TrB0880C). Enfin, nous notons que le gène codant pour la protéine STP1 est sous-exprimé dans cette condition.

Nous comptons quatre gènes associés aux pics DA impliqués dans les processus de production de cellulases. Nous retrouvons ici les gènes codant pour deux facteurs de transcription majeurs dans l'induction de la production enzymatique dont nous avons parlé en introduction, XYR1 et ACE3 mais aussi le gène codant pour CEL1A, une β -glucosidase intracellulaire impliquée dans l'induction des cellulases par le lactose (Xu et al. 2014) et le gène codant pour CRT1. Cette dernière est une protéine de transport transmembranaire permettant l'import de cellodextrines et de lactose et dont l'expression basale facilite le début de l'induction (Havukainen et al. 2020).

Nous notons également que le gène codant pour la protéine transporteur STP1 a un marquage enrichi par H3K4me3 en lactose bien qu'étant sous-exprimé. Il a été démontré que lorsqu'en culture sur cellulose, les mutants $\Delta stp1$ de *Trichoderma reesei* arborent une augmentation significative de l'activité des β -glucosidases et des cellulases extracellulaires (Zhang et al. 2013). Ainsi, la protéine STP1 semble avoir un effet répresseur sur l'induction des gènes majeurs de cellulases et hemicellulases et donc un rôle antagoniste à celui de CRT1 (Havukainen et al. 2020).

Quant à la protéine putative cfa14D, il s'agit d'un transporteur membranaire appartenant à la superfamille MFS et a été décrite comme exprimé précocement lors de l'induction de la production de cellulases sur sophorose (Schmoll et al. 2004).

Enfin, le gène codant pour la protéine putative de l'ADN méthyltransférase DIM2 est également associé à un pic DA. Néanmoins, celui-ci se trouve sur la partie 3' du gène et après vérification des données de transcriptomique, nous nous sommes rendu compte que les *reads* au locus du pic étaient anti-sens et qu'en réalité un transcrit ne correspondant pas au gène était généré.

Un orthologue du gène codant pour ARI9 est présent. Chez *A. thaliana*, celle-ci est impliquée dans la réponse aux stress environnementaux par la modification des protéines par ubiquitination, les menant à leur dégradation (Mladek, Guger et Hauser 2003). Deux autres orthologues de gènes de protéines putatives impliquées dans les

réponses aux stress sont aussi présents. La protéine humaine Hsp70 (« *Heat shock protein family* ») qui sert à la stabilisation de sa protéine substrat jusqu'à l'amélioration des conditions environnementales et évite sa dégradation (Tavaria et al. 1996) ainsi que l'orthologue de actfB chez *A. oryzae* conférant aux conidies une résistance aux stress oxydatifs et thermiques (Sakamoto et al. 2008).

Pour conclure, nous retrouvons dans les gènes marqués différenciellement par H3K4me3 en lactose un groupe fonctionnel de gènes dédiés à la production de cellulases. Ainsi, la présence des deux facteurs de transcription ACE3 et XYR1 est particulièrement satisfaisante et montre une activation bien ciblée du système de production cellulolytique. Ceci est accentué par la présence de trois autres gènes également impliqués dans l'induction de la synthèse ou dans le transport de sucres. De plus, l'apposition différentielle de la marque est ici couplée à une expression différentielle des gènes. Pour finir, un autre groupement de gènes engagés dans la réponse aux stress environnementaux sont également différenciellement marqués par H3K4me3 bien qu'ils ne soient pas positivement différenciellement exprimés dans cette condition.

Le Tableau 23 ci-dessous donne la liste des 19 gènes associés aux pics différenciellement apposés en glucose ainsi que des informations sur les analyses différentielles de ChIP Seq et de transcriptomique.

ID-TFW	Protéine	Commentaire	Log2FC ChIP Seq	Adj p-value ChIP Seq	Log2FC Expression	RPKM Expression Glucose
TrA0951C	DDP1 (S. cerevisiae)	Putative Polyphosphate phosphatase	-9,49	5,79E-11	-0,85	55,31
TrC1011W	RAD5 (N. crassa)	Putative DNA repair and recombination protein	-2,01	1,17E-08	-0,15	1034,24
TrB1672W	VPS34 (S. cerevisiae)	Putative Vacuolar protein sorting-associated protein	-5,05	8,42E-10	-0,08	30,38
TrC1210C	BAR1 (S. cerevisiae)	Putative α -factor (ppg1) protease	-9,73	1,01E-11	0,08	44,18
TrB1686C	AXL2 (S. cerevisiae / C. albicans)	Putative plasma membrane	-5,61	8,90E-03	2,48	1,70
TrG0599W	Putative Galactose-1-phosphate uridylyltransferase		-5,27	1,39E-03	0,82	7,16
TrA0066W	Putative Glycosyltransferase family 2		-10,53	7,73E-16	-1,45	381,04
TrB0866C	Putative Glycosyltransferase		-5,40	1,28E-12	-0,36	74,17
TrF0711C	Putative Zn(2)Cys6 transcription factor (fungal specific)		-9,23	8,57E-10	-0,41	14,06
TrC1010W	Putative Trypsin		-2,01	1,17E-08	-0,73	929,03
TrE0285C	Putative Dnase I		-9,20	2,00E-09	-1,58	44,08
TrF0392C	Putative FAD/NAD(P)-binding domain protein		-10,33	7,88E-15	-1,67	136,30
TrB0613W	Putative DUF1275 family protein		-9,11	3,34E-09	-0,79	7,25
TrE1290W	Putative DUF500 family protein		-10,69	7,16E-16	-1,35	198,96
TrF0881C	Putative PH domain protein		-3,05	2,69E-13	-2,00	50,99
TrA1796W	Unknown function		-9,82	1,09E-12	-1,87	90,39
TrE1163W	Unknown function		8,57	1,36E-06	-0,10	33,16
TrG0124W	Unknown function		-9,72	3,53E-12	-0,50	16,64
TrB1683-2W	Unknown function		-7,38	1,20E-03	-0,25	11,01

Tableau 23 : Gènes associés aux pics différentiellement apposés dans la condition glucose chez la souche QM6a.

Les logFC et les q-values de l'analyse différentielle du ChIP Seq sont donnés ainsi que les logFC de l'analyse différentielle du transcriptome et le nombre de reads par gènes (normalisé en RPKM). Les gènes en gras sont ceux touchés par un événement d'apparition de la marque en glucose. Les logFC en rouge indiquent les gènes surexprimés dans cette condition tandis que les logFC en vert indique les gènes sous exprimés dans cette condition.

Au total, 12 des 19 gènes associés aux pics DA présentent une marque dans la condition glucose mais pas dans la condition lactose. Les six autres gènes DA présentent une marque dans la condition lactose significativement moins importante que dans la condition glucose.

Dans cette condition, tous les gènes possèdent un nombre de *reads* suffisamment important pour être considérés comme exprimés (RPKM > 10), à l'exception des gènes codant pour les protéines putatives AXL2, de la Galactose-1-phosphate uridylyltransférase et de la DUF1275. De plus, seul le gène codant pour une protéine putative contenant un domaine PH est considéré comme surexprimé en glucose. Le reste des gènes ne possèdent pas une différence de transcription entre conditions suffisamment importante pour être considérés comme différentiellement exprimés. Il apparaît donc ici que la présence différentielle de H3K4me3 sur un gène n'entraîne pas automatiquement son expression différentielle.

Parmi les 19 gènes associés aux pics différentiellement apposés en glucose, deux gènes de CAZymes codant pour des glycosyltransférases putatives dont une glycosyltransférase putative de la famille GT2 sont présents. Les gènes de cette famille sont spécifiquement exprimés en conditions de non-induction (mais ne sont pas exprimées en lactose, ou sophorose (données IFPEN non publiées)). Nous retrouvons aussi comme lié à un pic DA le gène codant pour la protéine putative galactose-1-phosphate uridylyltransférase qui fait partie de la voie de Leloir et permet la dégradation du galactose. Le galactose étant un des deux monomères composant le lactose, nous nous attendrions à retrouver ce gène lié à un pic DA plutôt en lactose. De fait, ce gène n'est pas considéré comme exprimé dans la condition glucose.

Également, le gène codant pour l'orthologue de la protéine DDP1 est présent. Chez *S. cerevisiae*, cette protéine est une polyphosphate phosphatase hydrolysant les molécules d'inositol pyrophosphates et les polyphosphates de dianosine, permettant le maintien de nombreux procédés cellulaires importants ou vitaux tels que la réponse aux stress, la maintenance des télomères, la réparation de l'ADN ou encore la synthèse de la paroi cellulaire. Ceci va de pair avec la présence de l'orthologue du gène codant pour la protéine putative RAD5 chez *N. crassa* impliqué dans la réparation des coupures double brin de l'ADN.

Un autre élément intéressant à commenter est la présence du gène orthologue de la protéase BAR1 du α -facteur ppg1 dont l'absence impacte fortement la reproduction à cause de la perte en efficacité de détection de partenaires chez *S. cerevisiae* (Barkai, Rose et Wingreen 1998). Aussi, nous notons le marquage du gène l'orthologue pour la protéine AXL2 chez *Saccharomyces cerevisiae* et *C. albicans*, servant à l'organisation des tissus et à la polarisation des axes de croissance (CGD data base, Gene ID : C4_04170C_A). Pourtant, ce gène n'est pas exprimé, ni en condition glucose, ni en condition lactose (RPKM < 10).

Globalement, l'expérience en fed-batch glucose de la QM6a engendre le

marquage différentiel par H3K4me3 de gènes variés. Nous notons plusieurs gènes impliqués dans les mécanismes d'adaptation et de résilience au stress engendré par les conditions de culture. Ceci est notamment visible par la présence des orthologues de *S. cerevisiae* de gènes codant pour les protéines DDP1 et RAD5. Pareillement, nous retrouvons des gènes liés au développement et à la reproduction du champignon avec une présence de la marque sur les gènes des protéines putatives de BAR1 et AXL2 de chez *S. cerevisiae*. Aucun gène directement décrit comme étant en lien avec la production de cellulases n'est présent. De plus, si l'apparition de la marque est le phénomène majoritaire dans les pics DA de cette condition cela n'est pas lié à l'expression différentielle des gènes touchés. Nous remarquons également que la transcription des gènes est, en moyenne, moins importante ici qu'en lactose. Nous pouvons donc nous interroger sur l'impact de cette apparente ouverture chromatinienne différentielle et le rôle que celle-ci possède sur les mécanismes de régulation d'expression.

VI.2. Analyse différentielle de la marque H3K4me3 chez la souche RutC30

De la même façon que pour la QM6a, l'analyse de la RutC30 a été effectuée avec l'outil DiffBind et a suivi les mêmes étapes de préparation des échantillons. Ainsi, 3300 pics ont été retenus pour le consensus de pics et la longueur modèle du sommet retenu est de 833 pb.

Nous avons gardé les mêmes valeurs de q-value et de logFC pour les critères de filtration qu'auparavant. Un Volcanoplot de l'analyse a été généré afin de visualiser la distribution des données (Figure 59).

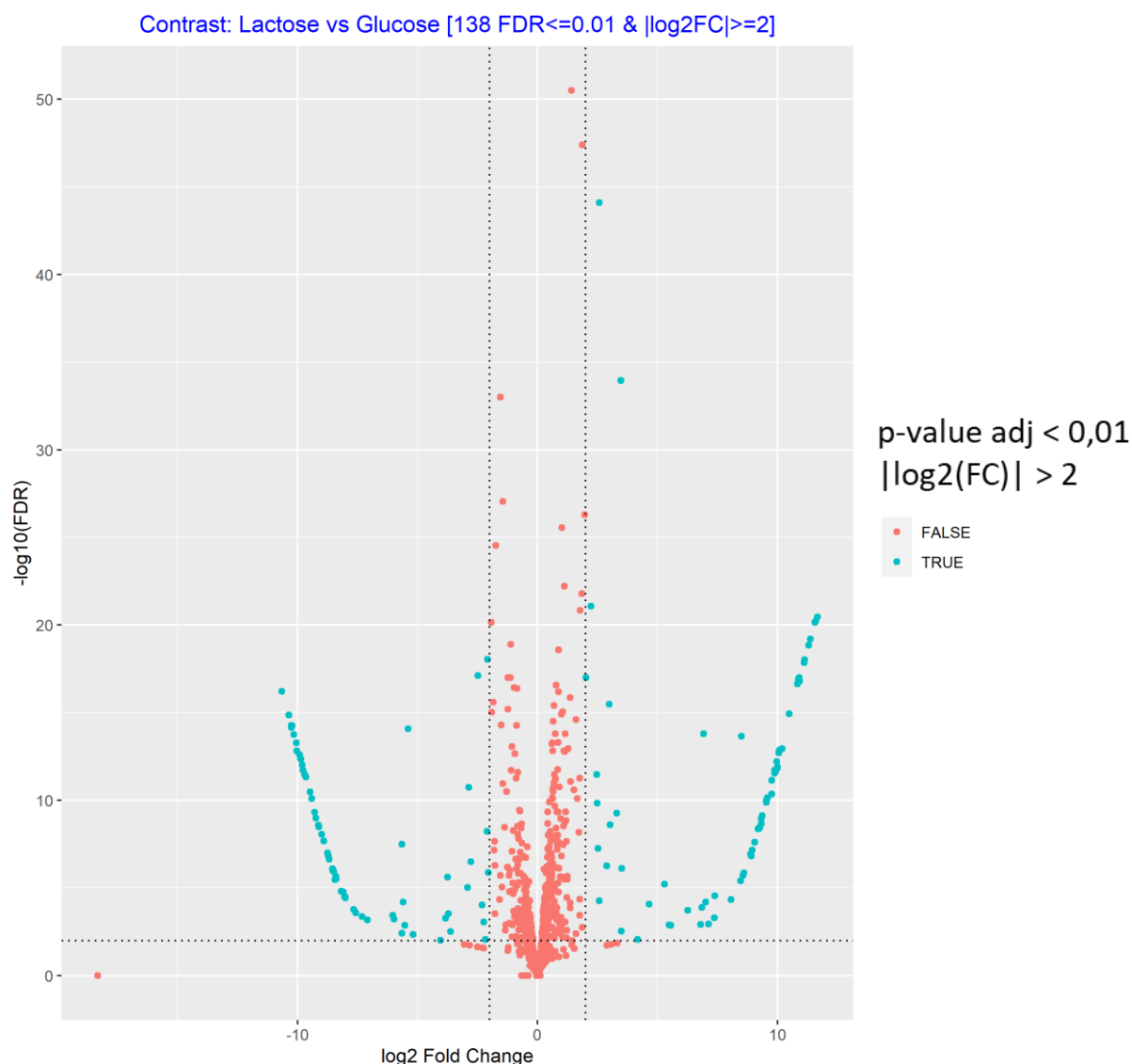


Figure 59: Volcano plot de l'analyse différentielle sur les données de la souche RutC30.

Les pics en bleu-vert sont ceux ayant passé avec succès les filtres sur la q-value et le fold change. Ceux en orange ont été écartés pour le reste de l'analyse.

Ainsi, 73 en lactose et 65 sommets DA en glucose ont été conservés pour la suite de l'analyse. Après rapprochement des sommets et des pics entiers, nous avons trouvé 70 gènes associés aux pics DA en lactose et 76 gènes associés à ces pics DA en glucose.

Pour la condition lactose, le Tableau 24 renseigne la liste des 70 gènes touchés par un pic différentiellement apposé. Parmi eux, nous comptons 51 gènes provenant d'un phénomène d'apparition de la marque en lactose alors qu'elle est absente en

glucose. Hormis le gène codant pour la protéine de DIM2 et trois gènes aux fonction non définies (Tr1720W, TrA1955C et TrB1619W), tous les gènes sont considérés comme exprimés (RPKM> 10). Aussi, 25 gènes sont surexprimés dans cette condition et un gène à la fonction inconnue est sous-exprimé.

ID-IFPEN	Protéine	Commentaire	Log2FC ChIP	Adj p-value ChIP	Log2FC Expression	RPKM Expression Lactose
TrB0496C	XYL1	D-xylose reductase	5,30	6,12E-06	2,47	844,97
TrC1552C	BXL1	Glycoside hydrolase family 3 CAZyme	5,48	1,29E-03	2,79	1062,26
TrA0210C	AES1	Carbohydrate esterase family	10,20	1,13E-13	4,69	5090,00
TrB1462W	LAD1	L-arabinitol 4-dehydrogenase	10,91	9,96E-18	3,14	69,81
TrB0038W	Putative Glycosyltransferase family 15	CAZyme	11,57	7,04E-21	3,38	191,68
TrA1911W	Putative Glycoside hydrolase family 18	CAZyme	10,01	1,24E-12	3,10	309,10
TrD0360W	Putative Glycoside hydrolase family 18	CAZyme	9,88	2,89E-12	2,16	315,85
TrE0671C	Putative Auxiliary Activities enzyme family 11	CAZyme	9,28	3,26E-09	4,63	4483,04
TrE0539W	VIB1	NDT80 Transcription factor	8,61	1,38E-06	1,39	114,02
TrF1042C	ClbR2 (<i>A. nidulans</i>)	Putatif Transcription factor (fungal specific)	10,06	1,79E-13	-0,21	48,16
TrE0494C	SxIR	Transcriptional repressor of GH11 xylanase genes	10,88	1,23E-17	1,00	111,09
TrE0589W	sgr-31 (<i>N. crassa</i>)	Putative Transcription factor (fungal specific)	9,54	1,08E-10	1,20	19,17
TrA0260C	tah-11 (<i>N.crassa</i>)	Putative Transcription factor (fungal specific)	11,60	6,07E-21	1,85	56,38
TrB1269C	Putative Zn(2)Cys6 transcription factor (fungal specific)		9,88	1,93E-12	0,59	17,44
TrC0680C	Putative Transcription factor		9,32	2,12E-09	1,38	58,74
TrD0036W	TrSTR1	MFS pentose transporter	11,67	3,45E-21	3,54	1227,10
TrB0814W	cdt-2 (<i>N. crassa</i>)	Putative MFS cellodextrin transporter	3,47	1,11E-34	2,08	35,08
TrB1268W	Putative MFS monosaccharide transporter		9,88	1,93E-12	1,31	95,58

ID-IPEN	Protéine	Commentaire	Log2FC ChIP	Adj p-value ChIP	Log2FC Expression	RPKM Expression Lactose
TrA0863W	Putative MFS monosaccharide transporter		9,05	2,38E-08	0,62	21,80
TrG0041C	GIT1 (<i>S. cerevisiae</i>)	Putative MFS transporter	9,90	2,60E-12	0,75	146,67
TrA1166C	Putative Inositol- 1,4,5-trisphosphate 5- phosphatase		6,27	1,90E-04	-0,26	15,58
TrD0533W	Putative INO80 subunit 2 IES2 (<i>S. cerevisiae</i>)		2,47	3,38E-12	1,11	129,06
TrB0908W	DIM2 (<i>N. crassa</i>)	Putative DNA methyltransferase	3,00	3,36E-16	-1,17	1,27
TrB1509W	ADY2 / ATO2 (<i>S. cerevisiae</i>)	Putative Acetate permease	2,88	5,56E-07	1,96	838,34
TrB0738W	Putative Succinyl- CoA:3-ketoacid- coenzyme A transferase		6,92	1,57E-14	1,53	59,44
TrD1298C	Putative Thiamine triphosphatase		10,09	1,33E-13	1,69	160,16
TrF1155W	Putative FAD/NAD(P) dependent oxidoreductase		8,46	3,93E-06	-0,10	561,79
TrB1312C	Putative Clavaminatase synthase		8,57	1,90E-06	1,36	32,52
TrG0156C	Putative Class I glutamine amidotransferase-like		8,07	4,55E-05	0,18	88,14
TrG0424W	MET16 (<i>S. cerevisiae</i>)	Putative 3'- phosphoadenylsulfate reductase	3,30	5,32E-10	1,33	68,56
TrB1170W	MET10 (<i>S. cerevisiae</i>)	Putative Assimilatory sulfite reductase alpha subunit	11,37	6,33E-20	0,65	70,51
TrG0629W	Putative NAD(P)H- dependent quinone reductase		11,30	1,41E-19	5,15	601,19
TrC0374C	Putative Quinoprotein alcohol dehydrogenase-like domain protein		7,38	2,83E-05	1,87	23,40
TrC1314C	Putative FMN-binding split barrel domain protein		2,53	5,50E-08	1,48	125,75
TrA0865C	Putative Ferric reductase transmembrane component		11,12	1,39E-18	6,48	718,40
TrC0741W	Putative GPI anchored protein		7,01	6,35E-05	1,51	1121,28
TrB0277C	Putative GPCR PTH11 type receptor gpr-25		7,37	5,11E-04	2,02	27,13
TrG0359C	Putative CFEM domain protein		10,49	1,14E-15	1,66	205,05

ID-IPEN	Protéine	Commentaire	Log2FC ChIP	Adj p-value ChIP	Log2FC Expression	RPKM Expression Lactose
TrA0030W	TauD (<i>E. coli</i>)	Putative Taurine catabolism dioxygenase	9,92	2,13E-12	4,59	174,09
TrF0767C	Putative Phospholipase D active site protein		2,49	1,48E-10	3,32	141,38
TrA1003W	AMA1 (<i>S. cerevisiae</i>)	Putative Activator of meiotic anaphase promoting complex (APC/C)	9,20	4,19E-09	1,33	37,62
TrA1310C	TRA1 (<i>M. grisea</i>)	Putative Spore adhesion transcription factor	9,76	7,20E-12	0,62	34,38
TrC0974C	IST1 (<i>S. cerevisiae</i>)	Putative Protein involved in the multivesicular body sorting pathway	4,65	8,32E-05	0,72	43,32
TrA1459C	Putative RNA-binding domain protein		2,02	1,01E-17	0,74	181,39
TrB0276W	Putative Flavin-containing monooxygenase		3,03	2,46E-09	0,95	139,33
TrC1463W	Putative Flavin-containing monooxygenase		4,17	8,46E-03	1,31	108,39
TrB1629W	Putative Nuclear migration protein		3,49	2,92E-03	0,24	6,88
TrF0098C	Putative Periplasmic binding protein-like II domain protein		2,24	8,17E-22	1,32	130,98
TrB1580C	Putative LIMBR1 domain protein		6,86	1,31E-04	0,36	17,29
TrA0578C	Putative MFS transporter		10,92	1,51E-17	1,73	35,52
TrB0334W	Putative MFS transporter		8,94	6,98E-08	2,32	149,67
TrE0940C	Putative MFS transporter		9,76	4,40E-11	1,51	105,23
TrE1138C	Putative MFS transporter		10,84	2,30E-17	5,10	562,54
TrB0195W	Unknown function		6,80	1,20E-03	1,64	366,89
TrB1119C	Unknown function		7,13	1,14E-03	1,55	182,60
TrE0436W	Unknown function		8,92	1,51E-07	2,28	248,87
TrA1909W	Unknown function		3,52	7,66E-07	2,75	309,94
TrB1579W	Unknown function		6,86	1,31E-04	2,26	96,29
TrC0325W	Unknown function		11,13	9,38E-19	1,59	76,22
TrC0967C	Unknown function		2,58	8,27E-45	2,40	926,25
TrE0036C	Unknown function		9,98	1,59E-12	4,43	125,75
TrE0124C	Unknown function		2,58	5,42E-05	0,16	10795,66
TrE1139W	Unknown function		9,97	6,16E-13	4,85	226,51
TrG0270W	Unknown function		9,36	7,51E-10	1,12	25,39
TrB0545C	Unknown function		9,89	2,53E-12	3,44	209,31
TrA1195W	Unknown function		5,56	1,33E-03	1,00	75,57
TrA1728C	Unknown function		9,20	4,18E-09	3,64	106,57
TrA1720W	Unknown function		9,58	6,91E-11	0,98	8,86

ID-IPEN	Protéine	Commentaire	Log2FC ChIP	Adj p-value ChIP	Log2FC Expression	RPKM Expression Lactose
TrA1955C	Unknown function		10,06	2,00E-13	-2,93	0,24
TrG0748C	Unknown function		8,88	1,18E-07	1,32	17,52

Tableau 24: Gènes associés aux pics différentiellement apposés dans la condition lactose chez la souche RutC30.

Les logFC et les q-values de l'analyse différentielle du ChIP Seq sont donnés ainsi que les logFC de l'analyse différentielle du transcriptome et le nombre de reads par gènes (normalisé en RPKM). Les gènes en gras sont ceux touchés par un événement d'apparition de la marque en glucose. Les logFC en rouge indiquent les gènes surexprimés dans cette condition tandis que les logFC en vert indique les gènes sous exprimés dans cette condition.

Nous dénombrons huit gènes codant pour des enzymes impliquées dans le catabolisme des carbohydrates notamment ceux codant pour la glycoside hydrolase BXL1, la protéine AES1 appartenant à la famille des carbohydrate esterases et la L-arabinitol 4-dehydrogenase LAD1. En plus de ceux-là s'ajoute le gène codant pour la D-xylose reductase XYL1 mais aussi des transporteurs de sucre dont le transporteur de xylose STR1 (Sloothaak et al. 2016) et l'orthologue chez *N. crassa* du gène codant pour le transporteur de cellodextrine CDT-2. La suppression de ce dernier entraîne un défaut de croissance et une chute dans l'expression des gènes de cellulases et d'hémicellulases chez le champignon (Cai et al. 2014). Aussi, un gène putatif codant pour une auxiliaire d'activité au CAZymes de la famille 11, possédant de ce fait une activité LPMO est présent.

Également, huit facteurs de transcription dont trois directement liés à la production ou la régulation de cellulases et d'hémicellulases sont à noter. Nous trouvons donc ici le gène codant pour la protéine VIB, facteur de transcription ayant été démontré comme essentiel dans la production de cellulases chez *Trichoderma reesei* grâce notamment à son travail en coopération avec Xyr1 (Chen, Xiuzhen et al. 2021). L'orthologue du gène putatif *clbR2* codant la protéine ClbR, impliquée dans l'induction des cellulases et des xylanases chez *A. nidulans* et *N. crassa* est aussi touché par un pic DA (Benocci et al. 2017; Coradetti et al. 2012). Enfin, nous comptons le gène codant pour la protéine SxIR (« *Specialized Xylanase Regulator* ») malgré son rôle antagoniste à la production et à l'activité des xylanases (Liu et al. 2017).

De plus, deux gènes directement liés à l'épigénétique ressortent comme associés à des pics différentiellement marqués. Ainsi, l'orthologue du gène de *N. crassa* codant pour l'ADN méthyltransférase DIM2 est présent bien que son expression montre une transcription plus importante en glucose. Aussi, nous retrouvons l'orthologue du gène codant pour la protéine putative IES2, impliquée dans le complexe remodeleur de la chromatine INO80, permettant le glissement des nucléosomes mais aussi la réparation et la transcription de l'ADN chez *S. cerevisiae* (Shen et al. 2000). Toutefois, ce gène présente aussi un rôle potentiel intéressant dans la production de cellulases à travers son interaction avec les gènes liés à l'inositol (Schmoll et al. 2004; Ford et al. 2007).

A ce propos, le gène codant pour la protéine putative inositol-1,4,5-trisphosphate 5-phosphatase est également présent ainsi que le gène codant pour la protéine putative de transport du glycerophosphoinositol GIT1 chez *Saccharomyces cerevisiae* (Patton-Vogt et Henry 1998). Ces deux gènes prennent part à de nombreux procédés cellulaires à travers le métabolisme de l'inositol et il a été démontré que leur activation n'est possible que par ouverture de la chromatine par INO80 au niveau du promoteur du gène *ino1* (Ford et al. 2007), gène majeur dans la synthèse de l'inositol phosphate.

De manière générale, nous notons un biais vers les gènes codant des protéines touchant à la production d'hémicellulases (BXL1, LAD1, XYL1, STR1, CDT-2, SxIR, CbIR2) ainsi que vers de nombreuses glycosides hydrolases et glycosyltransférases. Nous retrouvons également le gène codant pour la protéine putative IES2, impliquée dans le remodelage de la chromatine et présentant des perspectives intéressantes quant à son rôle dans la production de cellulases. Le marquage différentiel avec H3K4me3 de tous ces gènes est en grande majorité corrélé avec leur surexpression en lactose ou, pour le moins, avec leur expression active. Tout ceci tendrait à confirmer le lien entre apparition de la marque H3K4me3, transcription des gènes touchés et production cellulolytique chez cette souche.

Le Tableau 25 ci-dessous donne la liste des 76 gènes associés aux pics différentiellement apposés en glucose ainsi que des informations sur les analyses différentielles de ChIP Seq et de transcriptomique.

ID-IFPEN	Protéine	Commentaire	Log2FC ChIP Seq	Adj p-value ChIP Seq	Log2FC Expression	RPKM Expression Glucose
TrA0063C	Putative Glycoside hydrolase family 12	CAZyme	-5,99	5,88E-04	-2,60	207,61
TrD1261W	Putative Glycoside hydrolase family 16	CAZyme	-9,87	4,49E-13	-1,54	271,58
TrG0415W	Putative Glycoside Hydrolase family 16	CAZyme	-9,00	8,52E-09	-2,02	125,94
TrD1393W	Putative Glycoside hydrolase family 31	CAZyme	-8,69	2,26E-07	-2,41	190,46
TrE0752W	Putative Glycosyltransferase family 90	CAZyme	-5,52	1,38E-03	-0,20	24,38
TrB0866C	Putative Glycosyltransferase		-2,08	8,91E-19	0,45	133,88
TrG0713C	Putative Gluconokinase		-2,86	1,85E-11	-0,26	142,12
TrE0310C	RES-1 (<i>N. crassa</i>)	Putative Transcription factor involved in ER stress response and cellulase synthesis	-8,75	1,01E-07	-0,59	58,74
TrD1238W	PMH29	Transcription factor involved in cellulase activity	-10,23	5,66E-15	-0,14	80,16

ID-IPFEN	Protéine	Commentaire	Log2FC ChIP Seq	Adj p-value ChIP Seq	Log2FC Expression	RPKM Expression Glucose
TrA1580C	FlbC (<i>A. nidulans</i>)	Putative Conidiation transcription factor involved in cellulase and xylanase production	-10,36	1,34E-15	-1,74	36,59
TrA0983W	CDC25 (<i>S. pombe</i>)	Putative Membrane bound guanine nucleotide exchange factor	-8,91	2,03E-08	-1,20	47,58
TrB1167W	PCL1 (<i>S. cerevisiae</i>)	Essential for the control of the cell cycle at the G1/S (start) transition	-2,31	9,15E-05	-1,32	113,42
TrA0404W	BSN1 (<i>B. bassiana</i>)	Putative Positive regulator of mitotic exit	-5,19	4,43E-03	-1,27	30,71
TrE0773W	Asr1 (<i>S. cerevisiae</i>)	Putative ubiquitin ligase involved in protein ubiquitination	-8,71	1,47E-07	0,22	46,55
TrA1605C	SET6 (<i>S. pombe</i>)	Putative Histone-lysine N-methyltransferase	-8,09	1,69E-05	-1,15	13,85
TrC0899W	TRM1 (<i>S. cerevisiae</i>)	Putative tRNA (guanine-N2-)-methyltransferase	-8,41	2,09E-06	-0,45	29,36
TrA0985W	NGT1 (<i>C.albicans</i>)	Putative MFS N-acetylglucosamine (GlcNAc)-specific transporter	-7,32	4,31E-04	-2,09	367,06
TrB1672W	VPS34 (<i>S. cerevisiae</i>)	Putative Vacuolar protein sorting-associated protein	-8,54	8,21E-07	-0,96	19,88
TrE0620W	CLF1 (<i>S. cerevisiae</i>)	Putative pre-mRNA splicing factor	-3,72	2,91E-04	0,09	29,95
TrG0349C	IRC24 (<i>S. cerevisiae</i>)	Putative Benzil reductase	-2,05	1,33E-06	-0,14	53,44
TrE1025C	SOK1 (<i>S. cerevisiae</i>)	Putative protein involved in cAMP-mediated signaling	-5,59	6,41E-05	-2,04	48,13
TrG0604C	IDC2 (<i>P. anserina</i>)	Putative Small secreted protein involved in cell non-autonomous signaling of fruiting body development	-8,52	1,05E-06	-3,79	16,98
TrB1225W	PpoA/B (<i>A. nidulans</i>)	Putative Linoleate diol synthase	-7,57	2,56E-04	-1,81	343,94
TrC0942W	EMC4 (<i>S. cerevisiae</i>)	Putative Transmembrane complex required for efficient folding of proteins in the ER component	-2,92	9,51E-06	0,26	78,80

ID-IFPEN	Protéine	Commentaire	Log2FC ChIP Seq	Adj p-value ChIP Seq	Log2FC Expression	RPKM Expression Glucose
TrE0749W	Putative Peptidyl-prolyl cis-trans isomerase		-5,52	1,38E-03	-0,26	66,87
TrC0941C	Putative Subunit of the AP-2 complex (F. graminearum)	adaptor complex that functions in vesicle mediated transport	-2,92	9,51E-06	0,65	62,54
TrC1193W	TNA1 (<i>S. cerevisiae</i>)	Putative MFS nicotinic acid transporter	-10,26	5,34E-15	-1,82	291,71
TrD1010C	Putative MFS peptide transporter		-9,41	7,88E-11	-3,56	594,43
TrC0031C	Putative MFS transporter		-9,12	3,19E-09	-1,61	0,59
TrD0323C	Putative Iron regulated transporter		-10,04	1,45E-13	-1,85	231,53
TrE0753C	PAF1 (<i>S. cerevisiae</i>)	Putative Paf1p complex that binds to and modulates the activity of RNA polymerases I and II component	-5,52	1,38E-03	-0,07	34,06
TrE0750C	Putative 40S ribosomal protein S12 RSP12 (<i>S. cerevisiae</i>)		-5,52	1,38E-03	-0,96	2287,44
TrE0751W	Putative 40S ribosomal protein S22 RPS22B (<i>S. cerevisiae</i>)		-5,52	1,38E-03	-0,11	2420,66
TrD0187W	L29 (<i>N. neoformans</i>)	Putative Ribosomal protein	-9,13	2,60E-09	-2,02	160,35
TrE0774W	Putative Mitochondrial ribosomal protein of the large subunit MRPL17 (<i>S. cerevisiae</i>)		-8,71	1,47E-07	-0,14	40,28
TrD1190W	Putative dsRNA-binding domain protein		-7,08	6,55E-04	-0,93	51,76
TrE0754C	RER1 (<i>S. cerevisiae</i>)	Putative Protein involved in retention of membrane proteins (COPI)	-5,52	1,38E-03	0,56	163,30
TrD0622W	Putative SWI-SNF protein	Involved in chromatin remodeling	-5,65	3,71E-03	-0,50	198,86
TrE1237C	Putative SKG6/AXL2 alpha-helix transmembrane domain protein		-10,26	7,05E-15	-0,90	166,52
TrC1390C	Putative GPCR PTH11 type receptor		-3,75	2,36E-06	0,09	146,30
TrA1155W	PRO41 (<i>E.festucae</i>)	Putative ER membrane protein	-9,24	1,05E-09	-1,63	97,78
TrC0279W	PMA2 (<i>S. pombe</i>)	Plasma membrane H ⁺ -ATPase	-2,10	5,86E-09	-0,17	853,52
TrC1554W	Putative Metalloprotease		-2,77	3,14E-07	-0,54	10,13
TrC0415W	Putative Secreted aspartic protease		-6,04	3,64E-04	-1,19	82,49

ID-IFPEN	Protéine	Commentaire	Log2FC ChIP Seq	Adj p-value ChIP Seq	Log2FC Expression	RPKM Expression Glucose
TrA0840C	PPN1 (<i>S. cerevisiae</i>)	Putative Endopolyphosphatase	-5,38	8,15E-15	-0,85	15,91
TrD0194C	Putative Alpha/beta-hydrolase		-10,05	5,22E-14	-1,32	143,67
TrB0065W	Putative Betaine aldehyde dehydrogenase		-2,24	8,79E-04	-2,17	127,86
TrA1866C	NoxR (<i>E.festuca</i>)	Putative NADPH oxidase activator involved in Nox1 et Nox2 regulation	-8,49	1,12E-06	-2,06	33,58
TrD0358C	Putative Phosphoribosylaminoimidazole-succinocarboxamide synthase		-2,48	7,75E-18	-0,61	10,94
TrG0644C	Putative Homeobox protein		-9,11	3,14E-09	-0,36	13,67
TrA1455C	RfeD (<i>A. fumigatus</i>)	Putative Transcription factor	-9,91	2,56E-13	-0,38	57,14
TrA0609C	Putative C2H2 Transcription factor		-8,02	3,70E-05	-0,97	65,89
TrA1630C	Putative SAM and PH domain protein		-2,17	8,46E-03	-0,93	56,31
TrB0354W	MCO8 (<i>S. cerevisiae</i>)	Putative Mitochondrial protein	-4,04	9,39E-03	0,29	544,08
TrA1255W	Unknown function		-10,66	6,00E-17	-0,38	22,44
TrA1848C	Unknown function		-9,71	3,33E-12	-1,56	453,42
TrB1195C	Unknown function		-9,48	3,28E-11	-1,47	20,87
TrB0355C	Unknown function		-4,04	9,39E-03	-0,09	17,07
TrB1605W	Unknown function		-9,65	4,73E-12	-2,17	23,37
TrC0728W	Unknown function		-10,15	1,75E-14	0,14	17,48
TrD0264W	Unknown function		-9,87	4,21E-13	-1,19	135,43
TrE1210C	Unknown function		-3,63	3,07E-03	-0,52	73,80
TrF1025W	Unknown function		-9,78	1,93E-12	-3,39	12,23
TrG0101C	Unknown function		-7,66	1,68E-04	-0,55	3341,87
TrG0124W	Unknown function		-8,39	3,00E-06	-1,89	19,30
TrG0364W	Unknown function		-9,81	9,55E-13	-3,69	333,33
TrD0341C	Unknown function		-8,44	3,29E-06	-2,13	104,00
TrA1296C	Unknown function		-5,65	3,21E-08	0,00	390,90
TrB1168C	Unknown function		-2,31	9,15E-05	-1,94	49,59
TrG1064W	Unknown function		-8,01	3,34E-05	-0,53	524,02
TrG0187C	Unknown function		-9,13	2,60E-09	NA	1895,41
TrE0619C	Unknown function		-3,72	2,91E-04	3,60	NA
TrC0269W	Unknown function		-8,19	1,51E-05	-0,87	1,38
TrD0899W	Unknown function		-3,84	5,26E-04	-0,46	36,20
TrA1512W	Unknown function		-9,30	4,71E-10	-3,60	NA
TrF1115C	Unknown function		-8,04	3,07E-05	-4,53	NA

Tableau 25 : Gènes associés aux pics différentiellement apposés dans la condition glucose chez la souche RutC30.

Les logFC et les q-values de l'analyse différentielle du ChIP Seq sont donnés ainsi que les logFC de l'analyse différentielle du transcriptome et le nombre de reads par gènes (normalisé en RPKM). Les gènes en gras sont ceux touchés par un événement d'apparition de la marque en glucose. Les logFC en rouge indiquent les gènes surexprimés dans cette condition tandis que les logFC en vert indique les gènes sous exprimés dans cette condition.

Sur l'ensemble des gènes DA, 54 d'entre eux proviennent d'un phénomène d'apparition de la marque alors qu'elle est absente en lactose. Parmi tous les gènes associés à un pic DA, seul un codant pour un transporteur MFS putatif (TRC0031C) et un gène sans fonction déterminée (TrG0187C) présentent un niveau d'expression trop faible en glucose (RPKM<10) et ne sont donc pas considérés comme exprimés. Pour finir, 16 gènes, dont six ne possèdent pas de fonction connue à ce jour, possèdent une expression différentielle positive dans cette condition.

Sur le total de 76 gènes, six gènes putatifs de CAZymes ont été identifiés : deux gènes de la famille 16 des glucoside hydrolases (GH16), codant pour des putatives endo- β -1,3-D-glucosidases, un gène de la famille GH12 codant pour une putative endo- β -1,4-glucanase, et un gène de la famille GH31, codant pour une putative α -glycosidase. Nous en observons deux autres codant pour des putatives glycosyltransférases (TrB0866C et TrE0752W). Parmi eux, à l'exception des gènes TrA0063C (GH12) et TrB0866C (glycosyltransférase), le marquage par H3K4me3 est spécifique au glucose et est absent en lactose. Il est associé, en majorité, à une surexpression de ces gènes et, dans tous les cas, à une transcription active.

Nous comptons cinq gènes ou orthologues de gènes de plus codant pour des facteurs de transcription dont PMH29 et l'orthologue de chez *N. crassa* de RES-1. Il a été montré chez *Trichoderma reesei* que la surexpression de PMH29 entraîne une augmentation 1,5 fois plus importante de la production de cellulases chez la souche recombinante que chez la souche parentale lorsque cultivées en milieu liquide riche contenant du lactose (Häkkinen et al. 2014). Nous nous attendrions donc à trouver ce gène en lactose. D'un autre côté, chez *Neurospora crassa* la suppression du gène codant pour RES-1 résulte en une augmentation de 30% de la production de cellulases en comparaison avec la production de la souche sauvage lorsque cultivé en présence de cellulose (Fan et al. 2015).

Également, quatre orthologues de gènes codant pour des protéines impliquées dans le cycle cellulaire, en particulier dans l'initiation et la terminaison des cycles mitotiques, sont présents : CDC25 (*S. pombe*), PCL1 (*S. cerevisiae*), BSN1 (*B. bassiana*) et FlbC (*P. oxalicum*). Toutefois, nous notons que le dernier code pour la protéine putative du facteur de transcription FlbC impliqué dans la conidiation mais est aussi reconnu comme un activateur essentiel de la production de cellulases et xylanases chez *P. oxalicum* (Yao et al. 2016).

Quatre gènes en lien avec l'épigénétique sont aussi à remarquer, notamment le gène codant pour la protéine SET6 chez *S. pombe* possédant un domaine SET et MYN, et étant connue pour être une histone lysine méthyltransférase (ID : set6 / SPBP8B7.07c.)

Nous trouvons aussi l'orthologue du gène codant pour l'ubiquitine ligase putative Asr1 chez *S. cerevisiae*, inhibant la transcription des gènes situés dans les régions subtélomériques grâce à sa capacité d'ubiquitination de l'ARN polymérase (McCann et al. 2016). Chez *C. albicans*, le transporteur MFS NGT1, spécifique de la N-acetylglucosamine (GlcNAc), est responsable de l'apparition des hyphes et de la filamentation à travers des mécanismes résultant en l'acétylation des histones ou le repositionnement des nucléosomes au niveau des promoteurs des gènes HSG (« *Hyphal Specific Genes* ») (Hanumantha Rao, Paul et Ghosh 2021). Pour finir, un orthologue d'un gène codant pour une putative protéine appartenant au complexe remodeleur de la chromatine SWI/SNF est aussi marqué différemment.

Six autres orthologues de gènes impliqués dans le cycle cellulaire sont également présents, parmi lesquels nous comptons l'orthologue de *paf1* et quatre orthologues de gènes codant pour des protéines liées au complexe ribosomal (MRPL17, L29, S12 RSP12 et S22 RPS22B). Chez *S. cerevisiae*, la protéine de *paf1* forme un complexe avec les ARN polymérases I et II et est essentielle à l'élongation transcriptionnelle (SGD data base. Gene ID : PAF1 / YBR279W).

Pour résumer, nous observons le marquage différentiel de plusieurs gènes de CAZymes ainsi que de deux facteurs de transcription (PMH29 et FlbC) jouant un rôle inducteur dans la production de cellulases chez *T. reesei* et *P. oxalicum*. Nous retrouvons également l'élément régulateur de la production de cellulases RES-1 de chez *N. crassa*. En plus de cela, plusieurs gènes impliqués dans des mécanismes de régulation épigénétique se trouvent en lien avec H3K4me3. Nous notons que la quasi-totalité des gènes sont exprimés confirmant ainsi leur disponibilité pour la transcription à travers l'ouverture de la chromatine.

VII.COMPARAISON ENTRE LES SOUCHES DES RESULTATS DES ANALYSES DIFFERENTIELLES DES CHIP SEQ

Tout d'abord, nous notons que bien que les données pour les ChIP Seq aient été générées par deux expérimentateurs différents pour chaque souche et que les séquençages ont été réalisés par deux plateformes différentes nous obtenons des résultats robustes et analogues. Ainsi, nous retrouvons un nombre de pics dans les 6 échantillons IP très proches et un nombre de pics dans les ensembles consensus des deux souches également similaires (3185 pics pour la QM6a et 3300 pour la RutC30). Pour rappel, au sein d'une souche l'ensemble de gènes consensus rassemble,

conditions confondues, les gènes touchés par un pic H3K4me3 qui est présent dans au moins trois échantillons IP sur six. Les longueurs calculées du sommet des pics pour les analyses différentielles sont également équivalentes puisque nous avons obtenu 850 pb pour la QM6a et 833 pb pour la RutC30.

Dans les deux souches, nous retrouvons des corrélations importantes entre échantillons d'une même condition tandis que la corrélation inter-condition est beaucoup moins marquée. La variabilité des échantillons chez les deux souches est en majorité due à la condition dans laquelle ils ont été générés (Figure 41 et Figure 45).

Au regard de ces résultats robustes et après visualisation sur l'outil IGV des données, nous avons pu procéder avec confiance à la comparaison qualitative des résultats entre les souches indépendantes des analyses différentielles des ChIP Seq. Nous n'avons pas procédé à une analyse différentielle supplémentaire entre les souches du fait des translocations de la RutC30. Pour cela, il aurait fallu utiliser un caryotype unique et commun aux deux souches.

Nous avons vérifié si les principaux acteurs de la production enzymatique étaient touchés par H3K4me3 et étaient présents dans les ensembles de gènes consensus. Il s'est en effet avéré que les gènes codant pour ACE3, BglR, Xyr1, CRT1 et Hap2/3/5 étaient touchés par la marque dans les deux souches mais que cependant le gène pour CRE1 n'était présent que chez la QM6a comme attendu.

Nous avons utilisé l'outil BioVenn (Hulsen, Vlieg et Alkema 2008) afin de comparer les différentes listes de gènes obtenues durant les analyses des deux souches afin de déterminer si les paysages épigénétiques varient beaucoup d'une souche à l'autre au sein d'une même condition et identifier des mécanismes épigénétiques communs.

Nous avons commencé par comparer les deux listes des ensembles de gènes consensus. Il apparaît ainsi que 89% du total des gènes des deux consensus sont en commun entre la QM6a et la RutC30. D'un autre côté, 146 gènes sont spécifiques à la QM6a et 326 à la RutC30. Il est frappant de voir qu'au sein des gènes spécifiques de la RutC30 nous retrouvons 79% des gènes touchés par un pic DA en lactose de cette souche. De même pour la condition glucose, nous retrouvons également une proportion notable (34%) des gènes différentiellement associés à un pic, bien que non majoritaires. Nous notons la différence avec la QM6a qui n'affiche qu'un seul gène associé à un pic DA lactose dans sa liste de gènes spécifiques. Ce score s'améliore en glucose où nous comptons tout de même 11 gènes associés à un pic DA appartenant à la liste de gènes caractéristiques de la QM6a (Figure 60).

Il semble donc que les gènes DA de la RutC30 soient en majorité spécifiques à la souche et ne soient pas touchés par la marque chez la QM6a. En revanche, les gènes associés à un pic DA en lactose chez cette dernière sont aussi compris dans le consensus de la RutC30.

Nous remarquons que dans les deux conditions, la souche hyper-productrice présente près de quatre fois plus de gènes associés à un pic DA que la souche sauvage. Parmi les gènes touchés par des pics DA chez la RutC30, nous retrouvons de nombreux gènes associés à la production de cellulases mais contrairement à la QM6a les principaux acteurs de la production enzymatiques ne sont pas présents (i.e chez la QM6a : Xyr, ACE3, CRT1, BGL2) mais plutôt d'autres gènes dont le rôle est secondaire (XYL1, AES1, STR1, BXL1, ClbR).

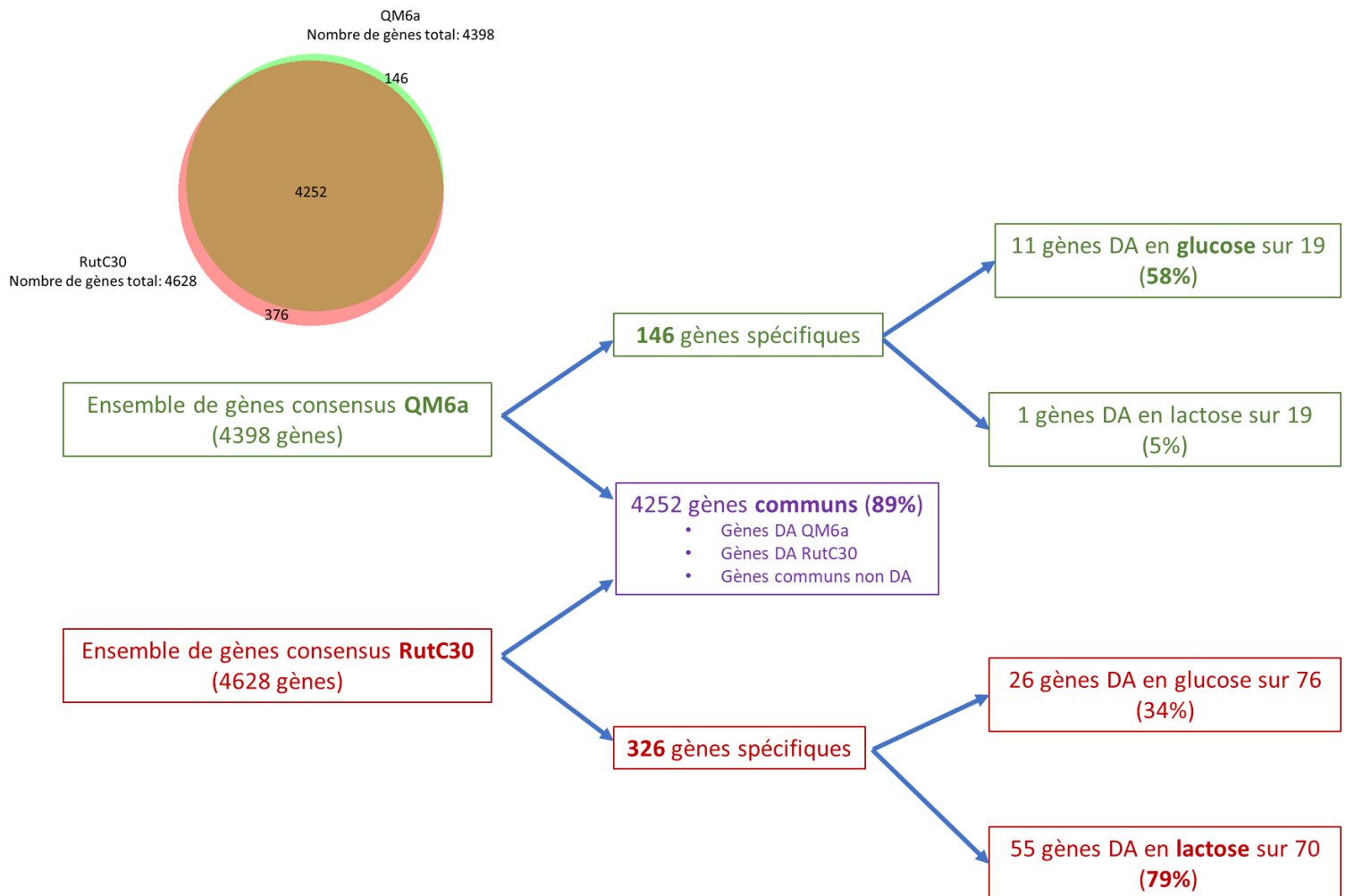


Figure 60 : Récapitulatif de la distribution des ensembles consensus de gènes associés à un pic H3K4me3 chez les deux souches.

89% des gènes sont partagés entre les deux souches (en violet). Nous comptons dans ces gènes communs, des gènes associés à un pic DA chez la QM6a et/ou chez la RutC30 mais aussi des gènes en commun entre les deux qui sont associés à un pic H3K4me3 sans qu'il soit différentiel. Il se trouve que 146 gènes sont spécifiques à la QM6a (vert) et ne se trouvent pas chez la RutC30. Parmi ces gènes nous en comptons 11 appartenant au groupe de gènes associés à un pic DA en glucose, ce qui représente un total de 58% des gènes associés à un pic DA dans cette condition (19 au total). Un seul gène associé à un pic DA en lactose fait partie de ces gènes spécifiques, les 18 autres sont dans le groupe des gènes communs avec la RutC30. Enfin, 326 gènes sont spécifiques à la RutC30 (rouge). Nous retrouvons 79% de la totalité des gènes associés

à un pic DA en lactose dans ce groupe spécifique de la souche hyper-productrice. Egalement, une part non négligeable (34%) des gènes associés à un pic DA en glucose appartiennent à ces gènes particuliers de la RutC30.

Ajouter à cela, après croisement des listes de gènes touchés par les pics DA de la QM6a et de la RutC30 pour les deux conditions, il ressort que très peu d'entre eux sont en fait partagés entre les deux souches (Figure 61). En effet, pour la condition lactose, seul le gène codant pour la protéine DIM2 est en commun entre la QM6a et la RutC30. Pour la condition glucose, trois gènes sont partagés entre les deux souches. Un gène codant pour une putative glycosyltransferase n'ayant pas encore de famille déterminée (TrB0866C), un orthologue de gène codant pour la protéine vacuolaire VPS34 chez *S. pombe* et un dernier ne possédant pas encore de fonction connue (TrG0124W). Ceci témoigne une fois de plus d'une forte différenciation entre les deux souches des mécanismes épigénétiques mis en place lors de l'induction de cellulases.

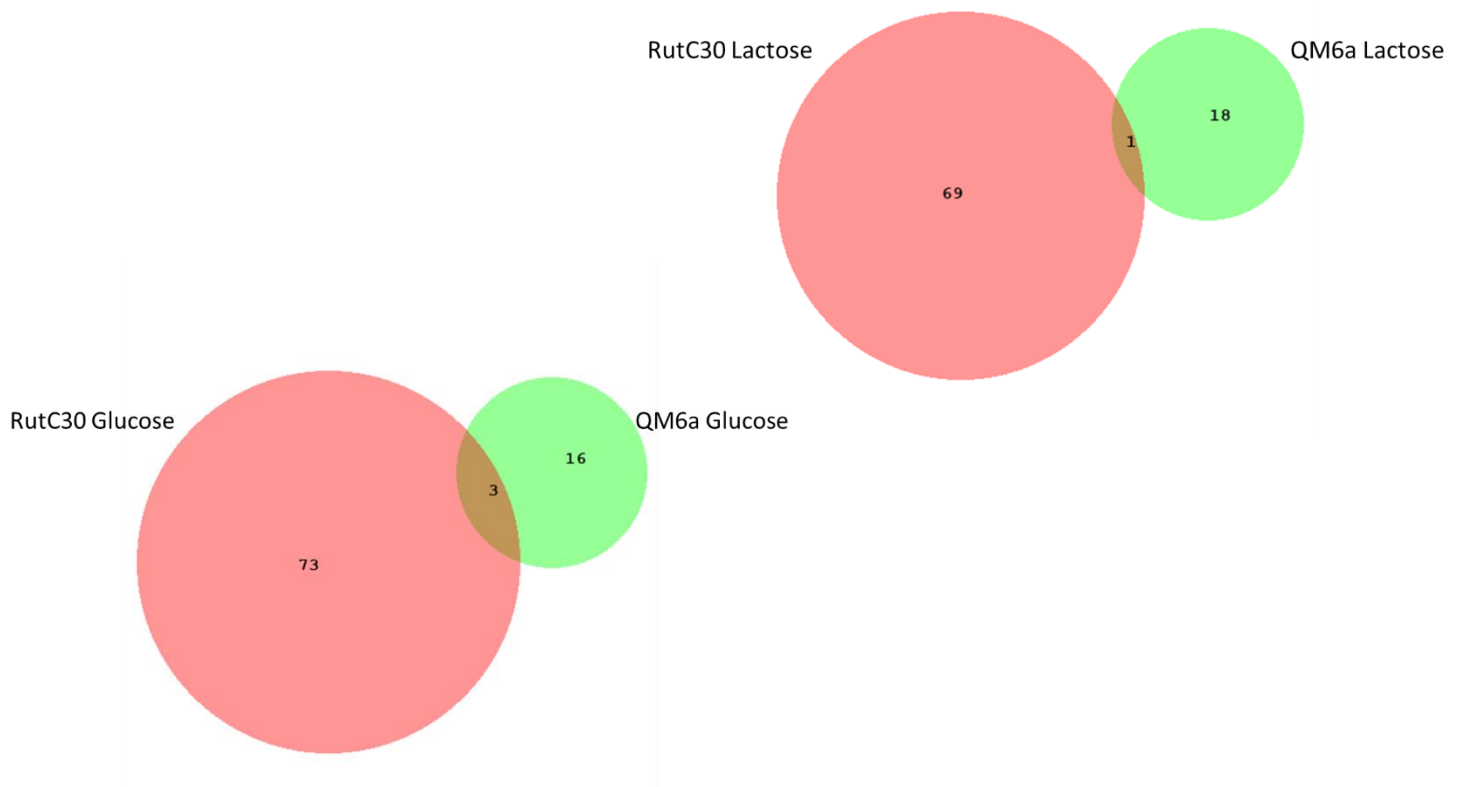


Figure 61 : Comparaison des gènes associés à des pics DA chez la RutC30 et chez la QM6a dans les deux conditions d'étude.

Chez les deux souches la transcription active des gènes touchés par des pics DA est la norme mais il est néanmoins difficile de tirer des conclusions sur le rôle joué par l'apposition différentielle de la marque et la surexpression de ceux-ci.

DISCUSSION

Dans ce premier chapitre, l'objectif principal était de générer pour la première fois chez *Trichoderma reesei* une cartographie des marques épigénétiques H3K4me3 (associée à l'euchromatine), H3K9me3 (associée à l'hétérochromatine constitutive) et H3K27me3 (associée à l'hétérochromatine facultative) dans deux conditions expérimentales (induction et non induction de cellulases) et sur deux souches de référence qui sont la souche sauvage QM6a et la souche hyper-productrice RutC30. Cette étude avait pour but de déterminer s'il existait un remodelage de ces marques lors de l'induction de cellulases et si elles avaient été impactées par les mutagénèses résultants en l'hyper-productivité de la RutC30 afin de mettre en lumière les différences existantes entre le système épigénétique sauvage et le muté. Pour répondre à ces interrogations, des expériences de ChIP Seq ont été réalisées afin d'identifier la localisation de ces marques le long des génomes. Pour finir, des analyses transcriptomiques des deux souches sont venues compléter les informations génétiques et épigénétiques de cette étude. Pour des raisons de temps liées à des difficultés d'analyses bioinformatiques, nous avons décidé de nous consacrer exclusivement sur la distribution de la marque H3K4me3.

Avec la cartographie de H3K4me3, l'intérêt était de générer une base de travail par la délimitation des régions euchromatiniennes pouvant être utilisées pour de futures études mais aussi mettre en lumière les premiers principes de fonctionnement des mécanismes épigénétiques lors de l'induction de cellulases chez *Trichoderma reesei*.

Concernant la distribution générale de la marque, il est apparu que celle-ci est similaire chez les deux souches avec des enrichissements au niveau des régions latérales des centromères et des appauvrissements au niveau subtélomérique. Chez la RutC30, il s'est avéré que les translocations ne génèrent aucun changement de distribution de la marque en comparaison avec la QM6a et que les jonctions n'affichent aucune « cicatrice ». Tout ceci pousse à penser que ces dernières ne sont donc pas impliquées dans l'hyper-productivité de la souche RutC30 et que de plus, contrairement à la souche QM9978 citée en introduction, aucun gène n'ayant été décrit comme impliqué dans la production enzymatique n'est proche des sites de coupure et n'a pu être influencé par ceux-ci. Afin de répondre définitivement à cette question, une étude des marques épigénétiques sur une souche possédant les mutations de la RutC30 sans les translocations serait nécessaire.

Après observation globale, les zones plus petites enrichies en marque le long des chromosomes sont identiques chez la QM6a et chez la RutC30 et ceci est appuyé par les 4252 gènes associés aux deux consensus de pics qui sont en commun sur les 4398 et 4628 gènes des consensus de la QM6a et de la RutC30, respectivement. Les secteurs déplétés en H3K4me3 sont également délimités chez les deux souches et rassemblent

plus de la moitié des gènes. Si ces zones rassemblent de nombreux gènes de cellulases nous n'avons pour l'instant pas déterminé si ces phénomènes de déplétion sont dus à la fonction des gènes ou aux zones sur lesquelles ils se trouvent (régions subtelomériques ou régions adjacentes à forte densité en marque H3K4me3). Toutefois, une certaine proportion de gènes non transcrits se trouvent aussi dans ces régions. Il se peut donc que ces dernières soient touchées à la fois par des marques épigénétiques répressives telles que H3K27me3 ou H3K9me3 mais aussi par des marques de transcription active telles que H3K36me3 ou H3K9ac. Il sera intéressant d'analyser plus profondément les données déjà générées.

Pour l'analyse différentielle des ChIP Seq de la souche QM6a il s'avère que peu de gènes sont associés à des pics DA puisque nous retrouvons seulement 19 gènes dans les deux conditions. En lactose, nous retrouvons comme attendu les gènes de l'induction de cellulases avec les acteurs principaux de l'activation : les facteurs de transcriptions activateurs des gènes de cellulases *xyr1* et *ace3*, et le transporteur de lactose *crt1*. Après analyse du transcriptome, il s'avère que ces derniers sont également surexprimés. Néanmoins, d'autres gènes jouant un rôle dans l'induction des cellulases ne sont pas touchés par H3K4me3 bien qu'ils soient transcrits ou même, pour certains, surexprimés en lactose. Après vérification de l'analyse différentielle, il s'avère que la QM6a est en mesure d'initier une réponse transcriptionnelle complète puisque nous retrouvons bien surexprimés en lactose de nombreux gènes liés à la production enzymatique ainsi que plusieurs gènes de cellulases. Il semble donc que chez la souche QM6a, l'ouverture chromatinienne associée à H3K4me3 se fasse simplement sur les gènes inducteurs les plus importants mais pas sur les gènes de cellulases *per se* ni sur les gènes aux rôles plus secondaires dans l'induction. La présence d'une autre marque d'euchromatine sur ces loci est donc probable ou que leur régulation soit purement transcriptionnelle.

En glucose, il n'a pas été possible d'identifier de groupe fonctionnel parmi les gènes associés à un pic DA et nous dénombrons une majorité de gènes non caractérisés. De plus, contrairement à ce qui pouvait être attendu pour cette condition, nous ne retrouvons pas le gène de la répression catabolique *cre1* ni surexprimé, ni touché par un pic DA. Après vérification, il se trouve que ce dernier est en fait touché par la marque H3K4me3 et est également exprimé de façon basale dans les deux conditions et ne ressort donc pas après analyses différentielles. Il semble donc que même dans un milieu inducteur, la souche QM6a soit sous l'emprise d'une régulation sous-jacente qui se traduit par une ouverture locale de la chromatine et une transcription basale de ce gène particulier.

Chez la RutC30, l'analyse différentielle des ChIP Seq révèle 70 et 76 gènes associés à des pics DA en condition lactose et glucose, respectivement. Dans le contexte d'induction enzymatique, nous retrouvons parmi les gènes touchés par les pics DA, plusieurs gènes liés à l'expression des hémicellulases (BXL1, XYL1, LAD1) mais aussi deux facteurs de transcription (VIB1 et l'orthologue de ClbR) et deux transporteurs de

sucres (STR1 et CDT-2) connus pour leur implication dans la production enzymatique. Ces gènes sont touchés par un phénomène d'apparition de la marque alors qu'ils ne le sont pas dans la condition glucose. Il est intéressant de noter que ces gènes sont majoritairement tournés vers la production d'hémicellulases (BXL1, XYL1, LAD1, STR1, ClbR). Toutefois, les gènes des acteurs principaux de l'induction tels que Xyr1, ACE3, CRT1, BglR ne sont pas associés à un pic DA bien qu'étant surexprimés. Il s'avère que ces gènes sont en fait touchés de manière indiscriminée par H3K4me3 dans les deux conditions et sont donc écartés lors de l'analyse différentielle. Il semblerait donc que contrairement à la QM6a, la RutC30 possède, grâce à H3K4me3, une ouverture basale de la chromatine au niveau de ces loci particuliers quelle que soit la condition. De cette façon, les gènes concernés sont en fait disponibles immédiatement pour transcription et ne nécessitent pas un mécanisme ouverture/fermeture de la chromatine tel que visible chez les gènes cités plus haut ou encore comme ce que nous avons vu chez la souche sauvage. Une hypothèse serait qu'au moment de l'induction, une/des marque(s) de transcription active telle(s) que H3K36me3 ou une acétylation soi(en)t déposée(s) sur les gènes déjà touchés par H3K4me3 et permet(tent) ainsi une transcription positive plus importante et en temps réel. Une régulation transcriptionnelle facilitée par l'ouverture constitutive de la chromatine est aussi une possibilité à conserver. Tout ceci est appuyé par les résultats de transcription montrant que les gènes touchés par la marque H3K4me3 sont significativement plus exprimés que ceux ne l'étant pas et tend ainsi à valider le lien entre présence de la marque et transcription active. Il apparaît donc ici que la RutC30 possède un « temps d'avance » grâce à cette action stable sur les gènes majeurs de l'induction cellulolytique. Une autre hypothèse serait que cet avantage lui permette ensuite une ouverture chromatinienne spécifique sur d'autres gènes inducteurs (particulièrement les gènes de production d'hémicellulases) lorsque nécessaire. Ceci participe peut-être à ces capacités hyper-productrices par une amélioration de ces capacités réactionnelles.

Au regard de ces informations, les résultats obtenus lors de la comparaison des analyses différentielles des ChIP Seq des deux souches en condition lactose sont tout à fait dans la continuité des observations précédentes. Il s'est avéré que les gènes associés aux pics DA de la RutC30 lui sont en majorité (79%) spécifiques et ne se retrouvent pas dans le consensus de gènes de la QM6a (Figure 61). Au contraire, seul 1 gène associé à un pic DA en lactose chez la QM6a lui est spécifique, le reste se trouve en commun avec les gènes consensus de la RutC30. Ces résultats suggèrent que le schéma de régulation épigénétique de la RutC30 est une évolution et non une rupture complète par rapport à celui de la QM6a. La marque H3K4me3 a donc bien été influencée par les mutations menant à l'hyper-productivité de la RutC30 et joue peut-être un rôle positif dans celle-ci en conjonction avec d'autres mutations comme la troncation de *cre1* ou l'interaction avec une autre marque épigénétique.

De façon remarquable, le seul point commun entre les deux listes de gènes associés aux pics DA dans la condition lactose chez les deux souches est le gène codant pour l'ADN méthyltransférase DIM2, responsable de la méthylation de l'ADN et

spécifique de l'hétérochromatine constitutive. Après vérification, la marque H3K4me3 est présente sur la partie terminale du gène mais au début d'un transcript anti-sens (*Dim2-AS*). Les transcrit antisens sont une classe de molécules d'ARN qui sont transcrites à partir du brin opposé d'un gène codant et qui sont au moins partiellement complémentaires de l'ARN sens. En cette qualité, ils peuvent participer au contrôle de l'expression du gène localisé en sens. En observant le niveau d'expression de *Dim2-AS*, il apparaît clairement plus exprimé en lactose qu'en glucose dans les deux souches. La présence de la marque H3K4me3 pourrait donc permettre l'activation de l'expression de *Dim2-AS* et permettre la répression de ce gène en présence de lactose. Une expérience supplémentaire pour vérifier l'implication de ce transcrit dans la régulation de *dim2* serait l'insertion ectopique d'une nouvelle copie du gène avec un terminateur différent afin d'éliminer *Dim2-AS* et examiner l'expression de *dim2* en fed-batch lactose.

Pour les résultats en glucose, il est plus difficile de tirer des conclusions claires au sujet des mécanismes communs mis en place par l'une ou l'autre des souches. Les deux possèdent un groupe de gènes qui leur est spécifique mais ceux-ci rassemblent majoritairement des gènes du cycle cellulaire ou aux fonctions encore inconnues. Les trois gènes associés à des pics DA qu'elles partagent ne permettent pas non plus de formuler des hypothèses quant à leur implication sur la vie du champignon. Néanmoins, le peu de gènes associés à des pics DA communs entre les deux souches, même dans cette condition, démontre qu'il existe des différences basales importantes qu'il serait certainement digne d'intérêt d'élucider. Un premier pas dans ce sens serait la détermination des fonctions des gènes spécifiques à l'une ou l'autre des souches.

Pour pousser plus loin, il serait intéressant de terminer les analyses sur les marques H3K9me3 et H3K27me3 et de les mettre au regard des résultats présentés dans ce manuscrit. De plus, la réalisation de cartographies de H3K36me3 ou H3K9ac serait une addition de valeur à ces données notamment si elles sont réalisées au cours de cinétiques longues afin de délimiter avec précision quels sont les moments clés d'ouverture/fermeture de la chromatine et transcription active des gènes. De cette façon, nous pourrions aussi déterminer si ces marques sont bien impliquées dans la transcription des gènes qui ne sont pas touchés par H3K4me3. De plus, durant cette thèse, des données de séquençage bisulfite ont été générées, il serait donc opportun de les intégrer au reste de résultats afin d'obtenir un point de vue systémique plus ample de l'épigénétique de *Trichoderma reesei*.

CHAPITRE 2

La deuxième partie du travail de cette thèse a consisté en la construction d'une série de mutants de *Trichoderma reesei* pour des gènes impliqués dans la mise en place de marques épigénétiques afin d'étudier leur rôle dans la régulation de la production enzymatique de ce champignon. Ces mutants ont été construits dans les deux souches QM6a et RutC30. L'utilisation de la souche sauvage nous apporte un éclairage sur le rôle des gènes dans un fond génétique sauvage aux régulations strictes de l'induction de cellulases. La souche hyper-productrice RutC30 permet d'évaluer l'effet additif d'un fond génétique muté et tourné vers la production enzymatique. Cela nous donnera la possibilité de vérifier si les gènes visés peuvent présenter un intérêt industriel. Au total, cinq gènes ont été sélectionnés pour une étude de leur fonction par invalidation : *dim5*, *dim2*, *kmt6*, *hp1* et *rid*. Les protéines DIM5 et KMT6 sont des méthyltransférases apposant respectivement les marques H3K9me3 (hétérochromatine constitutive) et H3K27me3 (hétérochromatine facultative). La protéine DIM2, est une ADN méthyltransférase de maintien alors que RID est une ADN méthyltransférase *de novo*. Enfin, HP1 est une protéine lectrice des marques H3K9me3 et responsable du recrutement de DIM2. Ces gènes correspondent donc chacun à un niveau de contrôle épigénétique chez le champignon et sont à l'origine des marques principales de la fermeture de la chromatine. Ils font partie de mécanismes connus chez d'autres champignons mais n'ont pas encore été caractérisés chez *Trichoderma reesei*. Leur invalidation permettra de faire la lumière sur le rôle de ces différents acteurs chez *T. reesei*.

Les gènes ont été inactivés via le mécanisme de recombinaison homologue permettant leur remplacement par une cassette d'expression du gène *hph*, conférant la résistance à l'hygromycine. A ceci, nous avons couplé la technologie CRISPR-Cas9 pour une efficacité accrue de l'insertion au bon locus. Des étapes d'isolement et de purification des transformants puis une validation génétique de l'insertion de la cassette d'invalidation au site du gène ciblé ont permis une sélection des mutants portant les modifications génétiques visées. Ensuite, plusieurs expériences de phénotypage sur milieu liquide inducteur ont apporté des informations quant à l'impact des délétions sur la production enzymatique des transformants et ont permis de détecter ceux possédant les caractéristiques les plus remarquables par rapport à leur souche parentale et présentant un intérêt dans le cadre de notre étude. Des phénotypes sur boîtes de Petri contenant différents milieux et soumises à différentes conditions environnementales ont apporté des renseignements sur la croissance et le développement de ces mutants.

RESULTATS

I. CONSTRUCTIONS DES MUTANTS DANS LES SOUCHES QM6A ET RUTC30

I.1. Identification des gènes à invalider

Les orthologues des gènes *dim5* (protéine XP_957479.2), *dim2* (protéine XP_959891.1), *hp1* (protéine XP_957632.1), *kmt6* (protéine XP_965043.2) et *rid* (protéine AAM27408.1) chez *N. crassa* ont été identifiés chez *T. reesei* sur fungidb (<https://fungidb.org/fungidb/app>). Les comparaisons des annotations structurales de la version 2 disponible sur la base de données du Joint Genome Institute et de celle aimablement fournie par TF Wang (Li, Chen et Wang 2018), montre des différences sur la taille et/ou sur le nombre d'exons des gènes *dim5*, *dim2*, *kmt6* et *hp1*. Les annotations ont été révisées grâce à l'utilisation de données transcriptomiques générées par Pirayre et al. (2020) et l'alignement avec les orthologues respectifs de *N. crassa*. Les domaines protéiques principaux ont été déterminés sur le logiciel Geneious Prime avec l'outil Interproscan. Les nouvelles séquences nucléotidiques sont disponibles en annexe VI. La Figure 62 ci-dessous renseigne sur les domaines protéiques de ces gènes chez *Trichoderma reesei* ainsi que sur leur phylogénie.

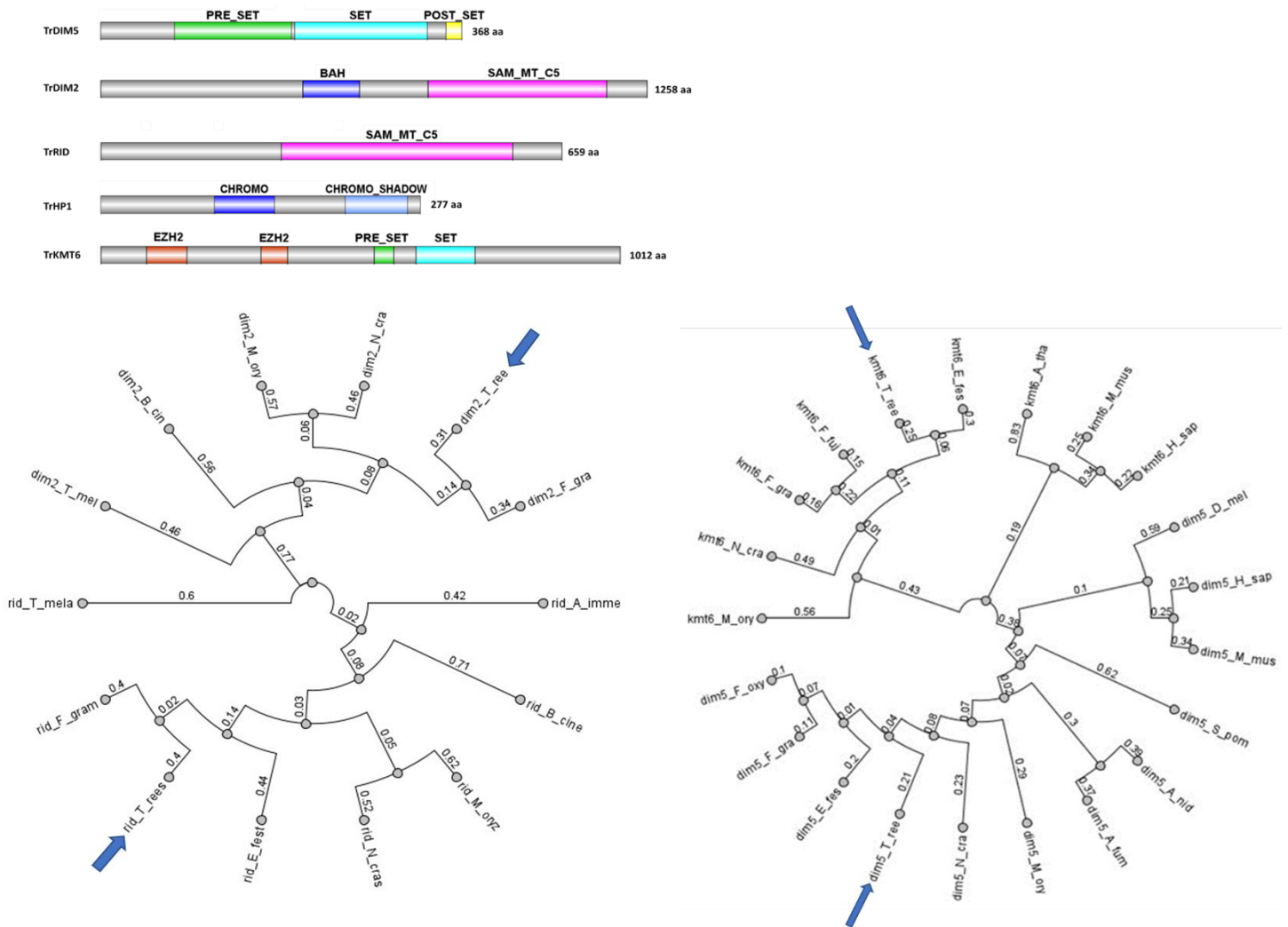


Figure 62: Domaines protéiques et arbres phylogénétiques des protéines des gènes *dim5* et *kmt6* et des gènes *dim2* et *rid*.

Les flèches bleues indiquent les gènes de *T. reesei*. A gauche est représenté l'arbre des orthologues de *dim2* et *rid* et à droite celui de *kmt6* et *dim5*. Les organismes représentés sont les suivants : *T. reesei* (*T_ree*), *T. melanosporum* (*T_mel*), *N. crassa* (*N_cra*), *F. graminearum* (*F_gra*), *M. oryzae* (*M_ory*), *B. cinerea* (*B_cin*), *A. nidulans* (*A_nid*), *S. pombe* (*S_pom*), *F. oxysporum* (*F_oxy*), *H. sapiens* (*H_sap*), *D. melanogaster* (*D_mel*), *M. musculus* (*M_mus*), *E. festucae* (*E_fes*), *A. fumigatus* (*A_fum*), *A. thaliana* (*A_tha*), *F. fujikuroi* (*F_fuj*), *A. immersus* (*A_imm*).

Les domaines pre-SET (IPR001214), SET (IPR001214) et post-SET (IPR003616) de la protéine TrDIM5 sont hautement conservés et sont caractéristiques des histones méthyltransférases et permettent la triméthylation des résidus lysine 9 de l'histone H3. Le domaine SET couplé aux domaines EZH2 de TrKMT6 sont également des domaines retrouvés chez les histones méthyltransférases et contribuent à la méthylation de H3K27me3. Les ADN méthyltransférases TrDIM2 et TrRID partagent un site d'ADN méthyltransférase SAM (IPR001525) caractéristique. En plus de cela, un domaine BROMO-Adjacent BAH (IPR001025) chez TrDIM2 est spécialisé dans les interactions protéines-protéines et agit spécifiquement sur l'extinction de gènes. Pour finir, les

domaines CHROMO (IPR023780) et CHROMO-SHADOW (IPR00825) de TrHP1 sont fréquemment associés aux protéines impliquées dans le remodelage de la chromatine via la reconnaissance de résidus lysines méthylés.

L'arbre phylogénétique (Figure 62) nous permet d'observer l'évolution et les relations entre protéines homologues. Nous remarquons que les deux groupes d'ADN méthyltransférases proviennent du même ancêtre comme les deux groupes d'histones méthyltransférases descendent d'une protéine commune. Toutefois, les homologues de KMT6 chez *Homo sapiens*, *Mus musculus* et *Arabidopsis thaliana* semblent être plus proches au niveau évolutif des homologues DIM5 que des protéines KMT6 des autres espèces. Comme attendu, les protéines homologues de HP1 sont les plus distinctes des quatre autres groupes.

1.2. Construction des mutants et validation PCR

Les souches QM6a et RutC30 ont été transformées avec des cassettes d'inactivation (Annexe VII) contenant le gène *hph* confèrent la résistance à l'hygromycine, flanqué de part et d'autre de 200 pb correspondant aux régions flanquantes du gène ciblé (Figure 63). L'intégration des cassettes par recombinaison homologue a été facilitée par l'utilisation de la technologie CRISPR-Cas9.

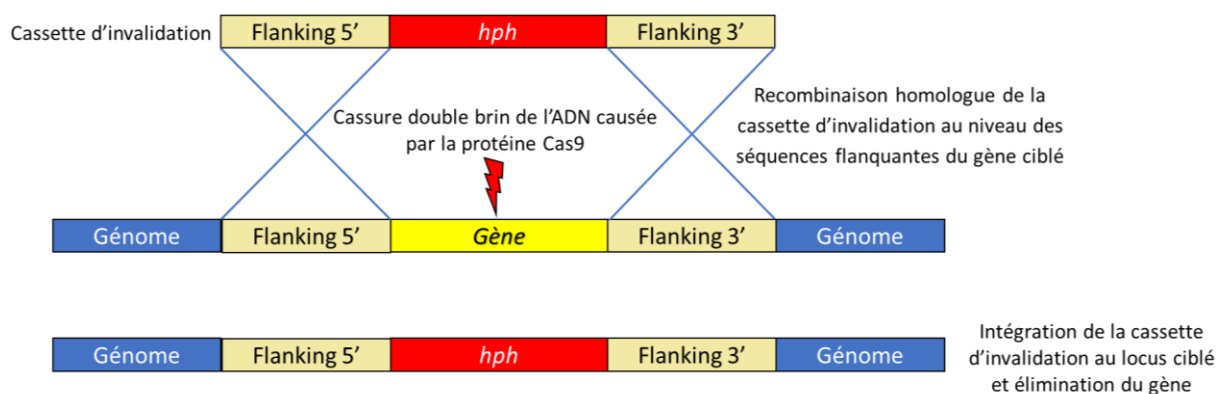


Figure 63: Représentation simplifiée de la recombinaison homologue de la cassette d'inactivation au niveau des gènes ciblés lors de la transformation effectuée avec l'aide de la technologie CRISPR-Cas9.

Les séquences flanquantes situées de part et d'autre du gène de résistance à l'hygromycine sont identiques à celles situées de part et d'autre du locus visé. Elles vont permettre un échange des deux gènes et ainsi l'intégration de la cassette d'inactivation au locus.

À la suite de la transformation, trois étapes de purification par isolement à partir de conidies (uninucléées chez *Trichoderma*) sont nécessaires à l'isolement de mutants homocaryotiques durant lesquelles des mutants premièrement isolés pour leur résistance à l'hygromycine peuvent finalement s'avérer être faux positifs. Les

hypothèses les plus probables pour expliquer ce phénomène sont :

- Les thalles obtenus sont hétérocaryotiques (mélange de noyaux ayant intégré le gène de résistance et de noyaux sauvages) car les protoplastes peuvent potentiellement contenir plusieurs noyaux
- Une intégration transitoire du gène de résistance s'est produite et a été éliminée lors de la réplication de l'ADN

La Figure 64 ci-dessous récapitule le taux de mutants uninucléés récupérés après les trois étapes de purification. Le taux de transformants conservés varie selon les mutants construits mais il est néanmoins notable que le nombre de mutants issus de la RutC30 est globalement plus élevé que celui des mutants issus de la QM6a. Ceci pourrait être dû à un protocole de transformation plus adapté à cette souche mais aussi à une fragilité plus grande de la structure du génome de la RutC30 causée par les nombreuses mutations et remaniements génétiques que la souche a subi et qui pourraient désormais faciliter l'intégration d'ADN exogène.

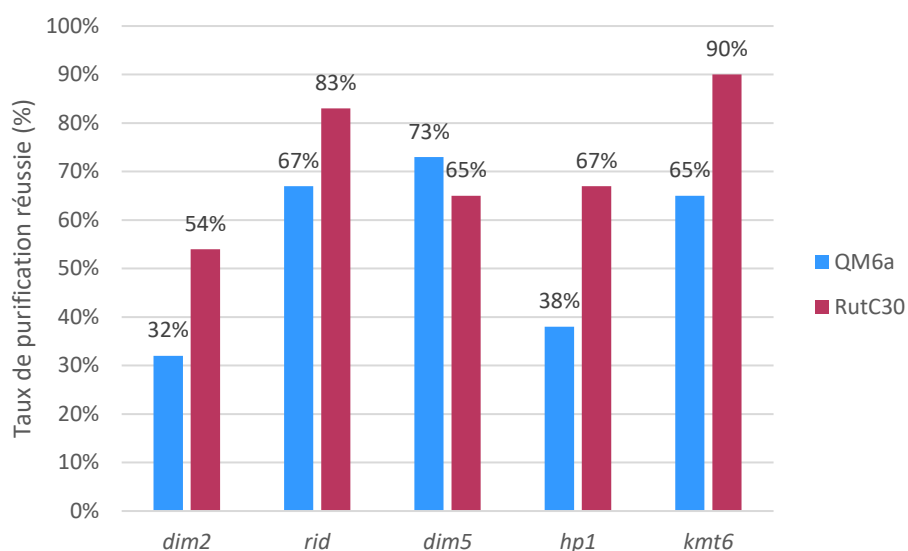


Figure 64: Taux de mutants purifiés pour chacun des gènes invalidés chez les deux souches .

Après les trois cycles de nettoyage génétique le nombre d'individus initialement sélectionnés après transformation était de 48 sauf pour les gènes *dim2* et *rid* chez la souche QM6a, pour lesquels nous avons repiqués seulement 41 et 42 clones, respectivement.

Après isolement des mutants nous avons procédé à une vérification génétique des mutants par PCR afin de s'assurer de la bonne intégration de la cassette d'inactivation (Figure 65). Pour cela, deux étapes de contrôle ont été réalisées. Une première étape de vérification de la délétion du gène visé est faite grâce à une PCR en duplex, c'est-à-dire avec deux paires d'amorces dans la même réaction. La première paire amplifie une petite partie du gène URA3 situé dans le génome de *T. reesei* et

n'ayant pas été touché par la transformation. Il s'agit d'un contrôle positif de la réaction qui permet de vérifier que la PCR a fonctionné correctement et que l'ADN extrait est utilisable. La seconde paire va cibler le gène qui devrait avoir été supprimé. Une bande de présence pour le gène de URA3 devrait donc apparaître alors qu'aucune bande ne devrait être visible pour le gène ciblé si celui-ci a bien été invalidé (Figure 65). Ceci permet de faire une sélection des mutants n'ayant plus le gène visé et donc a priori ayant bien intégré la cassette d'inactivation au bon locus.

Dans une seconde étape, des amplifications ont été réalisées avec des amorces s'hybridant de part et d'autre des régions d'homologie utilisées pour la recombinaison. Pour ce faire, des amorces ont été placées à l'extérieur et à l'intérieur des cassettes afin de vérifier le locus d'intégration (Figure 65).

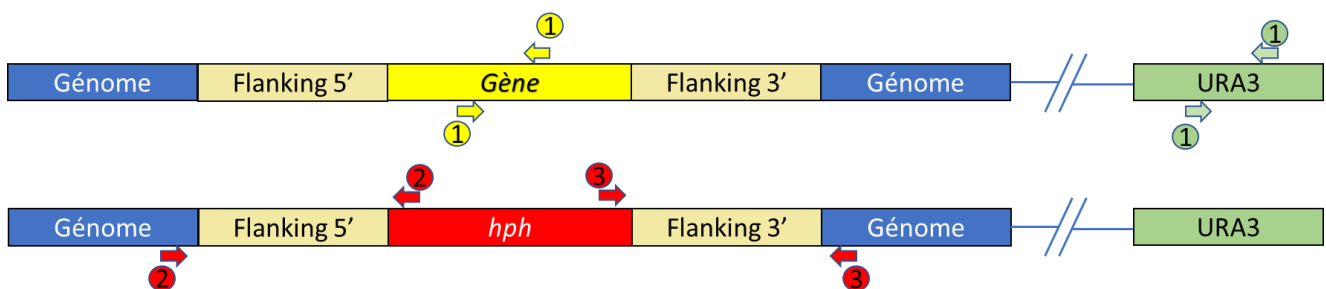


Figure 65: Représentation schématique des régions amplifiées par les trois PCR de validation.

(1) La PCR 1 permet de vérifier la disparition du gène. Si une bande apparaît, la délétion n'a pas fonctionné. D'un autre côté, l'amplification dans le gène URA3 sert de témoin positif. (2) et (3) Les PCR 2 et 3 permettent de vérifier l'insertion correcte de la cassette d'inactivation aux extrémités 5' et 3', respectivement.

Enfin, les PCR en 5' et 3' de l'insertion ont été réalisées chez les mutants ne possédant plus le gène ciblé. La Figure 66 et la Figure 67 ci-dessous présentent les résultats obtenus pour les transformants $\Delta dim5$ pour chacune de ces étapes. Le reste des gels pour l'ensemble des mutants sont disponibles en Annexe XIX.

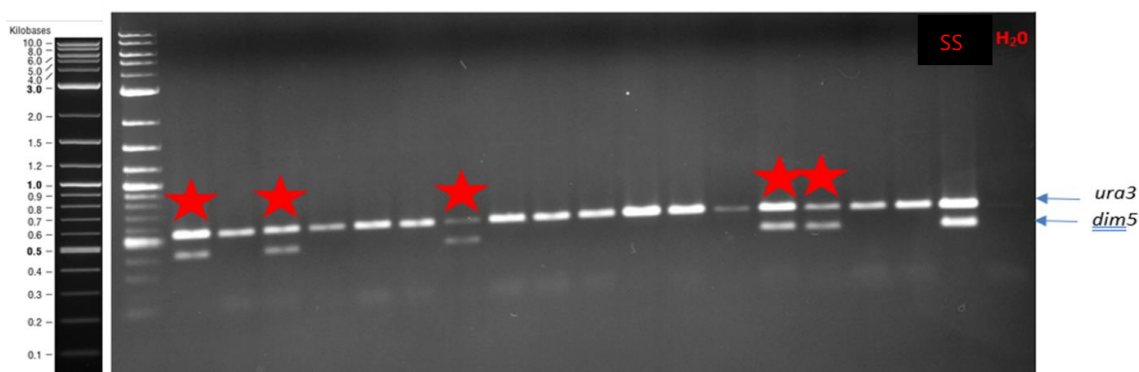


Figure 66: Gel de vérification de la PCR en duplex de l'invalidation du gène *dim5* chez la souche *RutC30* transformée.

Sur la gauche, le marqueur moléculaire permet de confirmer la taille attendue des

amplifications. Les transformants marqués par une étoile rouge ne sont pas validés car ils possèdent encore le gène *dim5*. Deux contrôles ont été fait afin de vérifier que les bandes visibles correspondent à ce que nous recherchions : une réaction pour la souche sauvage (SS) et une sur de l'eau (H₂O).

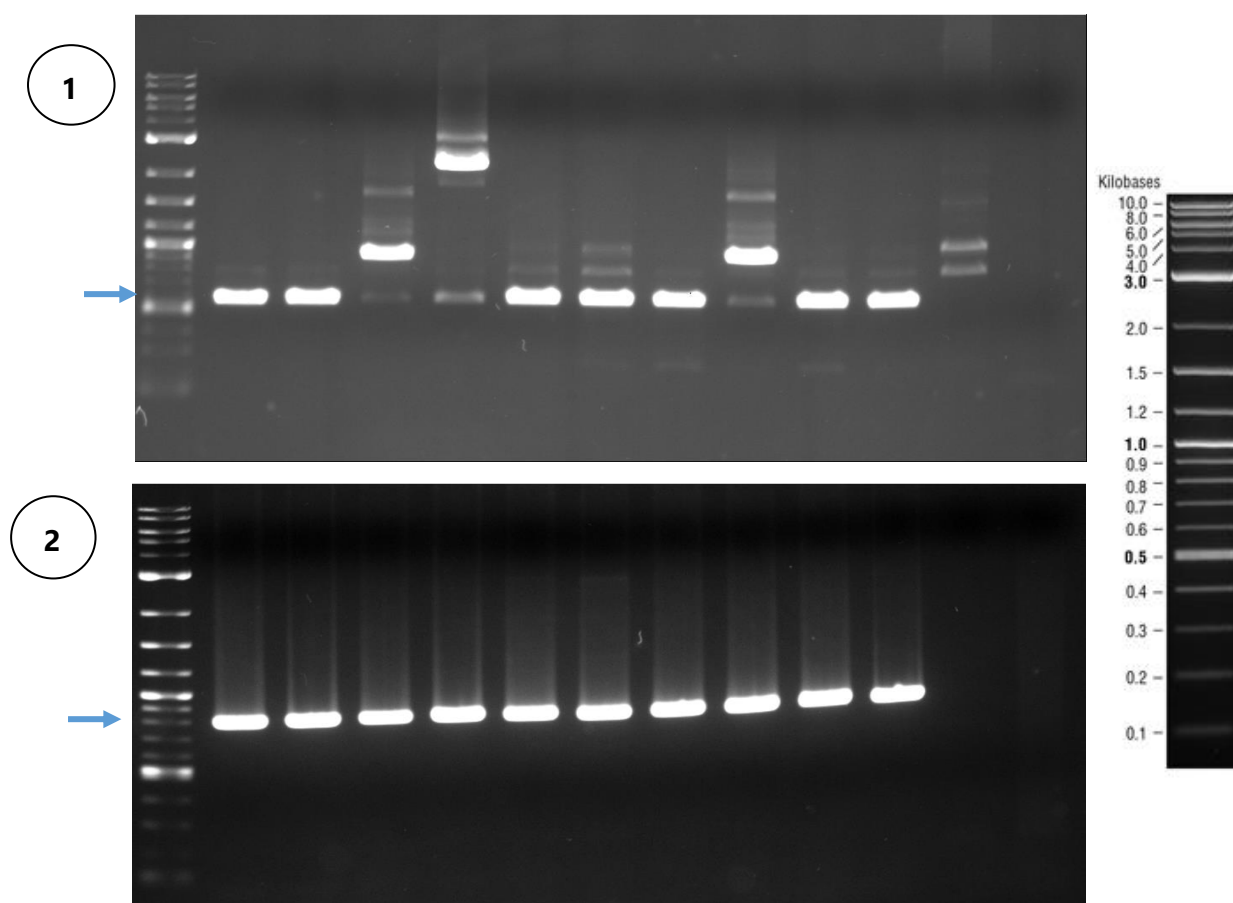


Figure 67: Amplifications PCR au locus du gène *dim5* chez la souche RutC30.

(1) Vérification de l'insertion correcte de la cassette d'inactivation à l'extrémités 5' (taille attendue de 579 bp indiquée par la flèche bleue) tandis que (2) a été réalisé à l'extrémité 3' (taille attendue de 788 bp indiquée par la flèche bleue). Les mutants ont été gardés dans le même ordre lors de la dépose sur gel.

Nous pouvons observer que le côté 3' de l'insertion est parfait pour tous les mutants mais que le côté 5' présente plusieurs profils de recombinaisons différents.

Le Tableau 26 ci-dessous récapitule le nombre de mutants validés pour chaque type de modification chez les deux souches. Bien qu'un nombre conséquent de transformants uninucléés aient été purifiés pour les mutants $\Delta kmt6$ dans les deux souches parentales, aucun clone n'a pu être validé génétiquement (Annexe XIX) à cause de l'impossibilité de valider les abords de l'insertion avec, soit la présence de nombreuses bandes soit une absence totale d'amplification. Après plusieurs tentatives infructueuses d'amplifications du locus entier de *kmt6* (jusqu'à 1kb de part et d'autre de l'insertion théorique pour une longueur totale de 5 kb d'amplicon) chez les mutants

de la QM6a et de la RutC30, nous avons déduit que l'hypothèse la plus probable était celle d'une insertion imparfaite de la cassette d'inactivation au locus ciblé avec remaniements génétiques et éventuellement une délétion seulement partielle du gène. Six tentatives de transformation et plus d'une centaine de clones ont été testés par PCR mais aucun résultat positif n'a pu être généré. Nous avons donc décidé de ne pas pousser plus loin notre étude afin de nous consacrer aux mutants des autres gènes.

Souches	Gènes invalidés	Transformants sélectionnés	Transformants uninucléés purifiés	Transformants validés par PCR	Taux de transformation
QM6a	<i>dim2</i>	41	13	7	17%
	<i>dim5</i>	48	35	23	48%
	<i>kmt6</i>	48	31	0	0%
	<i>hp1</i>	48	18	7	15%
	<i>rid</i>	42	28	19	45%
RutC30	<i>dim2</i>	48	26	8	17%
	<i>dim5</i>	48	31	16	33%
	<i>kmt6</i>	48	43	0	0%
	<i>hp1</i>	48	32	13	27%
	<i>rid</i>	48	40	20	42%

Tableau 26: Récapitulatif du nombre de transformants uninucléés validés par PCR et du taux de transformation final pour chaque souche.

Avec ces résultats, il apparaît finalement que les deux souches possèdent sensiblement le même taux de transformation validées alors que la RutC30 possédait un nombre au départ de mutants uninucléés purifiés plus important.

II. PRODUCTION ENZYMATIQUE DES MUTANTS VALIDES PAR PCR ET CHOIX DES CLONES ETUDIES

Après validation génétique par PCR, nous avons décidé de faire une sélection phénotypique des mutants afin d'identifier ceux que nous garderions pour la suite de l'étude. Ainsi, des cultures liquides sur milieu inducteur, contenant un mélange de lactose et cellulose, ont été réalisées sur les mutants durant sept jours afin de déterminer l'impact de la délétion de chaque gène sur la production de cellulases. Ce temps de culture long a pour but de limiter les variabilités liées à l'inoculum. La quantité d'enzymes présentes dans le milieu a été mesurée par dosage de Bradford et nous a permis de sélectionner les mutants les plus intéressants pour notre problématique.

Dans la suite de ce manuscrit nous ferons référence aux mutants par l'initiale de la souche parentale dont l'individu est issu, puis $\Delta gene$ pour la mutation qu'il a subie

et enfin son numéro d'identifiant. Par exemple, l'individu muté *QΔdim5* 39 provient de la souche sauvage QM6a, a subi la transformation sur le gène *dim5* et possède l'identifiant 39.

La Figure 68 rassemble l'ensemble des mesures de la concentration en protéines cellulolytiques dans les surnageants de culture des mutants issus de la souche QM6a. Sept clones indépendants ont été testés pour la mutation *Δdim2*, 17 pour *Δrid*, sept pour *Δhp1* et 12 pour *Δdim5*. Pour chacun d'eux, deux cultures indépendantes ont été effectuées ainsi que pour les souches parentales. Pour chacune des cultures, deux mesures ont été réalisées successivement et sont représentées par les barres d'erreur sur le graphique.

Concernant la production enzymatique de la souche parentale QM6a, une faible quantité de cellulases est sécrétée avec une concentration moyenne sur les quatre cultures à 0,12 g.L⁻¹. La production est homogène entre les cultures avec un écart-type à 0,01. Ces résultats sont conformes à ceux attendus pour une performance de la souche sauvage.

En comparaison avec la QM6a, peu de différences sont observées entre la production des transformants *QΔdim2* mais les productions des clones *QΔrid* et *QΔhp1* sont en moyenne plus faibles que celle de la souche parentale. Au sein des groupes de transformants *QΔdim2*, *QΔrid* et *QΔhp1* la production enzymatique est stable entre clones indépendants et peu de variabilités sont observées entre les cultures d'un même mutant.

Les mutants *QΔdim5*, quant à eux, affichent une différence plus marquée avec la production enzymatique de la QM6a. De plus, des variations plus marquées des concentrations en cellulases entre clones (0,030 g.L⁻¹ d'écart-type au sein du groupe) mais également entre cultures d'un même individu sont apparues (0,021 g.L⁻¹ d'écart-type entre réplicats d'un même clone). Les concentrations très faibles en enzymes dans le milieu rendent néanmoins difficile une conclusion sur l'hétérogénéité des résultats bien que la suppression du gène responsable de la marque H3K9me3 semble avoir entraîné une certaine hétérogénéité dans la production enzymatique en comparaison aux autres mutations.

Après test statistique de Wilcoxon sur les groupes de transformants, par gène, il ne s'avère néanmoins qu'aucun d'entre eux n'affiche une différence significative avec la QM6a dans la production enzymatique (qvalue>0,05).

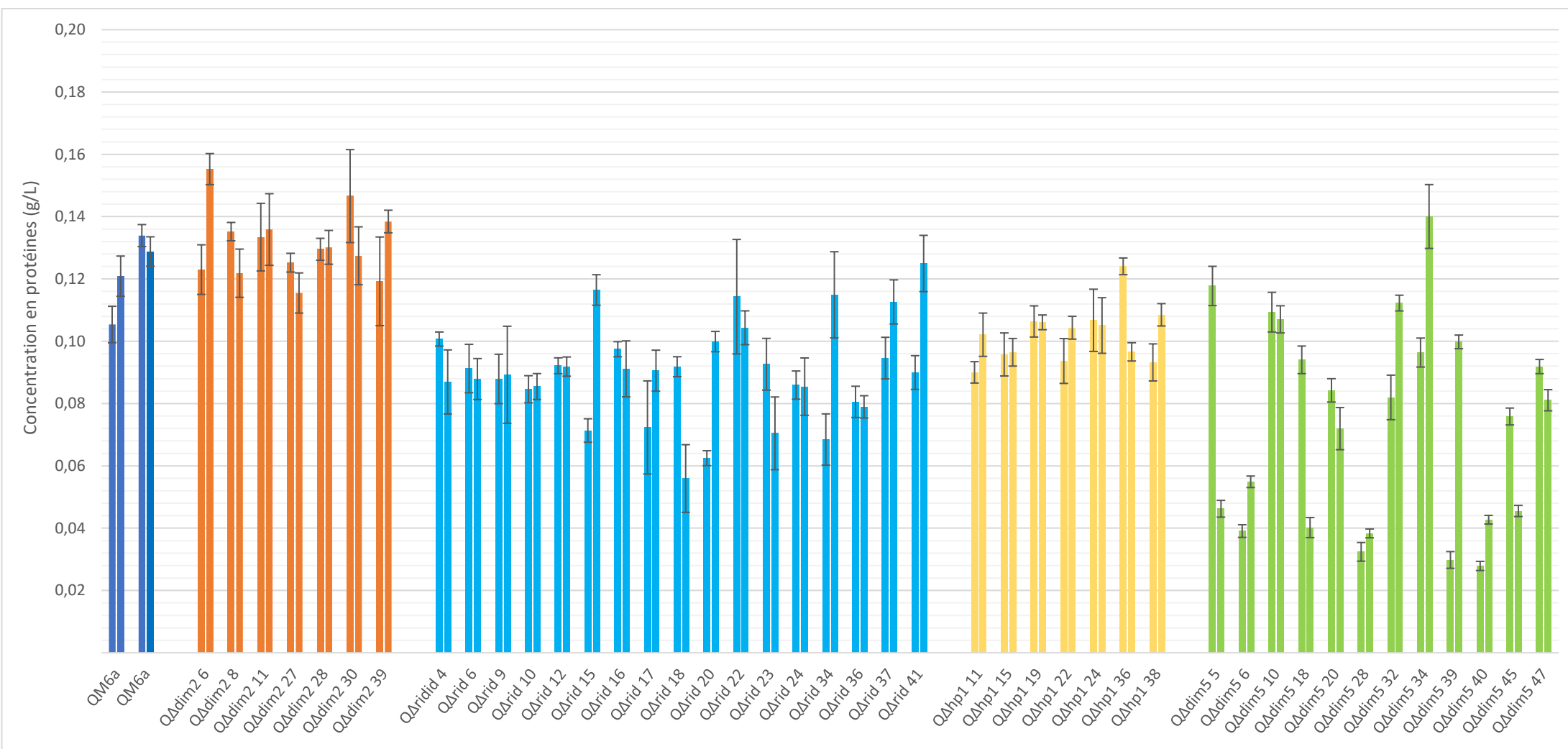


Figure 68: Production enzymatique chez les différents clones issus de la souche QM6a.

Les mesures de la concentration en protéines cellulolytiques dans les surnageants de culture des mutants issus de la souche QM6a ont été réalisées sur des clones indépendants pour chacun gènes délétés. Pour chacun des clones (axe des abscisses), deux cultures ont été réalisées et leur

concentration en protéines extracellulaires sont présentées en g.L^{-1} en axe des ordonnées. Pour chacune des cultures, deux mesures successives ont été réalisées afin de s'assurer de la précision de la mesure et ont servi à représenter les barres d'erreurs.

La Figure 69 rassemble comme précédemment les mutants issus de la souche parentale RutC30. Sept clones ont été testés pour la mutation *RΔdim2*, 18 pour *RΔrid*, 13 pour *Δhp1* et 15 pour *Δdim5*.

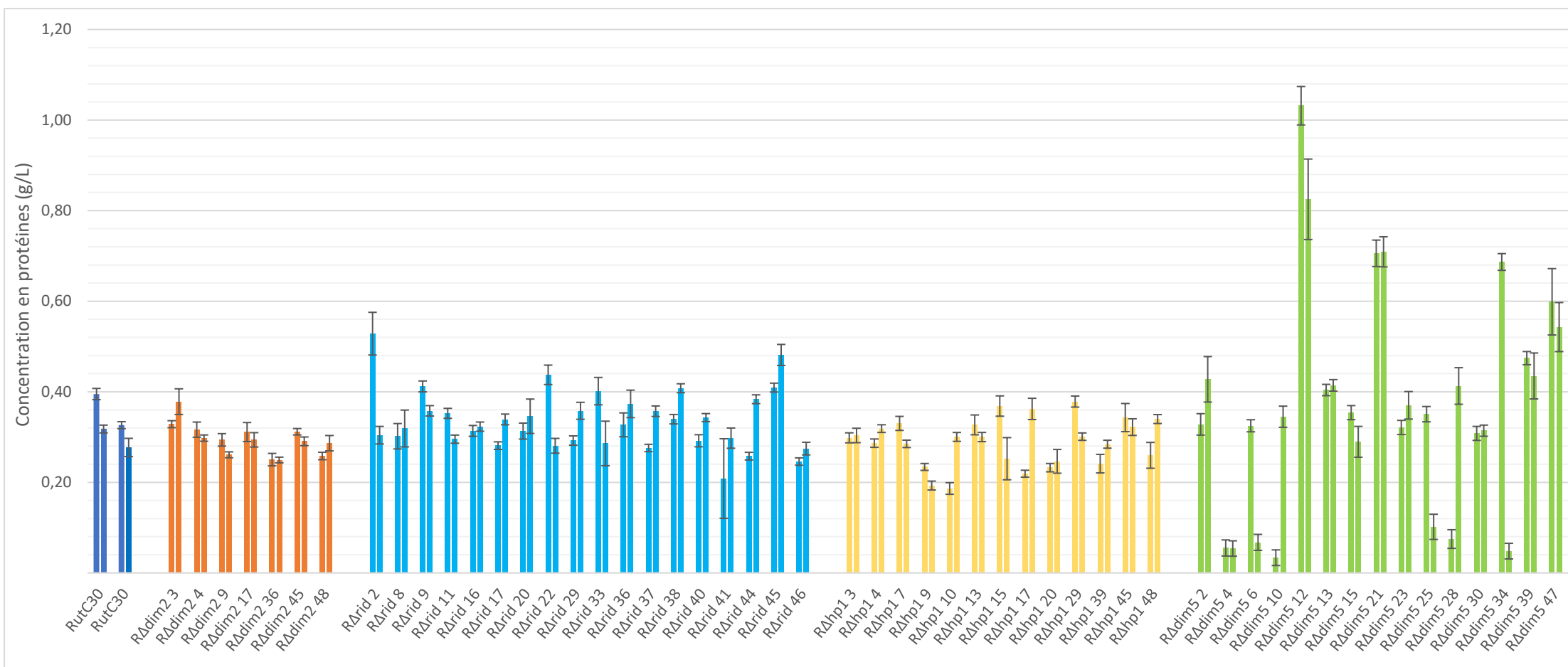


Figure 69: Production enzymatique chez les différents clones issus de la souche RutC30.

Les mesures de la concentration en protéines cellulolytiques dans les surnageants de culture des mutants issus de la souche RutC30 ont été réalisées sur des clones indépendants pour chacun gènes délétés. Pour chacun des clones (axe des absisses), deux cultures ont été réalisées et leur concentration en protéines extracellulaires sont présentées en g.L^{-1} en axe des ordonnées. Pour chacune des cultures, deux mesures successives ont été réalisées afin de s'assurer de la précision de la mesure et ont servi à représenter les barres d'erreurs.

Concernant la souche parentale RutC30, la production enzymatique moyenne est aux alentours de $0,33 \text{ g.L}^{-1}$, ce qui est une valeur attendue pour ce type d'expérience.

Au sein des groupes de mutants *RΔdim2*, *RΔrid* et *RΔhp1*, la production enzymatique entre clones est généralement stable ainsi qu'entre réplicats de cultures et n'affichent à première vue pas de différence significative avec la souche parentale RutC30.

D'un autre côté, les clones *RΔdim5* sécrètent des quantités d'enzymes très variables d'un individu à l'autre ($0,22 \text{ g.L}^{-1}$ d'écart-type entre clones) et dans certains cas des différences importantes entre cultures d'un même clone (en moyenne $0,11 \text{ g.L}^{-1}$ d'écart-type entre réplicats de culture). En plus de cela, les individus *RΔdim5 12*, *RΔdim5 21* et *RΔdim5 47* présentent des concentrations en protéines cellulolytiques bien plus importantes que celle de leur souche parentale.

L'hétérogénéité présentée par les mutants du gène *dim5* chez les deux souches QM6a et RutC30 suggère une dérégulation des mécanismes de production enzymatique chez ces clones tandis que l'apparition de plusieurs profils de production chez des transformants indépendants évoque le rôle central de ce gène pour la stabilité du phénotype de production des souches. Néanmoins, après un test statistique de Wilcoxon, il apparaît qu'en moyenne, aucun groupe de mutants ne produit de façon significativement différente de la RutC30 ($q\text{value} > 0,05$). En effet, chez les mutants *RΔdim5*, la production importante de certains transformants est contrebalancée par la production très faible des autres, résultants en une moyenne de production sur l'ensemble du groupe proche de celle de la souche mère. Les autres groupes de mutants sont visiblement très proches de la souche parentale, cela étant également confirmé par le test statistique.

À la suite de ces résultats, nous avons décidé de n'approfondir ici que l'étude des mutants *Δdim5*. Néanmoins, d'autres expériences ont également été menées sur les mutants *Δhp1* et sont actuellement sujettes à une publication en cours de rédaction.

Bien qu'en moyenne le groupe de mutants *RΔdim5* ne se différencie pas significativement de la RutC30, l'hétérogénéité présentée par les transformants témoigne d'un certain impact sur la synthèse de protéines cellulolytiques. La sélection des clones pour la suite de l'étude s'est faite sur les phénotypes de production enzymatique homogène/hétérogène entre réplicats techniques ainsi que sur la capacité d'hyperproduction. Ainsi, pour la souche QM6a, les mutants *QΔdim5 10*, *QΔdim5 20*, *QΔdim5 39* et *QΔdim5 45* ont été choisis, les deux premiers présentant des profils homogènes entre réplicats techniques tandis que les deux derniers possédant plus de variabilité. Pour la souche RutC30, les mutants *RΔdim5 12*, *RΔdim5 21*, *RΔdim5 28*, *RΔdim5 34* et *RΔdim5 39* ont été sélectionnés. Les clones *RΔdim5 12* et *RΔdim5 21* ont été choisis pour leur capacité accrue à produire des cellulases, *RΔdim5*

28 et *RΔdim5* 34 pour leur phénotype hétérogène et enfin *RΔdim5* 39 pour sa stabilité de production.

III. VERIFICATION GENETIQUE DES INDIVIDUS $\Delta DIM5$ ET DISPARITION DE LA MARQUE H3K9ME3

III.1. Vérification par Southern blot

Afin de vérifier qu'il n'existe pas de remaniements ectopiques aux alentours de l'insertion et si une autre copie de la cassette n'a pas été insérée ailleurs dans le génome, nous avons effectué une vérification par Southern blot sur les individus choisis parmi les mutants $\Delta dim5$ présentés auparavant.

Les enzymes de restrictions BamHI et EcoRI ont été choisies afin de permettre de distinguer la souche sauvage des souches mutantes. La Figure 70 ci-dessous représente les sites de coupures théoriques chez les souches sauvages et mutées ainsi que l'emplacement des sondes montrées ici en rouge sur la souche originale.

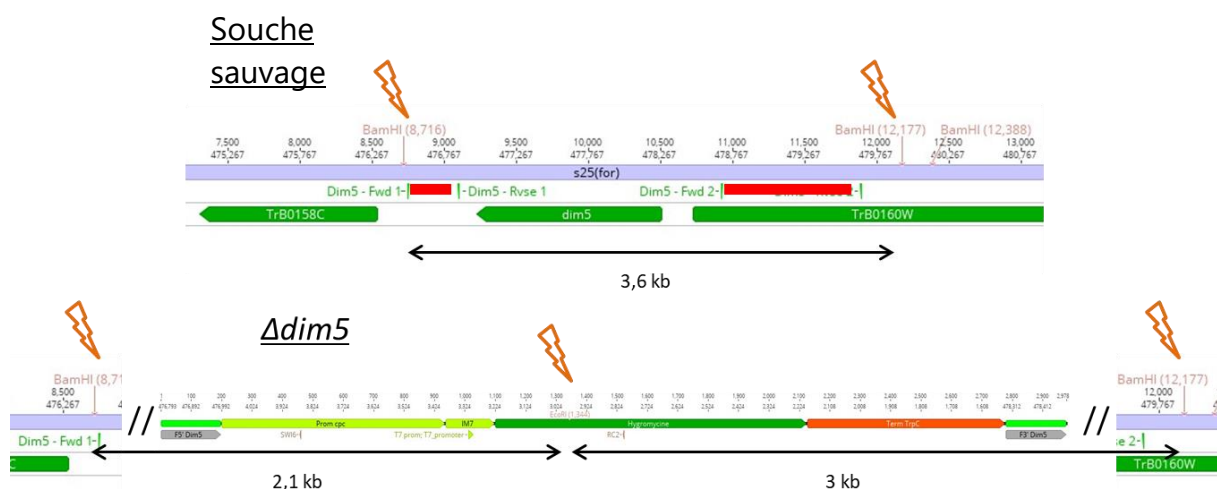


Figure 70: Sites de restriction choisis pour les mutants $\Delta dim5$.

Les enzymes utilisées sont BamHI et EcoRI. En haut, le profil de la souche sauvage comportant le gène *dim5* : une seule bande de 3,6 kb sera visible. En bas, la présence d'un site de restriction EcoRI dans la cassette d'hygromycine (à la place de *dim5*) permettra de révéler 2 bandes de 2,1 kb et 3 kb. Les rectangles rouges sur la souche sauvage représentent les sondes utilisées pour la révélation.

La Figure 71 ci-dessous présente les résultats obtenus : sur la membrane de gauche, la QM6a et les mutants qui en sont issus et sur celle de droite, la RutC30 et ses mutants. Sur la membrane QM6a, il apparaît clairement que le mutant *QΔdim5* 45 a

subi une insertion annexe à la transformation initiale, augmentant la taille du fragment inséré. Les résultats du mutant *QΔdim5 20* ne sont pas visibles à cause d'une bulle présente entre le gel d'agarose et la membrane au moment du transfert. Les mutants *QΔdim5 10* et *QΔdim5 39*, en revanche, affichent les tailles de bande attendues caractérisant une insertion propre sans évènement ectopique important.

Concernant la membrane RutC30, seul le mutant *RΔdim5 21* semble posséder une insertion tierce.

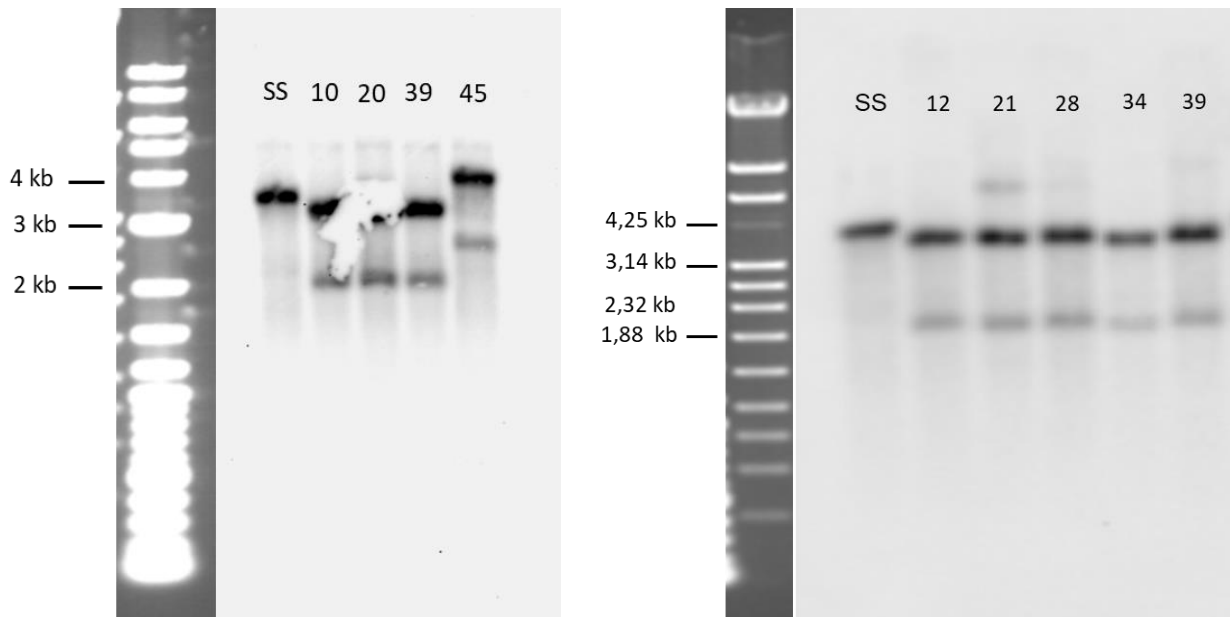


Figure 71: Révélation de la membrane de Southern blot pour les mutants $\Delta dim5$ de la souche QM6a (à gauche) et RutC30 (à droite).

La bande à 3,6 kb des souches sauvages (SS) est visible sur la partie gauche des deux gels, les mutants avec une bande à 3 kb et 2,1 kb sont sur la droite.

Ainsi, les mutants *QΔdim5 10*, *QΔdim5 39*, *RΔdim5 12*, *RΔdim5 28*, *RΔdim5 34* et *RΔdim5 39* ont été validés par Southern blot. Nous avons choisi d'éliminer les mutants *QΔdim5 10* et *RΔdim5 34* afin de ne conserver qu'un petit nombre de clones à étudier. Pour ce qui est du choix entre les clones *QΔdim5 10* et *QΔdim5 39* nous avons orienté notre choix sur le phénotype de production le plus disruptif par rapport à la souche parentale. Le phénotype de *RΔdim5 34* étant semblable à celui de *RΔdim5 28*, nous avons conservé uniquement ce dernier.

Finalement, nous conservons donc les mutants *QΔdim5 39*, *RΔdim5 12*, *RΔdim5 28* et *RΔdim5 39* pour la suite de notre étude.

III.2. Western blot

Afin de vérifier que la suppression du gène *dim5* ait bien entraîné la disparition de la marque H3K9me3, nous avons effectué un Western blot sur les cinq mutants sélectionnés ainsi que sur leur souche parentale (Figure 72).

Pour ce faire, nous avons extrait les protéines nucléaires totales des quatre mutants ainsi que des deux souches parentales. Nous avons ensuite effectué le Western blot en utilisant l'anticorps primaire contre la marque H3K9me3 et un anticorps secondaire couplé HRP (Péroxydase de raifort) pour la détection.

Les histones canoniques telle que l'histone H3 ont une masse moléculaire comprise entre 10 et 15 kDa environ (Bertin et Mangenot 2008). Nous nous attendions donc à voir apparaître une bande entre ces deux valeurs pour les souches QM6a et RutC30 mais pas pour les mutants.

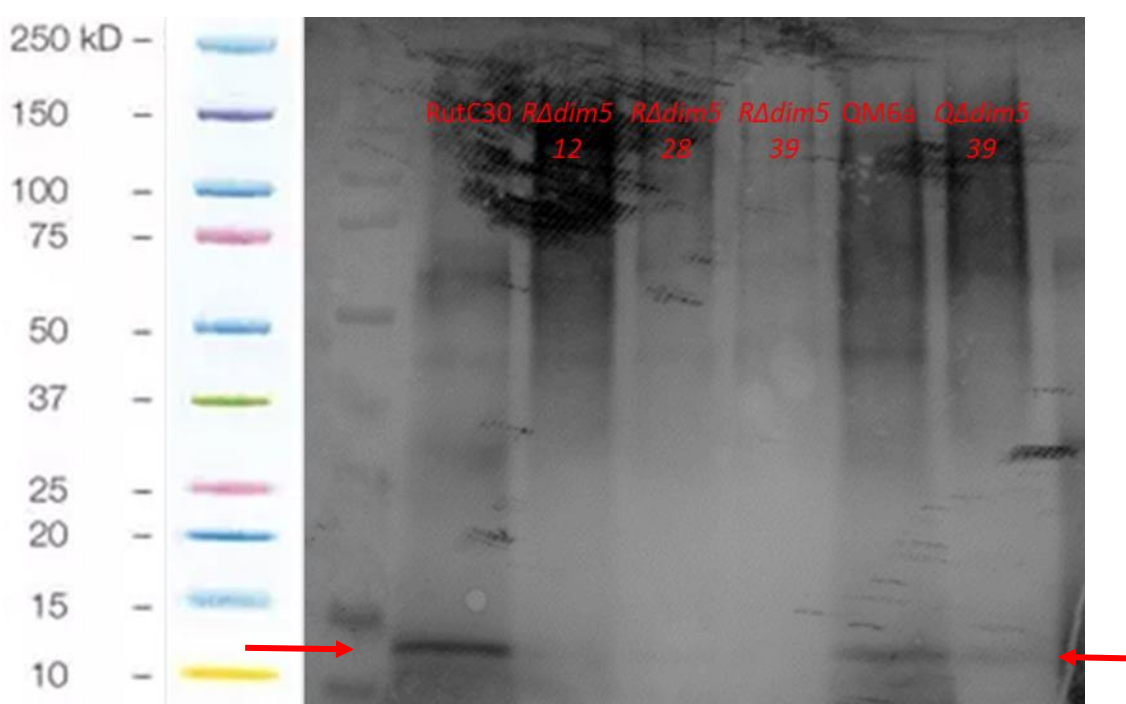


Figure 72: Western blot réalisé sur les souches parentales RutC30 et QM6a ainsi que sur leurs mutants respectifs.

Les flèches rouges indiquent la taille à laquelle nous attendions la bande de l'histone portant H3K9me3 chez les souches parentales.

Comme attendu, une bande située entre 10 et 15kDa est présente chez les souches parentales mais nous observons aussi une légère coloration au même niveau chez les mutants, en particulier chez *QΔdim5* 39 (Figure 72). La suppression du gène

dim5 entraîne donc bien en grande partie la suppression de H3K9me3 mais il semble néanmoins que des résidus soient toujours présents, en moindre quantité.

Une vérification de la présence de l'histone H3 avec une autre marque que H3K9me3 - H3K4me3 par exemple - n'ayant pas été touchée par la mutation aurait été une vérification précieuse en plus pour confirmer que l'absence de marque chez les mutants est bien due à la disparition de H3K9me3 et non pas à l'absence totale d'histone. Malheureusement, nous n'avons pas eu le temps de finaliser ces expériences.

III.3. Complémentation

Pour finir, nous avons procédé à la complémentation des quatre mutants afin de s'assurer que les caractéristiques sauvages pouvaient être récupérées par réinsertion du gène *dim5*. Cela permet de s'assurer que le phénotype observé chez le mutant est bien dû à l'invalidation du gène visé et non pas à une mutation ectopique tierce ou une dérégulation des gènes adjacents à l'insertion

Un plasmide contenant le gène *dim5* avec son promoteur et son terminateur ainsi que le gène conférant la résistance à la généticine a été utilisé pour une insertion ectopique du gène d'intérêt. Trois cycles de purification des transformants sur généticine ont été réalisés pour s'assurer de l'homogénéité génétique des individus. Trois milieux avec différentes sources de carbone sur lesquels les mutants $\Delta dim5$ possèdent un phénotype très marqué par rapport à leur souche parentale ont été utilisés afin de déterminer si les transformants ont été complémentés avec succès.

Après réinsertion ectopique du gène *dim5*, les phénotypes du transformant complémenté $Q\Delta dim5$ 39 (Figure 73) sur les trois milieux contenant du glucose, du lactose ou du sorbitol ont été entièrement récupérés et sont identiques à celui de la souche sauvage dans les mêmes conditions. Nous ne présentons ici qu'un seul transformant complémenté mais nous avons obtenu plus de trois clones aux phénotypes sauvages.

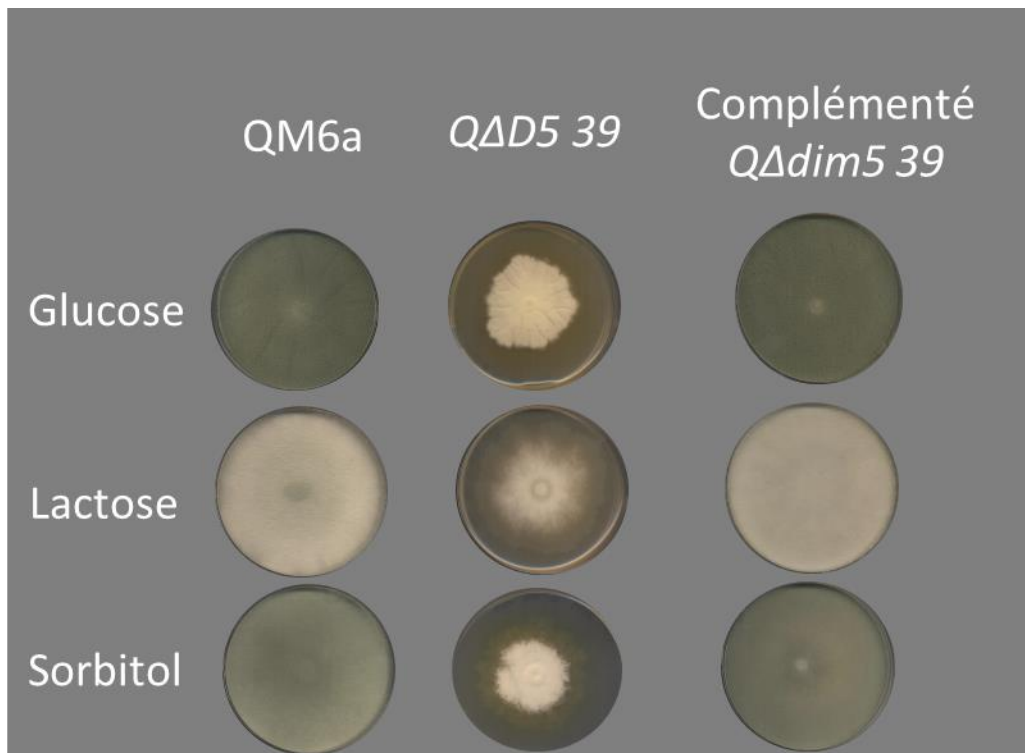


Figure 73: Complémentation du mutant *QΔdim5 39*.

Tests phénotypiques effectués sur milieu contenant comme source de carbone du glucose, du lactose ou du sorbitol.

Nous considérons donc que la complémentation du mutant *QΔdim5 39* est complète et que les phénotypes observés sont bien dus à la délétion du gène *dim5*.

Les tests de complémentation des mutant *RΔdim5* sont présentés en Figure 74

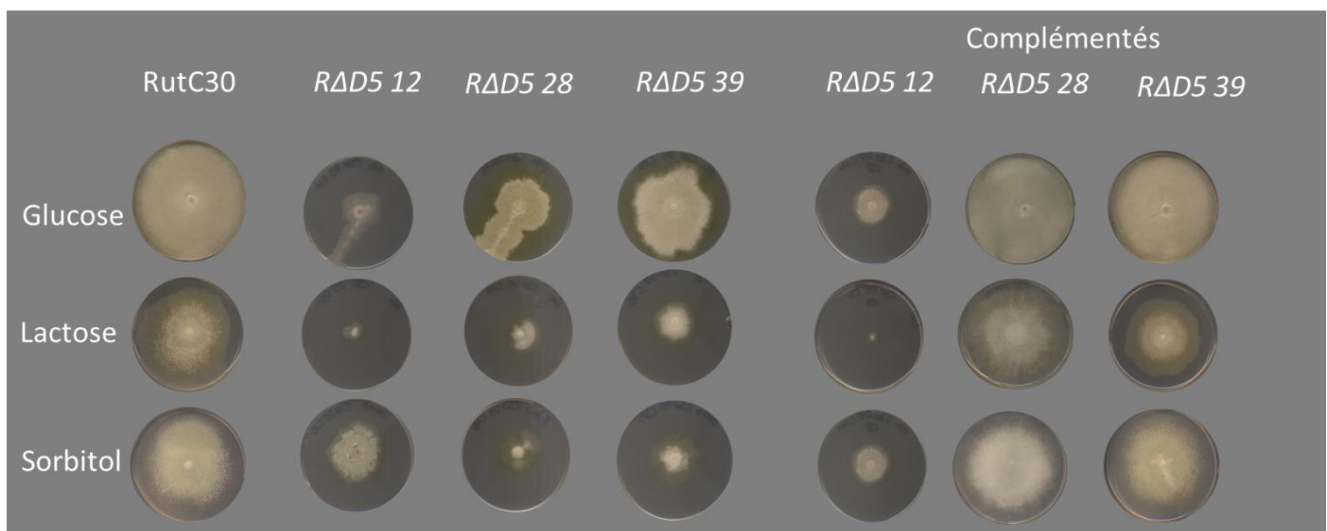


Figure 74: Complémentation des mutants *RΔdim5 12*, *RΔdim5 28* et *RΔdim5 39*.

Le mutant *RΔdim5 28* a pu être complémenté avec succès sur les trois milieux avec les trois sources de carbone différentes et nous avons été en mesure d'isoler au moins trois transformants complémentés. Le mutant *RΔdim5 39* a également pu être complémenté mais seulement un transformant possédant des phénotypes sauvages sur les trois milieux a pu être isolé. Nous avons d'autre part obtenu plusieurs mutants avec des phénotypes partiellement sauvages notamment avec de légères différences de morphologie sur le milieu sorbitol. En revanche, aucun transformant *RΔdim5 12* n'ont pu être complémentés. Comme sur la Figure 74, la soixantaine de clones testés issus des deux tentatives de complémentation arboraient le même phénotype que le mutant d'origine.

Après cette étape de complémentation des mutants *dim5*, la suite de la caractérisation se fera donc avec les trois mutants *QΔdim5 39*, *RΔdim5 28* et *RΔdim5 39*. Nous avons été dans l'obligation d'écarter le mutant *RΔdim5 12* après échec des complémentations.

III.4. Séquençage du génome du mutant *RΔdim5 12*

A la suite des problèmes de complémentation rencontrés pour le mutant *RΔdim5 12*, nous avons décidé de le faire séquencer afin de déterminer si une mutation ponctuelle non détectable par Southern blot était présente dans son génome et pouvait donc éclaircir les raisons pour lesquelles la complémentation n'avait pas fonctionné.

Il s'est avéré en effet qu'une substitution nucléotidique résultant en un changement d'un acide glutamique en acide aspartique avait eu lieu au niveau d'un gène à la fonction encore non définie (Figure 75). Par souci de confidentialité, puisque ce gène est potentiellement responsable d'une augmentation de la production d'enzymes chez ce mutant, nous le nommerons « X ». Ce mutant est donc écarté de notre étude puisque cette mutation est peut-être responsable du phénotype hyper-

producteur que nous observons. Des expériences complémentaires sont prévues afin de comprendre mieux l'hyper-productivité de ce mutant.

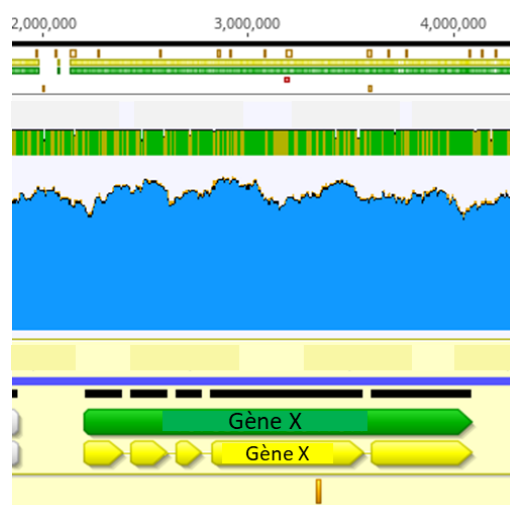


Figure 75 : Changement de la base G en base C au niveau du codon GAG du gène X du mutant *RΔdim5 12*.

IV. CARACTERISATION DES INDIVIDUS *ΔDIM5* RETENUS

IV.1. Production enzymatique des clones *Δdim5*

Après sept jours sur milieu liquide inducteur contenant un mélange de lactose et cellulose, nous avons mesuré la concentration en protéines sécrétées des mutants et des souches parentales grâce au dosage de Bradford. Ce temps long de culture assure une uniformisation de la croissance du champignon et de minimiser la variabilité due à l'inoculum.

Afin de caractériser plus finement la production enzymatique des mutants sélectionnés, nous avons mené deux analyses supplémentaires sur les surnageants de culture utilisés auparavant pour la quantification de la concentration en protéines cellulolytiques. Les activités enzymatiques de CBH1 et de la β -glucosidase CEL3A ont été respectivement mesurées par dosages de détection spécifique avec le pNPL (p-Nitrophenyl β -D-Lactopyranoside) et le pNPG (p-Nitrophenyl β -D-Glucopyranoside). L'activité spécifique correspond au nombre de nanomoles de p-nitrophénol (pNPL ou pNPG) hydrolysés par seconde et par milligramme de protéines présentes dans le surnageant, qui correspondent dans ces conditions chez *Trichoderma reesei* à plus de 90% d'enzymes cellulolytiques (Herpoël-Gimbert et al. 2008). De cette façon, nous obtenons une indication sur la composition du cocktail d'enzymes sécrétées, indépendamment de la production totale présente dans le milieu. Nous pouvons ainsi déterminer s'il existe des variations dans la composition du cocktail enzymatique d'un réplicat et d'un individu à l'autre.

Deux cultures ont été effectuées pour les souches parentales et quatre cultures pour les mutants. Une des cultures pour le mutant *RΔdim5 28* n'a pas pu être utilisée après un problème de croissance, nous ne présenterons donc que trois cultures pour ce mutant. Deux mesures ont été effectuées sur chacune des cultures et sont représentées par les barres d'erreur sur les graphiques de concentration en protéines et sur les graphiques d'activité enzymatique. Nous avons choisi de ne pas représenter des résultats moyennés des réplicats afin d'observer plus précisément les potentielles variations de productivité ou d'activité au sein d'un même clone.

La Figure 76 renseigne sur la concentration et les activités spécifiques des enzymes présentes dans le surnageant des cultures de la souche QM6a et du mutant *QΔdim5 39*. Chez la souche sauvage, la concentration en enzymes cellulolytiques est faible et il existe peu de différences entre les deux cultures. Les quantités en protéines extracellulaires mesurées ici correspondent à celles normalement attendues pour ce type d'expérience avec cette souche. Le mutant *QΔdim5 39* quant à lui, est un peu plus hétérogène sur sa production enzymatique. En effet, le réplicat 1 est nettement inférieur aux trois autres et affiche une production quasi nulle. Les trois autres réplicats sont légèrement moins productifs que la QM6a.

Les activités enzymatiques de la β -glucosidase CEL3A et de CBH1 chez la QM6a sont similaires entre les deux cultures. Nous notons que l'activité β -glucosidase est bien moins importante que celle de CBH1, ce qui est également attendu étant donné que cette dernière forme la majorité du cocktail enzymatique de *Trichoderma reesei*. Le mutant *QΔdim5 39* surprend par une activité β -glucosidase particulièrement divisée avec d'un côté les réplicats 1 et 2 aux alentours de 31 nkat.mg⁻¹ d'enzyme et de l'autre les réplicats 3 et 4 avec des activités inférieures à 5 nkat.mg⁻¹ d'enzyme. L'activité de CBH1 est hétérogène sans que l'on puisse distinguer un patron de profil.

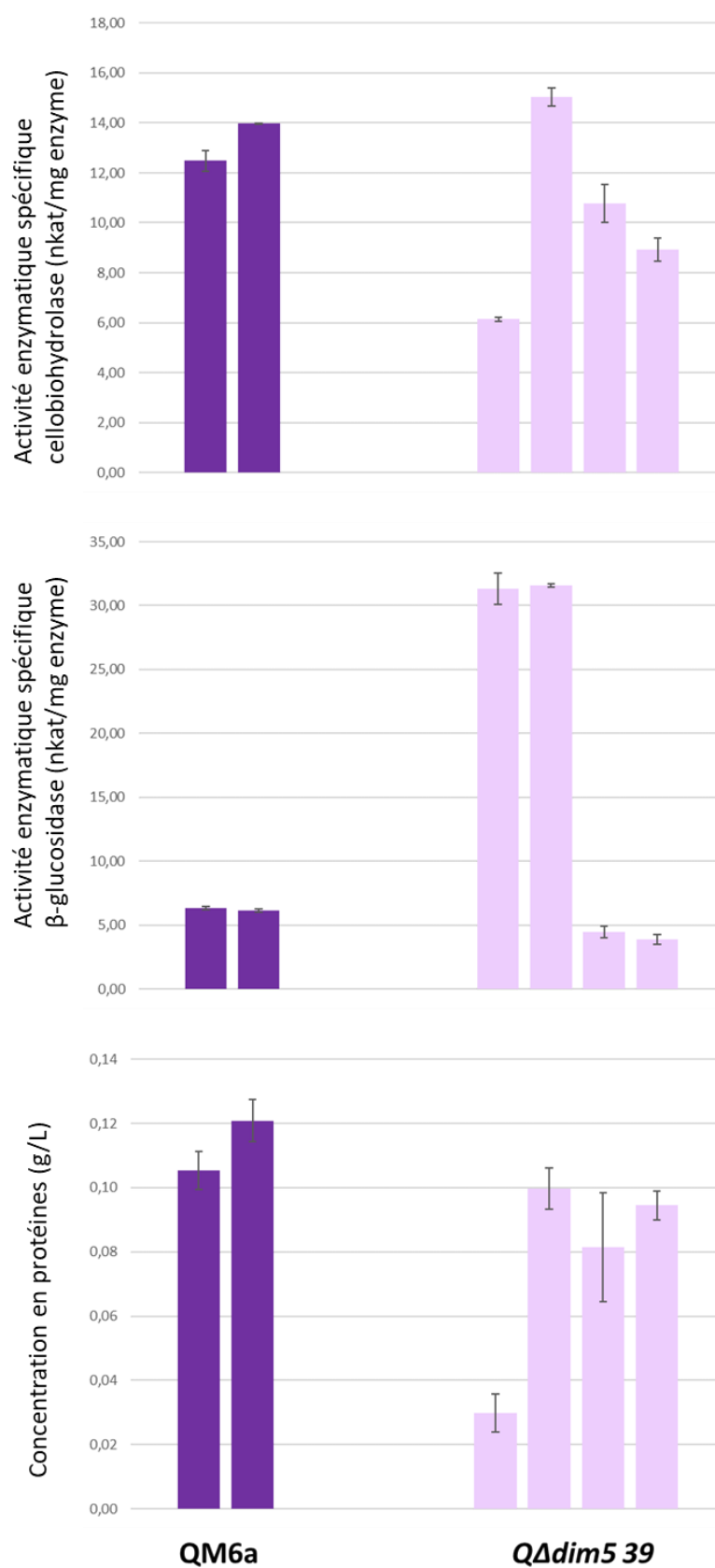


Figure 76: Résultats des dosages de quantité d'enzymes produites et de leurs activités enzymatiques spécifiques chez la souche parentale QM6a et le mutant QΔdim5 39.
 Les activités enzymatiques sont les activités mesurées pour 1 mg d'enzyme.

Les résultats de concentration en protéines et des activités spécifiques des enzymes cellulolytiques chez la souche RutC30 et ses deux mutants sont présentés en Figure 77. La production enzymatique de la RutC30 est légèrement différente entre les deux cultures (réplicat 1 à 0,39 g.L⁻¹ et réplicat 2 à 0,32 g.L⁻¹) mais reste dans les deux cas dans des valeurs attendues pour cette souche. Les réplicats du mutant *RΔdim5 28* sont très hétérogènes et présentes des productions globalement inférieures à celle de sa souche parentale. Le mutant *RΔdim5 39* est lui aussi très hétérogène dans ses résultats avec une production plus importante chez les réplicats 1 et 2 que chez les réplicats 3 et 4. Il est intéressant de noter que les cultures de la RutC30 et des réplicats 1 et 2 ont été effectués au même moment et dans le même incubateur tandis que les réplicats 3 et 4 ont été fait ensemble. Il s'agit donc peut-être ici d'une variabilité due à la manipulation.

L'activité β-glucosidase chez la souche RutC30 est similaire entre réplicats de culture et se trouve aux alentours de 4 knat.mg⁻¹ d'enzyme. A nouveau nous retrouvons un profil hétérogène chez le mutant *RΔdim5 28* avec une activité du réplicat 1 très différente des deux autres. L'activité de ce premier réplicat est au même niveau que ceux de la RutC30 tandis que les réplicats 2 et 3 sont nettement supérieurs. Pour le mutant *RΔdim5 39*, les quatre cultures ont des résultats d'activité relativement similaire aux alentours de 15 nkat.mg⁻¹ d'enzyme, presque 3 fois supérieur à l'activité présente chez la RutC30.

Concernant l'activité de CBH1, la RutC30 possède une activité similaire entre ses deux cultures, en moyenne à 24 nkat.mg⁻¹ d'enzyme. Encore une fois l'activité du mutant *RΔdim5 28* est très hétérogène entre ses trois réplicats tandis que le mutant *RΔdim5 39* affiche des profils d'activités très marqués entre les réplicats 1 et 2 contre réplicats 3 et 4.

Globalement, une forte hétérogénéité des résultats est visible chez les trois mutants. Il semble donc que la suppression du gène *dim5* ait entraîné une instabilité phénotypique avec des modifications sur les concentrations d'enzymes extracellulaires mais aussi sur les proportions de celles-ci dans le cocktail puisque les activités spécifiques varient fortement d'une culture à l'autre. Nous notons que pour les deux mutants de la souche RutC30, l'activité β-glucosidase est majoritairement augmentée.

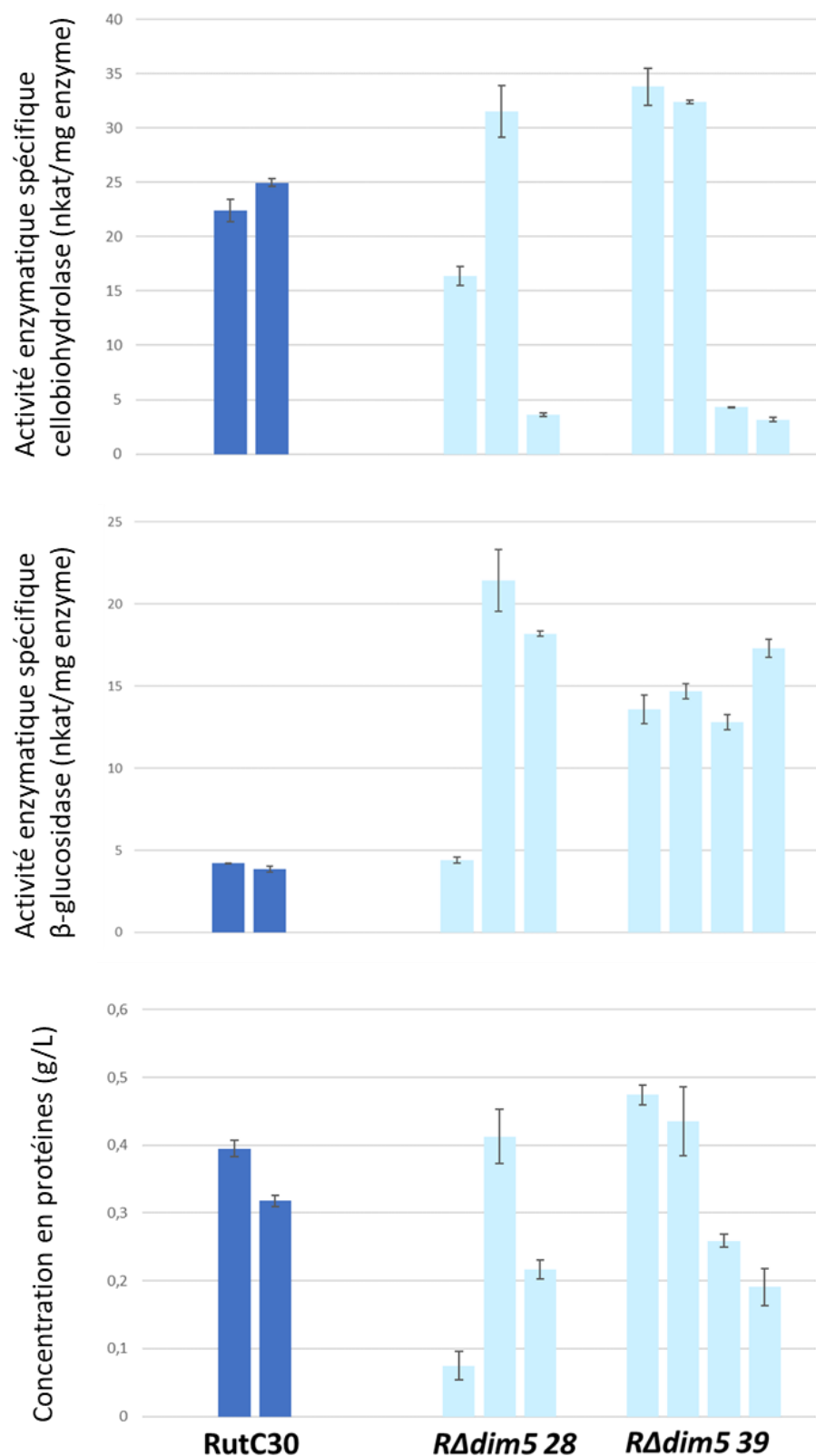


Figure 77: Résultats des dosages de quantité d'enzymes produites et de leurs activités enzymatiques spécifiques chez la souche parentale RutC30 et les mutants RΔdim5 28 et RΔdim5 39.

Les activités enzymatiques sont les activités mesurées pour 1 mg d'enzyme.

IV.2. Phénotypage sur milieux solides

À la suite des mesures d'activités enzymatiques effectuées sur les surnageants de cultures liquides, nous avons procédé à une étude du phénotype des trois mutants après sept jours d'incubation sur différents milieux solides. Un test sur quatre milieux minimums différant par leur source de carbone (lactose, glucose et sorbitol) a permis d'évaluer la capacité des mutants à pousser sur celles-ci, tandis que des essais sur milieu oxydatif (ménadione 30 μ M) et sur milieu générant une pression osmotique (NaCl 0.5M) ont permis d'observer leur croissance face à ces stress. Enfin, des tests sur milieu riche mais en conditions variables de lumière et de température ont permis d'observer la réceptivité des mutants aux stimuli extérieurs.

Outres les expériences d'exposition lumineuse, l'ensemble des boîtes ont été conservées à l'obscurité pendant les tests afin de minimiser les facteurs extérieurs à ceux que nous souhaitons mesurer. Également, toutes les expériences se sont déroulées à 30°C sauf celles étudiant l'impact des températures. Les boîtes de Pétri utilisées font 9 cm de diamètre.

Ainsi, dans la Figure 78 ci-dessous nous renseignons les résultats obtenus pour les différentes sources de carbone et pour les deux stress. Un impact général peut se voir sur l'ensemble des mutants avec des croissances et des morphologies variées propres à chaque individu mais il est également intéressant de noter l'éventail de textures de mycéliums que génèrent les différents milieux chez une même souche. Les deux réplicats techniques effectués pour chaque milieu et chaque mutant sont disponibles en Annexe XX.

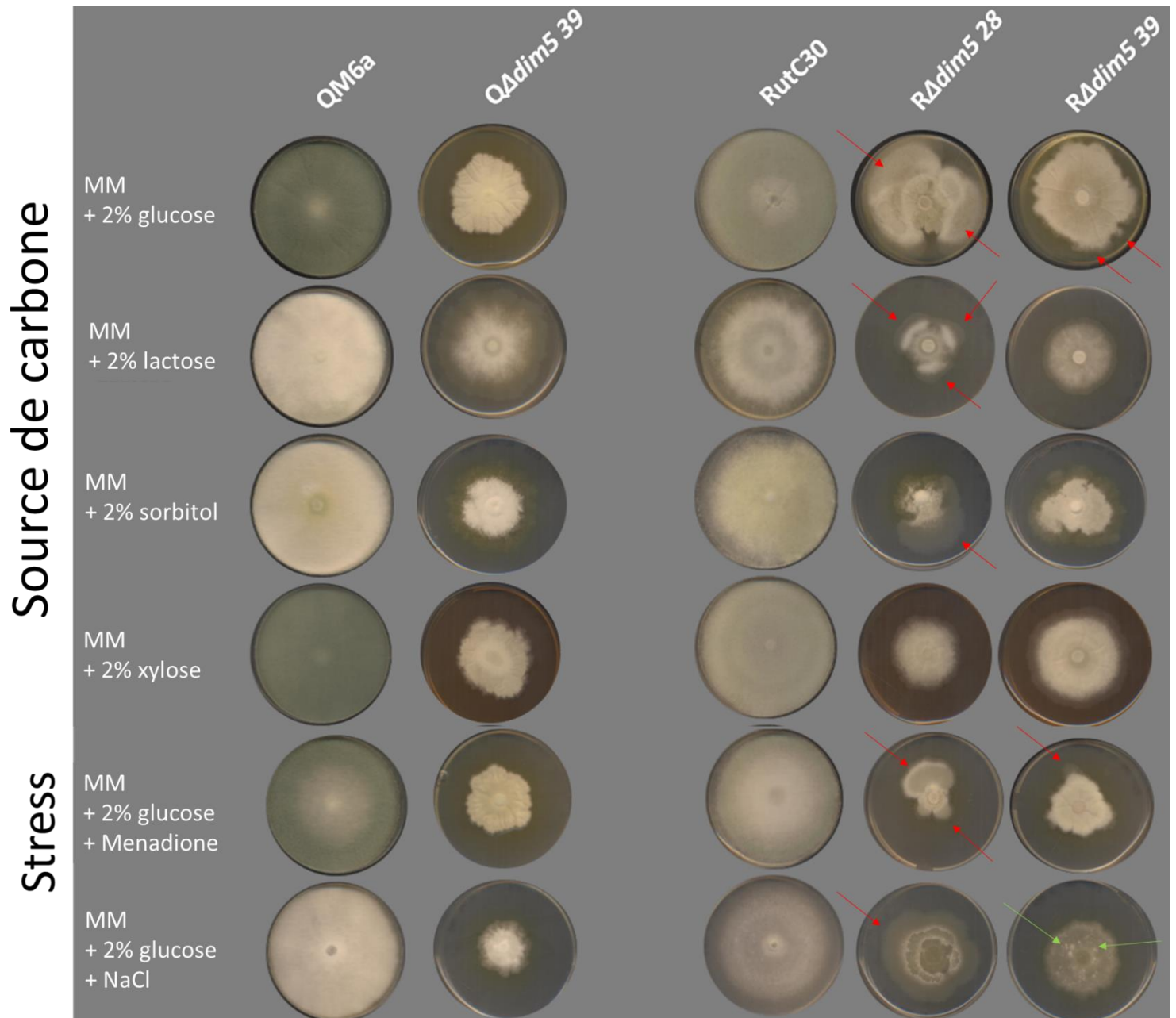


Figure 78 : Résultats des phénotypages de l'ensemble des mutants $\Delta dim5$ sur milieux solides contenant diverses sources de carbone et sur milieux provoquant un stress oxydatif (ménadione) et un stress osmotique (NaCl).

Les flèches rouges indiquent les excroissances apparaissant chez certains mutants. Les flèches vertes indiquent des points de concentration de sporulation plus importante.

Dans les conditions testant les différentes sources de carbone, la souche parentale QM6a a envahi entièrement les boîtes de Petri. Le mycélium est fin et recouvert d'une importante couche de conidies. Sur milieux contenant du glucose ou du xylose le champignon est vert alors que sur milieux contenant du lactose ou du sorbitol il apparaît comme blanc.

Le mutant *QΔdim5 39* arbore quant à lui des phénotypes plus variables sur ces substrats. En effet, nous notons un déficit de croissance important sur les milieux contenant du glucose, du sorbitol et du xylose (3 à 4 cm de diamètre environ). Sur milieu contenant du lactose ce déficit reste léger (8 cm de diamètre) en comparaison avec la QM6a. Les textures de mycélium sont aussi très différentes d'un milieu à l'autre. Sur milieux contenant du glucose et du xylose, il apparaît comme caoutchouteux et ses contours sont légèrement dentelés alors que sur milieux lactose et sorbitol le mycélium est plus fin, aérien et légèrement translucide sur la partie externe. Également sur ces deux milieux, la conidiation est centrale, dense et bien délimitée alors qu'elle est très faible sur les deux autres.

Sur l'ensemble des quatre milieux présentés en Figure 78, la souche parentale RutC30 a entièrement envahi les boîtes et possède une conidiation dense. La coloration est moins marquée que chez la QM6a et est verdâtre sur tous les milieux.

Chez les mutants RutC30, *RΔdim5 28* se caractérise principalement par un déficit de croissance important sur la partie principale du mycélium et l'apparition d'excroissances anarchiques envahissantes sur les parties périphériques du champignon (indiquées par les flèches rouges sur la figure). Ces bourgeonnements possèdent une forte conidiation sur les milieux avec du glucose et du lactose tandis qu'ils en sont complètement dénués sur milieu contenant du sorbitol. Sur milieu contenant du xylose, *RΔdim5 28* possède une morphologie similaire à celle de sa souche parentale mais avec une croissance amoindrie. D'un autre côté, *RΔdim5 39* possède le même phénotype de départ que *RΔdim5 28* avec un déficit de croissance un peu moins important mais surtout sans les excroissances.

Ensuite, deux milieux induisant un stress pour le champignon ont été étudiés : un milieu contenant de la ménadione et un milieu contenant du NaCl, provoquant respectivement un stress oxydatif par génération de radicaux hydroxyle $\cdot\text{OH}$ dans le réticulum endoplasmique et un stress osmotique par concentration importante de sel dans le milieu (Figure 78).

La souche sauvage QM6a ne présente pas de déficit de croissance ni de morphologie particulière face aux stress induits, mis à part un gradient de coloration sur milieu contenant de la ménadione. Sur ce même milieu, le mutant *QΔdim5 39* présente une perte de croissance importante avec un diamètre d'environ 3 cm. Également, le mycélium est plus épais et dense que celui de sa souche parentale et n'arbore pas de conidiation. Nous remarquons que le phénotype exprimé ici est le même que celui observé sur milieu avec du glucose. Il semble donc que la ménadione n'a aucun effet additionnel sur *QΔdim5 39* que le milieu de base.

Sur milieu contenant du NaCl, *QΔdim5 39* est également déficitaire au niveau de sa croissance (environ 2 cm de diamètre) et le mycélium n'est pas identique sur toute la surface. En son centre le champignon est plus dense et épais tandis que le réseau d'hyphe se fait plus fin et translucide sur les bords. Une conidiation centrale blanche

est présente à l'instar de celle de la souche QM6a. Le stress infligé par la présence de NaCl dans le milieu semble plus impactant sur le mutant que ne l'est la ménadione puisque le phénotype sur ce milieu n'est pas le même que celui observé sur le milieu de base avec 2% de glucose.

La souche parentale RutC30 ne présente pas de déficit de croissance sur les deux milieux stressants. En condition contenant du ménadione une légère coloration verte des conidies est visible sur les bords de la boîte alors que le centre reste blanc. Sur milieu contenant du NaCl, il semble que la conidiation c'est fait concentriquement et qu'un cercle plus dense est présent au milieu de la boîte. Les conidies sont restées blanches.

Sur milieu avec de la ménadione, les deux mutants *RΔdim5 28* et *RΔdim5 39* possèdent les mêmes phénotypes que sur milieu avec du glucose avec simplement une croissance moins importante. Nous retrouvons donc, comme chez la QM6a, un impact mineur de la ménadione sur le champignon.

Sur milieu NaCl, les deux mutants possèdent un mycélium très fin et translucide avec une croissance amoindrie en comparaison avec la RutC30. De nouveau, *RΔdim5 28* se caractérise par l'apparition d'un bourgeonnement qui envahit la boîte tandis que *RΔdim5 39* n'en possède quasiment pas. En revanche, ce dernier possède des points localisés de forte densité en conidies (indiqués par des flèches vertes sur la figure) qui ne sont pas présents chez l'autre mutant.

Le reste des phénotypes a été effectué sur milieu riche PDA et en conditions environnementales variables. Ainsi, nous présentons en Figure 79 les boîtes ayant été exposées à une lumière constante ou bien à l'obscurité totale durant sept jours.

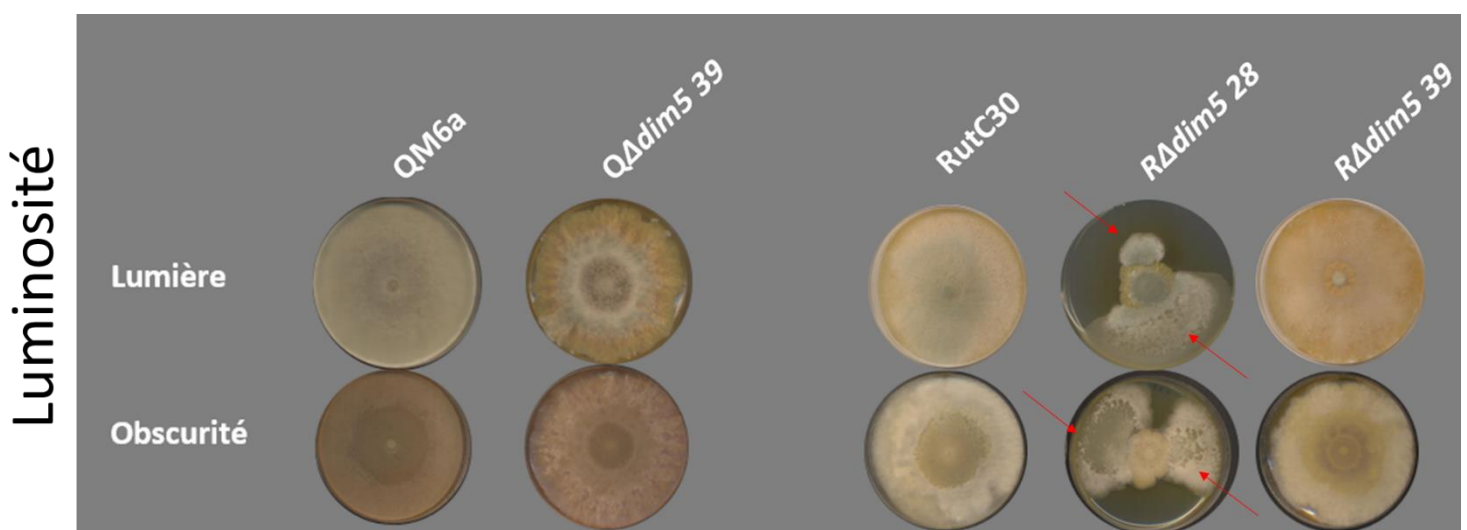


Figure 79 : Résultats des phénotypes sur milieu PDA avec ou sans exposition à la lumière.
Les flèches rouges indiquent les excroissances apparaissant chez certains mutants.

Comme attendue, la souche sauvage QM6a présente une conidiation légèrement colorée lorsque exposée à la lumière. Le duvet de sporulation est également plus épais et homogène que lorsque la souche est gardée à l'obscurité. Un disque exempt de spores est notamment visible au centre de la boîte gardée dans le noir. Concernant le mutant QM6a en condition lumineuse, nous notons l'apparition d'un anneau de sporulation vert et dense dans le centre de la boîte puis une disparition totale des conidies. Dans l'obscurité, le mutant a également développé un anneau de conidiation central plus dense et incolore qui se diffuse vers le reste de la boîte. Quasiment aucun déficit de croissance n'est visible, néanmoins le mycélium apparaît plus dense que celui de la souche mère.

Concernant la souche RutC30, en condition lumineuse nous notons la présence d'une sporulation dense et homogène sur l'ensemble du milieu bien que la coloration verte soit répartie de façon inégale sur la boîte. Dans l'obscurité, nous pouvons observer une absence de conidiation dans la région centrale de la boîte tandis que les pourtours sont recouverts d'une couche importante de spores.

Le mutant *RΔdim5 39* possède des phénotypes similaires dans les deux conditions présentées et ne possède que peu de différence avec sa souche parentale. Un léger déficit de croissance est visible en condition d'obscurité et la légère coloration verte présente sur la RutC30 n'est pas présente.

D'un autre côté, dans les deux conditions, le mutant *RΔdim5 28* affiche dans un premier temps un fort ralentissement de croissance avec une sporulation très centralisée. Le mycélium est très dense avec des bordures dentelées arrondies. Puis, encore une fois, des poussées anarchiques hypertrophiques localisées sur une partie du mycélium se déclarent. Nous notons que celles-ci sont recouvertes d'un épais duvet de sporulation et que contrairement à la partie centrale, le mycélium semble se rapprocher de la texture de la souche mère.

Ainsi, dans ces conditions, les deux mutants RutC30 n'affichent pas du tout le même phénotype et ne possèdent que peu de points communs contrairement aux conditions étudiées auparavant.

Pour finir, la Figure 80 présente les résultats obtenus après une exposition à différentes températures sur milieu riche PDA. Nous rappelons que la température de croissance optimale du champignon est de 30°C.

La souche sauvage QM6a n'est pas affectée par une température de 34°C mais montre une plus faible croissance à 37°C tout en conservant une importante conidiation. A 30°C et 34°C, une absence de conidiation au centre des boîtes est observable. En comparaison, le mutant *QΔdim5 39* est impacté dès 34°C, avec un léger retard de développement et un pourtour en dents de scie. L'apparition d'un bourgeonnement de mycélium comme observé auparavant chez les mutants de la RutC30 est également clairement visible. A 30°C et 34°C, le mutant est couvert d'un

duvet de sporulation dense. En revanche, à 37°C *QΔdim5 39* est chétif, le réseau mycélien est fin et la conidiation est sporadique et inégale.

La souche RutC30 pousse sur l'ensemble de la boîte à 30°C et arbore une conidiation dense sur la totalité du mycélium. Un déficit de croissance est observable à 34°C mais la conidiation est maintenue. En revanche, un déficit de croissance plus important à 37°C est visible avec un diamètre du champignon d'à peine 2 cm et aucune sporulation n'est présente à cette température.

Sur milieu à 30°C, *RΔdim5 28* a encore une fois très peu poussé et possède un mycélium dense et caoutchouteux puis, l'apparition d'une poussée hypertrophique envahi la quasi-totalité de la boîte. D'un autre côté, *RΔdim5 39* est moins impacté et ne diffère de la souche sauvage que par une conidiation moins importante et un mycélium légèrement plus dense. Ainsi, sur cette condition exempte de stress et sur milieu riche, les deux mutants arborent des caractéristiques très différentes alors que sur milieux à 34°C et 37°C, les deux champignons se rejoignent vers un phénotype commun.

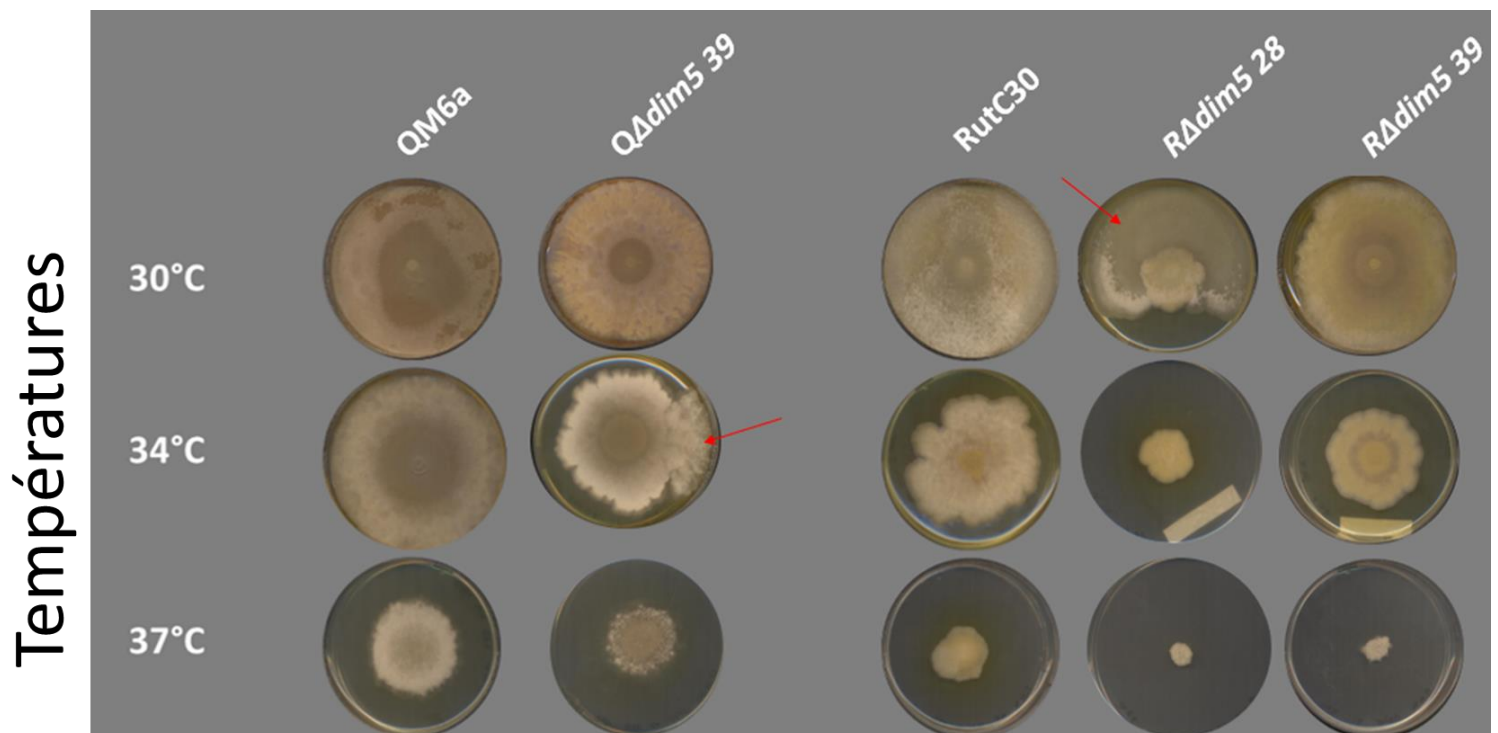


Figure 80 : Résultats des phénotypes sur milieu PDA exposés à différentes températures.

Les flèches rouges indiquent les excroissances apparaissant chez certains mutants.

Enfin, nous notons que les mutants répondent différemment aux stimuli appliqués et génèrent des phénotypes variables en fonction de la souche parentale dont ils ont été tirés.

Ainsi, *QΔdim5 39* génère des phénotypes altérés par rapport à la QM6a mais présente également de nombreuses similitudes avec la souche sauvage. De fait, les différences entre souche parentale et mutant apparaissent lorsque l'environnement n'est pas optimum ou génère un stress (milieu minimum, stress par pression osmotique ou stress de température).

De la même façon, les mutants de la RutC30 possèdent des similitudes de base lorsque l'environnement n'est pas optimum avec néanmoins la présence d'instabilités phénotypiques chez *RΔdim5 28* qui le démarque de *RΔdim5 39*. Bien que possédant le même génotype ces deux mutants arborent donc des variations phénotypiques importantes caractérisées par l'apparition d'instabilité chez l'un deux.

IV.3. Stabilité sur le temps

Au vu des expériences de phénotypage présentées auparavant et de l'hétérogénéité entre mutants, nous avons pensé que la délétion de *dim5* avait provoqué chez les clones un remodelage du paysage épigénétique non homogène, en particulier chez les mutants de la RutC30, résultants en des « épimutations ». Les différences observées entre *RΔdim5 28* et *RΔdim5 39* seraient ainsi dues à des différences de redistribution des marques telles que par exemple eH3K27me3, H3K4me3 ou H3K36me3 le long du génome à la suite de l'élimination du gène. Ceci pourrait également expliquer l'instabilité de *RΔdim5 28*. Nous avons donc émis l'hypothèse qu'après un temps d'adaptation, cette distribution s'homogénéiserait chez les mutants et les phénotypes seraient alors semblables. Concernant *QΔdim5 39* nous nous attendions à peu de variations entre le début et la fin de l'expérience étant donné les résultats relativement homogènes pour trois réplicats sur quatre lors des mesures de concentration enzymatique présentés auparavant. De même, la présence d'un seul événement de bourgeonnement hétérogène sur les expériences en milieux solides indique une certaine stabilité de la souche en comparaison des autres.

Afin de tester cette hypothèse nous avons ensemencé les cinq individus sur milieux liquides riches contenant différentes combinaisons de sources de carbone :

- Lactose + Cellulose : L + C
- Lactose : L
- Glucose + Cellulose : G + C
- Glucose : G

Nous avons choisi d'étudier le lactose et le glucose car nous avons observé que l'impact de ces sucres est fort sur les phénotypes d'hétérogénéité des mutants dans les expériences sur milieu solide. Nous voulions ainsi les placer dans des environnements sur lesquels nous pourrions induire ces caractéristiques. Ensuite, l'ajout de la cellulose

induit la production de cellulases, même sur milieu glucose, et permet donc de suivre la production enzymatique.

Tous les sept jours durant cinq semaines, nous avons repiqué les individus sur milieu frais. Nous avons réalisé deux cultures pour chaque individu et chaque milieu. Les mesures de concentration en enzymes présentes dans le surnageant ont été effectuées après une semaine de culture et après la cinquième semaine de repiquages consécutifs grâce à un dosage de Bradford. Pour chaque culture, deux mesures de la concentration ont été réalisées.

Chez la souche QM6a et le mutant *QΔdim5 39* (Figure 81) après une semaine d'expérience, les concentrations en enzymes cellulolytiques des surnageants sont très faibles sur les quatre milieux. Les concentrations sont globalement équivalentes entre souche sauvage et mutant, avec une valeur maximale atteinte par la QM6a à 0,08 g.L⁻¹. Nous notons qu'en milieu contenant du glucose la production enzymatique est, comme attendue, moins importante que sur les milieux contenant du lactose ou du glucose et de la cellulose.

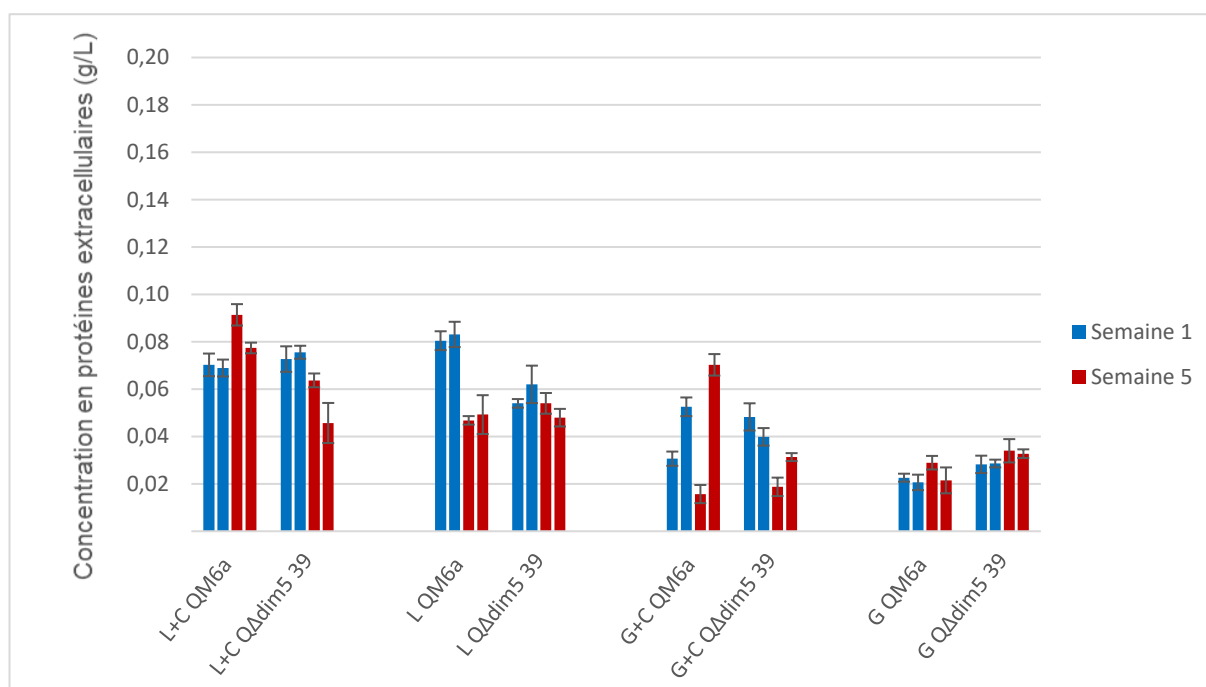


Figure 81 : Concentration en protéines extracellulaires dans les surnageants de culture de la souche QM6a et du mutant QΔdim5 39.

Pour chaque culture, deux mesures de la concentration ont été réalisées et sont utilisées pour représenter l'écart-type. La première semaine de repiquage sur milieux frais est représentée en bleu et la cinquième semaine en rouge.

Après cinq semaines de repiquages consécutifs, les concentrations enzymatiques de la QM6a sur les deux milieux L+C et G sont restées égales à la première semaine d'expérience. En revanche, sur milieux L et G+C la production a varié avec une concentration enzymatique diminuée presque de moitié sur milieu L et avec deux cultures aux productions très différentes sur milieu G+C. Pour le mutant, nous retrouvons des concentrations en protéines extracellulaires identiques à celles de la première semaine sur les milieux L et G mais des productions qui apparaissent légèrement diminuées sur milieux L+C et G+C. Cependant, il est important de ne pas oublier que dans la gamme de concentration enzymatique mesurée ici, les quantités étant très faibles, des différences minimales peuvent faire rapidement varier la production du simple au double. Avec une différence maximale de $0,03 \text{ g.L}^{-1}$ nous considérons donc que ces dernières productions sont restées stables à travers le temps.

En outre, il apparaît que la production du mutant varie peu au cours du temps comme nous l'attendions mais il est néanmoins difficile de pousser plus loin les conclusions étant donné les faibles quantités sécrétées et l'instabilité des résultats obtenus pour la QM6a.

Chez la souche RutC30 et ses deux mutants (Figure 82), la première semaine d'expérience nous donne des concentrations en protéines extracellulaires très diverses. En effet, si la quantité d'enzymes produites par la RutC30 est relativement stable sur tous les milieux, les mutants affichent dans certains cas des différences importantes ($0,13 \text{ g.L}^{-1}$ de différence entre les deux réplicats de culture du mutant *RΔdim5 28* sur milieu L+C et $0,20 \text{ g.L}^{-1}$ de différence entre les deux réplicats de culture du mutant *RΔdim5 39* sur milieu L). Il apparaît clairement que le milieu L+C est le milieu le plus inducteur de la production enzymatique ($0,30 \text{ g.L}^{-1}$ en moyenne sur les six cultures de ce milieu, toutes souches confondues) et que le milieu G est le moins ($0,10 \text{ g.L}^{-1}$ en moyenne sur les six cultures de ce milieu, toutes souches confondues). Il est plus difficile de tirer des conclusions sur les deux autres milieux L et G+C puisqu'ils induisent une production moyenne sur les six cultures de $0,18 \text{ g.L}^{-1}$ et $0,21 \text{ g.L}^{-1}$, respectivement.

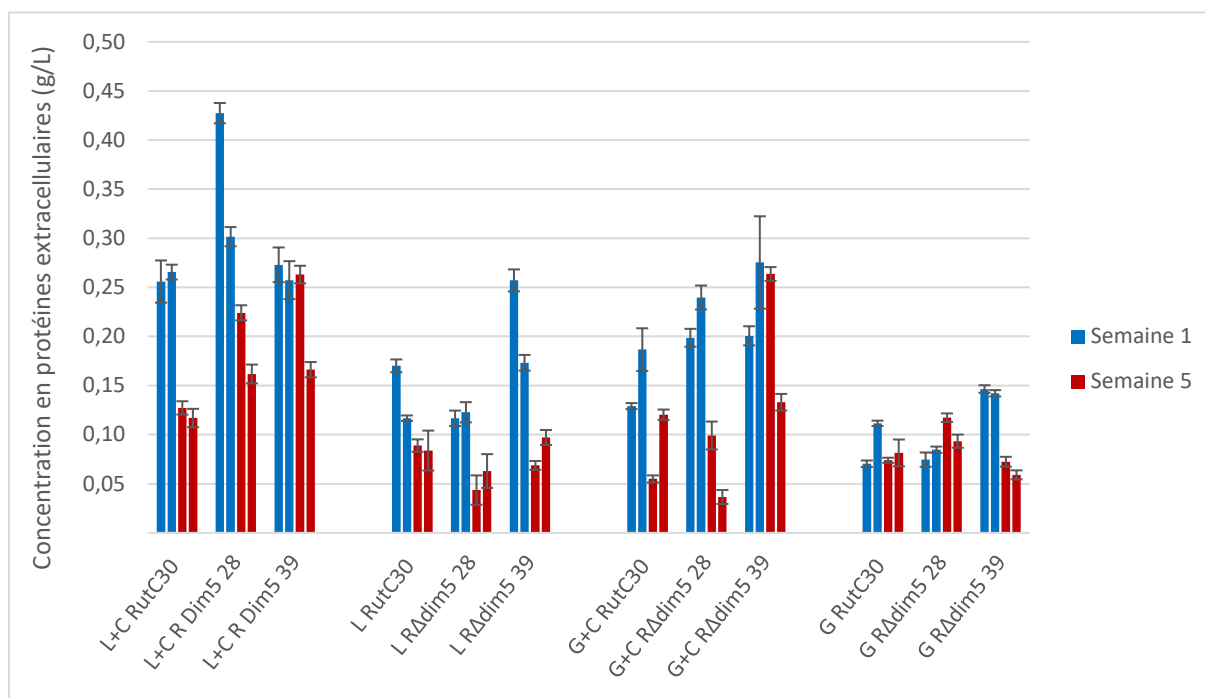


Figure 82 : Concentration en protéines extracellulaires des surnageants de culture de la souche RutC30 et des mutants RΔdim5 28 et RΔdim5 39.

Après cinq semaines d'expérience une diminution générale de la production est observée à l'exception du réplicat 1 sur milieu L+C et réplicat 1 sur milieu G+C du mutant RΔdim5 39. Sur milieu L et sur milieu G, les différences entre RutC30 et mutants ont été gommées et les productions sont maintenant minimales chez les trois individus.

Finalement, cette expérience n'aboutit pas sur les résultats escomptés puisque nous souhaitions démontrer une homogénéisation de la production des mutants RutC30 tandis que la production de la souche parentale devait rester stable à travers le temps. Il est donc difficile de tirer des conclusions étant donné les variabilités fortes observées au sein des cultures après cinq semaines d'expériences. De plus, la diminution de la production de la souche parentale ne nous autorise pas non plus à exploiter les données au maximum puisque notre contrôle n'est pas stable dans les conditions d'expérimentations et ne nous permet pas une comparaison fiable. Notre hypothèse selon laquelle la distribution de la marque H3K9me3 allait s'homogénéiser chez les deux mutants après plusieurs générations et entraîner une homogénéisation du phénotype de production n'est donc pas validée.

V. ANALYSES TRANSCRIPTOMIQUE

À la suite de ces phénotypages, nous avons décidé d'effectuer un test en fioles alimentées afin d'observer le comportement des mutants RutC30 en condition de fed-batch lactose. Pour rappel du Chapitre 1 de ce manuscrit, ces expériences de fioles alimentées se déroulent en deux temps. Dans un premier temps, le champignon pousse en batch sur milieu riche contenant du glucose afin de maximiser la quantité de biomasse produite puis, dans un second temps, la source de carbone est introduite petit à petit. Dans ce contexte précis, nous introduisons du lactose afin d'activer l'induction de la production de cellulases. Des transcriptomes ont été réalisés et nous souhaitons voir s'il existe des différences entre l'expression des gènes de la souche parentale RutC30 et ses deux mutants dans des conditions de production enzymatique mais également si nous observons des variations entre les deux transformants. Étant donné la diversité phénotypique observée jusqu'à présent, nous nous attendions en effet à voir apparaître des différences entre *RΔdim5 28* et *RΔdim5 39*. Si jusqu'à présent nous avons considéré l'hétérogénéité entre les deux réplicats de culture pour un même mutant, ici, ces différences ne sont pas prises en compte afin d'étudier l'expression centrale de chacun des mutants et pouvoir faire ressortir les profils individuels au cœur de *RΔdim5 28* et *RΔdim5 39* par rapport à la RutC30. Ainsi, nous sommes conscients que l'analyse de cette expérience ne permet pas une visualisation complète de l'expression génétique des mutants mais elle nous autorise néanmoins une première visualisation des dérégulations génétiques résultant de l'absence de *dim5* et de H3K9me3.

En Figure 83 ci-dessous, nous présentons les résultats obtenus par dosage de Bradford pour la concentration en enzymes cellulolytiques sécrétées dans le milieu ainsi que la quantité de biomasse sèche produite au bout de 48h d'expérience en fed-batch lactose. Deux cultures pour chaque souche ont été réalisées et deux mesures successives pour chaque culture ont été faites et sont représentées ici par les barres d'erreur.

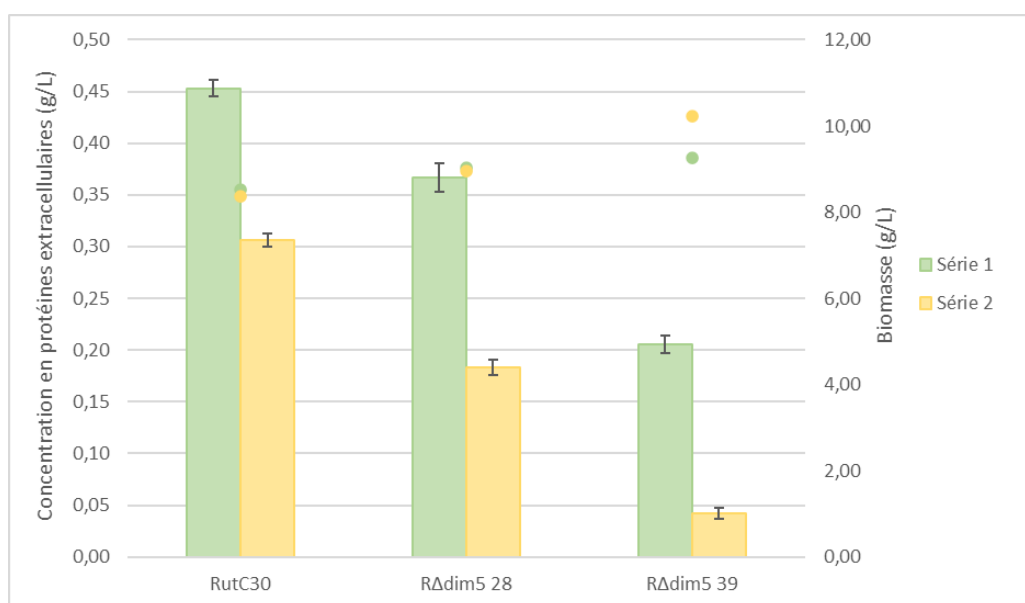


Figure 83: Concentration en protéines extracellulaires et biomasse sèche en condition de fed-batch lactose chez la souche RutC30 et ses deux mutants.

Les concentrations en protéines extracellulaires de la première série d'expériences sont représentées par les histogrammes en vert clair tandis que les concentrations en protéines extracellulaires de la deuxième série sont représentées par les histogrammes en jaune. Les points de la même couleur représentent la quantité de biomasse sèche dans l'un ou l'autre des réplicats de la souche après 48h d'expérience de fed-batch.

Chez la souche RutC30, alors qu'attendue comme stable pour ces expériences, la concentration en enzymes cellulolytiques entre les deux cultures n'est ici pas similaire ($0,15 \text{ g.L}^{-1}$ de différence entre les deux réplicats de culture). Toutefois, ces concentrations restent dans une gamme acceptable pour cette souche et nous avons décidé de garder cette expérience. Une production hétérogène est également observée chez les deux mutants et nous notons que la production des réplicats de la série 1 des trois individus est plus élevée que la production des réplicats de la série 2. Les trois réplicats 1 ayant été effectués tous ensemble puis les trois réplicats 2, il apparaît certainement ici un biais dû à la manipulation. Néanmoins, nous remarquons que la quantité de biomasse produite par les trois souches et sur chacun des réplicats ne varie que très peu contrairement aux concentrations enzymatiques. La variabilité de la production ne dépend donc pas ici d'une différence dans la croissance du champignon.

Afin de minimiser les erreurs dues aux manipulations et identifier les tendances principales d'expression nous avons choisi de considérer comme différentiellement exprimés (DE) les gènes passant un seuil de adjpvalue de 0,005 et un logFC de 2,5.

La Figure 84 ci-dessous expose le nombre de gènes différentiellement exprimés chez les mutants *RΔdim5* 28 et *RΔdim5* 39 en comparaison avec la souche parentale RutC30 en condition d'induction fed-batch lactose.

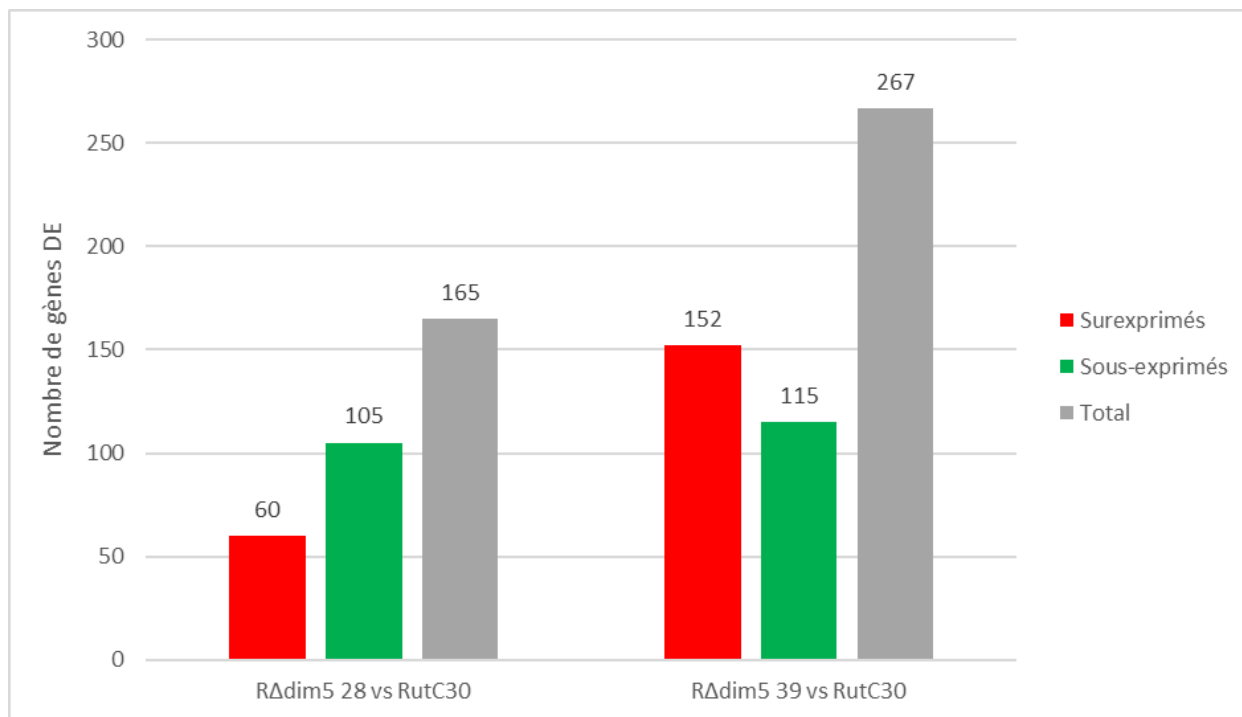


Figure 84 : Gènes différentiellement exprimés chez les mutants *RΔdim5* 28 et *RΔdim5* 39 en comparaison avec la souche parentale RutC30.

Il apparaît ici que, par rapport à la RutC30, un total 165 gènes sont différentiellement exprimés en lactose chez le mutant *RΔdim5* 28 et 267 chez *RΔdim5* 39. Au niveau des expressions, les tendances sont inversées chez les deux clones avec un nombre plus important de gènes sous-exprimés (105) que surexprimés (60) en lactose chez le mutant *RΔdim5* 28 et vice versa (115 et 152 respectivement) chez *RΔdim5* 39.

Les diagrammes de Venn de la Figure 85 représentent les profils d'ensembles des gènes DE des mutants l'un par rapport à l'autre.

Gènes surexprimés chez les mutants

Gènes sous-exprimés chez les mutants

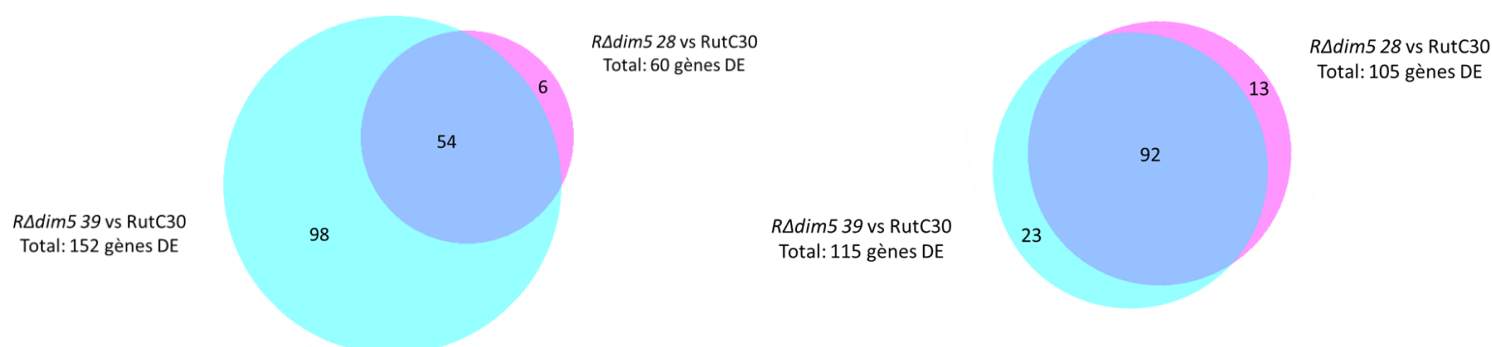


Figure 85 : Diagrammes de Venn des ensembles des gènes différentiellement exprimés chez les mutants RutC30 en condition lactose.

Il apparaît ici que l'ensemble des gènes sous-exprimés en lactose chez les deux mutants est largement partagé entre les deux individus. D'un autre côté, si *RΔdim5 28* partage la quasi-totalité de ces gènes surexprimés en lactose avec *RΔdim5 39*, ce n'est pas réciproque et ce dernier se distingue par 98 gènes qui lui sont uniques.

Parmi les gènes surexprimés en lactose par rapport à la RutC30 et en commun entre les deux mutants, nous retrouvons le gène orthologue de chez *A. niger* de la putative β -glucosidase EGLC et le gène codant pour le facteur de transcription pMH29. D'un autre côté, l'orthologue du gène codant pour la protéine putative GAL80 chez *S. cerevisiae*, jouant un rôle régulateur sur l'expression des gènes du métabolisme du galactose (Yun et al. 1991) est également commun entre les deux clones. Les gènes codant pour Ypr2/SO3, SOR5, SOR6 et SOR8 appartenant au cluster SOR responsable de la synthèse de sorbicillinoïdes sont aussi surexprimés en lactose chez *RΔdim5 28* et *RΔdim5 39*. Nous retrouvons aussi plusieurs gènes liés au stress avec notamment le gène codant pour l'orthologue chez *S. cerevisiae* du facteur de transcription RPN4, impliqué dans l'expression des gènes de protéasomes (Karpov et al. 2022). Nous comptons aussi l'orthologue putatif du gène codant pour UGA1 chez *S. cerevisiae*, gamma-aminobutyrate (GABA), requise pour la tolérance au stress oxydatif et l'utilisation de l'azote (Coleman et al. 2001) ainsi que deux autres gènes putatifs, l'un, d'une GABA perméase et l'autre de la gamma-glutamyltranspeptidase ECM38 (impliqué dans la dégradation du glutathione) (Lussier et al. 1997; Kumar, Sharma et Bachhawat 2003).

Chez le mutant *RΔdim5 39* spécifiquement, nous retrouvons trois groupes fonctionnels de gènes distincts. Le premier rassemble des gènes impliqués dans la réponse aux stress environnementaux et en particulier dans la détoxification des milieux. En effet, cinq gènes codant pour des putatives glutathione S-transférases et une putative aldéhyde réductase de l'aflatoxine B1, jouant un rôle dans l'élimination

des mycotoxines (Morel et al. 2009; Pócsi, Prade et Penninckx 2004) sont surexprimés spécifiquement chez ce mutant. Le gène codant pour la protéine HSP23 mais aussi celui du précurseur putatif de la peroxydase du cytochrome c (Kaya, Mariotti et Gladyshev 2017) et l'orthologue de la protéine SFK1 chez *S. cerevisiae* (Butcher et Schreiber 2003) sont tous spécifiques de *RΔdim5 39* et prennent part à la réponse induite par des stress de température, des stress oxydatifs ou bien encore osmotiques. Ceci représente un arsenal important et bien spécifique pour répondre aux contraintes imposées par le milieu. Ensuite, le second groupe compte plusieurs gènes impliqués dans la synthèse de métabolites secondaires dont les toxines. Les gènes des polyketide synthases PKS1S et PKS2S impliquées dans la synthèse de citrinine, un gène putatif impliqué dans la voie métabolique des trichothécènes mais aussi l'orthologue putatif de la ketoreductase StcE impliquée dans la synthèse de la stérigmatocystine chez *A. nidulans* (Yu et al. 2004) ainsi qu'une putative réductase de l'acide norsolorinique, précurseur de la stérigmatocystine, sont surexprimés en lactose chez ce mutant. Un orthologue putatif d'une synthase de la lovastatine est également présent. Enfin, parmi le dernier groupe fonctionnel nous comptons de nombreux gènes de CAZymes notamment des glycosides hydrolases et des glycosyltransférases. Cinq familles (familles 16, 28, 54, 71 et 92) de glycosides hydrolases sont représentées ici pour un total de six gènes putatifs en plus du gène codant pour la α -L-arabinofuranosidase 2 (ABF2). Trois orthologues de gènes putatifs pour les familles 1 et 32 de glycosyltransférases et le gène pour l'endoglucanase EGL4/CEL61A sont aussi surexprimés.

Les six gènes surexprimés en lactose spécifiques à *RΔdim5 28* sont un orthologue du gène codant pour la protéine putative de Mep2 (un transporteur d'ammonium exprimé en réponse à une déficience d'azote dans le milieu) (Boeckstaens, André et Marini 2007) ainsi qu'un orthologue du gène codant pour la protéine putative PAR32, régulatrice des transporteurs membranaires d'ammonium. Nous comptons également un facteur de transcription et une NADH déhydrogénase en plus de deux gènes dont la fonction n'est pas encore connue.

Concernant les gènes communs entre les deux mutants sous-exprimés en lactose par rapport à la RutC30, nous retrouvons le gène du facteur de transcription RXE1, démontré comme un nouveau régulateur de XYR1, activateur principal de la production de cellulases chez *T. reesei* (Wang et al. 2019). Également, la β -glucosidase CEL3B et le gène de CLF1, un facteur impliqué dans la régulation de l'activité des protéases sécrétées, sont sous-exprimés. Le gène *ooc1*, exprimé lorsque *T. reesei* pousse sur milieu contenant de la cellulose et dans l'obscurité est également sous-exprimé chez les mutants dans cette condition (Schmoll et Kubicek 2005). De plus, de nombreux gènes liés à la conidiation et/ou à la formation de la paroi cellulaire sont réprimés tels que ceux codant pour VOSA, CMP1, HFB2 et HFB3, CHI18-rel1, PhiA, MedA et AZF1 (Puyesky et al. 1999; Pirayre et al. 2020; Gravelat et al. 2010; Chung et al. 2011). Ce dernier est le sujet d'un brevet déposé par l'IFP Energies nouvelles en 2016 grâce à sa surexpression constitutive améliorant la production d'enzymes cellulolytiques. De

même, plusieurs gènes putatifs RGS (« *Regulators of G protein Signaling* ») jouant un rôle majeur dans la signalisation et la transmission des signaux extérieurs aux cellules sont ici sous-exprimés. Enfin, le gène *dim5* est sous-exprimé par rapport à la RutC30, ce qui est un résultat attendu étant donné l'absence de celui-ci dans le génome des mutants.

Parmi les gènes spécifiques à l'un ou l'autre des deux mutants, nous ne retrouvons pas de groupe fonctionnel particulier. Principalement, il s'agit de gènes impliqués dans la réparation de l'ADN et la recombinaison homologue ainsi que des protéines transmembranaires et des gènes impliqués dans l'homéostasie.

En conclusion, il apparaît donc ici que les mutants *RΔdim5* possèdent un nombre important de gènes différentiellement exprimés par rapport à la souche RutC30 dans les conditions d'expérimentation. Il est intéressant de noter que les gènes principaux de l'induction de production de cellulases ne font pas partie de cette liste, indiquant donc qu'ils sont aussi bien exprimés chez les mutants que chez leur souche mère. D'un autre côté, la surexpression de gènes appartenant au cluster impliqué dans la synthèse de sorbicillinoïdes est aussi signe de l'activation du métabolisme secondaire chez les deux mutants. Les résultats les plus marquants sont néanmoins ceux concernant les gènes spécifiques à *RΔdim5 39* où nous retrouvons trois groupes fonctionnels de gènes bien distincts. De nombreuses CAZymes sont retrouvées mais aussi plusieurs gènes impliqués dans la réponse aux stress environnementaux et à la production de mycotoxines. D'un autre côté, les gènes sous-exprimés chez *RΔdim5 28* et *RΔdim5 39* sont majoritairement liés à la conidiation, à la paroi cellulaire et à la réception/transmission des signaux environnementaux. Pas de différence significative dans les gènes spécifiques à l'un ou à l'autre des mutants n'est présente, peut-être signe que la diversification de phénotypes observés chez eux se fait plus sur l'expression que sur l'inhibition des gènes.

DISCUSSION

Dans ce chapitre, nous souhaitons vérifier si la production de cellulases chez *Trichoderma reesei* étaient sous contrôle de mécanismes épigénétiques jusqu'alors inexplorés. S'il a été démontré auparavant qu'il existe un lien entre épigénétique et production de cellulases chez *Trichoderma* (Introduction bibliographique Section II.6) il s'est avéré que celui-ci se fait principalement grâce à la modulation de la structure locale de la chromatine au niveau des promoteurs des gènes. A ce jour et à notre connaissance, une seule étude sur la délétion d'un gène principal de l'épigénétique a été faite sur *Trichoderma reesei* et touche uniquement à *dim5* et la production de sorbicillinoïdes (Wang et al. 2022). Nous avons donc décidé d'étudier l'impact que pouvait avoir la délétion des gènes *kmt6*, *dim5*, *dim2*, *hp1* et *rid* sur les souches QM6a et RutC30. Ces gènes sont tous garants de l'hétérochromatine (facultative pour *kmt6* et constitutive pour les autres) et jouent donc un rôle important dans la régulation transcriptionnelle du génome. De nombreuses études existent sur chez d'autres champignons et ont démontré chez eux l'importance de ces gènes pour, entre autres, la croissance, la production de métabolites secondaires, la stabilité génomique, ou encore le développement (Connolly, Smith et Freitag 2013; Carlier et al. 2021; Ren et al. 2021; Tamaru et Selker 2001; Singer et Selker 1995; Möller et al. 2021). Dans notre cas, l'étude chez les deux souches de ces gènes avait pour objectif principal de déterminer si l'un d'entre eux avait une influence sur la production de cellulase.

Le premier résultat dont nous souhaiterions discuter est celui concernant les mutants $\Delta kmt6$. S'il a été démontré chez plusieurs organismes tels que *N. crassa* (Jamieson et al., 2013), *E. festucae* (Chujo and Scott, 2014) ou bien encore *F. graminearum* (Connolly et al., 2013) que la délétion de *kmt6* n'entraîne pas la mort de ces champignons, il semblerait que chez *Podospora anserina* la délétion de ce gène nuit fortement à la croissance du champignon et provoque des effets adverses sévères sur l'ensemble de son cycle de vie (Carlier et al. 2021). Ici, le taux final d'isolement est très bon pour les deux souches et les étapes de purification nous assurent que les individus isolés soient bien homocaryotiques. De nombreux transformants ont été validés pour la délétion du gène, pourtant aucun d'entre eux ne s'est avéré génétiquement correctement modifié puisqu'il a été impossible de valider les abords de l'insertion. La résistance à l'hygromycine des mutants mais l'impossibilité de cartographier le locus entier suggère des insertions multiples avec remaniements génétiques (apparition de nombreuses bandes de différentes tailles chez les mutants pour les régions adjacentes de l'insertion et absence chez le contrôle). Également, le fait qu'aucun des transformants n'aient été validés après six transformations indépendantes laisse penser que la délétion de *kmt6* est en fait létale chez *T. reesei* et que nous n'avons peut-être qu'imparfaitement supprimé le gène (apparition d'une bande à une hauteur inattendue sur les PCR de vérification de délétion du gène). Cette hypothèse est soutenue par (Studt et al. 2016) qui se sont trouvés dans l'impossibilité de générer chez *Fusarium fujikuroi* des mutants $\Delta kmt6$ homocaryotiques viables

malgré une insertion de la cassette d'inactivation. De plus, leurs tentatives de substitution de la lysine K9 par une arginine se sont également révélées infructueuses et suggèrent le rôle crucial que joue H3K27me3 chez ce champignon. Nous émettons donc pour la première fois ici l'hypothèse que *T. reesei* ne soit pas en mesure d'y survivre. Il serait intéressant de tenter comme Studt et al. (2016) une substitution du résidu H3K27 pour un résidu H3R27 afin de vérifier cette hypothèse. Il aurait également pu être utile dans notre cas d'effectuer un Western blot pour H3K27me3 pour déterminer si l'insertion imparfaite de la cassette avait tout de même participé à la diminution de la marque. Nous aurions dans ce cas pu éventuellement continuer une étude de ces mutants. Dans le cas contraire, l'utilisation d'ARN interférents auraient également pu être une solution afin de diminuer la traduction de la protéine TrKMT6.

Il a été démontré auparavant chez de nombreux champignons que l'hétérochromatine constitutive est, entre autres, garante de l'intégrité génomique, la croissance, le développement, la pathogénicité, l'inhibition de clusters de métabolites secondaires etc... (Ren et al. 2021; Rountree et Selker 1997; Freitag et Selker 2005; Chan, Henderson et Jacobsen 2005; Reyes-Dominguez et al. 2010; Bok et al. 2009). Cette conformation de la chromatine dépend majoritairement de la présence de deux marques épigénétiques qui sont H3K9me3 et la méthylation de l'ADN 5mC. Pour rappel, H3K9me3 est apposé sur les anciens transposons par la protéine histone méthyltransférase DIM5 et est ensuite reconnu pour la protéine lectrice HP1 qui va permettre le dépôt de la méthylation de l'ADN par l'ADN méthyltransférase DIM2. RID1 d'un autre côté, maintient l'intégrité du génome par méthylation de l'ADN des nouveaux transposons après passage du système RIP lors de la reproduction sexuée. A cause de cette importance pour des fonctions primordiales, nous nous attendions donc à voir des phénotypes de production marqués par l'inactivation des gènes responsables de l'apparition de ces marques et de la protéine lectrice. Pourtant, il semble qu'à première vue, l'absence des trois gènes *dim2*, *hp1* et *rid* n'impacte que peu *Trichoderma reesei* sur le phénotype étudié et que la majorité des perturbations sont concentrées par les mutants $\Delta dim5$. Ceci suggère donc que la méthylation de l'ADN chez *Trichoderma reesei* n'est pas aussi importante pour la croissance et la production enzymatique du champignon que l'est H3K9me3. Une autre hypothèse serait qu'il existe une redondance fonctionnelle et qu'une autre ADN méthyltransférase prenne le relais comme DNMT5 puisque chez les ascomycètes la combinaison d'ADN méthyltransférases DIM2 + DNMT5 + RID1 est une des plus fréquente avec DIM2 + RID1. Cependant, aucune étude n'a démontré à ce jour la présence du gène codant pour cette protéine dans le génome de *Trichoderma*.

Ainsi, nous avons décidé de nous concentrer exclusivement sur les mutants $\Delta dim5$ générés. Une forte hétérogénéité, en particulier chez les mutants *dim5* de la RutC30, est observée entre individus indépendants issus de la même transformation. Cette variabilité importante se retrouve également entre deux réplicats techniques de culture pour un même mutant. En milieux liquides, ceci se traduit par des changements et des variations dans la concentration en protéines extracellulaires mais aussi par une

modification qualitative du cocktail enzymatique qui est ici visible par les variations d'activité spécifique de CBH1 et des β -glucosidases par rapport aux activités du cocktail de la RutC30 parentale. De manière globale, une augmentation sur l'activité des β -glucosidases est visible chez les mutants de la souche hyper-productrice, suggérant une augmentation de leur proportion dans le cocktail. Les résultats sont plus partagés chez la QM6a avec des variations moins marquées sur la production en protéines des réplicats 2,3 et 4 mais une différence forte au niveau de l'activité β -glucosidase entre les réplicats 1,2 et 3,4. Il est possible que la différence de concentration protéique observée chez le réplicat 1 soit due à une proportion de CBH1 moins importante dans le cocktail puisque cela se répercute visiblement sur son activité.

En milieux solides, au-delà d'une croissance et d'une sporulation altérées, l'hétérogénéité est particulièrement visible chez le mutant *R Δ dim5 28* qui arbore des régions de croissance intense autour d'un mycélium avec un développement plus modeste et avec un aspect similaire à celui du mutant *R Δ dim5 39*. Bien que ces excroissances soient impactantes au niveau visuel, le mutant *R Δ dim5 28* ne se distingue que très peu au niveau transcriptomique du mutant *R Δ dim5 39*. En effet, ce premier mutant ne possède qu'un nombre restreint de gènes différentiellement exprimés et la grande majorité d'entre eux sont également présents chez *R Δ dim5 39*. Il est intéressant de noter que leur fonction tourne principalement autour de la production de sorbicilline (surexpression) et de la conidiation (sous-expression). Wang et al. (2022), ont également observé chez la souche hyper-productrice QM9414 de *T. reesei* que la suppression de *TrDIM5* entraîne une augmentation de la production de sorbicillinoïdes et une conidiation restreinte. D'un autre côté, si le mutant *R Δ dim5 39* arbore des phénotypes altérés mais sans excroissances hypertrophiques, c'est au niveau de la transcriptomique qu'il se démarque, notamment par un ensemble de gènes surexprimés qui lui sont spécifiques et qui sont dédiés à la production de métabolites secondaires comme la lovastatine ou les mycotoxines comme la stérigmatocystine, la citrinine ou le trichothécènes mais aussi par des groupes de gènes de CAZymes et de réponse aux stress environnementaux. Cette production *de novo* de métabolites secondaires après suppression du gène *dim5* a déjà été observée chez le genre *Fusarium* avec notamment la synthèse de la mycotoxine FB1 chez *F. verticillioides* et l'aurofusarine chez *F. graminearum*. Chez *A. nidulans*, (Reyes-Dominguez et al. 2010) ont démontré que la suppression de l'orthologue de *dim5* résulte en une expression accrue des gènes *aflR* et *stcO* du cluster de la stérigmatocystine ainsi qu'une dérépression des clusters de la pénicilline et de la terrequinone A. Il est intéressant de voir que la marque H3K9me3 est, au-delà de son rôle dans le maintien strict de l'intégrité du génome, également impliquée dans la régulation de métabolites secondaires. Il pourrait donc être intéressant d'utiliser le mutant *R Δ dim5 39* pour la caractérisation de ces derniers chez *T. reesei*.

Etant donné l'absence d'explications génétiques à ces variations entre mutants, nous avons émis l'hypothèse de la présence d'hétérogénéités au niveau de la nouvelle distribution des marques épigéniques le long des chromosomes. En leur potentielle

qualité d'épimutations, celles-ci devraient être réversibles et se transmettre de générations en générations. Nous pensions donc que cinq semaines de repiquages successifs pour les deux mutants résulteraient en une reprogrammation identique du paysage épigénétique et engendreraient finalement un phénotype stable et commun. Toutefois, cette expérience s'est révélée inconclusive étant donné l'instabilité des souches de références dans les milieux testés après cinq semaines d'expérimentations.

Ainsi l'expérience ayant pour but de valider la présence d'épimutations réversibles n'étant pas concluante et bien que nous ayons procédé avec soin à l'isolement du transformant à partir d'une unique spore homocaryotique et aux vérifications génétiques, il serait tout de même opportun d'effectuer un nouveau séquençage indépendant des mutants et particulièrement des parties hétérogènes du mutant *RΔdim5 28*. Ensuite, il serait intéressant d'effectuer une analyse transcriptomique sur les deux parties distinctes du mycélium afin d'en comprendre les différences d'expression génétiques. Nous nous attendrions à voir apparaître deux profils distincts avec d'un côté une expression similaire à celle du mutant *RΔdim5 39* pour la partie centrale et une expression autre pour la partie envahissante. En effet, nous soupçonnons une redistribution de la marque H3K27me3 inégale au niveau des régions déplétées de H3K9me3 comme déjà observé chez *Zymoseptoria tritici* (Möller et al. 2019) et chez les cellules de mammifères (Basenko et al. 2015) et ainsi, la création de deux populations distinctes de cellules avec des profils d'expressions spécifiques. Sur les parties affichant une forte croissance, l'impact de la délétion de *dim5* a peut-être été contrecarrée par l'invasion de H3K27me3 sur les zones essentielles au maintien de la stabilité du génome et autres régions importantes, lui conférant ainsi un phénotype plus adapté au milieu tandis que la partie centrale est comparable à *RΔdim5 39*. Ceci serait en accord avec (Torres-Garcia et al. 2020) qui ont prouvé que sous certaines conditions, l'hétérochromatine forme des épimutations afin de favoriser la plasticité et la résilience des cellules pour une meilleure adaptation aux environnements défavorables. Afin de vérifier ces hypothèses, la réalisation de nouvelles expériences de ChIP Seq pour générer des cartographies sur l'une et l'autre des parties du mycélium devraient être effectuées au moins pour la marque H3K27me3. Également, des données de ChIP Seq sur le mutant *RΔdim5 39* renseigneraient sur l'état chromatinien de ce dernier et nous pourrions faire une comparaison avec le mutant *RΔdim5 28* pour mieux comprendre d'où peuvent provenir leurs différences phénotypiques. Enfin, nous possédons déjà les données de ChIP Seq de H3K9me3 pour la souche RutC30, notre souche de départ. Ces mêmes tests pourraient être effectués sur des mutants de la QM6a (*QΔdim5 39* et un autre mutant à valider génétiquement) afin de vérifier si le terrain génétique déjà très modifié de la RutC30 aggrave l'instabilité générale du paysage épigénétique et l'hétérogénéité phénotypique.

Néanmoins, les variations fortes observées pour la production enzymatique entre réplicats de culture d'un même mutant et ce, même chez *RΔdim5 39* restent encore inexpliquées. Une possibilité serait l'intoxication du champignon par sa propre production, jusqu'alors inhibée, de métabolites secondaires et de composés

organiques volatils (COV). Une étude du sécrétome de ce mutant et des tests sur la toxicité du surnageant de culture pourrait nous éclairer sur la nuisibilité de celui-ci. Toutefois, les tests de poids sec fait sur les cultures indiquent que dans notre cas il n'existe pas un lien direct entre croissance de la biomasse, production enzymatique et expression différentielle des gènes du métabolisme secondaire. Enfin, nous pourrions effectuer des analyses de chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC/MS) à l'instar de (Torál et al. 2021), afin d'identifier et quantifier les COV émis par nos mutants et l'impact potentiel que ces derniers ont sur la production enzymatique.

CONCLUSION GENERALE

Dans un contexte de production de bioéthanol de seconde génération, les industriels du secteur ont mené plusieurs campagnes d'amélioration génétique de *T. reesei* afin de diminuer les coûts de l'étape d'hydrolyse enzymatique. Jusqu'à présent, l'ingénierie génétique a permis la création de nouvelles souches dont la production d'enzymes cellulolytiques est optimisée. Désormais, au-delà du contrôle génétique, les recherches se tournent aussi vers les modifications épigénétiques de la chromatine et leur implication dans la synthèse enzymatique.

En effet, les mécanismes épigénétiques mis en place lors de l'induction de la synthèse de cellulases restent chez *T. reesei* largement inconnus bien que des recherches au laboratoire et dans la littérature suggèrent un lien entre dynamique chromatinienne et production enzymatique. Ainsi, la compréhension des phénomènes épigénétiques ayant lieu lors de la synthèse enzymatique permettrait de lever un verrou supplémentaire et de faire encore un pas vers la maîtrise de la production.

Mon travail de thèse a ainsi consisté en l'étude et la compréhension des mécanismes épigénétiques de *T. reesei* intervenant lors de la production de cellulases. Pour ce faire, deux axes de travail ont été développés. Le premier a consisté en la génération d'une cartographie de la marque H3K4me3, liée à l'ouverture chromatinienne, dans une condition de production enzymatique vs une condition de non-production et dans deux contextes génétiques différents. Le second axe a consisté en la création de mutants des gènes épigénétiques liés aux marques centrales de l'hétérochromatine dans les deux souches de *T. reesei* étudiées dans le premier volet de la thèse.

La première partie de mon travail a donc été consacrée à la génération de données omiques dans un environnement de production et de non-production de cellulases dans un but de capturer les informations à l'échelle du transcriptome et de l'épigénome de *Trichoderma reesei* afin de les mettre au regard des génomes des deux souches de références utilisées qui sont l'isolat sauvage QM6a et la souche hyper-productrice RutC30. La technique de ChIP Seq tout juste mise en place à IFP Energies nouvelles a servi à récolter les données épigénétiques désirées. Les méthodes bioinformatiques spécifiques à leur analyse ont été développés pour la première fois au laboratoire durant cette thèse. D'un autre côté, huit transcriptomes au total ont été générés, deux pour chaque condition et chaque souche. Les résultats de l'analyse différentielle du transcriptome de la souche RutC30 dans les conditions étudiées ont été déjà connus et ont servi de référentiels qualité pour les expériences. Enfin, une réannotation manuelle complète du génome de la souche QM6a a été réalisée au niveau de l'équipe afin d'utiliser les informations génomiques les plus précises possibles.

La cartographie de H3K4me3 le long des sept chromosomes de la QM6a et de la RutC30 n'a mis à jour aucune différence macroscopique entre les souches. Les translocations de la RutC30 n'ont subi aucune modification et n'arborent pas de « cicatrice » au niveau des zones de jonction suggérant ainsi qu'elles ne sont pas impliquées dans l'hyper-productivité de la RutC30. Des zones enrichies en marques H3K4me3 au niveau péricentromérique et sur certaines zones des bras des chromosomes sont également retrouvées chez les deux souches et pourraient être des cibles privilégiées pour de futures transformations. D'un autre côté, les zones communes déplétées en H3K4me3 rassemblent à la fois des gènes surexprimés et des gènes sous-exprimés en condition lactose (condition d'activation de la production enzymatique) suggérant ainsi la cohabitation de marques épigénétiques activatrices et répressives de la transcription. Après analyse différentielle individuelle pour chacune des souches des résultats de ChIP Seq, il s'avère que l'on retrouve en condition lactose les gènes majeurs de la production de cellulases différenciellement marqués par H3K4me3 chez QM6a, contrairement à la RutC30. Conjointement aux analyses différentielles de transcriptomique, les résultats mettent en évidence la présence d'un système épigénétique et transcriptionnel d'activation de la production enzymatique différent entre les deux souches avec une évolution de celui de la RutC30 vers une diversité plus importante que celui de la QM6a.

Dans un second temps, l'invalidation des gènes liés à la présence d'hétérochromatine *dim5*, *hp1*, *dim2*, *kmt6* et *rid* a permis de mettre en valeur le rôle que joue ce premier et par conséquent H3K9me3 dans la stabilité de la production enzymatique. Les phénotypages sur milieu inducteur liquide et sur différents milieux solides montrent, particulièrement chez la RutC30, des variabilités fortes au niveau de la synthèse de cellulases mais aussi au niveau de la morphologie et la croissance du champignon. De plus, l'analyse différentielle transcriptomique a permis de mettre en avant chez le mutant *RΔdim5* 39 la surexpression de nombreux gènes liés au métabolisme secondaire, notamment aux mycotoxines, qui ne sont pas surexprimés chez l'individu *RΔdim5* 28. La disparition de ce gène entraîne donc un éventail de réponses chez les individus isolés. Une de mes hypothèses sur les causes de cette hétérogénéité était la nouvelle distribution inégale de marques épigénétiques tierces le long des chromosomes entraînant de ce fait des phénotypes variables. L'uniformisation de cette distribution était donc envisagée après cinq semaines de culture sur quatre milieux suscitant ces hétérogénéités. Néanmoins, les expériences se sont révélées inconcluantes à cause de l'instabilité sur le temps observée chez les souches parentales utilisées comme contrôles.

Enfin, un autre résultat intéressant de ce chapitre porte sur l'invalidation du gène *Δkmt6* et l'incapacité de récolter des mutants validés génétiquement. La possibilité que ce gène soit en fait essentiel à la survie de *T. reesei* est ici avancée et reste encore à prouver.

Pour la suite de ces recherches, il serait opportun de faire le lien entre les deux

chapitres de cette thèse en exploitant les résultats de ChIP Seq de la marque H3K9me3 générés sur les deux souches QM6a et RutC30 au premier chapitre. Ceci permettrait ainsi de découvrir les zones ensuite déplétées chez les mutants générés au chapitre II et éventuellement de mieux comprendre les raisons des hétérogénéités des individus. De plus, la réalisation de ChIP Seq sur H3K4me3 et H3K27me3 sur les mutants $\Delta dim5$ étudiés dans le second axe ainsi que H3K36me2/3 et H3K9ac sur l'ensemble des individus étudiés dans cette thèse, permettrait de cette façon une comparaison des paysages chromatinien avec les souches parentales et apporterait une réponse à l'hypothèse formulée sur la nouvelle distribution inégale des principales marques le long des génomes.

BIBLIOGRAPHIE

Aghcheh, Razieh Karimi; Kubicek, Christian P. (2015) Epigenetics as an emerging tool for improvement of fungal strains used in biotechnology. In : *Applied microbiology and biotechnology*, vol. 99, n° 15, p. 6167–6181. DOI: 10.1007/s00253-015-6763-2.

Albright, Jessica C.; Henke, Matthew T.; Soukup, Alexandra A.; McClure, Ryan A.; Thomson, Regan J.; Keller, Nancy P.; Kelleher, Neil L. (2015) Large-scale metabolomics reveals a complex response of *Aspergillus nidulans* to epigenetic perturbation. In : *ACS chemical biology*, vol. 10, n° 6, p. 1535–1541. DOI: 10.1021/acscchembio.5b00025.

Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 51, p. 786–794. DOI: 10.1073/pnas.51.5.786.

C. David. Allis, Thomas. Jenuwein, Danny. Reinberg, Marie-Laure. Caparros, éd. (2015) *Epigenetics*. 2nd edition. Cold Spring Harbor (N.Y.) : Cold Spring Harbor laboratory press.

Amabile, Angelo; Migliara, Alessandro; Capasso, Paola; Biffi, Mauro; Cittaro, Davide; Naldini, Luigi; Lombardo, Angelo (2016) Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. In : *Cell*, vol. 167, n° 1, 219–232.e14. DOI: 10.1016/j.cell.2016.09.006.

Antonieto, Amanda Cristina Campos; Nogueira, Karoline Maria Vieira; Paula, Renato Graciano de; Nora, Luísa Czamanski; Cassiano, Murilo Henrique Anzolini; Guazzaroni, Maria-Eugenia et al. (2019) A Novel Cys2His2 Zinc Finger Homolog of AZF1 Modulates Holocellulase Expression in *Trichoderma reesei*. In : *mSystems*, vol. 4, n° 4. DOI: 10.1128/mSystems.00161-19.

Arents, G.; Moudrianakis, E. N. (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, n° 24, p. 11170–11174. DOI: 10.1073/pnas.92.24.11170.

Aro, N.; Saloheimo, A.; Ilmén, M.; Penttilä, M. (2001) ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. In : *The Journal of biological chemistry*, vol. 276, n° 26, p. 24309–24314. DOI: 10.1074/jbc.M003624200.

Aro, Nina; Ilmén, Marja; Saloheimo, Anu; Penttilä, Merja (2003) ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. In : *Applied and environmental microbiology*, vol. 69, n° 1, p. 56–65. DOI: 10.1128/AEM.69.1.56-65.2003.

Aro, Nina; Pakula, Tiina; Penttilä, Merja (2005) Transcriptional regulation of plant cell wall degradation by filamentous fungi. In : *FEMS Microbiology Reviews*, vol. 29, n° 4, p. 719–739. DOI: 10.1016/j.femsre.2004.11.006.

Atanasova, Lea (2014) *Ecophysiology of Trichoderma in Genomic Perspective: Biotechnology and Biology of Trichoderma*: Elsevier, p. 25–40.

Bailey, Andy M.; Alberti, Fabrizio; Kilaru, Sreedhar; Collins, Catherine M.; Mattos-Shipley, Kate de; Hartley, Amanda J. et al. (2016) Identification and manipulation of the pleuromutilin gene cluster from *Clitopilus passeckerianus* for increased rapid antibiotic production. In : Scientific reports, vol. 6, p. 25202. DOI: 10.1038/srep25202.

Bannister, Andrew J.; Kouzarides, Tony (2011) Regulation of chromatin by histone modifications. In : Cell research, vol. 21, n° 3, p. 381–395. DOI: 10.1038/cr.2011.22.

Barkai, N.; Rose, M. D.; Wingreen, N. S. (1998) Protease helps yeast find mating partners. In : Nature, vol. 396, n° 6710, p. 422–423. DOI: 10.1038/24760.

Barry, C.; Faugeron, G.; Rossignol, J. L. (1993) Methylation induced premeiotically in *Ascobolus*. Coextension with DNA repeat lengths and effect on transcript elongation. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 90, n° 10, p. 4557–4561. DOI: 10.1073/pnas.90.10.4557.

Basenko, Evelina Y.; Sasaki, Takahiko; Ji, Lexiang; Prybol, Cameron J.; Burckhardt, Rachel M.; Schmitz, Robert J.; Lewis, Zachary A. (2015) Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 112, n° 46, E6339-48. DOI: 10.1073/pnas.1511377112.

Bekunda, M.; Palm, C. A.; Fraiture, C. de; Leadley, P.; Maene, Luc; Martinelli, L. A. et al. (2009) Biofuels in developing countries. In : Robert Warren Howarth et S. Bringezu, coord.: Biofuels : Biofuels: Environmental Consequences and Interactions with Changing Land Use. Proceedings of the Scientific Committee on Problems of the Environment (SCOPE) International Biofuels Project Rapid Assessment, 22-25 September 2008, Gummersbach, Germany. Ithaca, N.Y.: Cornell University, p. 249–269.

Benjamini, Yoav; Hochberg, Yosef (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. In : Journal of the Royal Statistical Society: Series B (Methodological), vol. 57, n° 1, p. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Benocci, Tiziano; Aguilar-Pontes, Maria Victoria; Kun, Roland Sándor; Lubbers, Ronnie J. M.; Lail, Kathleen; Wang, Mei et al. (2019) Deletion of either the regulatory gene *ara1* or metabolic gene *xki1* in *Trichoderma reesei* leads to increased CAZyme gene expression on crude plant biomass. In : Biotechnology for biofuels, vol. 12, p. 81. DOI: 10.1186/s13068-019-1422-y.

Benocci, Tiziano; Aguilar-Pontes, Maria Victoria; Zhou, Miaomiao; Seiboth, Bernhard; Vries, Ronald P. de (2017) Regulators of plant biomass degradation in ascomycetous fungi. In : Biotechnology for biofuels, vol. 10, p. 152. DOI: 10.1186/s13068-017-0841-x.

Berson, Amit; Nativio, Raffaella; Berger, Shelley L.; Bonini, Nancy M. (2018) Epigenetic Regulation in Neurodegenerative Diseases. In : Trends in neurosciences, vol. 41, n° 9, p. 587–598. DOI: 10.1016/j.tins.2018.05.005.

Berteaux-Lecellier, V.; Zickler, D.; Debuchy, R.; Panvier-Adoutte, A.; Thompson-Coffe, C.; Picard, M. (1998) A homologue of the yeast *SHE4* gene is essential for the transition between the syncytial and

cellular stages during sexual reproduction of the fungus *Podospora anserina*. In : The EMBO journal, vol. 17, n° 5, p. 1248–1258. DOI: 10.1093/emboj/17.5.1248.

Bertin, Aurélie; Mangenot, Stéphanie (2008) Structure et dynamique de la particule coeur de nucléosome. In : Med Sci (Paris), vol. 24, n° 8-9, p. 715–719. DOI: 10.1051/medsci/20082489715.

Betina, V.; Zajacová, J. (1978) Inhibition of photo-induced *Trichoderma viride* conidiation by inhibitors of RNA synthesis. In : Folia microbiologica, vol. 23, n° 6, p. 460–464. DOI: 10.1007/bf02885576.

Bewick, Adam J.; Hofmeister, Brigitte T.; Powers, Rob A.; Mondo, Stephen J.; Grigoriev, Igor V.; James, Timothy Y. et al. (2019) Diversity of cytosine methylation across the fungal tree of life. In : Nature ecology & evolution, vol. 3, n° 3, p. 479–490. DOI: 10.1038/s41559-019-0810-9.

Bikard, David; Jiang, Wenyan; Samai, Poulami; Hochschild, Ann; Zhang, Feng; Marraffini, Luciano A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. In : Nucleic acids research, vol. 41, n° 15, p. 7429–7437. DOI: 10.1093/nar/gkt520.

Bischof, Robert H.; Ramoni, Jonas; Seiboth, Bernhard (2016) Cellulases and beyond. The first 70 years of the enzyme producer *Trichoderma reesei*. In : Microbial cell factories, vol. 15, n° 1, p. 106. DOI: 10.1186/s12934-016-0507-6.

Blum, Matthias; Chang, Hsin-Yu; Chuguransky, Sara; Grego, Tiago; Kandasaamy, Swaathi; Mitchell, Alex et al. (2021) The InterPro protein families and domains database: 20 years on. In : Nucleic acids research, vol. 49, n° D1, D344-D354. DOI: 10.1093/nar/gkaa977.

Bochner, B. R.; Gadzinski, P.; Panomitros, E. (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. In : Genome research, vol. 11, n° 7, p. 1246–1255. DOI: 10.1101/gr.186501.

Bodinaku, Ina; Shaffer, Jason; Connors, Allison B.; Steenwyk, Jacob L.; Biango-Daniels, Megan N.; Kastman, Erik K. et al. (2019) Rapid Phenotypic and Metabolomic Domestication of Wild *Penicillium* Molds on Cheese. In : mBio, vol. 10, n° 5. DOI: 10.1128/mBio.02445-19.

Boeckstaens, Mélanie; André, Bruno; Marini, Anna Maria (2007) The yeast ammonium transport protein Mep2 and its positive regulator, the Npr1 kinase, play an important role in normal and pseudohyphal growth on various nitrogen media through retrieval of excreted ammonium. In : Molecular microbiology, vol. 64, n° 2, p. 534–546. DOI: 10.1111/j.1365-2958.2007.05681.x.

Bok, Jin Woo; Chiang, Yi-Ming; Szewczyk, Edyta; Reyes-Dominguez, Yazmid; Davidson, Ashley D.; Sanchez, James F. et al. (2009) Chromatin-level regulation of biosynthetic gene clusters. In : Nature chemical biology, vol. 5, n° 7, p. 462–464. DOI: 10.1038/nchembio.177.

Bok, Jin Woo; Keller, Nancy P. (2004) LaeA, a regulator of secondary metabolism in *Aspergillus* spp. In : Eukaryotic cell, vol. 3, n° 2, p. 527–535. DOI: 10.1128/ec.3.2.527-535.2004.

Butcher, Rebecca A.; Schreiber, Stuart L. (2003) A small molecule suppressor of FK506 that targets the mitochondria and modulates ionic balance in *Saccharomyces cerevisiae*. In : *Chemistry & biology*, vol. 10, n° 6, p. 521–531. DOI: 10.1016/s1074-5521(03)00108-x.

Cai, Pengli; Gu, Ruimeng; Wang, Bang; Li, Jingen; Li Wan; Tian, Chaoguang; Ma, Yanhe (2014) Evidence of a critical role for cellobiose transporter 2 (CDT-2) in both cellulose and hemicellulose degradation and utilization in *Neurospora crassa*. In : *PloS one*, vol. 9, n° 2, e89330. DOI: 10.1371/journal.pone.0089330.

Cambareri, E. B.; Jensen, B. C.; Schabach, E.; Selker, E. U. (1989) Repeat-induced G-C to A-T mutations in *Neurospora*. In : *Science (New York, N.Y.)*, vol. 244, n° 4912, p. 1571–1575. DOI: 10.1126/science.2544994.

Cameron, Rosannah C.; Duncan, Elizabeth J.; Dearden, Peter K. (2013) Biased gene expression in early honeybee larval development. In : *BMC genomics*, vol. 14, p. 903. DOI: 10.1186/1471-2164-14-903.

Cao, Yanli; Zheng, Fanglin; Wang, Lei; Zhao, Guolei; Chen, Guanjin; Zhang, Weixin; Liu, Weifeng (2017) Rce1, a novel transcriptional repressor, regulates cellulase gene expression by antagonizing the transactivator Xyr1 in *Trichoderma reesei*. In : *Molecular microbiology*, vol. 105, n° 1, p. 65–83. DOI: 10.1111/mmi.13685.

Cao, Yanli; Zheng, Fanglin; Zhang, Weixin; Meng, Xiangfeng; Liu, Weifeng (2019) *Trichoderma reesei* XYR1 recruits SWI/SNF to facilitate cellulase gene expression. In : *Molecular microbiology*, vol. 112, n° 4, p. 1145–1162. DOI: 10.1111/mmi.14352.

Carlier, F.; Li, M.; Maroc, L.; Debuchy, R.; Souaid, C.; Noordermeer, D. et al. (2021) Loss of EZH2-like or SU(VAR)3-9-like proteins causes simultaneous perturbations in H3K27 and H3K9 tri-methylation and associated developmental defects in the fungus *Podospora anserina*. In : *Epigenetics & chromatin*, vol. 14, n° 1, p. 22. DOI: 10.1186/s13072-021-00395-7.

Carter, G. L.; Allison, D.; Rey, M. W.; Dunn-Coleman, N. S. (1992) Chromosomal and genetic analysis of the electrophoretic karyotype of *Trichoderma reesei*: mapping of the cellulase and xylanase genes. In : *Molecular microbiology*, vol. 6, n° 15, p. 2167–2174. DOI: 10.1111/j.1365-2958.1992.tb01390.x.

Castle, A.; Speranzini, D.; Rghei, N.; Alm, G.; Rinker, D.; Bissett, J. (1998) Morphological and molecular identification of *Trichoderma* isolates on North American mushroom farms. In : *Applied and environmental microbiology*, vol. 64, n° 1, p. 133–137.

Chan, Simon W-L; Henderson, Ian R.; Jacobsen, Steven E. (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. In : *Nature reviews. Genetics*, vol. 6, n° 5, p. 351–360. DOI: 10.1038/nrg1601.

Chang, C. H.; Luse, D. S. (1997) The H3/H4 tetramer blocks transcript elongation by RNA polymerase II in vitro. In : *The Journal of biological chemistry*, vol. 272, n° 37, p. 23427–23434. DOI: 10.1074/jbc.272.37.23427.

Chen, Xiangsong; Zhou, Dao-Xiu (2013) Rice epigenomics and epigenetics. Challenges and opportunities. In : Current opinion in plant biology, vol. 16, n° 2, p. 164–169. DOI: 10.1016/j.pbi.2013.03.004.

Chen, Xiuzhen; Song, Bingran; Liu, Minglu; Qin, Lina; Dong, Zhiyang (2021) Understanding the Role of *Trichoderma reesei* Vib1 in Gene Expression during Cellulose Degradation. In : Journal of fungi (Basel, Switzerland), vol. 7, n° 8. DOI: 10.3390/jof7080613.

Chen, Yumeng; Wang, Wei; Liu, Pei; Lin, Aibo; Fan, Xingjia; Wu, Chuan et al. (2021) The novel repressor Rce2 competes with Ace3 to regulate cellulase gene expression in the filamentous fungus *Trichoderma reesei*. In : Molecular microbiology. DOI: 10.1111/mmi.14825.

Chernin, Leonid; Chet, Ilan (2002) Microbial Enzymes in Biocontrol of Plant Pathogens and Pests: In: Enzymes in the Environment: Activity, Ecology, and Applications. (R. Burns, and R. Dick, eds.), Marcel Dekker, Inc., New York, p. 171–225.

Cherry, Joel R.; Fidantsef, Ana L. (2003) Directed evolution of industrial enzymes. An update. In : Current Opinion in Biotechnology, vol. 14, n° 4, p. 438–443. DOI: 10.1016/S0958-1669(03)00099-5.

Chung, Da-Woon; Greenwald, Charles; Upadhyay, Srijana; Ding, Shengli; Wilkinson, Heather H.; Ebbole, Daniel J.; Shaw, Brian D. (2011) acon-3, the *Neurospora crassa* ortholog of the developmental modifier, *medA*, complements the conidiation defect of the *Aspergillus nidulans* mutant. In : Fungal genetics and biology : FG & B, vol. 48, n° 4, p. 370–376. DOI: 10.1016/j.fgb.2010.12.008.

Coleman, Sean T.; Fang, Tung K.; Rovinsky, Sherry A.; Turano, Frank J.; Moye-Rowley, W. Scott (2001) Expression of a Glutamate Decarboxylase Homologue Is Required for Normal Oxidative Stress Tolerance in *Saccharomyces cerevisiae*. In : The Journal of biological chemistry, vol. 276, n° 1, p. 244–250. DOI: 10.1074/jbc.M007103200.

Connolly, Lanelle R.; Smith, Kristina M.; Freitag, Michael (2013) The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. In : PLoS genetics, vol. 9, n° 10, e1003916. DOI: 10.1371/journal.pgen.1003916.

Coradetti, Samuel T.; Craig, James P.; Xiong, Yi; Shock, Teresa; Tian, Chaoguang; Glass, N. Louise (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 109, n° 19, p. 7397–7402. DOI: 10.1073/pnas.1200785109.

Danecek, Petr; Bonfield, James K.; Liddle, Jennifer; Marshall, John; Ohan, Valeriu; Pollard, Martin O. et al. (2021) Twelve years of SAMtools and BCFtools. In : GigaScience, vol. 10, n° 2. DOI: 10.1093/gigascience/giab008.

Derntl, Christian; Mach-Aigner, Astrid R.; Mach, Robert L. (2014) *Trichoderma* Proteins with Disruption Activity on Cellulosic Substrates: Biotechnology and Biology of *Trichoderma*: Elsevier, p. 309–317.

Dillon, Shane C.; Zhang, Xing; Trievel, Raymond C.; Cheng, Xiaodong (2005) The SET-domain protein superfamily. Protein lysine methyltransferases. In : *Genome biology*, vol. 6, n° 8, p. 227. DOI: 10.1186/gb-2005-6-8-227.

Doench, John G.; Fusi, Nicolo; Sullender, Meagan; Hegde, Mudra; Vaimberg, Emma W.; Donovan, Katherine F. et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. In : *Nature biotechnology*, vol. 34, n° 2, p. 184–191. DOI: 10.1038/nbt.3437.

Drabik, Dušan; Ciaian, Pavel; Pokrivčák, Ján (2016) The effect of ethanol policies on the vertical price transmission in corn and food markets. In : *Energy Economics*, vol. 55, p. 189–199. DOI: 10.1016/j.eneco.2016.02.010.

Druzhinina, Irina; Kubicek, Christian P. (2005) Species concepts and biodiversity in *Trichoderma* and *Hypocrea*. From aggregate species to species clusters? In : *Journal of Zhejiang University. Science. B*, vol. 6, n° 2, p. 100–112. DOI: 10.1631/jzus.2005.B0100.

Druzhinina, Irina S.; Kopchinskiy, Alexey G.; Kubicek, Eva M.; Kubicek, Christian P. (2016) A complete annotation of the chromosomes of the cellulase producer *Trichoderma reesei* provides insights in gene clusters, their expression and reveals genes required for fitness. In : *Biotechnology for biofuels*, vol. 9, p. 75. DOI: 10.1186/s13068-016-0488-z.

Druzhinina, Irina S.; Kubicek, Christian P. (2017) Genetic engineering of *Trichoderma reesei* cellulases and their production. In : *Microbial biotechnology*, vol. 10, n° 6, p. 1485–1499. DOI: 10.1111/1751-7915.12726.

Du, Jiamu; Johnson, Lianna M.; Jacobsen, Steven E.; Patel, Dinshaw J. (2015) DNA methylation pathways and their crosstalk with histone methylation. In : *Nature reviews. Molecular cell biology*, vol. 16, n° 9, p. 519–532. DOI: 10.1038/nrm4043.

Dubey, Akanksha; Jeon, Junhyun (2017) Epigenetic regulation of development and pathogenesis in fungal plant pathogens. In : *Molecular plant pathology*, vol. 18, n° 6, p. 887–898. DOI: 10.1111/mpp.12499.

Dumesic, Phillip A.; Homer, Christina M.; Moresco, James J.; Pack, Lindsey R.; Shanle, Erin K.; Coyle, Scott M. et al. (2015) Product binding enforces the genomic specificity of a yeast polycomb repressive complex. In : *Cell*, vol. 160, n° 1-2, p. 204–218. DOI: 10.1016/j.cell.2014.11.039.

Edgar, Robert C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. In : *Nucleic acids research*, vol. 32, n° 5, p. 1792–1797. DOI: 10.1093/nar/gkh340.

Eveleigh, E.; Montenecourt, B. S. (1979) Increasing Yields of Extracellular Enzymes. In : D. Perlman, coord.: *Advances in Applied Microbiology*, 25^e tome: Elsevier (25).

Fan, Feiyu; Ma, Guoli; Li, Jingen; Liu, Qian; Benz, Johan Philipp; Tian, Chaoguang; Ma, Yanhe (2015) Genome-wide analysis of the endoplasmic reticulum stress response during lignocellulase production in *Neurospora crassa*. In : *Biotechnology for biofuels*, vol. 8, p. 66. DOI: 10.1186/s13068-015-0248-5.

Fargione, Joseph; Hill, Jason; Tilman, David; Polasky, Stephen; Hawthorne, Peter (2008) Land clearing and the biofuel carbon debt. In : Science (New York, N.Y.), vol. 319, n° 5867, p. 1235–1238. DOI: 10.1126/science.1152747.

Fischle, Wolfgang; Tseng, Boo Shan; Dormann, Holger L.; Ueberheide, Beatrix M.; Garcia, Benjamin A.; Shabanowitz, Jeffrey et al. (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. In : Nature, vol. 438, n° 7071, p. 1116–1122. DOI: 10.1038/nature04219.

Ford, Jason; Odeyale, Oluwafemi; Eskandar, Antonious; Kouba, Nafila; Shen, Chang-Hui (2007) A SWI/SNF- and INO80-dependent nucleosome movement at the INO1 promoter. In : Biochemical and biophysical research communications, vol. 361, n° 4, p. 974–979. DOI: 10.1016/j.bbrc.2007.07.109.

Foreman, Pamela K.; Brown, Doug; Dankmeyer, Lydia; Dean, Ralph; Diener, Stephen; Dunn-Coleman, Nigel S. et al. (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. In : The Journal of biological chemistry, vol. 278, n° 34, p. 31988–31997. DOI: 10.1074/jbc.M304750200.

Freitag, Michael (2017) Histone Methylation by SET Domain Proteins in Fungi. In : Annual review of microbiology, vol. 71, p. 413–439. DOI: 10.1146/annurev-micro-102215-095757.

Freitag, Michael; Selker, Eric U. (2005) Controlling DNA methylation: many roads to one modification. In : Current opinion in genetics & development, vol. 15, n° 2, p. 191–199. DOI: 10.1016/j.gde.2005.02.003.

Freitag, Michael; Williams, Rebecca L.; Kothe, Gregory O.; Selker, Eric U. (2002) A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 99, n° 13, p. 8802–8807. DOI: 10.1073/pnas.132212899.

Freitag M. (2014) Fungal Chromatin and Its Role in Regulation of Gene Expression. In : Karl Esser et Minou. Nowrousian, coord.: The mycota. A comprehensive treatise on fungi as experimental systems for basic and applied research, 13^e tome. 2nd edition. Berlin etc.: Springer, Chap. 5.

Gardiner-Garden, M.; Frommer, M. (1987) CpG Islands in vertebrate genomes. In : Journal of Molecular Biology, vol. 196, n° 2, p. 261–282. DOI: 10.1016/0022-2836(87)90689-9.

Geysens, Steven; Pakula, Tiina; Uusitalo, Jaana; Dewerte, Isabelle; Penttilä, Merja; Contreras, Roland (2005) Cloning and characterization of the glucosidase II alpha subunit gene of *Trichoderma reesei*: a frameshift mutation results in the aberrant glycosylation profile of the hypercellulolytic strain Rut-C30. In : Applied and environmental microbiology, vol. 71, n° 6, p. 2910–2924. DOI: 10.1128/AEM.71.6.2910-2924.2005.

Gladyshev, Eugene (2017) Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. In : Microbiology spectrum, vol. 5, n° 4. DOI: 10.1128/microbiolspec.FUNK-0042-2017.

Gladyshev, Eugene; Kleckner, Nancy (2014) Direct recognition of homology between double helices of DNA in *Neurospora crassa*. In : Nature communications, vol. 5, p. 3509. DOI: 10.1038/ncomms4509.

Gladyshev, Eugene; Kleckner, Nancy (2017) DNA sequence homology induces cytosine-to-thymine mutation by a heterochromatin-related pathway in *Neurospora*. In : Nature genetics, vol. 49, n° 6, p. 887–894. DOI: 10.1038/ng.3857.

Gottschling, Daniel E.; Aparicio, Oscar M.; Billington, Barbara L.; Zakian, Virginia A. (1990) Position effect at *S. cerevisiae* telomeres. Reversible repression of Pol II transcription. In : Cell, vol. 63, n° 4, p. 751–762. DOI: 10.1016/0092-8674(90)90141-z.

Gravelat, Fabrice N.; Ejzykowicz, Daniele E.; Chiang, Lisa Y.; Chabot, Josée C.; Urb, Mirjam; Macdonald, K. Denyese et al. (2010) *Aspergillus fumigatus* MedA governs adherence, host cell interactions and virulence. In : Cellular microbiology, vol. 12, n° 4, p. 473–488. DOI: 10.1111/j.1462-5822.2009.01408.x.

Gressel, J. B.; Hartmann, K. M. (1968) Morphogenesis in *Trichoderma*. Action spectrum of photoinduced sporulation. In : Planta, vol. 79, n° 3, p. 271–274. DOI: 10.1007/BF00396034.

Grewal, Shiv I. S.; Jia, Songtao (2007) Heterochromatin revisited. In : Nature reviews. Genetics, vol. 8, n° 1, p. 35–46. DOI: 10.1038/nrg2008.

Grognet, Pierre; Timpano, Hélène; Carlier, Florian; Aït-Benkhalil, Jinane; Berteaux-Lecellier, Véronique; Debuchy, Robert et al. (2019) A RID-like putative cytosine methyltransferase homologue controls sexual development in the fungus *Podospora anserina*. In : PLoS genetics, vol. 15, n° 8, e1008086. DOI: 10.1371/journal.pgen.1008086.

Gusakov, Alexander V. (2011) Alternatives to *Trichoderma reesei* in biofuel production. In : Trends in biotechnology, vol. 29, n° 9, p. 419–425. DOI: 10.1016/j.tibtech.2011.04.004.

Gutter, Y. (1957) Effect of light in sporulation of *Trichoderma viride*. In : Bulletin of the Research Council of Israel, vol. 5, p. 273–286.

Häkkinen, Mari; Valkonen, Mari J.; Westerholm-Parvinen, Ann; Aro, Nina; Arvas, Mikko; Vitikainen, Marika et al. (2014) Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production. In : Biotechnology for biofuels, vol. 7, n° 1, p. 14. DOI: 10.1186/1754-6834-7-14.

Hane, J. K.; Williams, A. H.; Taranto, A. P.; Solomon, P. S.; Oliver, R. P. (2015) Repeat-Induced Point Mutation: A Fungal-Specific, Endogenous Mutagenesis Process. In : Marco A. van den Berg et Karunakaran Maruthachalam, coord.: Genetic transformation systems in fungi. Cham: Springer (Fungal biology).

Hanumantha Rao, Kongara; Paul, Soumita; Ghosh, Swagata (2021) N-acetylglucosamine Signaling: Transcriptional Dynamics of a Novel Sugar Sensing Cascade in a Model Pathogenic Yeast, *Candida albicans*. In : Journal of fungi (Basel, Switzerland), vol. 7, n° 1. DOI: 10.3390/jof7010065.

Harkki, Anu; Mäntylä, Arja; Penttilä, Merja; Muttillainen, Susanna; Bühler, Rolf; Suominen, Pirkko et al. (1991) Genetic engineering of *Trichoderma* to produce strains with novel cellulase profiles. In : *Enzyme and microbial technology*, vol. 13, n° 3, p. 227–233. DOI: 10.1016/0141-0229(91)90133-u.

Harman, Gary E.; Kubicek, C. P. (1998) *Trichoderma And Gliocladium*, Volume 2 : CRC Press.

Havukainen, Sami; Valkonen, Mari; Koivuranta, Kari; Landowski, Christopher P. (2020) Studies on sugar transporter CRT1 reveal new characteristics that are critical for cellulase induction in *Trichoderma reesei*. In : *Biotechnology for biofuels*, vol. 13, p. 158. DOI: 10.1186/s13068-020-01797-7.

He, Ronglin; Ma, Lijuan; Li, Chen; Jia, Wendi; Li, Demao; Zhang, Dongyuan; Chen, Shulin (2014) TrpC1, a pH response transcription regulator, is involved in cellulase gene expression in *Trichoderma reesei*. In : *Enzyme and microbial technology*, vol. 67, p. 17–26. DOI: 10.1016/j.enzymtec.2014.08.013.

He, Yuehui; Michaels, Scott D.; Amasino, Richard M. (2003) Regulation of flowering time by histone acetylation in *Arabidopsis*. In : *Science (New York, N.Y.)*, vol. 302, n° 5651, p. 1751–1754. DOI: 10.1126/science.1091109.

Heard, Edith; Martienssen, Robert A. (2014) Transgenerational epigenetic inheritance: myths and mechanisms. In : *Cell*, vol. 157, n° 1, p. 95–109. DOI: 10.1016/j.cell.2014.02.045.

Hellman, Asaf; Chess, Andrew (2007) Gene body-specific methylation on the active X chromosome. In : *Science (New York, N.Y.)*, vol. 315, n° 5815, p. 1141–1143. DOI: 10.1126/science.1136352.

Henikoff, Steven (1990) Position-effect variegation after 60 years. In : *Trends in Genetics*, vol. 6, p. 422–426. DOI: 10.1016/0168-9525(90)90304-o.

Henikoff, Steven; Ahmad, Kami (2005) Assembly of variant histones into chromatin. In : *Annual review of cell and developmental biology*, vol. 21, p. 133–153. DOI: 10.1146/annurev.cellbio.21.012704.133518.

Herpoël-Gimbert, Isabelle; Margeot, Antoine; Dolla, Alain; Jan, Gwénaél; Mollé, Daniel; Lignon, Sabrina et al. (2008) Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. In : *Biotechnology for biofuels*, vol. 1, n° 1, p. 18. DOI: 10.1186/1754-6834-1-18.

Holliday, R.; Pugh, J. E. (1975) DNA modification mechanisms and gene activity during development. In : *Science*, vol. 187, n° 4173, p. 226–232.

Holstege, F. C.; Jennings, E. G.; Wyrick, J. J.; Lee, T. I.; Hengartner, C. J.; Green, M. R. et al. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. In : *Cell*, vol. 95, n° 5, p. 717–728. DOI: 10.1016/S0092-8674(00)81641-4.

Honda, Shinji; Lewis, Zachary A.; Huarte, Maite; Cho, Lucy Y.; David, Larry L.; Shi, Yang; Selker, Eric U. (2010) The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. In : *Genes & development*, vol. 24, n° 5, p. 443–454. DOI: 10.1101/gad.1893210.

Honda, Shinji; Selker, Eric U. (2008) Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in *Neurospora crassa*. In : *Molecular and cellular biology*, vol. 28, n° 19, p. 6044–6055. DOI: 10.1128/MCB.00823-08.

Hood, Michael E.; Katawczik, Melanie; Giraud, Tatiana (2005) Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. In : *Genetics*, vol. 170, n° 3, p. 1081–1089. DOI: 10.1534/genetics.105.042564.

Huang, Ruirui; Ding, Qiangqiang; Xiang, Yanan; Gu, Tingting; Li, Yi (2016) Comparative Analysis of DNA Methyltransferase Gene Family in Fungi. A Focus on Basidiomycota. In : *Frontiers in plant science*, vol. 7, p. 1556. DOI: 10.3389/fpls.2016.01556.

Huff, Jason T.; Zilberman, Daniel (2014) Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. In : *Cell*, vol. 156, n° 6, p. 1286–1297. DOI: 10.1016/j.cell.2014.01.029.

Hulsen, Tim; Vlieg, Jacob de; Alkema, Wynand (2008) BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. In : *BMC genomics*, vol. 9, p. 488. DOI: 10.1186/1471-2164-9-488.

Ilmén, M.; Thrane, C.; Penttilä, M. (1996) The glucose repressor gene *cre1* of *Trichoderma*. Isolation and expression of a full-length and a truncated mutant form. In : *Molecular & general genetics : MGG*, vol. 251, n° 4, p. 451–460. DOI: 10.1007/BF02172374.

Ivanova, Christa; Ramoni, Jonas; Aouam, Thiziri; Frischmann, Alexa; Seiboth, Bernhard; Baker, Scott E. et al. (2017) Genome sequencing and transcriptome analysis of *Trichoderma reesei* QM9978 strain reveals a distal chromosome translocation to be responsible for loss of *vib1* expression and loss of cellulase induction. In : *Biotechnology for biofuels*, vol. 10, n° 1, p. 962. DOI: 10.1186/s13068-017-0897-7.

Jamieson, Kirsty; Rountree, Michael R.; Lewis, Zachary A.; Stajich, Jason E.; Selker, Eric U. (2013) Regional control of histone H3 lysine 27 methylation in *Neurospora*. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, n° 15, p. 6027–6032. DOI: 10.1073/pnas.1303750110.

Jamieson, Kirsty; Wiles, Elizabeth T.; McNaught, Kevin J.; Sidoli, Simone; Leggett, Neena; Shao, Yanchun et al. (2016) Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. In : *Genome research*, vol. 26, n° 1, p. 97–107. DOI: 10.1101/gr.194555.115.

Jeon, Junhyun; Choi, Jaeyoung; Lee, Gir-Won; Park, Sook-Young; Huh, Aram; Dean, Ralph A.; Lee, Yong-Hwan (2015) Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. In : *Scientific reports*, vol. 5, p. 8567. DOI: 10.1038/srep08567.

Jourdiere, Etienne; Baudry, Lyam; Poggi-Parodi, Dante; Vicq, Yoan; Koszul, Romain; Margeot, Antoine et al. (2017) Proximity ligation scaffolding and comparison of two *Trichoderma reesei* strains genomes. In : *Biotechnology for biofuels*, vol. 10, p. 151. DOI: 10.1186/s13068-017-0837-6.

Jourdiere, Etienne; Poughon, Laurent; Larroche, Christian; Monot, Frédéric; Ben Chaabane, Fadhel (2012) A new stoichiometric miniaturization strategy for screening of industrial microbial strains: application to cellulase hyper-producing *Trichoderma reesei* strains. In : *Microbial cell factories*, vol. 11, p. 70. DOI: 10.1186/1475-2859-11-70.

Karimi-Aghchegh, Razieh; Bok, Jin Woo; Phatale, Pallavi A.; Smith, Kristina M.; Baker, Scott E.; Lichius, Alexander et al. (2013) Functional analyses of *Trichoderma reesei* LAE1 reveal conserved and contrasting roles of this regulator. In : *G3 (Bethesda, Md.)*, vol. 3, n° 2, p. 369–378. DOI: 10.1534/g3.112.005140.

Karpov, D. S.; Spasskaya, D. S.; Tutyaeva, V. V.; Karpov, V. L. (2022) Rpn4p without the DNA-Binding Domain Provides *Saccharomyces cerevisiae* Resistance to Oxidative Stress and Cycloheximide. In : *Molekuliarnaia biologiya*, vol. 56, n° 4, p. 628–641. DOI: 10.31857/S002689842204005X.

Kaya, Alaattin; Mariotti, Marco; Gladyshev, Vadim N. (2017) Cytochrome c peroxidase facilitates the beneficial use of H₂O₂ in prokaryotes. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, n° 33, p. 8678–8680. DOI: 10.1073/pnas.1710943114.

Koike, Hideaki; Aerts, Andrea; LaButti, Kurt; Grigoriev, Igor V.; Baker, Scott E. (2013) Comparative Genomics Analysis of *Trichoderma reesei* Strains. In : *Industrial Biotechnology*, vol. 9, n° 6, p. 352–367. DOI: 10.1089/ind.2013.0015.

Kredics, L.; Antal, Zsuzsanna; Dóczy, Ilona; Manczinger, L.; Kevei, F.; Nagy, Elisabeth (2003) Clinical importance of the genus *Trichoderma*. A review. In : *Acta microbiologica et immunologica Hungarica*, vol. 50, n° 2-3, p. 105–117. DOI: 10.1556/AMicr.50.2003.2-3.1.

Kubicek, Christian P.; Mikus, Marianna; Schuster, André; Schmoll, Monika; Seiboth, Bernhard (2009) Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. In : *Biotechnology for biofuels*, vol. 2, p. 19. DOI: 10.1186/1754-6834-2-19.

Kubicek, C., & Penttilä, M. (1998) Regulation of production of plant polysaccharide degrading enzymes by *Trichoderma*. In : Christian Peter Kubicek, Gary E. Harman et Kristen L. Ondik, coord.: *Trichoderma and Gliocladium*, 2^e tome. London, Bristol, PA: Taylor & Francis, p. 49–72.

Kuhls, K.; Lieckfeldt, E.; Samuels, G. J.; Kovacs, W.; Meyer, W.; Petrini, O. et al. (1996) Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, n° 15, p. 7755–7760. DOI: 10.1073/pnas.93.15.7755.

Kumar, Chitranshu; Sharma, Rakesh; Bachhawat, Anand Kumar (2003) Utilization of glutathione as an exogenous sulfur source is independent of gamma-glutamyl transpeptidase in the yeast *Saccharomyces cerevisiae*: evidence for an alternative glutathione degradation pathway. In : *FEMS microbiology letters*, vol. 219, n° 2, p. 187–194. DOI: 10.1016/S0378-1097(03)00059-4.

Lachner, Monika; Jenuwein, Thomas (2002) The many faces of histone lysine methylation. In : *Current Opinion in Cell Biology*, vol. 14, n° 3, p. 286–298. DOI: 10.1016/s0955-0674(02)00335-6.

Laetitia CHAN HO TONG (2017) Amélioration du champignon cellulolytique *Trichoderma reesei* par reproduction sexuée. Scientifique. Agro Paris Tech, Rueil Malmaison. IFP Energies nouvelles.

Landt, Stephen G.; Marinov, Georgi K.; Kundaje, Anshul; Kheradpour, Pouya; Pauli, Florencia; Batzoglou, Serafim et al. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. In : *Genome research*, vol. 22, n° 9, p. 1813–1831. DOI: 10.1101/gr.136184.111.

Langmead, Ben; Trapnell, Cole; Pop, Mihai; Salzberg, Steven L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. In : *Genome biology*, vol. 10, n° 3, R25. DOI: 10.1186/gb-2009-10-3-r25.

Le Crom, Stéphane; Schackwitz, Wendy; Pennacchio, Len; Magnuson, Jon K.; Culley, David E.; Collett, James R. et al. (2009) Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, n° 38, p. 16151–16156. DOI: 10.1073/pnas.0905848106.

Lee, Dong W.; Freitag, Michael; Selker, Eric U.; Aramayo, Rodolfo (2008) A cytosine methyltransferase homologue is essential for sexual development in *Aspergillus nidulans*. In : *PloS one*, vol. 3, n° 6, e2531. DOI: 10.1371/journal.pone.0002531.

Leslie, J. F.; Klein, K. K. (1996) Female fertility and mating type effects on effective population size and evolution in filamentous fungi. In : *Genetics*, vol. 144, n° 2, p. 557–567.

Lewis, Zachary A.; Honda, Shinji; Khlafallah, Tamir K.; Jeffress, Jennifer K.; Freitag, Michael; Mohn, Fabio et al. (2009) Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. In : *Genome research*, vol. 19, n° 3, p. 427–437. DOI: 10.1101/gr.086231.108.

Li, Heng; Handsaker, Bob; Wysoker, Alec; Fennell, Tim; Ruan, Jue; Homer, Nils et al. (2009) The Sequence Alignment/Map format and SAMtools. In : *Bioinformatics (Oxford, England)*, vol. 25, n° 16, p. 2078–2079. DOI: 10.1093/bioinformatics/btp352.

Li, Wan-Chen; Chen, Chia-Ling; Wang, Ting-Fang (2018) Repeat-induced point (RIP) mutation in the industrial workhorse fungus *Trichoderma reesei*. In : *Applied microbiology and biotechnology*, vol. 102, n° 4, p. 1567–1574. DOI: 10.1007/s00253-017-8731-5.

Li, Wan-Chen; Huang, Chien-Hao; Chen, Chia-Ling; Chuang, Yu-Chien; Tung, Shu-Yun; Wang, Ting-Fang (2017) *Trichoderma reesei* complete genome sequence, repeat-induced point mutation, and partitioning of CAZyme gene clusters. In : *Biotechnology for biofuels*, vol. 10, p. 170. DOI: 10.1186/s13068-017-0825-x.

Li, Xin-Liang; Skory, Christopher D.; Cotta, Michael A.; Puchart, Vladimir; Biely, Peter (2008) Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase

gene. In : *Applied and environmental microbiology*, vol. 74, n° 24, p. 7482–7489. DOI: 10.1128/AEM.00807-08.

Lichius, Alexander; Bidard, Frédérique; Buchholz, Franziska; Le Crom, Stéphane; Martin, Joel; Schackwitz, Wendy et al. (2015) Genome sequencing of the *Trichoderma reesei* QM9136 mutant identifies a truncation of the transcriptional regulator XYR1 as the cause for its cellulase-negative phenotype. In : *BMC genomics*, vol. 16, p. 326. DOI: 10.1186/s12864-015-1526-0.

Linke, Rita; Thallinger, Gerhard G.; Haarmann, Thomas; Eidner, Jasmin; Schreiter, Martina; Lorenz, Patrick et al. (2015) Restoration of female fertility in *Trichoderma reesei* QM6a provides the basis for inbreeding in this industrial cellulase producing fungus. In : *Biotechnology for biofuels*, vol. 8, p. 155. DOI: 10.1186/s13068-015-0311-2.

Liu, Kuimei; Dong, Yanmei; Wang, Fangzhong; Jiang, Baojie; Wang, Mingyu; Fang, Xu (2016) Regulation of cellulase expression, sporulation, and morphogenesis by velvet family proteins in *Trichoderma reesei*. In : *Applied microbiology and biotechnology*, vol. 100, n° 2, p. 769–779. DOI: 10.1007/s00253-015-7059-2.

Liu, Rui; Chen, Ling; Jiang, Yanping; Zhou, Zhihua; Zou, Gen (2015) Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. In : *Cell discovery*, vol. 1, p. 15007. DOI: 10.1038/celldisc.2015.7.

Liu, Rui; Chen, Ling; Jiang, Yanping; Zou, Gen; Zhou, Zhihua (2017) A novel transcription factor specifically regulates GH11 xylanase genes in *Trichoderma reesei*. In : *Biotechnology for biofuels*, vol. 10, p. 194. DOI: 10.1186/s13068-017-0878-x.

Liu, X. Shawn; Wu, Hao; Ji, Xiong; Stelzer, Yonatan; Wu, Xuebing; Czauderna, Szymon et al. (2016) Editing DNA Methylation in the Mammalian Genome. In : *Cell*, vol. 167, n° 1, 233-247.e17. DOI: 10.1016/j.cell.2016.08.056.

Love, Michael I.; Huber, Wolfgang; Anders, Simon (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. In : *Genome biology*, vol. 15, n° 12, p. 550. DOI: 10.1186/s13059-014-0550-8.

Luger, Karolin; Dechassa, Mekonnen L.; Tremethick, David J. (2012) New insights into nucleosome and chromatin structure. An ordered state or a disordered affair? In : *Nature reviews. Molecular cell biology*, vol. 13, n° 7, p. 436–447. DOI: 10.1038/nrm3382.

Lussier, M.; White, A. M.; Sheraton, J.; Di Paolo, T.; Treadwell, J.; Southard, S. B. et al. (1997) Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. In : *Genetics*, vol. 147, n° 2, p. 435–450. DOI: 10.1093/genetics/147.2.435.

Ma, Liang; Zhang, Jun; Zou, Gen; Wang, Chengshu; Zhou, Zhihua (2011) Improvement of cellulase activity in *Trichoderma reesei* by heterologous expression of a beta-glucosidase gene from *Penicillium decumbens*. In : *Enzyme and microbial technology*, vol. 49, n° 4, p. 366–371. DOI: 10.1016/j.enzmictec.2011.06.013.

Mach-Aigner, Astrid R.; Pucher, Marion E.; Steiger, Matthias G.; Bauer, Gudrun E.; Preis, Sonja J.; Mach, Robert L. (2008) Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. In : Applied and environmental microbiology, vol. 74, n° 21, p. 6554–6562. DOI: 10.1128/AEM.01143-08.

Maeshima, Kazuhiro; Hihara, Saera; Eltsov, Mikhail (2010) Chromatin structure. Does the 30-nm fibre exist in vivo? In : Current Opinion in Cell Biology, vol. 22, n° 3, p. 291–297. DOI: 10.1016/j.ceb.2010.03.001.

Malagnac, F.; Grégoire, A.; Goyon, C.; Rossignol, J. L.; Faugeron, G. (1999) Masc2, a gene from *Ascobolus* encoding a protein with a DNA-methyltransferase activity in vitro, is dispensable for in vivo methylation. In : Molecular microbiology, vol. 31, n° 1, p. 331–338. DOI: 10.1046/j.1365-2958.1999.01177.x.

Malagnac, Fabienne; Wendel, Birgit; Goyon, Christophe; Faugeron, Godeleine; Zickler, Denise; Rossignol, Jean-Luc et al. (1997) A Gene Essential for De Novo Methylation and Development in *Ascobolus* Reveals a Novel Type of Eukaryotic DNA Methyltransferase Structure. In : Cell, vol. 91, n° 2, p. 281–290. DOI: 10.1016/S0092-8674(00)80410-9.

Mandels, M.; Reese, E. T. (1957) Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals. In : Journal of bacteriology, vol. 73, n° 2, p. 269–278.

Manfrão-Netto, João Heitor C.; Mello-de-Sousa, Thiago M.; Mach-Aigner, Astrid R.; Mach, Robert L.; Poças-Fonseca, Marcio J. (2018) The DNA-methyltransferase inhibitor 5-aza-2-deoxycytidine affects *Humicola grisea* enzyme activities and the glucose-mediated gene repression. In : Journal of basic microbiology, vol. 58, n° 2, p. 144–153. DOI: 10.1002/jobm.201700415.

Mao, Xu-Ming; Xu, Wei; Li, Dehai; Yin, Wen-Bing; Chooi, Yit-Heng; Li, Yong-Quan et al. (2015) Epigenetic genome mining of an endophytic fungus leads to the pleiotropic biosynthesis of natural products. In : Angewandte Chemie (International ed. in English), vol. 54, n° 26, p. 7592–7596. DOI: 10.1002/anie.201502452.

Mao, Yuxin; Zhang, Zimei; Wong, Brian (2003) Use of green fluorescent protein fusions to analyse the N- and C-terminal signal peptides of GPI-anchored cell wall proteins in *Candida albicans*. In : Molecular microbiology, vol. 50, n° 5, p. 1617–1628. DOI: 10.1046/j.1365-2958.2003.03794.x.

Margeot, Antoine; Hahn-Hagerdal, Bärbel; Edlund, Maria; Slade, Raphael; Monot, Frédéric (2009) New improvements for lignocellulosic ethanol. In : Current Opinion in Biotechnology, vol. 20, n° 3, p. 372–380. DOI: 10.1016/j.copbio.2009.05.009.

Marie-Nelly, Hervé; Marbouty, Martial; Cournac, Axel; Flot, Jean-François; Liti, Gianni; Parodi, Dante Poggi et al. (2014) High-quality genome (re)assembly using chromosomal contact data. In : Nature communications, vol. 5, p. 5695. DOI: 10.1038/ncomms6695.

Markov, A. V.; Gusakov, A. V.; Kondratyeva, E. G.; Okunev, O. N.; Bekkarevich, A. O.; Sinitsyn, A. P. (2005) New effective method for analysis of the component composition of enzyme complexes from

Trichoderma reesei. In : Biochemistry. Biokhimiia, vol. 70, n° 6, p. 657–663. DOI: 10.1007/s10541-005-0166-4.

Martinez, Diego; Berka, Randy M.; Henrissat, Bernard; Saloheimo, Markku; Arvas, Mikko; Baker, Scott E. et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). In : Nature biotechnology, vol. 26, n° 5, p. 553–560. DOI: 10.1038/nbt1403.

McCann, Tyler S.; Guo, Yan; McDonald, W. Hayes; Tansey, William P. (2016) Antagonistic roles for the ubiquitin ligase Asr1 and the ubiquitin-specific protease Ubp3 in subtelomeric gene silencing. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 113, n° 5, p. 1309–1314. DOI: 10.1073/pnas.1518375113.

Melin, Petter; Schnürer, Johan; Wagner, E. Gerhart H. (2003) Characterization of *phiA*, a gene essential for phialide development in *Aspergillus nidulans*. In : Fungal genetics and biology : FG & B, vol. 40, n° 3, p. 234–241. DOI: 10.1016/s1087-1845(03)00108-7.

Mello-de-Sousa, Thiago M.; Gorsche, Rita; Rassinger, Alice; Poças-Fonseca, Marcio J.; Mach, Robert L.; Mach-Aigner, Astrid R. (2014) A truncated form of the Carbon catabolite repressor 1 increases cellulase production in *Trichoderma reesei*. In : Biotechnology for biofuels, vol. 7, n° 1, p. 129. DOI: 10.1186/s13068-014-0129-3.

Mello-de-Sousa, Thiago M.; Rassinger, Alice; Pucher, Marion E.; dos Santos Castro, Lilian; Persinoti, Gabriela F.; Silva-Rocha, Rafael et al. (2015) The impact of chromatin remodelling on cellulase expression in *Trichoderma reesei*. In : BMC genomics, vol. 16, p. 588. DOI: 10.1186/s12864-015-1807-7.

Mishra, Prashant K.; Baum, Mary; Carbon, John (2011) DNA methylation regulates phenotype-dependent transcriptional activity in *Candida albicans*. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 108, n° 29, p. 11965–11970. DOI: 10.1073/pnas.1109631108.

Mladek, Christina; Guger, Klaus; Hauser, Marie-Theres (2003) Identification and characterization of the ARIADNE gene family in *Arabidopsis*. A group of putative E3 ligases. In : Plant physiology, vol. 131, n° 1, p. 27–40. DOI: 10.1104/pp.012781.

Möller, Mareike; Habig, Michael; Lorrain, Cécile; Feurtay, Alice; Haueisen, Janine; Fagundes, Wagner C. et al. (2021) Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. In : PLoS genetics, vol. 17, n° 3, e1009448. DOI: 10.1371/journal.pgen.1009448.

Möller, Mareike; Schotanus, Klaas; Soyer, Jessica L.; Haueisen, Janine; Happ, Kathrin; Stralucke, Maja et al. (2019) Destabilization of chromosome structure by histone H3 lysine 27 methylation. In : PLoS genetics, vol. 15, n° 4, e1008093. DOI: 10.1371/journal.pgen.1008093.

Montenecourt, B. S.; Eveleigh, D. E. (1977) Preparation of mutants of *Trichoderma reesei* with enhanced cellulase production. In : Applied and environmental microbiology, vol. 34, n° 6, p. 777–782.

Montenecourt, Bland S.; Eveleigh, Douglas E. (1979) Selective Screening Methods for the Isolation of High Yielding Cellulase Mutants of *Trichoderma reesei*. In : Ross D. Brown et Lubo Jurasek, coord.: Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis, 181^e tome. WASHINGTON, D. C.: AMERICAN CHEMICAL SOCIETY (Advances in Chemistry), p. 289–301.

Morel, Mélanie; Ngadin, Andrew A.; Droux, Michel; Jacquot, Jean-Pierre; Gelhaye, Eric (2009) The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete *Phanerochaete chrysosporium*. In : Cellular and molecular life sciences : CMLS, vol. 66, n° 23, p. 3711–3725. DOI: 10.1007/s00018-009-0104-5.

Nakari-Setälä, T.; Penttilä, M. (1995) Production of *Trichoderma reesei* cellulases on glucose-containing media. In : Applied and environmental microbiology, vol. 61, n° 10, p. 3650–3655.

Neiman, A. M.; Mhaskar, V.; Manus, V.; Galibert, F.; Dean, N. (1997) *Saccharomyces cerevisiae* HOC1, a suppressor of *pkc1*, encodes a putative glycosyltransferase. In : Genetics, vol. 145, n° 3, p. 637–645. DOI: 10.1093/genetics/145.3.637.

Nikitin, Dmitri; Tosato, Valentina; Zavec, Apolonija Bedina; Bruschi, Carlo V. (2008) Cellular and molecular effects of nonreciprocal chromosome translocations in *Saccharomyces cerevisiae*. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 105, n° 28, p. 9703–9708. DOI: 10.1073/pnas.0800464105.

Nitta, Mikiko; Furukawa, Takanori; Shida, Yosuke; Mori, Kazuki; Kuhara, Satoru; Morikawa, Yasushi; Ogasawara, Wataru (2012) A new Zn(II)(2)Cys(6)-type transcription factor BglR regulates β -glucosidase expression in *Trichoderma reesei*. In : Fungal genetics and biology : FG & B, vol. 49, n° 5, p. 388–397. DOI: 10.1016/j.fgb.2012.02.009.

Oudet, P.; Gross-Bellard, M.; Chambon, P. (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. In : Cell, vol. 4, n° 4, p. 281–300. DOI: 10.1016/0092-8674(75)90149-X.

Paksa, Azadeh; Rajagopal, Jayaraj (2017) The epigenetic basis of cellular plasticity. In : Current Opinion in Cell Biology, vol. 49, p. 116–122. DOI: 10.1016/j.ceb.2018.01.003.

Park, Minhee; Keung, Albert J.; Khalil, Ahmad S. (2016) The epigenome. The next substrate for engineering. In : Genome biology, vol. 17, n° 1, p. 183. DOI: 10.1186/s13059-016-1046-5.

Patton-Vogt, J. L.; Henry, S. A. (1998) GIT1, a gene encoding a novel transporter for glycerophosphoinositol in *Saccharomyces cerevisiae*. In : Genetics, vol. 149, n° 4, p. 1707–1715. DOI: 10.1093/genetics/149.4.1707.

Perche, Pierre-Yves; Robert-Nicoud, Michel; Khochbin, Saadi; Vourc'h, Claire (2003) Différenciation du nucléosome: le rôle des variants de l'histone H2A. In : Med Sci (Paris), vol. 19, n° 11, p. 1137–1145. DOI: 10.1051/medsci/200319111137.

Perrin, Robyn M.; Fedorova, Natalie D.; Bok, Jin Woo; Cramer, Robert A.; Wortman, Jennifer R.; Kim, H. Stanley et al. (2007) Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by *LaeA*. In : PLoS pathogens, vol. 3, n° 4, e50. DOI: 10.1371/journal.ppat.0030050.

Persoon, C. (1794) *Disposita methodica fungorum*. In : Römer's Neues Magazin der Botanik, vol. 1, p. 81–128.

Pfannenstiel, Brandon T.; Keller, Nancy P. (2019) On top of biosynthetic gene clusters. How epigenetic machinery influences secondary metabolism in fungi. In : Biotechnology advances, vol. 37, n° 6, p. 107345. DOI: 10.1016/j.biotechadv.2019.02.001.

Pham, Kieu Thi Minh; Inoue, Yoshihiro; van Vu, Ba; Nguyen, Hanh Hieu; Nakayashiki, Toru; Ikeda, Ken-Ichi; Nakayashiki, Hitoshi (2015) MoSET1 (Histone H3K4 Methyltransferase in *Magnaporthe oryzae*) Regulates Global Gene Expression during Infection-Related Morphogenesis. In : PLoS genetics, vol. 11, n° 7, e1005385. DOI: 10.1371/journal.pgen.1005385.

Pirayre, Aurélie; Duval, Laurent; Blugeon, Corinne; Firmo, Cyril; Perrin, Sandrine; Jourdier, Etienne et al. (2020) Glucose-lactose mixture feeds in industry-like conditions: a gene regulatory network analysis on the hyperproducing *Trichoderma reesei* strain Rut-C30. In : BMC genomics, vol. 21, n° 1, p. 885. DOI: 10.1186/s12864-020-07281-8.

Pócsi, István; Prade, Rolf A.; Penninckx, Michel J. (2004) Glutathione, altruistic metabolite in fungi. In : Advances in microbial physiology, vol. 49, p. 1–76. DOI: 10.1016/S0065-2911(04)49001-8.

Poggi-Parodi, Dante; Bidard, Frédérique; Pirayre, Aurélie; Portnoy, Thomas; Blugeon, Corinne; Seiboth, Bernhard et al. (2014) Kinetic transcriptome analysis reveals an essentially intact induction system in a cellulase hyper-producer *Trichoderma reesei* strain. In : Biotechnology for biofuels, vol. 7, n° 1, p. 173. DOI: 10.1186/s13068-014-0173-z.

Portnoy, Thomas; Margeot, Antoine; Linke, Rita; Atanasova, Lea; Fekete, Erzsébet; Sándor, Erzsébet et al. (2011) The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. In : BMC genomics, vol. 12, p. 269. DOI: 10.1186/1471-2164-12-269.

Portnoy, Thomas; Margeot, Antoine; Seidl-Seiboth, Verena; Le Crom, Stéphane; Ben Chaabane, Fadhel; Linke, Rita et al. (2011) Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. In : Eukaryotic cell, vol. 10, n° 2, p. 262–271. DOI: 10.1128/EC.00208-10.

Puyesky, M.; Benhamou, N.; Noyola, P. P.; Bauw, G.; Ziv, T.; van Montagu, M. et al. (1999) Developmental regulation of *cmp1*, a gene encoding a multidomain conidiospore surface protein of *Trichoderma*. In : Fungal genetics and biology : FG & B, vol. 27, n° 1, p. 88–99. DOI: 10.1006/fgbi.1999.1134.

R Core Team (2022) R. R: A language and environment for statistical computing. : Foundation for Statistical Computing. En ligne : <https://www.R-project.org/>.

Reik, Wolf; Lewis, Annabelle (2005) Co-evolution of X-chromosome inactivation and imprinting in mammals. In : Nature reviews. Genetics, vol. 6, n° 5, p. 403–410. DOI: 10.1038/nrg1602.

Ren, Kang; Mou, Ya-Ni; Tong, Sen-Miao; Ying, Sheng-Hua; Feng, Ming-Guang (2021) DIM5/KMT1 controls fungal insect pathogenicity and genome stability by methylation of histone H3K4, H3K9 and H3K36. In : Virulence, vol. 12, n° 1, p. 1306–1322. DOI: 10.1080/21505594.2021.1923232.

Reyes-Dominguez, Yazmid; Bok, Jin Woo; Berger, Harald; Shwab, E. Keats; Basheer, Asjad; Gallmetzer, Andreas et al. (2010) Heterochromatic marks are associated with the repression of secondary metabolism clusters in *Aspergillus nidulans*. In : Molecular microbiology, vol. 76, n° 6, p. 1376–1386. DOI: 10.1111/j.1365-2958.2010.07051.x.

Ries, L.; Belshaw, N. J.; Ilmén, M.; Penttilä, M. E.; Alapuranen, M.; Archer, D. B. (2014) The role of CRE1 in nucleosome positioning within the *cbh1* promoter and coding regions of *Trichoderma reesei*. In : Applied microbiology and biotechnology, vol. 98, n° 2, p. 749–762. DOI: 10.1007/s00253-013-5354-3.

Riggs, A. D. (1975) X inactivation, differentiation, and DNA methylation. In : Cytogenetics and cell genetics, vol. 14, n° 1, p. 9–25. DOI: 10.1159/000130315.

Robinson, James T.; Thorvaldsdóttir, Helga; Winckler, Wendy; Guttman, Mitchell; Lander, Eric S.; Getz, Gad; Mesirov, Jill P. (2011) Integrative genomics viewer. In : Nature biotechnology, vol. 29, n° 1, p. 24–26. DOI: 10.1038/nbt.1754.

Robyn Peterson (2012) *Trichoderma reesei* RUT-C30--thirty years of strain improvement. In : Microbiology (Reading), Volume 158, Issue 1, 1 janvier 2012, p. 58–68, consulté le 04/10/21.

Rory Starkrory.stark@cruk.cam.ac.uk, Gord Browngdbzork@gmail.com (2017) DiffBind : Bioconductor.

Roth, S. Y.; Denu, J. M.; Allis, C. D. (2001) Histone acetyltransferases. In : Annual review of biochemistry, vol. 70, p. 81–120. DOI: 10.1146/annurev.biochem.70.1.81.

Rountree, M. R.; Selker, E. U. (1997) DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. In : Genes & development, vol. 11, n° 18, p. 2383–2395. DOI: 10.1101/gad.11.18.2383.

Rountree, M. R.; Selker, E. U. (2010) DNA methylation and the formation of heterochromatin in *Neurospora crassa*. In : Heredity, vol. 105, n° 1, p. 38–44. DOI: 10.1038/hdy.2010.44.

Sakamoto, Kazutoshi; Arima, Toshihide; Iwashita, Kazuhiro; Yamada, Osamu; Gomi, Katsuya; Akita, Osamu (2008) *Aspergillus oryzae* atfB encodes a transcription factor required for stress tolerance in conidia. In : Fungal genetics and biology : FG & B, vol. 45, n° 6, p. 922–932. DOI: 10.1016/j.fgb.2008.03.009.

Sandrine Augui et Edith Heard (2008) Inactivation du chromosome X. Comment une cellule sait compter jusqu'à deux X. In : Med Sci (Paris), vol. 24, n° Number 6-7, Juin-Juillet, 584 - 585, consulté le 4 octobre 2021.

Savas, Sevtap; Skardasi, Georgia (2018) The SWI/SNF complex subunit genes: Their functions, variations, and links to risk and survival outcomes in human cancers. In : Critical reviews in oncology/hematology, vol. 123, p. 114–131. DOI: 10.1016/j.critrevonc.2018.01.009.

Schmoll, Monika (2018) Regulation of plant cell wall degradation by light in *Trichoderma*. In : Fungal biology and biotechnology, vol. 5, p. 10. DOI: 10.1186/s40694-018-0052-7.

Schmoll, Monika; Esquivel-Naranjo, Edgardo Ulises; Herrera-Estrella, Alfredo (2010) *Trichoderma* in the light of day--physiology and development. In : Fungal genetics and biology : FG & B, vol. 47, n° 11, p. 909–916. DOI: 10.1016/j.fgb.2010.04.010.

Schmoll, Monika; Kubicek, Christian P. (2005) *ooc1*, a unique gene expressed only during growth of *Hypocrea jecorina* (anamorph: *Trichoderma reesei*) on cellulose. In : Current genetics, vol. 48, n° 2, p. 126–133. DOI: 10.1007/s00294-005-0585-1.

Schmoll, Monika; Zeilinger, Susanne; Mach, Robert L.; Kubicek, Christian P. (2004) Cloning of genes expressed early during cellulase induction in *Hypocrea jecorina* by a rapid subtraction hybridization approach. In : Fungal genetics and biology : FG & B, vol. 41, n° 9, p. 877–887. DOI: 10.1016/j.fgb.2004.06.002.

Schulz, W. A.; Steinhoff, C.; Florl, A. R. (2006) Methylation of endogenous human retroelements in health and disease. In : Current topics in microbiology and immunology, vol. 310, p. 211–250. DOI: 10.1007/3-540-31181-5_11.

Seiboth, Bernhard; Gamauf, Christian; Pail, Manuela; Hartl, Lukas; Kubicek, Christian P. (2007) The D-xylose reductase of *Hypocrea jecorina* is the major aldose reductase in pentose and D-galactose catabolism and necessary for beta-galactosidase and cellulase induction by lactose. In : Molecular microbiology, vol. 66, n° 4, p. 890–900. DOI: 10.1111/j.1365-2958.2007.05953.x.

Seiboth, Bernhard; Ivanova, Christa; Seidl-Seiboth, Verena (2011) *Trichoderma reesei*. A Fungal Enzyme Producer for Cellulosic Biofuels. In : Marco Aurelio Dos Santos Bernardes, coord.: Biofuel Production-Recent Developments and Prospects: InTech.

Seiboth, Bernhard; Karimi, Razieh Aghcheh; Phatale, Pallavi A.; Linke, Rita; Hartl, Lukas; Sauer, Dominik G. et al. (2012) The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*. In : Molecular microbiology, vol. 84, n° 6, p. 1150–1164. DOI: 10.1111/j.1365-2958.2012.08083.x.

Seidl, Verena; Seibel, Christian; Kubicek, Christian P.; Schmoll, Monika (2009) Sexual development in the industrial workhorse *Trichoderma reesei*. In : Proceedings of the National Academy of Sciences of the United States of America, vol. 106, n° 33, p. 13909–13914. DOI: 10.1073/pnas.0904936106.

Selker, Eric U.; Cambareri, Edward B.; Jensen, Bryan C.; Haack, Kenneth R. (1987) Rearrangement of duplicated DNA in specialized cells of *Neurospora*. In : Cell, vol. 51, n° 5, p. 741–752. DOI: 10.1016/0092-8674(87)90097-3.

Selker, Eric U.; Tountas, Nikolaos A.; Cross, Sally H.; Margolin, Brian S.; Murphy, Jonathan G.; Bird, Adrian P.; Freitag, Michael (2003) The methylated component of the *Neurospora crassa* genome. In : *Nature*, vol. 422, n° 6934, p. 893–897. DOI: 10.1038/nature01564.

Sharma, Shikhar; Kelly, Theresa K.; Jones, Peter A. (2010) Epigenetics in cancer. In : *Carcinogenesis*, vol. 31, n° 1, p. 27–36. DOI: 10.1093/carcin/bgp220.

Shaver, Scott; Casas-Mollano, J. Armando; Cerny, Ronald L.; Cerutti, Heriberto (2010) Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*. In : *Epigenetics*, vol. 5, n° 4, p. 301–312. DOI: 10.4161/epi.5.4.11608.

Shen, X.; Mizuguchi, G.; Hamiche, A.; Wu, C. (2000) A chromatin remodelling complex involved in transcription and DNA processing. In : *Nature*, vol. 406, n° 6795, p. 541–544. DOI: 10.1038/35020123.

Shida, Yosuke; Furukawa, Takanori; Ogasawara, Wataru (2016) Deciphering the molecular mechanisms behind cellulase production in *Trichoderma reesei*, the hyper-cellulolytic filamentous fungus. In : *Bioscience, biotechnology, and biochemistry*, vol. 80, n° 9, p. 1712–1729. DOI: 10.1080/09168451.2016.1171701.

Shogren-Knaak, Michael; Ishii, Haruhiko; Sun, Jian-Min; Pazin, Michael J.; Davie, James R.; Peterson, Craig L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. In : *Science* (New York, N.Y.), vol. 311, n° 5762, p. 844–847. DOI: 10.1126/science.1124000.

Shwab, E. Keats; Bok, Jin Woo; Tribus, Martin; Galehr, Johannes; Graessle, Stefan; Keller, Nancy P. (2007) Histone deacetylase activity regulates chemical diversity in *Aspergillus*. In : *Eukaryotic cell*, vol. 6, n° 9, p. 1656–1664. DOI: 10.1128/EC.00186-07.

Silar, Philippe (2019) Phenotypic instability in fungi. In : *Advances in applied microbiology*, vol. 107, p. 141–187. DOI: 10.1016/bs.aambs.2019.03.002.

Singer, M. J.; Marcotte, B. A.; Selker, E. U. (1995) DNA methylation associated with repeat-induced point mutation in *Neurospora crassa*. In : *Molecular and cellular biology*, vol. 15, n° 10, p. 5586–5597. DOI: 10.1128/mcb.15.10.5586.

Singer, M. J.; Selker, E. U. (1995) Genetic and epigenetic inactivation of repetitive sequences in *Neurospora crassa*: RIP, DNA methylation, and quelling. In : *Current topics in microbiology and immunology*, vol. 197, p. 165–177. DOI: 10.1007/978-3-642-79145-1_11.

Sloothaak, Jasper; Tamayo-Ramos, Juan Antonio; Odoni, Dorette I.; Laothanachareon, Thanaporn; Derntl, Christian; Mach-Aigner, Astrid R. et al. (2016) Identification and functional characterization of novel xylose transporters from the cell factories *Aspergillus niger* and *Trichoderma reesei*. In : *Biotechnology for biofuels*, vol. 9, p. 148. DOI: 10.1186/s13068-016-0564-4.

Springer, Nathan M.; Schmitz, Robert J. (2017) Exploiting induced and natural epigenetic variation for crop improvement. In : *Nature reviews. Genetics*, vol. 18, n° 9, p. 563–575. DOI: 10.1038/nrg.2017.45.

Steyaert, Johanna M.; Weld, Richard J.; Mendoza-Mendoza, Artemio; Stewart, Alison (2010) Reproduction without sex. Conidiation in the filamentous fungus *Trichoderma*. In : *Microbiology* (Reading, England), vol. 156, n° Pt 10, p. 2887–2900. DOI: 10.1099/mic.0.041715-0.

Strauss, Joseph; Reyes-Dominguez, Yazmid (2011) Regulation of secondary metabolism by chromatin structure and epigenetic codes. In : *Fungal genetics and biology : FG & B*, vol. 48, n° 1, p. 62–69. DOI: 10.1016/j.fgb.2010.07.009.

Strick, R.; Strissel, P. L.; Gavrilov, K.; Levi-Setti, R. (2001) Cation-chromatin binding as shown by ion microscopy is essential for the structural integrity of chromosomes. In : *The Journal of cell biology*, vol. 155, n° 6, p. 899–910. DOI: 10.1083/jcb.200105026.

Stricker, Astrid R.; Grosstessner-Hain, Karin; Würleitner, Elisabeth; Mach, Robert L. (2006) Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. In : *Eukaryotic cell*, vol. 5, n° 12, p. 2128–2137. DOI: 10.1128/EC.00211-06.

Studt, Lena; Rösler, Sarah M.; Burkhardt, Immo; Arndt, Birgit; Freitag, Michael; Humpf, Hans-Ulrich et al. (2016) Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*. In : *Environmental microbiology*, vol. 18, n° 11, p. 4037–4054. DOI: 10.1111/1462-2920.13427.

Sukumaran, Rajeev K.; Christopher, Meera; Kooloth-Valappil, Prajeesh; Sreeja-Raju, AthiraRaj; Mathew, Reshma M.; Sankar, Meena et al. (2021) Addressing challenges in production of cellulases for biomass hydrolysis: Targeted interventions into the genetics of cellulase producing fungi. In : *Bioresource Technology*, vol. 329, p. 124746. DOI: 10.1016/j.biortech.2021.124746.

Tamaru, H.; Selker, E. U. (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. In : *Nature*, vol. 414, n° 6861, p. 277–283. DOI: 10.1038/35104508.

Tavaria, Michael; Gabriele, Tim; Kola, Ismail; Anderson, Robin L. (1996) A hitchhiker's guide to the human Hsp70 family. In : *Cell Stress & Chaperones*, vol. 1, n° 1, p. 23. DOI: 10.1379/1466-1268(1996)001<0023:ahsgtt>2.3.co;2.

Toral, Laura; Rodríguez, Miguel; Martínez-Checa, Fernando; Montaña, Alfredo; Cortés-Delgado, Amparo; Smolinska, Agnieszka et al. (2021) Identification of Volatile Organic Compounds in Extremophilic Bacteria and Their Effective Use in Biocontrol of Postharvest Fungal Phytopathogens. In : *Frontiers in microbiology*, vol. 12, p. 773092. DOI: 10.3389/fmicb.2021.773092.

Torres-Garcia, Sito; Yaseen, Imtiyaz; Shukla, Manu; Audergon, Pauline N. C. B.; White, Sharon A.; Pidoux, Alison L.; Allshire, Robin C. (2020) Epigenetic gene silencing by heterochromatin primes fungal resistance. In : *Nature*, vol. 585, n° 7825, p. 453–458. DOI: 10.1038/s41586-020-2706-x.

van Vu, Ba; Pham, Kieu Thi Minh; Nakayashiki, Hitoshi (2013) Substrate-induced transcriptional activation of the MoCel7C cellulase gene is associated with methylation of histone H3 at lysine 4 in the rice blast fungus *Magnaporthe oryzae*. In : *Applied and environmental microbiology*, vol. 79, n° 21, p. 6823–6832. DOI: 10.1128/AEM.02082-13.

Viikari, L.; Vehmaanperä, J.; Koivula, A. (2012) Lignocellulosic ethanol. From science to industry. In : Biomass and Bioenergy, vol. 46, p. 13–24. DOI: 10.1016/j.biombioe.2012.05.008.

Vitikainen, Marika; Arvas, Mikko; Pakula, Tiina; Oja, Merja; Penttilä, Merja; Saloheimo, Markku (2010) Array comparative genomic hybridization analysis of *Trichoderma reesei* strains with enhanced cellulase production properties. In : BMC genomics, vol. 11, p. 441. DOI: 10.1186/1471-2164-11-441.

Wang, Lei; Liu, Jialong; Li, Xiaotong; Lyu, Xinxing; Liu, Zhizhen; Zhao, Hong et al. (2022) A histone H3K9 methyltransferase Dim5 mediates repression of sorbicillinoid biosynthesis in *Trichoderma reesei*. In : Microbial biotechnology, vol. 15, n° 10, p. 2533–2546. DOI: 10.1111/1751-7915.14103.

Wang, Lei; Lv, Xinxing; Cao, Yanli; Zheng, Fanglin; Meng, Xiangfeng; Shen, Yu et al. (2019) A novel transcriptional regulator RXE1 modulates the essential transactivator XYR1 and cellulase gene expression in *Trichoderma reesei*. In : Applied microbiology and biotechnology, vol. 103, n° 11, p. 4511–4523. DOI: 10.1007/s00253-019-09739-6.

Watson, James Dewey; Baker, Tania A.; Bell, Stephen P.; Gann, Alexander; Levine, Michael; Losick, Richard. (2014) Molecular biology of the gene. 7th edition. Boston, Paris etc., New York (N.Y.) : Pearson; Cold Spring Harbor laboratory press.

Weigel, Detlef; Colot, Vincent (2012) Epialleles in plant evolution. In : Genome biology, vol. 13, n° 10, p. 249. DOI: 10.1186/gb-2012-13-10-249.

Williams, Russell B.; Henrikson, Jon C.; Hoover, Ashley R.; Lee, Andriynn E.; Cichewicz, Robert H. (2008) Epigenetic remodeling of the fungal secondary metabolome. In : Organic & biomolecular chemistry, vol. 6, n° 11, p. 1895–1897. DOI: 10.1039/b804701d.

Xu, Jintao; Zhao, Guolei; Kou, Yanbo; Zhang, Weixin; Zhou, Qingxin; Chen, Guanjin; Liu, Weifeng (2014) Intracellular β -glucosidases CEL1a and CEL1b are essential for cellulase induction on lactose in *Trichoderma reesei*. In : Eukaryotic cell, vol. 13, n° 8, p. 1001–1013. DOI: 10.1128/EC.00100-14.

Yamamoto, Ryota; Sato, Genki; Amai, Takamitsu; Ueda, Mitsuyoshi; Kuroda, Kouichi (2022) Development of Artificial System to Induce Chromatin Loosening in *Saccharomyces cerevisiae*. In : Biomolecules, vol. 12, n° 8. DOI: 10.3390/biom12081138.

Yao, Guangshan; Li, Zhonghai; Wu, Ruimei; Qin, Yuqi; Liu, Guodong; Qu, Yinbo (2016) *Penicillium oxalicum* PoFlbC regulates fungal asexual development and is important for cellulase gene expression. In : Fungal genetics and biology : FG & B, vol. 86, p. 91–102. DOI: 10.1016/j.fgb.2015.12.012.

Yu, Jiujiang; Chang, Perng-Kuang; Ehrlich, Kenneth C.; Cary, Jeffrey W.; Bhatnagar, Deepak; Cleveland, Thomas E. et al. (2004) Clustered pathway genes in aflatoxin biosynthesis. In : Applied and environmental microbiology, vol. 70, n° 3, p. 1253–1262. DOI: 10.1128/AEM.70.3.1253-1262.2004.

Yun, S. J.; Hiraoka, Y.; Nishizawa, M.; Takio, K.; Titani, K.; Nogi, Y.; Fukasawa, T. (1991) Purification and characterization of the yeast negative regulatory protein GAL80. In : The Journal of biological chemistry, vol. 266, n° 2, p. 693–697.

Zeilinger, S.; Ebner, A.; Marosits, T.; Mach, R.; Kubicek, C. P. (2001) The Hypocrea jecorina HAP 2/3/5 protein complex binds to the inverted CCAAT-box (ATTGG) within the cbh2 (cellobiohydrolase II-gene) activating element. In : Molecular genetics and genomics : MGG, vol. 266, n° 1, p. 56–63. DOI: 10.1007/s004380100518.

Zeilinger, S.; Schmoll, M.; Pail, M.; Mach, R. L.; Kubicek, C. P. (2003) Nucleosome transactions on the Hypocrea jecorina (*Trichoderma reesei*) cellulase promoter cbh2 associated with cellulase induction. In : Molecular genetics and genomics : MGG, vol. 270, n° 1, p. 46–55. DOI: 10.1007/s00438-003-0895-2.

Zemach, Assaf; McDaniel, Ivy E.; Silva, Pedro; Zilberman, Daniel (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. In : Science (New York, N.Y.), vol. 328, n° 5980, p. 916–919. DOI: 10.1126/science.1186366.

Zhang, Weixin; Kou, Yanbo; Xu, Jintao; Cao, Yanli; Zhao, Guolei; Shao, Jing et al. (2013) Two major facilitator superfamily sugar transporters from *Trichoderma reesei* and their roles in induction of cellulase biosynthesis. In : The Journal of biological chemistry, vol. 288, n° 46, p. 32861–32872. DOI: 10.1074/jbc.M113.505826.

Zhang, Xiujun; Qu, Yinbo; Qin, Yuqi (2016) Expression and chromatin structures of cellulolytic enzyme gene regulated by heterochromatin protein 1. In : Biotechnology for biofuels, vol. 9, p. 206. DOI: 10.1186/s13068-016-0624-9.

Zhang, Yong; Liu, Tao; Meyer, Clifford A.; Eeckhoutte, Jérôme; Johnson, David S.; Bernstein, Bradley E. et al. (2008) Model-based analysis of ChIP-Seq (MACS). In : Genome biology, vol. 9, n° 9, R137. DOI: 10.1186/gb-2008-9-9-r137.

Zheng, Fanglin; Cao, Yanli; Wang, Lei; Lv, Xinxing; Meng, Xiangfeng; Zhang, Weixin et al. (2017) The mating type locus protein MAT1-2-1 of *Trichoderma reesei* interacts with Xyr1 and regulates cellulase gene expression in response to light. In : Scientific reports, vol. 7, n° 1, p. 17346. DOI: 10.1038/s41598-017-17439-2.

Annexes

Tous les milieux sont autoclavés 20 minutes à 120 °C sauf mention contraire.

I. Milieux pour cultures sur plaques

Milieux PD Agar et PD

	Pour 1 L
PD (Potatoes Dextrose)	24 g
PDA (Potatoes Dextrose Agar)	39g

Les antibiotiques sont ajoutés après stérilisation.

Milieux LB Agar et LB

	Pour 1 L
LB (Lysogeny Broth)	24 g
LBA (Lysogeny Broth Agar)	39g

Les antibiotiques sont ajoutés après stérilisation.

Milieu Minimum (MM) gélosé

	Pour 1L
KH ₂ PO ₄	5 g
(NH ₄)SO ₄	5 g
Acide maléique (100mM final)	11.6 g
Source de carbone	20 g
Agar	20 g
MgSO ₄ (1M)	5 mL
CaCl ₂ (1M)	5 mL

Les concentrations des composés ajoutés au milieu minimum sont les suivantes :

Glucose (2%), Lactose (2%), Sorbitol (2%), Cellulose (2%), Xylose (2%), AZCL (0,1%), NaCl (0,5M), Ménadione (30 µM), Calcofluor white (100 mM).

Milieu EJ (liquide)

	Pour 1L final
(NH ₄) ₂ SO ₄	5,6 g
K ₂ HPO ₄	4,4 g
MgSO ₄ , 7H ₂ O	0,3 g
CaCl ₂ , 2H ₂ O	0,15 g
Tampon BTCA	5,85 g
KOH en cristaux	3 g
Cornsteep (Solulys)	1,5 g
Solution d'oligos éléments	1 mL

Source de carbone :

Lactose monohydrate	12,5 g
Glucose monohydrate	13,7 g

Pour induction de production de cellulase :

TechnoCel BH200 (Fibres de 200 µm)	12,5 g
------------------------------------	--------

Ajuster le pH à 6,0 avec de la soude à 30%.

Stérilisation 30 min à 110°C.

II. Solutions pour formation de protoplastes, transformation et conservation

Solution CTS50 :

	Pour 1 L
Saccharose	13,7 g
Tris-HCl 1 M pH 7,5	10 mL
CaCl ₂ 1 M	5 mL
H ₂ O milliQ	qsp 1 litre

Le milieu est stérilisé par autoclavage (30 minutes à 110 °C) ou filtration.

Solution CTS10 :

	Pour 1 L
Saccharose	136,9 g
Tris-HCl 1 M pH 7,5	100 mL
CaCl ₂ 1 M	10 mL
H ₂ O milliQ	qsp 1 litre

Le milieu est stérilisé par autoclavage (30 minutes à 110 °C) ou filtration.

Tampon KPAm :

	Pour 1 L
KH ₂ PO ₄ PM 136.1 g/mol	3,4 g
(NH ₄) ₂ SO ₄ PM 132.1 g/mol	79,3 g
H ₂ O milliQ	qsp 1 litre

pH ajusté à 5,8. Stérilisation par filtration sur filtre de 0,22 µm.

Solution PEG :

	Pour 1 L
PEG 4000	60 g
CaCl ₂ 1M	5 mL
Tris HCl 1M pH 7.5	1 mL
H ₂ O milliQ	qsp 100 mL

III. Milieux et Solutions pour cultures en fioles alimentées

Milieu salin F45

Ce milieu est préparé pour un volume final de 1L après l'ajout de la solution de glucose concentrée donc le volume ici est ajusté à 950 mL.

	Pour 1 L Final
H ₃ PO ₄ à 85%	1 mL
CaCl ₂ ·2H ₂ O	150 mg
MgSO ₄ ·7H ₂ O	300 mg
Solution d'oligo-éléments	1 mL
(NH ₄) ₂ SO ₄	2,8 g
H ₂ O distillée	Qsp 950 mL

Milieu salin F45 complet

Dans 475 mL de milieu salin F45, ajouter:

Tampon dipotassium phthalate	4 g
Cornsteep	750 mg

Ajuster le pH à 5,4 avec de la soude à 30%. Stériliser à l'autoclave 20 min à 121°C.

Solution de glucose à 250 g.L⁻¹

Glucose monohydrate	27,5 g
H ₂ O distillée	Qsp 100 mL

Stériliser 30 min à 110°C.

Solution de lactose pour fed-batch à 53,3 g.L⁻¹

Lactose monohydrate	16 g
H ₂ O distillée	Qsp 300 mL

Stériliser 30 min à 110°C.

Solution d'urée à 250 g.L⁻¹

Urée	10 g
H ₂ O distillée	Qsp 40 mL

Stériliser par filtration sur filtre de 0,22 µm.

Solution d'oligo-éléments

	Pour 1 L
FeSO ₄ ·7H ₂ O	30 g
Co(NO ₃) ₂ ·6H ₂ O	9 g
MnSO ₄ ·7H ₂ O	6,4 g
ZnSO ₄ ·7H ₂ O	8,4 g
H ₃ PO ₄ 85 %	Jusqu'à pH 1,8

IV. Solutions pour l'expérience d'immunoprécipitation de la chromatine

Phosphate-buffered saline (PBS)

	Concentration finale (1X)	Pour 1 L
NaCl	137 mM	8 g
KCl	2.7 mM	0.2 g
Na ₂ HPO ₄	10 mM	1.44 g
KH ₂ PO ₄	1.8 mM	0.24 g

Ajuster le pH à 7,4.

Tampon de lyse (4°C)

	Concentration solution mère (M)	Concentration solution finale (M)	Vol. (mL)	Vol. (mL)
Hepes pH 7.5	1	0,05	0,750	2,500
NaCl	3	0,14	0,700	2,333
EDTA pH8	0,5	0,001	0,030	0,100
Triton X-100	10%	1%	1,500	5,000
Sodium deoxycholate (DOC)	10%	0,10%	0,150	0,500
H ₂ O	n/a	n/a	11,87	39,567
Vol. final			15 mL	50 mL

Extemporairement :

CaCl ₂	1	0,006	0,090	
Inhibiteurs de protéases	n/a	n/a	0,300	

Tampon de lyse + NaCl (4°C)

	Concentration solution mère (M)	Concentration solution finale (M)	Vol. (mL)	Vol. (mL)
Hepes pH 7.5	1	0,05	0,250	2,500
NaCl	3	0,5	0,833	8,333
EDTA pH8	0,5	0,001	0,010	0,100
Triton X-100	10%	1%	0,500	5,000
Sodium deoxycholate (DOC)	10%	0,10%	0,050	0,500
H ₂ O	n/a	n/a	3,357	33,567
Vol. final			5 mL	50 mL

Tampon LiCl (4°C)

	Concentration solution mère (M)	Concentration solution finale (M)	Vol. (mL)	Vol. (mL)
Tris-HCl pH 8	1	0,010	0,100	0,500
LiCl	3	0,250	0,833	4,167
IGEPAL	10%	0,50%	0,500	2,500
Sodium deoxycholate (DOC)	10%	0,50%	0,500	2,500
EDTA pH 8	0,5	0,001	0,020	0,100
Eau	n/a	n/a	8,047	40,233
Vol. final			10 mL	50 mL

TE (Tris-EDTA) (4 °C)

	Concentration solution mère (M)	Concentration solution finale (M)	Vol. (mL)	Vol. (mL)
Tris-HCl pH 8	1	0,05	0,250	0,750
EDTA pH 8	0,5	0,001	0,010	0,030
Eau	n/a	n/a	4,740	14,220
Vol. final			5 mL	15 mL

TES (Tris-EDTA-SDS) (Température ambiante)

	Concentration solution mère (M)	Concentration solution finale (M)	Vol. (mL)	Vol. (mL)
Tris-HCl pH 8	1	0,05	0,250	0,750
EDTA pH 8	0,5	0,01	0,100	0,300
SDS	10%	1%	0,500	1,500
Eau	n/a	n/a	4,150	12,450
Vol. final			5 mL	15 mL

Solution de TNE 2X

	Pour 1 L
Tris 1M pH8	10 mL
EDTA 1M	1 mL
NaCl	5,84 g
SDS 20%	100 mL
H ₂ O	889 mL

TNE 1X = 2.5 ml TNE 2X + 1 ml SDS 10 % + 1.5 ml H₂O = 8 extractions

V. Solutions pour Western blot et Dot blot

Lait

Blotting-Grade Blocker, nonfat dry milk (170-6404, Biorad).

TBS-T 10X

	Concentration finale	Pour 1L
Tris	0,2 M	24,2 g
NaCl	1,5 M	87,7 g
Tween 20	0,5 % v/v	5mL

Ajouter Tris et NaCl dans un peu d'eau distillée et ajuster jusqu'à 800mL.

Ajuster le pH à 7,5 avec HCl.

Ajouter H₂O jusqu'à 1L.

Autoclaver.

Ajouter Tween 20.

VI. Gènes

Nom du gène	Séquence du gène
<i>rid</i>	ATGTCCTCTCGGATGATGATCGACTCCTACGGTGACGAGGAGGTGCAATACTT GACGTCGCGAGATGTCGTTCCAAGGACAGATCGAGAGCCTCGTCATCGGCTA CCCTTGCCCGCACCACCGATGATGATAATGAAGAAGGCGAGGTGGAAGAGCT GGTGGAGCTGTTTGAACGCGAAATCATCGACCTCACCGAGCAGGAAATAGAG GCTGGTACACACGGTTCGATTCTGTCAGGAGAATTCTCCCTCGAAAGGTTTCG TGAGCCGCGCTCGGGGGCTCAGCTCTCTCGTGGCGACCTGATTCAGGTGAAAC ATGTCGAACTCGGGACTTACACCATCGACTTTGTCCAGATCAAGATCATTGCGA GCGATCGACAGGGTGACTTCAAGATCCGCGGAATTCCGTTTGTGCGGACCAG AAAAGTGTGGGCAAAGTCCAAAGAAGTGCAATGAGTTATGCATGATTCTTC ACATCCAGAGACGAGATGACGGACAAGAGCTGCCCCGAGTTCTTATCGATGTT CCCACTACCTCGATTGTCAGAAGAAGAGAACTTGTCATGACCAACGCAATCTA CCCAGAGCACTGTTACAACGGAGAGAAAACACTGATCTACGTGTCCAGAATGAAC AAGCGCAACAGAGGCTGGTTGAAGATTACGGAAGCTAGTCTGCCGCTGGAA GTTTACGATATACTTTACCTTGCAACGCCAGAGTCGTGCTACAAGGCCAGAAG AAGAAGTCCTAGAGCGCGTCCAAGCCGATGAGGTTCCAGTGTGAGATATCG TGCGTCTGACGAGGGGCTTTGCAACCAAGTGCGGGGGGGACGCATCAAGGG AGGGTCGTGGCCTTACAGAGAGTCGAGCATCATTGATCTGGAATCAAAGTCTC GTCATCAACGGCTCGCTCTGCAGGTAGTAGACGGATAGGGCAGAAAGTATAC ACTCTTCGACTCCTTCTGTGGAGCGGGGGGTGTATCTCGTGGCGCGCAGAGCT CTGGCTTCAAGGTCCAGTATGCCGTTGACAAGTGGGCCGATGTCTAGCAAACA TACCAGGCCAACTTTTCGCGACACTGAGCTATTCAGGCAGTCTATTGACGAGTTT ATTCGAGACACAAAGGACAGGCGGATGAGGGTCGACGTGTTGCATCTATCGC CGCCCTGTCAGTATTTCTCGCCGGCACACACGCATCAATCAGTCCACGACGATA CCAACATATTTGCGCTCTTTGGGTGCCTATCCCTGATTTTCAAGCTGCGGCCAC GCCTCATTACGCTTGAGCAAACGTTTGGGATTACACATGAGCGGCATCATCAC TACCTCAGGGCCCTCATCGGAGACTACACCAACTTGGATACTCTGTTTCGATGG AAGGTGGTTCGTCTCTGCACTTGGGGTTCGGCACAGGATAGAAAACGCCTCAT TATCCTTGCGGCTGCGCCGGGGGAGAAGCTGCCTCCTTTCCCTAAGCCAACTC ATTCCGAGAGCGGTGCTGGAGGCTTACAGCCATGGACGACCGTTGAGAAAGC

	GATTGGATCCATTGCGATTGGCGACGACTTACACAATCTGGACACCGTCAAAT ACCACCGACCCCTAGAGCACCCTTGACCCAAGCCGGTTGGCAGGCACAATT ACAACTGGAGGCAACGAAGTCTACTATCCCGACGGATCAAGAGACCTCACCAT TCGCGAGTACGCTTGCAATTCAGGGATTCCCAAGGTACCACAAGTTTATTGGAA CAAAAATTGCAATCAGGAGACAAATAGGCAACGCATTCCTCCGAATACAGTT CAGGTGCTGTATCGGCATCTGGAAGACTGGCTCCTGCAGCAGGATGGAATGA CGCGATACCAGCCTTTGCCCGACAGCGGACTGAATGTCGACCTCTTTGACAGC TACCTTCTGACGACTTCACGCGCTTACAAACGTCTTCGATGGCGCAGTCAGCT CCGGAGATGGGCGAGGACATTGTTGAACTGGCATACCCATGGGGAAGGTTTC GAAGTTCTCGGCACAGACGACGCGCCGTTGTTGACCTTACATGA
<i>hp1</i>	ATGGCACCCACGGGACGGAGAGGAAGAGCGATTGCTAAGAATGACGCAGCA CTCAGCGATGAAGAGGCCTCGGACTTTGAGAGCGCGCCGCCCCGTCAAGCCGT CGCGCAACTCGGCGTCCAGGTCCGTGAGCGATGCGGACGACGCGTCGACGAA GCCCCGAAGGAACGGACGGGCGCCGCAAGAAGGATCTGAAGGAGGTGGA CGACGATGTCGACATTGACGATGACGTTGACGTCAAGGATGCGGACGAGGAA GACGAGGAAGACGAAGAGGGTCTCGAAGATGACGTGTATGGCGAGCCGCC AGATGTTGCGTGCCATGAAGTCAGAATCTGACGGCTGTTTGAGGTTTATTGT CGAGGCGATCAAGAAGCACATGATTGACGAAGATGTATGGACAGATGACGCC TCTTTTACTGTTTGACAATGGCTAATGTGCGCGCAGGGAACGCTGAAATTCC ACGTCAAGTGGAAGGCTACGACAAGAAGTCGGACATGACATGGGAACCCG AGGAGAACCTCGTGTATGTGCAACACCGGTCACTGCTCAACAAAAAAAAAAAA GAAGAGAAAAAATACCGGGGCGACTGACTTTTGATGCGCTACAGGGAGTCTG CGGGCGAGATCCTCAGCGCTACCTCGAATCCATCGGCGGGCGAGAAAAGAT CTTTGAGGAGACGACGAGGCTTCGAGGGGCAAGAAGCGCGGCCGCACGAC GAGCGCAACGCCGGGCGCGACGAAGCGGTGAGGAGGAACGGAACGCACCC TGCCGACTCGGCACCCCGGCGACGGCGGACAAATGGAGCCCCCGCGGGGA TCGTGGGAGGACGAGATCCAGACGATTGACGCGTGCGAGGACGAGGGCAGC GGCCGGCTCGTAGTTTACCTTGTGTGGAAGAATGGCCGTAAGACGAAGCATG ACACGAAGGTCATTTACAAGAAGTGCCCGCAAAAGGTAAGGCGTCATGGATT TGGCCGTGGACTGTGTAGCTGACTTTTATCGCAGATGCTGCAATTCTACGAGC GCCACGTCAAGATCATCAGGGAGGAGAACAAAGGCCTTGACTGAGTAA
<i>dim5</i>	ATGGAGGGCCGTCACGAGACAGCACTTTTTCTTCCACGACCAGCCGGGTACGTA CGCCCATACAATTTACAACAACAGGGCCCCCTTTGAACACCTACACAAGAATCC CCCCTCAAAAAAAAAAAAAAAGACTGACATTCAAGGCCAGGAATACGCGGCC CAAAGGGAATCCTGCCACTTCTGCCAGCTCCGCGCCTTTCCGCGCACAAAGAC CATCCCCGTACCATCGTCAACGAGGTGACACGAGCCGTCATCCCGCCCACTT CCGCTTCGTCGACCGCATGGTCCTGCGCCGCGGCGTCGAGCCCCGCTAGGATA GCTTCCGACGCGGCTGCTCGTGCGAGTCCGACGACGCGTGCCAGTACACGGG CTGCCTGTGCCTCGCCGACCTGGACGACCAGGACAGCGGCAGCGACGAGGAC GACGACGAGGACGACGACGAGGGGAACGACGCCTTTCACATCGCCAACCAGC TCAATGGGGATGATTTGAGGGCCGCGGCCGCAAGAAGAAGAAGATGAAGA AGAAGGACAGGACGAAGAGAAAGGCGTACGCGTATCACTCGCACGGGGCCA AGGCGGGCCTGCTGCGCTCCAAGATGCTCAACTCCAAGGAGCCGCTGTACGA ATGCCATGCCGGTGTCTGCTGCTCCATGAGCTGCCCCAACCGGGTTGTCGAGC GAGGGCGAACCGTGCCGCTGCAGATATTCCGCACCCCCGATAGGGGCTGGGG TGAGTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCAGGGGGAGCATTGTGAAG CTGGGAGACATCAAGCCTAACGCCCTTGGGCCAGGAGTACACGCCAGGCGG CCATCAAAAAGGGCCAGTTGCTCGACCGCTACTACGGCGAAATCATCACCTCG GCCGAGGCCGACCGCCGGCGCACCGCCGCCGCTTCTCCAGCGCAAGGACG TCTACCTCTTCGCCCTGACAAGTTACCGACCCGGACTCGCTGGACGCGCGCC TCCGGGGCCCCGCGCTCGAGGTCGACGGCGAGTTCCAGTCGGGCCCCGACGCG

	CTTCATCAACCACTCGTGCGAGCCCAACCTGCGCATCTTTGCGCGCGTCGGCG ACCACGCCGACAAGCACATTCACGACCTGGCGCTCTTTGCCATCAGGGACATT CCCAGGGGGGAGGAGCTGACGTTTGATTACGTCGACGGGGTGATGACGGGG GACTTGCCGGGATGGAGGAGCAGGAGGCGCATGGCGAGATGGCCAAGTGT TTGTGCGGGAGCCGCAAGTGCAGGGGATATCTGTGGTGA
<i>dim2</i>	ATGGAGGACAGCTCGGACGATACCGCCAAGGAACCATCAGCTCTGACCGATG GCAATTTACAGAAAAAGACGGCGAAGCAGCGGCTGCTTTGGGTACCACTCT CACGGCGATTCAATCCTGGCAGACCTCCGTCTTGACCATCAATGATAACAACC GTTGAGCAACACACCCGTCTGACCATGAGTCGTTGCCATGTATAGACACTTTTG AAGAGACTCCGAGCTCTGACGAAACCGTTATTGAGCTCGTCACATCCGACCTG ATACCGCCTCCATCTCGTCTCGTCACATCATCACGAGTTCTACGCTCAAAGTCT CCGAGCTTGATGATGCTGTGTGTCAGACAGTCTTTGACAGAGACTTTTCGAGCA ACTTCACCGGACTGGGTACCTTCAACCGGGTCCTTGAGCGTTGAGATACCCGG AAGCGTTCAGCTTGATACCAAGGTCATGCTACGAGGGCTTTCAGCCTCCTGACG ACCTACCGATCTTACCAGAGTGGAAGGCTGTTGAGACTCTTCTCCAGCCTTGG CGGAAAGCACGCAAGATTTTGACGAGATCATGCTGGACGATTTTGCGATATAC CTTGATAAGGAGAGATGCTCACAGGAAATGCGATCTCTACACCAGCTCAATAC CAAGATTGGACACAACGACTTCTACTTTGACGGTGTTCTCCGCATTGGCGATAC CAGGACGTATGTGCGCCGATTCCCATCACCGCGGTTCCCATGGCAACTACG GCTCAATCTCTTATCACACTGTCCGAGACAACATCTGGCTGCAATCTCCTCTGA GCTACAAACGGGAGCTATATTACAGGTTGGGGAAGCCGGCCAAGGAATATGC ACGCTTCTTTTACCCTTCTTATGGGTGGCGGACCTGGCGAAGCACTTTGTGGA CTTCTCTCCATCATGGCGGAGAACGAGCGTAGTGTTTCCATACACCACTTTG ATCTACTTTTGAGCATGGATGAAGAAGAGCCACAAGGACTCGCCCGAGTTTT TCAGCTGGCTACGCCAACACCCGAGCGATGACTACCGGACTTCGATAGCCGCC AACATAGGGTTTCTCCATAAAGAGACTCTGGGGGTGTTGAACCATCGAGAGG CGTATTCCCATGCTATCTGGGCGGAGATCTGGGAATTCCAATCCTTTACATCAA TCACGAGCCAGCCCAAGACTGAGACTCCGCCGCCGACGATTGTTACGCAATAC ATATTTGACTGCTTTGAACATCTTCTTTTGGCGATCGCCTCAAGGTTGTCCCTC TGAGTTCACGGACCACGGAGCTCCGCAACAGCTTGATCCGACAGCGGCACCTA GAGCTTCCGTCAAAGCTGCACCAGACATCTAAGGACATCTCTACGGCGCCACA AGAGCGGATCAAGAATATCCGGCCAGGCGACACAATCTCAACCCATCGCGAC GACGCAGAGTCCGGCACATTGTGGAAACGAGAATTGTCCAAAGGCTTCTCTG ACGTGGACAGATGGTTTGCCTGCTGTCAGTCGGTTCACGTTGACAGAAAAGG GGTGCGTACCTTTGATGTGATATGGTATTACAGACCCGTGGACACGCTCTGCG GGTTGATGAAGTATCCGTGGAACAACGAGCTGTTTCTCTCAGACCATTGCTCTT GCTCCGAGAAAGCCAAGATCGAAGAAAGCGAGGTTCTCGGTGTCCATGGCGT TGACTTTTGGGGCACCTCAGCAACAAGCGCCGAGCTCTTTGCCGGCAGACCT ACCTTCAAGAAGAGAGGAGATGGGTGACACTGGATGGGAAGCACCTGCAAT GTAGGCACACTTCACCAATACCAGATGATGAAATGCCACTCGAGTATCAGCCA GGGAAGACACATCTGCTTCGTTTGAACCTGAAAAGTCCATTTCCGAGCCTTGT GAGTTTATGCACGCGTCTGATGAAGGAAGCCGAAGGTTGATTACTTCAGGCG ACTATTGCGGCGCCGCCAGGTTGACCCAAACGCTCGCGAGGCGCGGCCAAAT GAACTTGATACAGCGAGCAGACCTGCAAAGTCAGCAGGGATCGAATCATGG GCCCATTGTCAGTGAGATGCTTCCGAGATGGAGAGATGATCCCTACGCCGTAT GACCACGACGGCACGGGCAACCTCTTTTACATCACTCACCGACAGGTAGCCGC CAATGGACATTCCGTATATGTGCCGCTTGACGAAGCACCGTCCTCTCTGCAGC AAGGCTTTGATCCTCTCCAGAAGTTGCAGAAGCTTCGCGGGATGGATCTCTTC TGTGGAGGCGGTAACCTTGGCAGAGGGCTTGAAGACGGCGGAGGAATCGAG ATGAAATGGGCAATGACTACGACAGCAAGGCTATCCACACGTACATGGCCA ACGTCAAAAACCTCAAGACGTCCACCCGTTTCTGGGCTCCATTGACGACTTAC AGCGCCTTGCCATTCAAGGCGACTTCGCGGACAACGTACCCCTAATTGGAGAC

	GTTGATTTCTGTGTCGGTGGCAGCCCATGTCCTGGATTCTCACGCCTCACCAAC GACAAGACGACGGCAGCGCAAAGGAAGAACCAATCGCTTGTTGCGGCGTTTG CGTCCTTTGTTGATCTCTATCGGCCCAAATACGGCGTGTTGGAGAATGTACCTG GGATGGTCCACAAGAAGCACGACCGAGACCGAGACGTCTTCAGCCAGCTCAT CTGCGCCCTAGTCGGCTTGGGCTATCAGACGCAATTCTTCTACCTGGACGCATC GTCATGCGGGTCGCCGACGCGCAGATCACGCATCTTCATCGTCTTTGCTGCCC CGGGACTCGAGCTGCCCCACAAGCCGGTGCAGACTCACTCACACCCCCCTAAC ACAAGGGAGCATGGGATCGGGTGGCTGCCCCAACGGCCAGCACATGGCAAGC CGAGAAATGCCGGCAGCAACGCCCTTCAAGTTTGTCTCTGCGGAAGAAGCGA CCGCCGACCTGCCGCCATCTACGATGCCAAGCCCGATATATGTGTTCTTTCC CCGATCACCGCTCAACCTTGGGTATGTCAGACTTGCTGCGAGCTCGGATATCG GTCATACCGACACGACCATGGGGCATGAACTTTTCGCAAGCGTGGTTCGGGG AGAGGAAGGTCGGCGGCTCCAGCACGATGACGGCAGCCGAGCGCGAGTTCTT CATCGGGACGAAAGGCGAGGGCAAGCAGCCCATGTCGGTGCAGCCCTACTCA AATTCGTACGGAAGGATGTTCCCCAACAGGCTCTTTGAGACTGTGGTGACGAG GCAGACGCTGGGCGACGCGAAGAACGGGCGACTGCTCCACTGGCGAGAGAA TAGGCCCCGTGACCATCATGGAGGCGCGCAGGGCGCAGGGCTTCCTCGACGAA GACGTGCTGCTCGGGGACCCCCGACGCAGTACAGGATTGTAGGCAACAGCG TCGCGAGGCAGCCTGCTGTGGCTCTCGGTGTGACGTTTCGCGAAGCGTGGGT GGCGAGCCTGAAGAGGAACAGAGAGCTGCAGCGCGCAAGGAGCTGCGTTGC CGAGGAACATGGCGGATTCACGAACGGCGTTCCAGTGACTCATGAGTCCCTC GAGAAGATGGCGGACGGCATGGGCAGCATATCGAGAGATCAGCAGCTAAGG GCCTTTACGCCCCGAACGGACAGCACACTGCAGCAGTCGATGGATATCACGAT GGGCAGTAAACGCCCTCGTTCGGCCAACATGGTGATTGAGCTGCTCGAGTCCT CGAAACGACAGAGGCCGAACCCCTTTTGA
<i>kmt6</i>	ATGCCGCCAGCAACGCCATCACGAGAGCGTCATCGCCAGCCTGAAGGCGTGC AGACTCCCAAGACCCAGACGGTGCCCGTGACGGAGTGACTGTTTCGGACGCT TGCTAAGCAGCTCACGACGTTTAACCAGCAGGTTGGAAGGATCACGCACAG CTCGTCGCCCACGCCATTGAGAGCGCAAGGTGTCGAGAGTGCGGCACGCAAC ATGGCCTGGATCTCTCCGTGAGCTAACCACTACCCCAATCCGGAAATCAAG GGCAAGCAGATCAAAATCAAGTTCAAGGTCAGCCAGCACTGTTTGTTCACCAG CACCCGAGGTTTCGACGATCTTGTTTGAAATTCACGAATTGACCTCCCAACTCA GCAGCATAACAAGAACAAGAAGGAACAAAAGGACGTTTCGCTACAATGTGCAC TGCGTCAAGACCAACAAGCCTGCTGTTCCGTCTTACCGGTTTCACCATGTGGAA ATCAAGAAGAACGTCTTGATGCCAACACCATGCTGACGTTTGTACCCCATCTG CGGGATCTGGAGGGGTCCGAGGAGGCAAAGTACAACCTCTGGCTTAAGGAGT TGAGGACATTGACCTCAAATCAGGATTCAAGCCCATGAACCGCGATGAGAA GTTTGTTGCCATCATCCAGGGCGAGCGCGCTGCTACGCTCTCGCTCTACTTGG ACATGTGGCTCAAGGAGCTCAACATACCATCCTGCAACAAATCCCACTCATC AGATATATGGCGAGCCATGCCTCCGACGACGCCATAACGCCCCAGCAAAAGA CGGACATTATCAAATCCCATGGCCACGGGCACGCCATGACGCCAGAGGAGCA CAGGGCGGCAGAGCTCTTTACTGCCGCGTTTCAGTTGGCTTTCAAGGAAAATC AACCGCAGTTGAAGCAGATTGAGCTGCGCAAGGTGCTGTTCTCGACGAGTC GGTCGATAGTGCATGGACAACAAGCCCGCGGCCAAGGAAGTTTCCCTTGTC AGCAGGATAACGAGGACGAGGTTGTCGAGTCCAACCTTGCCACCTACTGCATT TTGGGATGCCTTATCTGCTTCAGCCATTCTGCGACCACGGCGAGTATGACATC AAGAATTGGAACGCACTTTTTCTTGTCATCCTGTGCACCCTTGTCGGACATT TTGAGGAGAAAGAGGATAGCGGCAGCAGGGGTAAACACGATTGCTGATGTG CCTTGTCGACGGAGATGCTATCGGCTGAAGATTCTCTGACCCGACGAGTGC GGTGGCAAAGCCGTGGACCGAGGATGAGCGCATGCTCTACGGACTCTGTAT ATCACGACTGATTTAGCAAGACAAAGATGGACCCGATCTGTCTTGCAACGGA GTTTCTGAACCGCAACTGCGATGAGGTATACGCAGAATTTAGTCTCTCAGCA

	TTGGCATACCTGGGCCTGACCTCACGGAACCGCCTCGACCGCGGAACATTTCA TGGTACGATCGCAAGCGGAAGATGCTTATTGGAGACTGGCAAGAACACACCA TACTCATGAGCATCAACGACGGGAAGTCTGGAACCATGCTCCACGACGGC CCCTGCGCCCCAAGATATGCAGCTGCGTCGATGCTGGGGTTCTGTGCGAGA GGTTTTGCGGGTGACAGAGGCCAACTGCGCTTACAAGTTTACAGGCTGTGCT TGCCACTCTCTCGGCAAGACATGTCTCCCGAAGCAAAAGGACAGGCCGTGTAT CTGCGTGCAACTCAACCGCGAGTGTGACCCCCAGCTGTGCGGTTCTTGTGGCG CTTTCGAACGTGCGGATCCCGCCAATGCAGAAGACGATATTTTGAAGCTCGGCG GGTTGCCAAAATTGCGATCTCCAGCGTGGGCGGCACAAGGCGCTCCTCCTCGG ACAAAGCCAGCTGGAAGGAGTTGGCTACGGACTCTTTACCGCCGAGGACATT GCTCAGGACGAGTTCATTGTGAGTATGTGGGAGAGCTCATCCCCACGACG AGGGCGTCCGCCGGAAGCGAGACGCGGCGACGTATTGATGAGACGTGCA ACATCTCATACTCTTTACTCTGCTTGAACACGAGGGCATCTGGGTGACGCC GCTATTTACGGCAACCTGAGCCGGTACATTAACACGCTTCAGAGCACGACAC CCGGGGATGCAACATTACGCCGCGGATCCTTTATGTCAATGGGGAGTTCCGGA TCAAGTTCACCGCATTGCGGGATATCAAGGCCGAGAGAAGAGCTCTTTTCAAC TACGGCGAAAACCTTCCCAACCTACCAAGAAGCTGCTGGATAACAAGTCCGG GGAGAAGGCGGAGGGCCCTAGAAAGGTGCGCCGACCAAGCGTGCCGAGAC TGGTGATCGAGTGGCACGCAAGGTGCCCAAGATAATAGAGCACAGAAGGC GGGGAGACCAAGTCAAAGAAGAAGAAGATGGCGGCAGCGGCGACGACTGT TACTACAACACGACTACGACCGTTCAGGCGGCAGTGGTGGCAGTCTCGGATG ACGATTCTGCGGCTGCCAGCCCATCAATAGAAATCTATACCCGCAAGCGAAAA CGCCACACGATGATTGATTCCGAGGAGGAAGAATACCATCCTCCTGGCACAGA CACCGCGTCGTCTGCAGCTGCACATGCCACGACCGAGGATGGGGAGCTTCTA GGGTCGAAGCGGCTGCGCAAGACGACGCGGTACAGAGAGGCCAAGGAGCCC GAGACACCAGCAGTCAAGACTCCGGGATCAGAGGAGCAGCAAAAGCCGCGG GGCAAGCGTGGAGGGGGCCCGACCCGGCAGCGGACGACCTAGAAAGCATCCT CGAGTCATCCCAAGGCCAAACAAAGAACCTCCGCAGCCTGCCATGCAGCCCT GCCCACTGATTCTGCTCAGGAGTGCAGCAGCTCGACGTGAATACCAAGTGCCT CTGCTCCCGCGCAGACGCCGACAACTACAACGTCCACGCCAAAGCAAAGCAG GACGGGCCGCGGAGAGGTGCTGAGGATGTTGTTGAAATCGAAGACAGCGA GGATCAGGACGTGGTGACGATGAGTAGGACATCGAGGAATAGACGGTTACC GGCGAAATTTAGGGAGGAGGCATAG
--	---

VII. Cassettes d'invalidation

Cassette d'invalidation ciblant les gènes des protéines:	Séquence de l'insert
RID	CAAATACTGGTGCATCGCAATTCATCAGCGCCAGCTTCCCCTTTGCCAGCTT CGAAGCGCTGCAGTTGGCCTACGCATTTGACAAGCTCTCCCTGTTCTGACTCT CTTCATCATCTCGAAAACCTTGATCGGCATATAACCATCTGCTGATCTGGATACA TCAGACTGCATCAGGGCTCTTCACTCTATGTCCATAAGGGTGGGGCAGTGCTA GTGTGTGTACCGACCCGAGGATTGGTGCTTTGCCAGAGCTCTACAGAATAG CGCGCGCATCCATATGTTAGTTCTGCAATTTTCTTGATCGGTGCTGTGACTCAT ACTTCCCCCTTTGGCTGGCCTTGCGGCAACCAATAAGAACGCACAGTGAAATC TTGCGGGTGGGGAGTGCATGGCGCCTGCATTGGCTTGGGGACGCGCACTGTC GCACACTTCCATCTGACCTTTCAGAAGGGTTTCGTGGTGGGCAAGGACCAACC GGTTGCGCGGCCGTGCGTGGGTGCCTCGCCCGGCACTGCCAGGGCCACTGCA GTGGCAGTTTGCTGCCTGATACAAAATCCTTCCCTCCGCCAGTTTTCCCTCTTT GACCTTCTTTCTCTTCTCTGCAACCAAATCCACCCTATCAAACCAAAACAGTGT CTCGACCGAGGTATCAACCTGAATCAGCAACATCGTAGCCAGCATTTGTCTCC GTCTCTGCAGAACCGAGGTTGCAAACATTATCCAGGCAACAGGGCACCAAC TCACTTCTTCGGCTTTCACCAATCGGTACAGCTCTTCTCAGAACTCGCGTCCGCA ACAGTTCTACGCTTCTCAGCACCTTCTTCAGCTTCAATCCTGAACACTCAGAAC CGCGCACAGCAGCGCCCTCCTGTTCCCTTGTTCCTCCAAAAGTACCGGTAGTATT TCGCACGGAAAGCAGGGCAACACGTAAGTGATATCTACTAGTTGTGACCGGC GCCTAGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACG ACTCACTATAGGAGGGCCACCATGTCGACTACTAACCTTCTTCTTTCTTCTACA GCTGAGATCACCGGTAGGAGGGCCATCATGAAAAAGCCTGAACTCACCGCGA CGTCTGTCGCGAAGTTTCTGATCGAAAAGTTGACAGCGTCTCCGACCTGATG CAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGC GTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGT TATGTTTATCGGCACTTTGCATCGGCCGCGTCCCGATTCCGGAAGTGCTTGAC ATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCGTGACAGGG TGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAACCCGT CGCGGAGCTCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGG TTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTT CATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACGTGATGGACG ACACCGTCAGTGCGTCCGTGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCC GAGGACTGCCCCGAAGTCCGGCACCTCGTGACGCGGATTTGCGCTCCAACAA TGTCCTGACGGACAATGGCCGCATAACAGCGGTCTTCACTGGAGCGAGGCG ATGTTGCGGGATTCCCAATACGAGGTGCGCAACATCTTCTTCTGGAGGCCGTG GTTGGCTTGATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAG CTTGACAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCA ACTCTATCAGAGCTTGTTGACGGCAATTCGATGATGCAGCTTGGGCGCAGG GTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACA AATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAAGTACTC GCCGATAGTGGAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATGAC CGGAGTTGAGACAAATGGTGTTCAAGGATCTCGATAAGATACGTTCAATTTGTCC AAGCAGCAAAGAGTGCCTTCTAGTGATTTAATAGCTCCATGTCAACAAGAATA AAACGCGTTTCGGGTTTACCTCTTCCAGATACAGCTCATCTGCAATGCATTAAT GCATTGGACCTCGCAACCCTAGTACGCCCTCAGGCTCCGGCGAAGCAGAAGA ATAGCTTAGCAGAGTCTATTTTCATTTTCGGGAGACGAGATCAAGCAGATCAA CGGTCGTCAAGAGACCTACGAGACTGAGGAATCCGCTCTTGCTCCACGCGAC TATATATTTGTCTCTAATTGTACTTTGACATGCTCCTCTTCTTACTCTGATAGCT

	TGACTATGAAAATTCCGTCACCAGCCCCTGGGTTTCGCAAAGATAATTGCACTGT TTCTTCCTTGAACTCTCAAGCCTACAGGACACACATTCATCGTAGGTATAAACCC TCGAAAATCATTCTACTAAGATGGGTATACAATAGTAACCATGCATGGTTGCC TAGTGAATGCTCCGTAACACCCAATACGCCGGCCGAAACTTTTTTACAACCTC CTATGAGTCGTTTACCCAGAATGCACAGGTACACTTGTTTAGAGGTAATCCTTC TTCTAGCTAGACACATCATGTTAGTCATGGAGGATGAATGACTTCCCATATTT TGCCTTACTTGTCTGGGTCGTGCTGCAGTAATGCCACATTCCTCTGCAGAGG TACTCTGCAGTAGTGGACGCAGAAGCTCAAAGGAACGGCAGCATTGCTCTGG CGATACGGATGCCCTTGAGTGGTACATGGGTAAAGAGAACACCTGT
HP1	CAGATCGATAAGCAGCGATAAGCCCCATCCCCAGGAAAAAGTCGGCTCACGT GACCCAGCGTCACTCGGCAACACCAAAGAGCTTCCGCCTGTCGCGTTTGCTTA TGCCCGGACAAAGCGTCAAGCGTGCGACCAGGTTGCGCCGCCGGACCTGCG AACCAGCTCAGCAATCGCCAGCCGACCCAAGAAGACCGCGACAGGTGGGGCA GTGCTAGTGTGTACCGACCCGCAGGATTGGTGCTTTGCCAGAGCTCTACA GAATAGCGCGCGCATCCATATGTTAGTTCTGCAATTTTCTGTATCGGTGCTGT GACTCATACTTCCCCCTTGGCTGGCCTTGCGGCAACCAATAAGAACGCACAG TGAAATCTTGCGGGTGGGGAGTGCATGGCGCCTGCATTGGCTTGGGGACGCG CACTGTCGCACTTCCATCTGACCTTTCAGAAGGGTTTCGTGGTGGGCAAGG ACCAACCGGTTGCGCGGCCGTGCGTGGGTGCCTCGCCCGGCACTGCCAGGGC CACTGCAGTGGCAGTTTGCTGCCTGATACAAAATCCTTCCCTCCGCCAGTTTT CCCTCTTGACCTTCCTTCTCTCTGCAACCAAATCCACCCTATCAAACCA AACAGTGTCTCGACCGAGGTATCAACCTGAATCAGCAACATCGTAGCCAGCAT TTGTCTCCGTCTCTGCAGAACCAGCGAGTTGCAAACATTATCCAGGCAACAGG GCACCAACTCACTTCTTCGGCTTTCACCAATCGGTACAGCTCTTCTCAGAACTC GCGTCCGCAACAGTTCTACGTTCTCTCAGCACCTTCTCAGCTTCAATCCTGAA CACTCAGAACCGCGCACAGCAGCGCCCTCTGTTCCCTGTTTCCCAAAAGTAC CGGTAGTATTTTCGCACGGAAAGCAGGGCAACACGTAAGTGATATCTACTAGTT GTGACCGGCGCCTAGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAG TATAATACGACTCACTATAGGAGGGCCACCATGTCGACTACTAACCTTCTTCTC TTTCTACAGCTGAGATCACCGGTAGGAGGGCCATCATGAAAAAGCCTGAACT CACCGCGACGTCTGTCGCGAAGTTTCTGATCGAAAAGTTGACAGCGTCTCCG ACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTA GGAGGGCGTGATATGTCCTGCGGGTAAATAGCTGCGCCGATGTTTCTACA AAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAA GTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCG TGACAGGGGTGTACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTT TGCAACCCGTCGCGGAGCTCATGGATGCGATCGCTGCGGCCGATCTTAGCCAG ACGAGCGGGTTTCGGCCATTTCGGACCGCAAGGAATCGGTCAATACACTACAT GGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAAC GTGATGGACGACACCGTCAGTGCGTCCGTGCGCGAGGCTCTCGATGAGCTGA TGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGACGCGGATTTT GGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCACTGACTG GAGCGAGGCGATGTTGCGGGATTCCCAATACGAGGTGCGCAACATCTTCTTCT GGAGGCCGTGGTTGGCTTGATGGAGCAGCAGACGCGCTACTTCGAGCGGAG GCATCCGGAGCTTGCAAGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTG GTCTTGACCAACTCTATCAGAGCTTGTTGACGGCAATTTTCGATGATGCAGCTT GGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCG GGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGT AGAAGTACTCGCCGATAGTGGAACCGACGCCCCAGCACTCGTCCGAGGGCA AAGGAATGACCGGAGTTGAGACAAATGGTGTTTCAGGATCTCGATAAGATACG TTCATTTGTCCAAGCAGCAAAGAGTGCTTCTAGTGATTTAATAGCTCCATGTC

	AACAGAATAAAACGCGTTTCGGGTTTACCTCTTCCAGATACAGCTCATCTGCA ATGCATTAATGCATTGGACCTCGCAACCCTAGTACGCCCTTCAGGCTCCGGCG AAGCAGAAGAATAGCTTAGCAGAGTCTATTTTCATTTTCGGGAGACGAGATCA AGCAGATCAACGGTCGTCAAGAGACCTACGAGACTGAGGAATCCGCTCTTGG CTCCACGCGACTATATATTTGTCTCTAATTGTACTTTGACATGCTCCTCTTCTTA CTCTGATAGCTTGACTATGAAAATTCCGTCACCAGCCCCCTGGGTTTCGCAAAGAT AATTGCACTGTTTCTTCTTGAACCTCTCAAGCCTACAGGACACACATTCATCGT AGGTATAAACCTCGAAAATCATTCCCTACTAAGATGGGTATACAATAGTAACCA TGCATGGTTGCCTAGTGAATGCTCCGTAACACCCAATACGCCGGCCGAACTT TTTTACAACCTCTCCTATGAGTCGTTTACCCAGAATGCACAGGTACACTTGTTTA GAGGTAATCCTTCTTTCTAGAGAACGACCTCTGAGACGGTGAAGCACTCTTCC CCATTGACTCTACCGCGTCACTGAAAAAAACAAAGAAAAAAACACTGCTGCTG GGATCAGACACGGCACGGAGGAGTGGATTCGTGATTTCTGCGACTGCTTCAT GTCCTATGGTCTCAAGACTGGCGTATATGGATCGCATCAAGCCTCGATTGTTT CTTGGATT
DIM5	ATAAAGATTGTCGAGTGGGCATTGACACGGGTATGGATATGTTGCTCTCTTG TCAGTTTGTATAAGAGAGCTTGACCGGCATTGATGAGAATGATTCACTACCTC GGAAACCCGCCACAAGCGGAATACAAGGGGAAAACGCATTGTACTAGGCCAT TAGCTAAAACCAAACAACGAGAAAGAGTCAGAATAGAATCAGGTGGGGCA GTGCTAGTGTGTGTACCGACCCGCAGGATTGGTGCTTGGCCAGAGCTCTACA GAATAGCGCGCATCCATATGTTAGTTCTGCAATTTCTTGATCGGTGCTGT GACTCATACTTCCCCCTTGGCTGGCCTTGCGGCAACCAATAAGAACGCACAG TGAAATCTTGCGGGTGGGGAGTGCATGGCGCCTGCATTGGCTTGGGGACGCG CACTGTGCGCACTTCCATCTGACCTTTCAGAAGGGTTTCGTGGTGGGCAAGG ACCAACCGGTTGCGCGGCCGTGCGTGGGTGCCTCGCCGGCACTGCCAGGGC CACTGCAGTGGCAGTTTGCTGCCTGATACAAAATCCTTCCCTCCGCCAGTTTT CCCTCTTGACCTTCTTCTCTTCTGCAACCAAATCCACCCTATCAAACCAA AACAGTGTCTCGACCGAGGTATCAACCTGAATCAGCAACATCGTAGCCAGCAT TTGTCTCCGTCTCTGCAGAACCAGCGAGTTGCAAACATTATCCAGGCAACAGG GCACCAACTCACTTCTTCGGCTTTCACCAATCGGTACAGCTCTTCTCAGAACTC GCGTCCGCAACAGTTCTACGTTCTCTCAGCACCTTCTTCACTTCAATCCTGAA CACTCAGAACC CGCACAGCAGCGCCCTCTGTTCCCTGTTTCCAAAAGTAC CGGTAGTATTTGCACGGAAAGCAGGGCAACACGTAAGTGATATCTACTAGTT GTGACCGGCGCCTAGTGTGACAATTAATCATCGGCATAGTATATCGGCATAG TATAATACGACTCACTATAGGAGGGCCACCATGTCGACTACTAACCTTCTTCTC TTTCTACAGCTGAGATCACCGGTAGGAGGGCCATCATGAAAAGCCTGAACT CACCGCGACGTCTGTCGGAAGTTTCTGATCGAAAAGTTGACAGCGTCTCCG ACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTA GGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACA AAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAA GTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCG TGCACAGGGTGTACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTT TGCAACCCGTCGCGGAGCTCATGGATGCGATCGCTGCGGCCGATCTTAGCCAG ACGAGCGGGTTGCGCCCATTCGGAACGCAAGGAATCGGTCAATACACTACAT GGCGTGATTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAAC GTGATGGACGACACCGTCAGTTCGTCGTCGCGCAGGCTCTCGATGAGCTGA TGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGACGCGGATTTT GGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCACTGACTG GAGCGAGGCGATGTTGCGGGATTCCCAATACGAGGTGCGCAACATCTTCTTCT GGAGGCCGTGGTTGGCTTGATGGAGCAGCAGACGCGCTACTTCGAGCGGAG GCATCCGGAGCTTGCAAGATCGCCGCGGCTCCGGGCGTATATGCTCCGCACTG

	GTCTTGACCAACTCTATCAGAGCTTG GTTGACGGCAATTCGATGATGCAGCTT GGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCG GGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGT AGAAGTACTCGCCGATAGTGGAACCGACGCCCCAGCACTCGTCCGAGGGCA AAGGAATGACCGGAGTTGAGACAAATGGTGTT CAGGATCTCGATAAGATACG TTCATTTGTCCAAGCAGCAAAGAGTGCCCTTCTAGTGATTTAATAGCTCCATGTC AACAAGAATAAAACGCGTTTCGGGTTTACCTCTTCCAGATACAGCTCATCTGCA ATGCATTAATGCATTGGACCTCGCAACCCTAGTACGCCCTCAGGCTCCGGCG AAGCAGAAGAATAGCTTAGCAGAGTCTATTTTCATTTTCGGGAGACGAGATCA AGCAGATCAACGGTCGTCAAGAGACCTACGAGACTGAGGAATCCGCTCTTGG CTCCACGCGACTATATATTTGTCTCTAATTGTACTTTGACATGCTCCTCTTTTA CTCTGATAGCTTGACTATGAAAATTCCGTCACCAGCCCCCTGGGTTTCGCAAAGAT AATTGCACTGTTTCTTCCTTGAACCTCTCAAGCCTACAGGACACACATTTCATCGT AGGTATAAACCTCGAAAATCATTCTACTAAGATGGGTATACAATAGTAACCA TGCATGGTTGCCTAGTGAATGCTCCGTAACACCCAATACGCCGGCCGAAACTT TTTTACAACCTCTCCTATGAGTCGTTTACCCAGAATGCACAGGTACACTTGTTTA GAGGTAATCCTTCTTTCTAGGGCTGGCTGAGTTGCTCACTTGACGGATGCGCC TGTGCTGTTGAGGGTCTTATGAGATGGAGTATCCGGTCATCAACCATAACCT CCTGAATCAGTCGTAAAGATAAATGGCTGTTCAATGGTGAAATTGTATCAAAT TAAAGTACGAGGCAGAACGGTAGTAACAGCACTGAGCCAGGACGCGTTATGA GAATGAAGC
DIM2	ACTCGAGGAGTGGTTGCTCTTTGAGGCTTCCATTCTGCTCATGCTTCTGGGATA TCCAATGGGAATCATATGTCCCTTTCTTTGGATATCAACTTGATCCATTTTGG CCATCGTCACAGATTCACGCTCTCGGTGAAGTTCGGTTGGCGATAGTTTCACTT TTCAGGTCCCTCGCATCCTGCGTTTATGGAGTGCCTCGGTGGGGCAGTGCTAG TGTGTGTACCGACCCGAGGATTGGTGCTTTGCCAGAGCTCTACAGAATAGC GCGCGCATCCATATGTTAGTTCTGCAATTTTCTGTATCGGTGCTGTGACTCAT ACTTCCCCCTTTGGCTGGCCTTGCGGCAACCAATAAGAACGCACAGTGAAATC TTGCGGGTGGGGAGTGATGGCGCCTGCATTGGCTTGGGGACGCGCACTGTC GCACACTTCCATCTGACCTTTCAGAAGGGTTTCGTGGTGGGCAAGGACCAACC GGTTGCGCGGCCGTGCGTGGGTGCCTCGCCCGGCACTGCCAGGGCCACTGCA GTGGCAGTTTGCTGCCTGATACAAAATCCTTCCCTCCGCCAGTTTTCCCTCTTT GACCTTCTTTTCTTCTCTGCAACCAAATCCACCCTATCAAACCAAACAGTGT CTCGACCGAGGTATCAACCTGAATCAGCAACATCGTAGCCAGCATTTGTCTCC GTCTCTGCAGAACCAGCGAGTTGCAAACATTATCCAGGCAACAGGGCACCAAC TCACTTCTTCGGCTTTCACCAATCGGTACAGCTCTTCTCAGAACTCGCGTCCGC AACAGTTCTACGCTTCTCAGCACCTTCTTCAGCTTCAATCCTGAACACTCAGA ACCGCGCACAGCAGCGCCCTCCTGTTCCCTTGTTCCTTCCAAAAGTACCGGTAGTA TTTCGCACGGAAAGCAGGGCAACACGTAAGTGATATCTACTAGTTGTGACCGG CGCCTAGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATAC GACTCACTATAGGAGGGCCACCATGTGCGACTACTAACCTTCTTCTTTCTTAC AGCTGAGATACCGGTAGGAGGGCCATCATGAAAAAGCCTGAACTCACCGCG ACGTCTGTGCGGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGAT GCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGG CGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCG TTATGTTTATCGGCACTTTGCATCGGCCGCGTCCCGATTCCGGAAGTGCTTGA CATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGG GTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCCGTGTTCTGCAACCC GTCGCGGAGCTCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCG GGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACTACATGGCGTGAT TTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACCTGTGATGGA

	CGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGG CCGAGGACTGCCCCGAAGTCCGGCACCTCGTGACGCGGATTTGCGCTCCAAC AATGTCCTGACGGACAATGGCCGCATAACAGCGGTCAATTGACTGGAGCGAGG CGATGTTGCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCG TGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGG AGCTTGCAAGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGAC CAACTCTATCAGAGCTTGGTTGACGGCAATTCGATGATGCAGCTTGGGCGCA GGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTGCGGGCGTACA CAAATCGCCCGCAGAAGCGCGGCGCTCTGGACCGATGGCTGTGTAGAAGTAC TCGCCGATAGTGGAACCGACGCCCCAGCACTCGTCGAGGGCAAAGGAATG ACCGGAGTTGAGACAAATGGTGTTCAAGATCTCGATAAGATACGTTCAATTTGT CCAAGCAGCAAAGAGTGCCTTCTAGTGATTTAATAGCTCCATGTCAACAAGAA TAAACGCGTTTCGGGTTTACCTCTTCCAGATACAGCTCATCTGCAATGCATTA ATGCATTGGACCTCGCAACCCTAGTACGCCCTTCAGGCTCCGGCGAAGCAGAA GAATAGCTTAGCAGAGTCTATTTTCATTTTCGGGAGACGAGATCAAGCAGATC AACGGTCGTCAAGAGACCTACGAGACTGAGGAATCCGCTCTTGGCTCCACGC GACTATATATTTGTCTCTAATTGTACTTTGACATGCTCCTCTTCTTACTCTGATA GCTTGACTATGAAAATTCCGTCACCAGCCCCTGGGTTTCGAAAGATAATTGCA CTGTTTCTTCTTGAAGCTCTCAAGCCTACAGGACACACATTCATCGTAGGTATA AACCTCGAAAATCATTCTACTAAGATGGGTATACAATAGTAACCATGCATGG TTGCCTAGTGAATGCTCCGTAACACCCAATACGCCGGCCGAAACTTTTTTACAA CTCTCCTATGAGTCGTTTACCCAGAATGCACAGGTACACTTGTTTAGAGGTAAT CCTTCTTCTAGAGATATGGGGTTCATATGCCTTGCTTCTCGATGAAGAGTATA TATATGTATATCAGCAGATACCTCGGGGATTGGATGGAAATCCTCTTGATTTA GCGGCCCTTTCTTGCTTTTTGGAGGATTGGATGGGTAGGGTGAGAATAAAACA AGGACACAGTAGATACATGGTGCTTGCTAAGTATATACTGCATATCAACTG
KMT6	GCCGCATAAACTCTGGGCCAATCACCAACAGCCCCAAGCCCAGCCAACCAGGA CTCGCGTACTCCAGGCTGGCCCTTTGTCCGAAACGCGATCTGAAAGGCATCTC GAATCCGCGATTCTCAGAATCTCAACTGCAGGATCAGGAGCCCTGCTGTTGC GCTGCGTTGGAGATGCCCTGCTGATTTTTCTTCTTACTAGGAGTGGAGAGCT GTCCGCGGGGACCTGAGCTGTTGCCTCCGTGGCGAGGAGACTGGTGGCGGCC TGCAGTTGGACCTGGAATCTTCAAGCCCGTCTTCTTCAGCCTCATGTGCATGA GTTGCTGCCGATCGTTTGCTGTGCGCCGAACGAGACGCTCCTTGACACCGGCT TGTTATCCTCGTGACTTCATTGTGCTTGCCATTGTTTCGCTTCGTTTTCACTACT TTCTCAGCGCACAATCCTTTCAGGCCTCTCATGCAGGCCCTTTTGCAGCCTC GGCCGTTCAATTTGCAGAGCTGGGTCTGCGGCTTCGTTTCGCCATCCCTTCCAGCC TCCCGCTTCCCGCCTGCTTCAGCAGAAAAGGGCGACGCTCACCACCTGCCGTG ACACAGACGGGGCGCAACGGAACAGACACGGATTGCTACCTGCACATACAGG CCAGGATGATGCCCCACGGACCGCCGCGGGTGATCGCAACGTCGGAGGAGG ATCCAAGCCACCTGGTCCGTCTAGATCTAGACTTGACTAGCGGATCTGGCGTC GCGGCGTCGACAAGAGGTCCCATAGCCCTCCTCCCTTGAGGAGAAGCTGCTC TTCAGCTCCTCCAGCGGCTTCCGCGACTTCAGCATCCCAGCCAGCGATCCCAG CAGCGGCAGCGGCAGCGGAGCCTGGACGCCCCTCCCTCTTCTGAGCGCGCG TCAGTTCCCCTGATCGACCGCGTGTCTGAAAATGCGCCGAGATGCCAGAGAC CAGGAAACAAGGCGTGGCATCCGCCACCAGCCCGGCCCTGAACGGTGGGGCA GTGCTAGTGTGTACCGACCCGCAGGATTGGTGCTTTGCCAGAGCTCTACA GAATAGCGCGCGCATCCATATGTTAGTTCTGCAATTTCTTGATCGGTGCTGT GACTCATACTTCCCCCTTGGCTGGCCTTGCGGCAACCAATAAGAACGCACAG TGAAATCTTGCGGGTGGGGAGTGATGGCGCCTGCATTGGCTTGGGGACGCG CACTGTCGCACACTTCCATCTGACCTTTCAGAAGGGTTTCGTGGTGGGCAAGG ACCAACCGGTTGCGCGGCCGTGCGTGGGTGCCTCGCCCGGCACTGCCAGGGC

	CACTGCAGTGGCAGTTTGCTGCCTGATACAAAATCCTTCCCTCCGCCAGTTTT CCCTCTTTGACCTTCCTTTCTTTCTCTGCAACCAAATCCACCCTATCAAACCAA AACAGTGTCTCGACCGAGGTATCAACCTGAATCAGCAACATCGTAGCCAGCAT TTGTCTCCGTCTCTGCAGAACCAGCGAGTTGCAAACATTATCCAGGCAACAGG GCACCAACTCACTTCTTCGGCTTTACCAATCGGTACAGCTCTTCTCAGAACTC GCGTCCGCAACAGTTCTACGCTTCCTCAGCACCTTCTTCAGCTTCAATCCTGAA CACTCAGAACC CGCACAGCAGCGCCCTCCTGTTCCCTTGTTCCCAAAGTAC CGGTAGTATTTTCGCACGGAAAGCAGGGCAACACGTAAGTGATATCTACTAGTT GTGACCGGCGCCTAGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAG TATAATACGACTCACTATAGGAGGGCCACCATGTCGACTACTAACCTTCTTCTC TTTCTACAGCTGAGATCACCGGTAGGAGGGCCATCATGAAAAAGCCTGAAC CACCGCGACGTCTGTGCGGAAGTTTCTGATCGAAAAGTTTCGACAGCGTCTCCG ACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTA GGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACA AAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAA GTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCG TGCACAGGGTGTACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTT TGCAACCCGTCGCGGAGCTCATGGATGCGATCGCTGCGGCCGATCTTAGCCAG ACGAGCGGGTTCGGCCATTTCGACCGCAAGGAATCGGTCAATACACTACAT GGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAAC GTGATGGACGACACCGTCAGTGCGTCCGTGCGCAGGCTCTCGATGAGCTGA TGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGACGCGGATTTT GGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCAATTGACTG GAGCGAGGCGATGTTTCGGGGATTCCCAATACGAGGTGCGCAACATCTTCTTCT GGAGGCCGTGGTTGGCTTGATGGAGCAGCAGACGCGCTACTTCGAGCGGAG GCATCCGGAGCTTGCAAGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTG GTCTTGACCAACTCTATCAGAGCTTGTTGACGGCAATTTTCGATGATGCAGCTT GGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCG GGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGT AGAAGTACTCGCCGATAGTGGAACCGACGCCCCAGCACTCGTCCGAGGGCA AAGGAATGACCGGAGTTGAGACAAATGGTGTTTCAGGATCTCGATAAGATACG TTCATTTGTCCAAGCAGCAAAGAGTGCCCTTCTAGTGATTTAATAGCTCCATGTC AACAAGAATAAAACGCGTTTCGGGTTTACCTCTTCCAGATACAGCTCATCTGCA ATGCATTAATGCATTGGACCTCGCAACCCTAGTACGCCCTTCAGGCTCCGGCG AAGCAGAAGAATAGCTTAGCAGAGTCTATTTTCATTTTCGGGAGACGAGATCA AGCAGATCAACGGTCGTCAAGAGACCTACGAGACTGAGGAATCCGCTCTTGG CTCCACGCGACTATATATTTGTCTCTAATTGTACTTTGACATGCTCCTCTTTTA CTCTGATAGCTTGACTATGAAAAATTCGTCACCAGCCCCCTGGGTTTCGAAAGAT AATTGCACTGTTTCTTCTTGAAGTCTCAAGCCTACAGGACACACATTCATCGT AGGTATAAACCTCGAAAATCATTCTACTAAGATGGGTATACAATAGTAACCA TGCATGGTTGCCTAGTGAATGCTCCGTAACACCCAATACGCCGGCCGAACTT TTTTACAACCTCCTATGAGTCGTTTACCCAGAATGCACAGGTACACTTGTTTA GAGGTAATCCTTCTTTCTAGGCGTAATAAAGGTCAATGTGTCAAGGGTCTTAG GGGGTTATGGAGTTGTACAGAAAAAGGGAGTCATTATTATTGATAGACTTCT CGCAAATGATGTAGGCCGAATTTGGTCCACTGTTGAACCAGGAAAACATGGA ATTGGGGGCAACCGAATGTGATGGCGTGTACCCTTGTGAGGAGTCGCAATTT GAAAGATGCCATGGACTCGGGATGGGGAAAGCTGCCAAACAGGATCAAGGC GATGACGAGCATGTGATGGAGCCCTTTTTTAAATGAGGACAAGGGCAGTTC GTACAAGTACATGGATCTGGAATGTGGCGCAGAGCCGAGTCGCTGCATTGGT CAATGCCCTGTCAGAAACAGTAAGGACAAGTAAACCTCGAAGAGGACGACGA GGTATGTGGTACGAAGCAGACATTCTCTAGGTAGGTAGATATGCATCATGAA GGATGTAATGGCCTATCAACACTGAAACCTGGAAGAACAACAAGGTCACTCA
--	--

	GTATCTGTAATATGTAAAGCTGCAAAATTGAGATTCTCACGTGCTGCCTTACCC AGGTACTCAGCATCCATCGTGATGGCACTTCAATTGATGCCGAGTCAAGCCA CACCACGTCGAGCTCTACCTTACCTTTTGACGCGGGGGCGGGGCGACTACCA GCCACAGAGACAGCTAACGATGACGTGGGCGATTTGACGCGGCTGGAAGCTG CACCAAGCTGAGGTACGAGCGCTTCGTACCTGACCTGACCTGGACCTGAACCC ACAGGCAGCAAACAGAACACTTTGCCTGCTTTAACTTGCAACTTGTCTCTGC TCTTTCTCTTCTTCATTCTACAACAACAACGACAACAACCAGAATTGCGATT CACGACACACACCGAGAGGCGTCGTCGCTGTCTACCTCTGACTTCGCCCATACT CTGAAATTGCCGTCATCGAGCTTCTCCGTTGACCCCGCCGAAAGAACACGAC AAATAGCAAACACAACAG
--	---

VIII. Amorces PCR, ARN guides et sondes marquées pour Southern blot

Nom	Fonction	Sens	Séquence 5' -> 3'
Fwd dim2 Int	Vérif. Disparition	forward	AACACCCGAGCGATGACTAC
Rvse dim2 Int	Vérif. Disparition	reverse	GTGGTCCGTGAACTCAGAGG
Fwd dim2 Ext	Vérif. extérieure Cassette 5'	forward	TACGACCTGGATCCGGACTT
Rvse dim2 Ext	Vérif. extérieure Cassette 3'	reverse	TGCAATTCGACTCGTCAGCT
Fwd dim5 Int	Vérif. Disparition	forward	CGACGAACTGGCCCTTTTG
Rvse dim5 Int	Vérif. Disparition	reverse	GAACGACGCCTTTCACATCG
Fwd dim5 Ext	Vérif. extérieure Cassette 5'	forward	GGGTTGGATGATGCGGTTTG
Rvse dim5 Ext	Vérif. extérieure Cassette 3'	reverse	GCACAGCTGGCTAGATCGAT
Fwd hp1 Int	Vérif. Disparition	forward	GGGCGACTGACTTTTGATGC
Rvse hp1 Int	Vérif. Disparition	reverse	ACCTTCGTGTCATGCTTCGT
Fwd hp1 Ext	Vérif. extérieure Cassette 5'	forward	GTGCCTCATCATCTCACCGT
Rvs hp1 Ext	Vérif. extérieure Cassette 3'	reverse	GTAGGTTGTCTGGTGGGCTC
Fwd kmt6 Int	Vérif. Disparition	forward	CTTGTCATCCTGTGCACCCT
Rvse kmt6 Int	Vérif. Disparition	reverse	TTGCAAGACAGATCGGGTCC
Fwd kmt6 Ext	Vérif. extérieure Cassette 5'	forward	AGGCACCTACCCGTCATTG
Rvse kmt6 Ext	Vérif. extérieure Cassette 3'	reverse	AGCAGCATAAACAGGGCCAT
Fwd rid1 Int	Vérif. Disparition	forward	TCGTGCTACAAGGCCAGAAG
Rvse rid1 Int	Vérif. Disparition	reverse	TGTCAACGGCATACTGGACC
Fwd rid1 Ext	Vérif. extérieure Cassette 5'	forward	CGGCCAACAGTGATTGCTTC
Rvs rid1 Ext	Vérif. extérieure Cassette 3'	reverse	CAATCAAAGTCCCCTCGGCT
Fwd ampli K7 dim2	Amplification cassette clonage	forward	GAGGCTTCCATTCTGCTCA
Rvse ampli K7 dim2	Amplification cassette clonage	reverse	CTTAGCCAAGCACCATGTATCT
Fwd ampli K7 dim5	Amplification cassette clonage	forward	ATAAAGATTGTCGAGTGGGC
Rvse ampli K7 dim5	Amplification cassette clonage	reverse	GCTTCATTCTCATAACGCGT
Fwd ampli K7 hp1	Amplification cassette clonage	forward	CATCCCCAGGAAAAAGTCGG
Rvse ampli K7 hp1	Amplification cassette clonage	reverse	TCCAAGAAAACAATCGAGGC

Fwd ampli K7 rid1	Amplification cassette clonage	forward	GCATCGCAATTCATCAGC
Rvse ampli K7 rid1	Amplification cassette clonage	reverse	GTGTTCTCTTTACCCATGTA
gRNA Rid1	ARN guide	forward	GATATCGTGCGTCTGACGAG
gRNA Kmt6	ARN guide	reverse	TCGTACCATGAAATGTTCCGCGG
gRNA Hp1	ARN guide	reverse	CGTCCTCCCACGATCCCGCGGGG
gRNA Dim5	ARN guide	reverse	GCCATCAGGGACATTCCCAG
gRNA Dim2	ARN guide	forward	GCAACAGCTTGATCCGACAGCGG
Dim5_Fwd_SFn_1	Sonde Southern blot	Forward	AGTTTCCCAAGCGAGGTGAA
Dim5_Rvse_Sfn_2	Sonde Southern blot	Reverse	TCATCAATGCCGGTCAAGCT
Dim5_Fwd_SFn_2	Sonde Southern blot	Forward	ATCGATCTAGCCAGCTGTGC
Dim5_Rvse_Sfn_2	Sonde Southern blot	Reverse	GCAAATGAATGGCAAGGCGA
Dim5_fwd	Amplification gene <i>dim5</i>	Forward	TTGTAAAACGACGGCCAGTGATTT ACACGTGCTCGTCAG
Dim5_rev	Amplification gène <i>dim5</i>	Reverse	TTCTTTCTAGCTGTGCGCTTTTGT GCAAG
G418_fwd	Amplification gène généticine	Forward	AAGCGCACAGCTAGAAAGAAGGA TTACCTCTAAACAAGTG
G418_rev	Amplification gène généticine	Reverse	TGCATGCCTGCAGGTCGACTGGTG GGGCAGTGCTAGTG

IX. Solutions pour Southern blot

Solution de dénaturation

	Concentration finale
NaOH	0,5 M
NaCl	1,5 M

Solution de neutralisation

	Concentration finale
Tris-HCl pH 7,5	0,5 M
NaCl	1,5 M

Solution SSC (20X)

		Concentration finale
NaCl	175,3 g	0,3 M
Sodium citrate	88,2 g	0,3 M

Dissoudre NaCl et sodium citrate dans 800mL d'eau.

Ajuster le pH à 7 avec HCl.

Ajuster jusqu'à 1L avec de l'eau.

Autoclaver.

X. Solution pour extraction d'histones

New Cell Lysis Buffer

	Concentration finale
Tris-HCl pH 7.5	50 mM
NaCl	150 mM
EDTA	5 mM
NP-40	0.5%
Triton	1%

Inhibiteur de protéases Roche (04693132001) : 1 capsule par millilitre de solution.

Nuclei Lysis Buffer

	Concentration finale
Tris-HCl pH 8	50 mM
EDTA	10 mM
SDS	0.5%

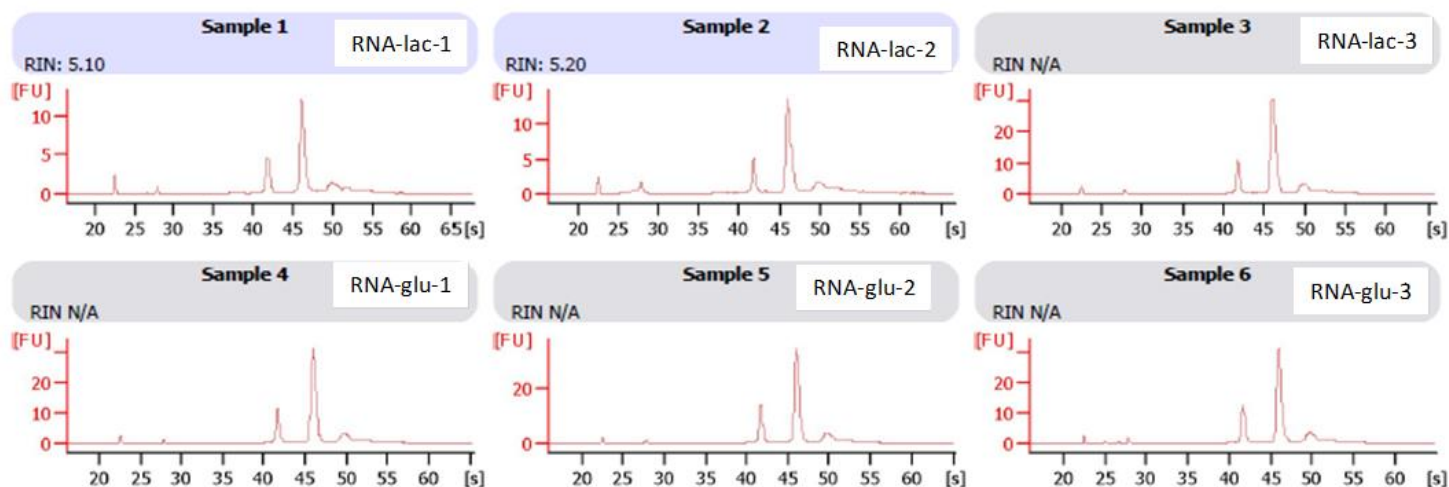
Inhibiteur de protéase Roche (04693132001) : 1 capsule par millilitre de solution.

XI. Cassette de complémentation pour les mutants *Δdim5*

Séquence de l'insert
GCGCCCAATACGCAAACCGCCTCTCCCCGCGGTTGGCCGATTCAATATGCAGCTGGCACGACAGGTTTCCCG ACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACA CTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCTATGACC ATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCCTGG CCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTTAATCGCCTTGCAGCACATCCCCCTT TCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGA ATGGCGCCTGATGCGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCGCATATGGTGCATCTCAGTAC AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCT TGTCTGCTCCCGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACC GTCATACCCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATA ATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTATTTTTCTAAATA CATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAT GAGTATTCAACATTTCCGTGTGCCCCCTATTCCCTTTTTGCGGCATTTTGCCTTCCTGTTTTGCTCACCCAGAA ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAAC AGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATG TGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGAC TTGGTTGAGTACTACCCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTG CCATAACCATGAGTGATAAACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGC TTTTTGCACAACATGGGGGATCATGTAACCTGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCA AACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAAGTGGCGAACTAC TACTCTAGCTTCCCGGCAACAATTAAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTC GGCCCTTCCGGCTGGCTGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCA GCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGAT GAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTGTGACACCAAGTTTACT CATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT

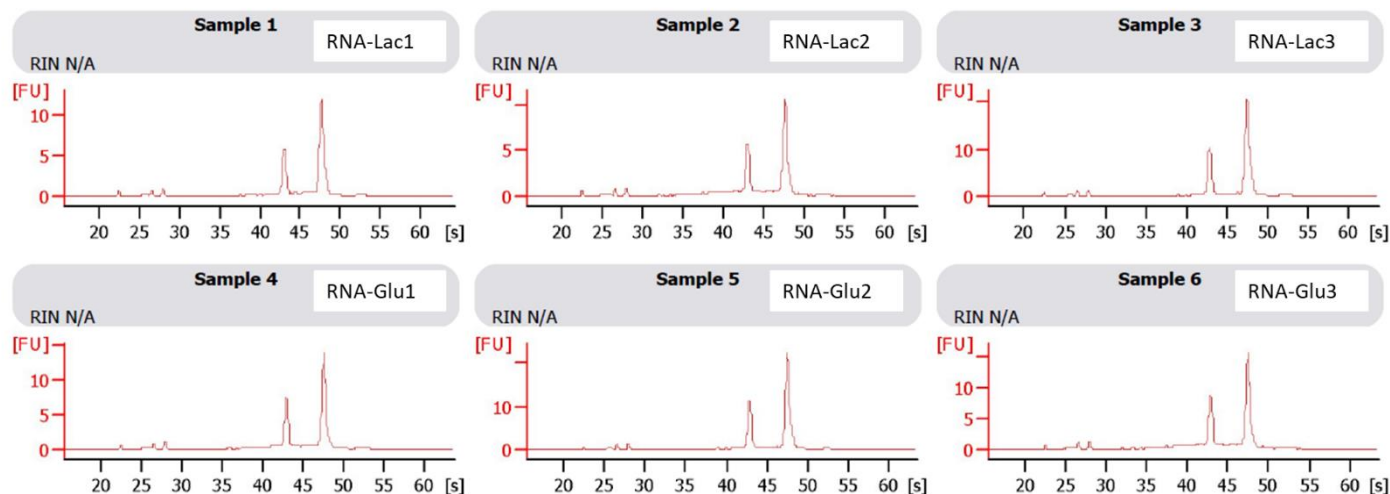
GAGGTTTATACCTACGATGAATGTGTGTCCTGTAGGCTTGAGAGTTCAAGGAAGAAACATGCAATTATCTTTGC
GAACCCAGGGCTGGTGACGGAATTTTCATAGTCAAGCTATCAGAGTAAAGAAGAGGAGCATGTCAAAGTACA
ATTAGAGACAAATATATAGTCGCGTGGAGCCAAGAGCGGATTCTCAGTCTCGTAGGTCTCTTGACGACCGTT
GATCTGCTTGATCTCGTCTCCCGAAAATGAAAATAGCTCTGCTAAGCTATTCTTCTCTTCGCCGGAGCCTGAAG
GCGTTACTAGGTTGCAGTCAATGCATTAATGCATTGCAGATGAGCTGTATCTGGAAGAGGTAAACCCGAAAAC
GCGTTTTATTCTTGTTGACATGGAGCTATTAAATCACTAGAAGGCACTCTTGCTGCTTGGACAAATGAACGTA
TCTTATCGAGATCCTGAACACCATTTGTCTCAACTCCGGCTAGCGAATTCTCGACCCTCAGGTCAGAAGAACTC
GTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGT
CAGCCCATTGCGCGCCAAGTTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACA
CCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAGGCATCGC
CATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCTCGCCTTGAGCCTGGCGAACAGTTCGGCTGGCGCGA
GCCCCTGATGTTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATG
CGATGTTTCGCTTGGTGGTGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGATTGCATCAGCCA
TGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCCCAATAGCA
GCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAACGCCGTCGTGGCCAGCCACGA
TAGCCGCGCTGCCTCGTCTTGCAAGTTCATTAGGGCACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGC
CCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCAGTCATAGCCGAATA
GCCTCTCCACCCAAGCGGCCGAGAACCTGCGTGCAATCCATCTTGTTCAATCATCCTCAGGGATGGCCCTCCT
ACCGGTGATCTCAGCTGTAGGAAAGAGAAGAAGGTTAGTAGTCGACATGGTGGCCCTCCTATAGTGAGTCGT
ATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAACACTAGGCGCCGGTCACAAGTAGTAGATATCA
CTTACGTGTTGAGAGGCGGCATGCGATAAGAGGTGTAATTACCTGAGAACATCTTGTTGCCCTGCTTTCCGTGC
GAAATACTACCGGTACTTTTGGGAAACAAGGGAACAGGAGGGCGCTGCTGTGCGCGGTTCTGAGTGTTACGG
ATTGAAGCTGAAGAAGGTGCTGAGGAAGCGTAGAACTGTTGCGGACGCGAGTCTGAGAAGAGCTGTACCG
ATTGGTGAAAGCCGAAGAAGTGAGTTGGTGCCCTGTTGCTGGATAATGTTTGCAACTCGCTGGTTCTGCAGA
GACGGAGACAAATGCTGGCTACGATGTTGCTGATTCAGGTTGATACCTCGGTCGAGACACTGTTTTGGTTTGA
TAGGGTGATTGTTGTTGTCAGAGAAGAGAAAGGAAGGTCAAAGAGGGAAAACTGGGCGGAGGGAAGGATTT
TGTATCAGGCAGCAAATGCCACTGCAGTGGCCCTGGCAGTGCCGGGCGAGGCACCCACGCACGGCCGCGCA
ACCGGTTGGTCCTTGCCACACGAAACCCTTCTGAAAGGTCAGATGGAAGTGTGCGACAGTGCGCGTCCCCA
AGCCAATGCAGGCGCCATGCACTCCCCACCCGCAAGATTCACTGTGCGTTCTTATTGGTTGCCGCAAGGCCAGC
CAAAGGGGGAAGTATGAGTCACAGCACCGATACAAGAAAATTGCAGAACTAACATATGGATGCGCGCGCTAT
TCTGTAGAGCTCTGGGCAAAGCACCAATCCTGCGGGTCGGTACACACACTAGCACTGCCCCACC

XII. Résultats bioanalyseur ARN QM6a



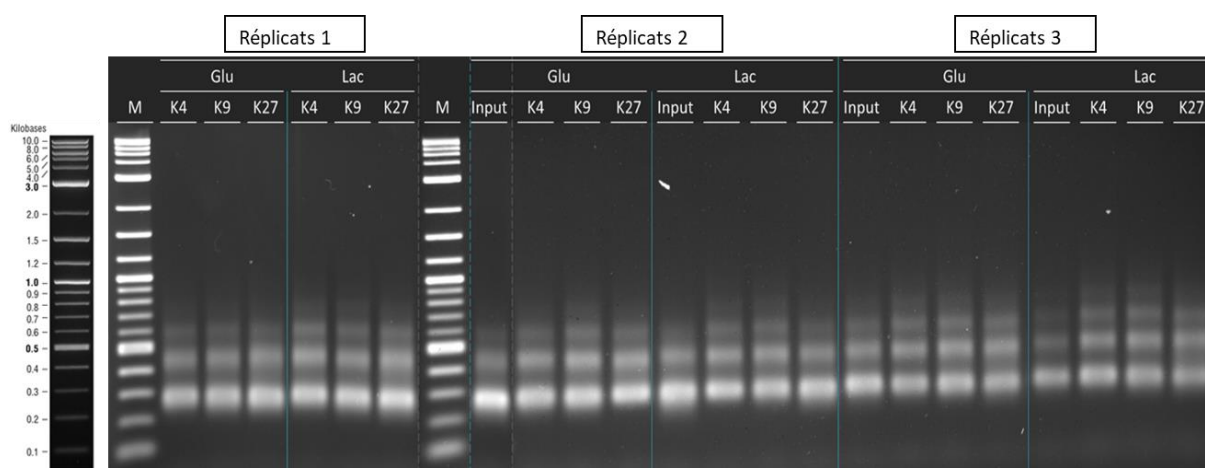
Les deux pics visibles sur les résultats de bioanalyseur sont le pic 18S et le pic 28S. Ces deux pics correspondent aux deux sous-unités ribosomales majeures, respectivement la petite (18S) et la grande (28S) sous-unité.

XIII. Résultats bioanalyseur ARN RutC30



Les deux pics visibles sur les résultats de bioanalyseur sont le pic 18S et le pic 28S. Ces deux pics correspondent aux deux sous-unités ribosomales majeures, respectivement la petite (18S) et la grande (28S) sous-unité.

XIV. Gel d'électrophorèse de la qualité de la librairie RutC30



Après amplification des fragments d'ADN immunoprécipités, chaque échantillon a été visualisé sur gel d'agarose 1,5% afin de s'assurer de la qualité de la fragmentation de la chromatine.

XV. Résultats d'alignement des *reads* pour les ChIP Seq des deux souches

Echantillon QM6a	Pourcentage <i>reads</i> mappés	Pourcentage <i>reads</i> éliminés pour non alignement	Pourcentage <i>reads</i> éliminés pour <i>multimapping</i>
Glu1_K4	88,18%	11,82%	0,00%
Glu2_K4	90,40%	9,60%	0,00%
Glu3_K4	79,44%	20,55%	0,00%
Lac1_K4	89,04%	10,96%	0,00%
Lac2_K4	88,37%	11,62%	0,00%
Lac3_K4	85,06%	14,93%	0,01%
Echantillon RutC30			
Glu1_K4	84,62%	15,33%	0,05%
Glu2_K4	84,56%	15,39%	0,04%
Glu3_K4	82,33%	17,62%	0,05%
Lac1_K4	84,26%	15,69%	0,05%
Lac2_K4	82,09%	17,87%	0,04%
Lac3_K4	82,97%	16,99%	0,04%

XVI. Résultats d'alignement des *reads* pour les ChIP Seq des deux souches après suppression des duplicats

Echantillon QM6a	Pourcentage duplicats dans échantillon mappé	Pourcentage reads totaux utilisés (sur reads de départ)
Glu1_K4	17,67%	72,60%
Glu2_K4	17,71%	74,39%
Glu3_K4	13,44%	68,76%
Lac1_K4	15,78%	74,99%
Lac2_K4	14,47%	75,59%
Lac3_K4	11,19%	75,55%
Echantillon RutC30		
Glu1_K4	22,58%	65,52%
Glu2_K4	17,13%	70,08%
Glu3_K4	13,53%	71,19%
Lac1_K4	13,68%	72,74%
Lac2_K4	16,37%	68,66%
Lac3_K4	14,24%	71,15%

XVII. Gènes touchés par H3K4me3 spécifiques à la QM6a

ID-IFPEN	Commentaire
TrA0066W	Putative Glycosyltransferase family 2
TrA0161W	Putative Platelet-activating factor acetylhydrolase
TrA0199C	Putative Ribosomal RNA methyltransferase RrmJ/FtsJ domain protein
TrA0768C	Putative Glutamate-1-semialdehyde aminotransferase
TrA0951C	Putative Polyphosphate phosphatase DDP1
TrA0991W	Putative Maltose permease
TRA1199W	Putative Tubulin beta chain 1
TrA1335W	Putative GTP-binding protein RBG1
TrA1449C	Putative S-adenosylmethionine-dependent methyltransferase TAE1
TrA1513C	Putative NAD-binding Rossmann fold oxidoreductase
TrA1600W	Putative S-adenosylmethionine-dependent methyltransferase MTQ2
TrA1606W	Putative Low-affinity copper transporter of the vacuolar membrane CTR2
TrA1658C	Putative Peptidase
TrA1803W	Putative Fatty acid elongation protein GNS1/SUR4
TrA1833C	Putative HIT/MYND zinc finger domain protein
TrA1871C	Putative Urate oxidase
TrA1929W	Putative Adenosine/AMP deaminase
TrA2009C	Putative WD40 repeat protein DGR2
TrB0268W	Putative ABC transporter
TrB0292C	Putative Oxidoreductase
TrB0612C	Putative Secreted oxalate decarboxylase
TrB0613W	Putative DUF1275 family protein
TrB0691C	Putative Integral membrane protein
TrB0777W	Putative Ankryn repeat protein
TrB1052W	Putative ARM repeat GTP-binding protein
TrB1255C	Putative MFS transporter
TrB1551C	Putative DUF59 family protein
TrB1601W	Putative RNA polymerase II mediator complex subunit (Middle) MED5
TrB1644W	Putative Serine/threonine protein kinase
TrB1825C	Putative Tropine dehydrogenase
TrC0078W	Putative Complex with Rad50p and Xrs2p (MRX complex) MRE11
TrC0477W	Putative Cytosolic phospholipase A2 zeta
TrC0559C	Putative Elongator complex subunit ELP4
TrC0563C	Putative Nucleoside diphosphatase
TrC0766W	Putative GPCR protein class XII (C-like) gpr-13
TrC0791W	Putative MFS transporter
TrC0877W	Putative Acyl-CoA N-acyltransferase
TrC0943C	Putative MFS transporter
TrC0951C	Putative Skeleton binding protein
TrC1010W	Putative Trypsin
TrC1011W	Putative DNA repair RAD5c / htf1
TrC1210C	Putative α -factor (ppg1) protease BAR1
TrC1539C	Putative Zn(2)Cys6 nitrate assimilation regulatory protein NirA
TrD0215W	Putative Mitochondrial carrier protein PET8
TrD0475C	Putative cAMP dependent protein kinase 2 catalytic subunit PKAC2

TrD0578C	Putative Cu ²⁺ /Zn ²⁺ superoxide dismutase SOD1
TrD0606W	Putative Integral membrane protein
TrD0701C	Putative Xylulokinase XKI1
TrD1077W	Putative N-acylethanolamine amidohydrolase
TrD1207W	Putative C2H2 transcription factor sgr-30
TrE0131C	Putative Lysophospholipase precursor PLB1
TrE0285C	Putative Dnase I
TrE0341W	Putative NADH cytochrome b5 reductase
TrE0661W	Putative Salicylate hydroxylase
TrE0662C	Putative MFS transporter
TrE0663W	Yellow pigment regulator 2 Ypr2
TrE0779W	Putative Cell wall biogenesis protein ECM15
TrE1267W	D-mannitol 2-dehydrogenase LXR1
TrE1290W	Putative DUF500 family protein
TrF0392C	Putative FAD/NAD(P)-binding domain protein
TrF0443W	Putative Protein required for thiolation of the uridine at the wobble position of Lys(UUU) and Glu(UUC) tRNAs NCS2
TrF0801C	Putative E3 ubiquitin-protein ligase ARI9
TrF0804W	Putative Heat shock protein family Hsp70
TrG0275W	Putative Serine hydrolase FSH2
TrG0873C	Putative Proline utilization protein PrnX-like
TrG0999W	Putative Gliotoxin biosynthesis protein GliK
TrG1035C	Putative NAD-dependent epimerase/dehydratase
TrG1048W	Putative protein of unknown function
TrG1073C	Putative protein of unknown function
TrG1107W	Putative protein of unknown function
TrG0470C	Putative protein of unknown function
TrG0496-2W	Putative protein of unknown function
TrG0510-2W	Putative protein of unknown function
TrG0551-2C	Putative protein of unknown function
TrG0625-2W	Putative protein of unknown function
TrF0831-2C	Putative protein of unknown function
TrF0893-2W	Putative protein of unknown function
TrF0941C	Putative protein of unknown function
TrG0236W	Putative protein of unknown function
TrF0583-2C	Putative protein of unknown function
TrF0774-2W	Putative protein of unknown function
TrE0859-2C	Putative protein of unknown function
TrE0872W	Putative protein of unknown function
TrE0925-2C	Putative protein of unknown function
TrE1090-2W	Putative protein of unknown function
TrE1145-2W	Putative protein of unknown function
TrE0675W	Putative protein of unknown function
TrE0711-2W	Putative protein of unknown function
TRE0165-2W	Putative protein of unknown function
TrD1385W	Putative protein of unknown function
TrD0782-2W	Putative protein of unknown function

TrD0806-2C	Putative protein of unknown function
TrD0544-2W	Putative protein of unknown function
TrD0556W	Putative protein of unknown function
TrC1041-2C	Putative protein of unknown function
TrC1183-3C	Putative protein of unknown function
TrC0922-2C	Putative protein of unknown function
TrC1221W	Putative protein of unknown function
TrC1341W	Putative protein of unknown function
TrC0227-2W	Putative protein of unknown function
TrC0329-2W	Putative protein of unknown function
TrC0683-2C	Putative protein of unknown function
TrB1408-2C	Putative protein of unknown function
TrB1441-2C	Putative protein of unknown function
TrB1453-2W	Putative protein of unknown function
TrB1496-2C	Putative protein of unknown function
TrB1550W	Putative protein of unknown function
TrC0535-2C	Putative protein of unknown function
TrA0148-2W	Putative protein of unknown function
TrB1683-2W	Putative protein of unknown function
TrB1156-2C	Putative protein of unknown function
TrB0845C	Putative protein of unknown function
TrB0395-2C	Putative protein of unknown function
TrA2056-2C	Putative protein of unknown function
TrA2063-2W	Putative protein of unknown function
TrB0191-2W	Putative protein of unknown function
TrA1874-2W	Putative protein of unknown function
TrA1568-2C	Putative protein of unknown function
TrA0293-2C	Putative protein of unknown function
TrA0302C	Putative protein of unknown function
TrA0644C	Putative protein of unknown function
TrA1741W	Putative protein of unknown function
TrB1407W	Putative protein of unknown function
TrA1723-2W	Putative protein of unknown function
TrA1788-2C	Putative protein of unknown function
TrA1338-2C	Putative protein of unknown function
TrA1368-2W	Putative protein of unknown function
TrA1035-2W	Putative protein of unknown function
TrA0922-2W	Putative protein of unknown function
TrA1333W	Putative protein of unknown function
TrA1431W	Putative protein of unknown function
TrB1659W	Putative protein of unknown function
TrB1619W	Putative protein of unknown function
TrA1796W	Putative protein of unknown function
TrB1828C	Putative protein of unknown function
TrB1880C	Putative protein of unknown function
TrC0116W	Putative protein of unknown function

TrG0911W	Putative protein of unknown function
TrD0209C	Putative protein of unknown function
TrC0582W	Putative protein of unknown function
TrD0300W	Putative protein of unknown function
TrD0618W	Putative protein of unknown function
TrE1184W	Putative protein of unknown function
TrF0012W	Putative protein of unknown function
TrF0013C	Putative protein of unknown function
TrB0509C	Putative protein of unknown function

XVIII. Gènes touchés par H3K4me3 spécifiques à la RutC30

ID-IFPEN	Commentaire
TrA0030W	Putative Taurine catabolism dioxygenase TauD
TrA0200C	Putative bZIP transcription factor similar to Cap2/Hap43 involved in Iron Homeostasis Gene Regulation.
TrA0210C	Carbohydrate esterase family AES1
TrA0362C	Putative MFS quinate transporter
TrA0397C	Putative CobW domain protein
TrA0418C	Putative Polyketide synthase PKS9
TrA0430C	Putative NAD-dependent aldehyde dehydrogenase
TrA0470C	Putative Glycoside hydrolase family 16
TrA0578C	Putative MFS transporter
TrA0597W	Putative DUF3433 family protein
TrA0666W	Putative Secreted lipase/esterase
TrA0671C	Putative Polyphosphatidylinositol phosphatase INP52
TrA0798W	Putative 3'-5' exoribonuclease RNH70
TrA0827C	Putative Mitochondrial delta-1-pyrroline-5-carboxylate dehydrogenase precursor PUT2
TrA0863W	Putative MFS L-fucose transporter
TrA0865C	Putative Ferric reductase transmembrane component
TrA0931W	Putative Asparaginase
TrA0983W	Putative Guanine nucleotide exchange factor
TrA1003W	Putative Meiotic fizzy-related protein
TrA1011W	Putative MutS family protein involved in meiosis MSH5
TrA1051W	Putative Histone deacetylase PHD1 / HOS2 / HosA
TrA1136C	Glycoside hydrolase family 20 NAG1
TrA1155W	Putative ER membrane protein PRO41
TrA1166C	Putative Inositol-1,4,5-trisphosphate 5-phosphatase
TrA1230C	Putative CPF (cleavage and polyadenylation factor) subunit MPE1
TrF1148C	Putative Small secreted protein IDC2-like
TrF1155W	Putative FAD/NAD(P) dependent oxidoreductase
TrB0038W	Putative Glycosyltransferase family 15
TrB0144W	Putative NAD(P)-binding Rossmann-fold domain protein
TrB0191W	Putative Secreted subtilisin
TrB0206W	Putative Transcriptional regulator of asexual and sexual reproduction MedA / acon-3
TrB0212C	Putative Amino acid transporter
TrB0277C	Putative GPCR PTH11 type receptor gpr-25
TrB0334W	Putative MFS transporter
TrB0352W	Putative MFS transporter
TrB0441W	Putative Short chain dehydrogenase
TrB0449W	Putative Aminotransferase
TrB0450W	Putative Secreted pepsin
TrB0454C	Putative FAD monooxygenase BNA4
TrB0496C	D-xylose reductase XYL1
TrB0567C	Putative Zn(2)Cys6 transcription factor (fungal specific) similar to cellulose degradation regulator CLR1
TrB0638W	Putative Secreted alpha/beta-hydrolase

TrB0673C	Putative Cystathionine gamma-synthase
TrB0719W	Putative Phospholipid:diacylglycerol acyltransferase LRO1
TrB0738W	Putative Succinyl-CoA:3-ketoacid-coenzyme A transferase
TrB0762C	Putative tRNA methyltransferase TRM10
TrB0960W	Putative Mitochondrial intermembrane space protein MDM35
TrB0984C	Putative Regulator involved in transcriptional control of spore-specific and metabolic genes VadA
TrB1181C	Putative Metalloprotease
TrB1225W	Putative Linoleate diol synthase PpoA / B
TrB1268W	Putative MFS transporter
TrB1269C	Putative Zn(2)Cys6 transcription factor (fungal specific)
TrB1278W	Putative FAD/NAD(P)-binding domain protein
TrB1285C	Putative Allantoinase DAL1
TrB1302C	Putative F-box protein
TrB1312C	Putative Clavamate synthase
TrB1388W	Putative MFS transporter
TrB1408C	Putative Glycerol-3-phosphate O-acyltransferase SCT1
TrB1419C	Putative ATP-dependent RNA helicase DBP5
TrB1428C	Putative Uracil-DNA glycosylase
TrB1462W	L-arabinitol 4-dehydrogenase LAD1
TrB1486C	Putative Ca ²⁺ -dependent lipid-binding protein CLB1/vesicle protein vp115/Granuphilin A
TrB1537C	Putative Glutaminy-peptide cyclotransferase
TrB1580C	Putative LIMBR1 domain protein
TrB1629W	Putative Nuclear migration protein
TrB1640C	Putative Cell division control protein 14
TrB1873C	Putative protein PAR32
TrC0040W	Putative Crossover junction endonuclease EME1
TrC0054W	Putative Zn(2)-C6 fungal-type DNA-binding domain protein
TrC0077W	Putative Aldo-keto reductase
TrC0112C	Putative Zn(2)Cys6 transcription factor OefC type (overexpression confers the "fluffy" phenotype)
TrE0157C	Putative Acyl-CoA N-acyltransferase
TrC0174W	C2H2 Transcription factor involved in conidiation and cellulases regulation BrIA / RXE1
TrC0220W	Putative DNA mismatch repair and crossing-over during meiotic recombination protein MLH3
TrC0270W	Putative O-acetylhomoserine MET17
TrC0271W	Putative Nonselective cation channel
TrC0279W	Plasma membrane H ⁺ -ATPase PMA2
TrC0329W	Putative S-adenosylmethionine-dependent methyltransferase TRM12
TrC0362W	Putative D-lactate dehydrogenase
TrC0374C	Putative Quinoprotein alcohol dehydrogenase-like domain protein
TrC0415W	Putative Secreted aspartic protease
TrC0436W	Putative Nitrogen assimilation transcription factor NIT4 / NirA
TrC0535C	Putative ATP-dependent helicase HRQ1
TrC0577W	Putative DNA-directed DNA polymerase.
TrC0624C	Putative Glycoside hydrolase family 31

TrC0680C	Putative Transcription factor
TrC0683C	Putative Mitochondrial division protein MDV1
TrC0741W	Putative GPI anchored protein
TrC0884C	Putative RING/U-box domain protein
TrC0889W	Putative Phosphate metabolism protein
TrC0899W	Putative tRNA methyltransferase TRM1
TrC0927W	Putative Acyl-CoA N-acyltransferase
TrC0933W	Putative Transcription factor of gluconeogenesis ERT1
TrC0974C	Putative Protein involved in the multivesicular body sorting pathway IST1
TrA1324C	Putative Selenoprotein domain protein
TrA1310C	Putative Spore adhesion transcription factor TRA1 (M. grisea)
TrA1399W	Putative MFS transporter
TrA1580C	Putative Conidiation transcription factor FlbC (ortholog A. nidulans) involved in cellulase and xylanase production (P. oxalicum)
TrA1598W	Putative Serine/threonine protein kinase
TrA1605C	Histone-lysine N-methyltransferase SMYD3
TrC1041C	Putative MATE family protein
TrC1084C	Putative RING/U-box domain protein
TrC1085C	Putative MFS transporter
TrC1117W	Putative Telomere binding factor Tay1
TrC1143W	Putative Class II abasic (AP) endonuclease involved in repair of DNA damage APN2
TrC1193W	Putative MFS nicotinic acid transporter TNA1
TrC1228C	Putative Mitochondrial pyrimidine nucleotide transporter RIM2
TrC1305W	Putative Glutathione-dependent formaldehyde-activating protein
TrC1327C	Putative YAE1 protein
TrC1443W	Putative Zn(2)Cys6 transcription factor (fungal specific)
TrC1463W	Putative Flavin-containing monooxygenase
TrC1552C	Glycoside hydrolase family 3 BXL1
TrD0036W	MFS pentose transporter TrSTR1
TrD0187W	Putative Ribosomal protein L29
TrD0194C	Putative Alpha/beta-hydrolase
TrD0323C	Putative Iron regulated transporter
TrD0360W	Putative Glycoside hydrolase family 18
TrD0414W	Putative Fungal specific cysteine rich domain protein
TrD0527C	Putative Phenol 2-monooxygenase
TrD0531W	Putative Chromo domain protein
TrD0533W	Putative INO80 subunit 2 IES2
TrD0543W	Putative ABC transporter
TrD0664W	Putative Cell division cycle protein CDC123
TrD0693W	Putative Ribonuclease H-like
TrD0715W	Secreted pepsin P7959
TrD0742C	Putative Protein disulfide-isomerase ERP38 / TigA
TrD0873W	Putative Protein tyrosine phosphatase pty-5 involved in conidiation
TrD0889C	Putative SNARE domain protein
TrD0896W	Putative Glycosyltransferase family 1
TrD0986C	Putative Laccase
TrD1096W	Putative Aldehyde dehydrogenase

TrD1176C	Putative NAD(P)-binding Rossmann-fold domain protein
TrD1190W	Putative dsRNA-binding domain protein
TrD1298C	Putative Adenylate cyclase
TrD1403C	Putative Mitochondrial tRNA:pseudouridine synthase PUS9
TrE0273C	Putative S-adenosylmethionine-dependent methyltransferase
TrE0310C	Putative Transcription factor involved in ER stress response and cellulase synthesis res-1 / ZfpA
TrE0484C	Putative Nucleosome assembly protein NAF2
TrE0494C	Zn(2)Cys6 transcriptional repressor of GH11 xylanase genes SxIR
TrE0539W	NDT80 Transcription factor VIB1
TrE0611C	Putative ATP-dependent DNA helicase SGS1 / qde-3
TrE0620W	Putative Galactose oxidase
TrE0638W	Putative GPCR receptor class VII (MG00532-like) / gpr-8
TrE0639W	Putative Glycoside hydrolase family 5
TrE0643W	Putative Ammonium permease MEPC
TrE0671C	Putative Auxiliary Activities enzyme family 11
TrE0711W	Putative Protein similar to <i>S. cerevisiae</i> SSU processome required for the biogenesis of 18s rRNA component UTP18
TrE0729C	Putative Nucleoside hydrolase
TrE0803W	Putative WD40 repeat protein
TrE0936W	Putative Set-domain histone methyltransferase
TrE0940C	Putative MFS transporter
TrE0945C	Putative Velvet domain protein VosA
TrE0964C	Putative Ornithine carbamoyltransferase ARG3
TrC0921W	Putative Lysophosphatidylcholine (lyso-PC) acyltransferase involved in cardiolipin pathway TAZ1
TrE1138C	Putative MFS sugar transporter (Fucose)
TrE1281W	Putative Mitochondrial proline oxidase PUT1
TrF0098C	Putative Periplasmic binding protein-like II domain protein
TrF0417W	Putative FAD monooxygenase
TrF0427C	Putative 66S pre-ribosomal particles component TIF6
TrF0537W	Putative Forkhead transcription factor FKH1
TrF0583C	Putative Mitochondrial translation optimization protein MTO1
TrF0643W	Putative Aminopeptidase
TrF0767C	Putative Phospholipase D active site protein
TrF0820C	Putative DNA polymerase delta subunit 4
TrF0831C	Putative Esa1p-associated factor EAF3
TrF1042C	Putative Zn(2)Cys6 transcription factor (fungal specific) similar to ClbR (<i>A. aculeatus</i>) involved in xylanase and cellulase production (<i>P. oxalicum</i>) / sah-3
TrA1630C	Putative SAM and PH domain protein
TrA1762W	Putative Diphthamide biosynthesis protein JJJ3
TrA1776W	Putative Prolyl aminopeptidase PapA
TrA1788C	Putative Glycoprotease QRI7
TrA1820C	Putative Carbohydrate Esterase family 9
TrA1856W	Putative Prephenate dehydrogenase [NADP+] TYR1
TrA1874W	Putative Zn(2)Cys6 transcription factor (fungal specific)
TrA1911W	Putative Glycoside hydrolase family 18

TrG0041C	Putative MFS transporter
TrG0156C	Putative Class I glutamine amidotransferase-like
TrG0261C	Putative Dam1 complex (aka DASH complex) subunit SPC19
TrG0359C	Putative CFEM domain protein
TrG0392C	Putative Catechol dioxygenase
TrG0430W	Putative Glycoside hydrolase family 47
TrG0473C	Putative Glycoside hydrolase family 92
TrG0489C	Putative Secreted cell wall protein
TrG0496W	Putative GPCR GprK-like receptor RgsC / MDM1
TrG0551C	Putative CCR4-NOT complex subunit NOT4 / MOT2
TrG0555C	Putative DUF159 family protein
TrG0629W	Putative NAD(P)H-dependent quinone reductase
TrG0706W	Putative Transcription factor involved in zinc homeostasis ZAP1 / ZNF658
TrG0711W	Putative Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex
TrG0806W	Putative Cysteine desulfurase
TrG0887C	Putative Cytosolic branched-chain amino acid aminotransferase BAT2
TrG0897W	Secreted tyrosinase TYR2
TrG1054W	Putative Glycosyltransferase family 59
TrG1083W	Putative DNA polymerase zeta subunit MUS26 / REV7
TrG1121C	Putative MFS sugar transporter BSC6
TrG0072W	Putative protein of unknown function
TrG0101C	Putative protein of unknown function
TrA1922W	Putative protein of unknown function
TrA2049W	Putative protein of unknown function
TrA1909W	Putative protein of unknown function
TrG0167W	Putative protein of unknown function
TrG0230C	Putative protein of unknown function
TrG1157W	Putative protein of unknown function
TrG0937W	Putative protein of unknown function
TrG0992W	Putative protein of unknown function
TrG0625C	Putative protein of unknown function
TrG0900W	Putative protein of unknown function
TrG0929W	Putative protein of unknown function
TrG0771W	Putative protein of unknown function
TrG0748C	Putative protein of unknown function
TrG0539W	Putative protein of unknown function
TrG0542C	Putative protein of unknown function
TrG0493C	Putative protein of unknown function
TrG0330W	Putative protein of unknown function
TrA1649C	Putative protein of unknown function
TrA1672W	Putative protein of unknown function
TrA1720W	Putative protein of unknown function
TrA1728C	Putative protein of unknown function
TrA1735W	Putative protein of unknown function
TrA1751C	Putative protein of unknown function
TrF0985W	Putative protein of unknown function

TrF1005W	Putative protein of unknown function
TrF0850W	Putative protein of unknown function
TrF0872C	Putative protein of unknown function
TrF0935C	Putative protein of unknown function
TrF0941W	Putative protein of unknown function
TrF0951C	Putative protein of unknown function
TrF0983C	Putative protein of unknown function
TrF0773C	Putative protein of unknown function
TrF0791W	Putative protein of unknown function
TrF0798W	Putative protein of unknown function
TrF0430W	Putative protein of unknown function
TrF0439W	Putative protein of unknown function
TrF0731W	Putative protein of unknown function
TrE1284C	Putative protein of unknown function
TrE0971W	Putative protein of unknown function
TrE0990C	Putative protein of unknown function
TrE1119C	Putative protein of unknown function
TrE1139W	Putative protein of unknown function
TrE1145W	Putative protein of unknown function
TrE1210C	Putative protein of unknown function
TrE1211C	Putative protein of unknown function
TrE1213W	Putative protein of unknown function
TrE1214W	Putative protein of unknown function
TrE1235W	Putative protein of unknown function
TrF0630W	Putative protein of unknown function
TrC1329W	Putative protein of unknown function
TrC1330C	Putative protein of unknown function
TrC1393C	Putative protein of unknown function
TrC1548C	Putative protein of unknown function
TrG0187C	Putative protein of unknown function
TrD0264W	Putative protein of unknown function
TrD0341C	Putative protein of unknown function
TrD0366W	Putative protein of unknown function
TrD0458C	Putative protein of unknown function
TrD0568W	Putative protein of unknown function
TrD0626W	Putative protein of unknown function
TrD0676W	Putative protein of unknown function
TrD0738C	Putative protein of unknown function
TrD0775C	Putative protein of unknown function
TrD0830W	Putative protein of unknown function
TrD0851C	Putative protein of unknown function
TrD0937W	Putative protein of unknown function
TrD0954C	Putative protein of unknown function
TrD1141C	Putative protein of unknown function
TrD1112W	Putative protein of unknown function
TrD1215C	Putative protein of unknown function

TrD1267W	Putative protein of unknown function
TrD1383W	Putative protein of unknown function
TrD1397W	Putative protein of unknown function
TrE0613W	Putative protein of unknown function
TrE0614W	Putative protein of unknown function
TrE0616W	Putative protein of unknown function
TrE0619C	Putative protein of unknown function
TrE0602C	Putative protein of unknown function
TrE0512W	Putative protein of unknown function
TrE0023W	Putative protein of unknown function
TrE0036C	Putative protein of unknown function
TrE0061C	Putative protein of unknown function
TrE0082W	Putative protein of unknown function
TrE0102C	Putative protein of unknown function
TrE0195W	Putative protein of unknown function
TrE0288W	Putative protein of unknown function
TrE0358C	Putative protein of unknown function
TrE0428C	Putative protein of unknown function
TrE0436W	Putative protein of unknown function
TrE0471W	Putative protein of unknown function
TrE0675C	Putative protein of unknown function
TrE0790C	Putative protein of unknown function
TrE0909C	Putative protein of unknown function
TrE0962C	Putative protein of unknown function
TrE0934W	Putative protein of unknown function
TrE0804W	Putative protein of unknown function
TrA0082C	Putative protein of unknown function
TrA0198C	Putative protein of unknown function
TrA0388C	Putative protein of unknown function
TrA0454W	Putative protein of unknown function
TrA0503W	Putative protein of unknown function
TrA0644W	Putative protein of unknown function
TrA0667C	Putative protein of unknown function
TrA0665W	Putative protein of unknown function
TrA0704W	Putative protein of unknown function
TrA0712C	Putative protein of unknown function
TrA0737C	Putative protein of unknown function
TrA0784C	Putative protein of unknown function
TrA0793W	Putative protein of unknown function
TrA0131C	Putative protein of unknown function
TrA0458C	Putative protein of unknown function
TrA0866W	Putative protein of unknown function
TrA0906W	Putative protein of unknown function
TrA1025W	Putative protein of unknown function
TrA1088C	Putative protein of unknown function
TrA1125C	Putative protein of unknown function

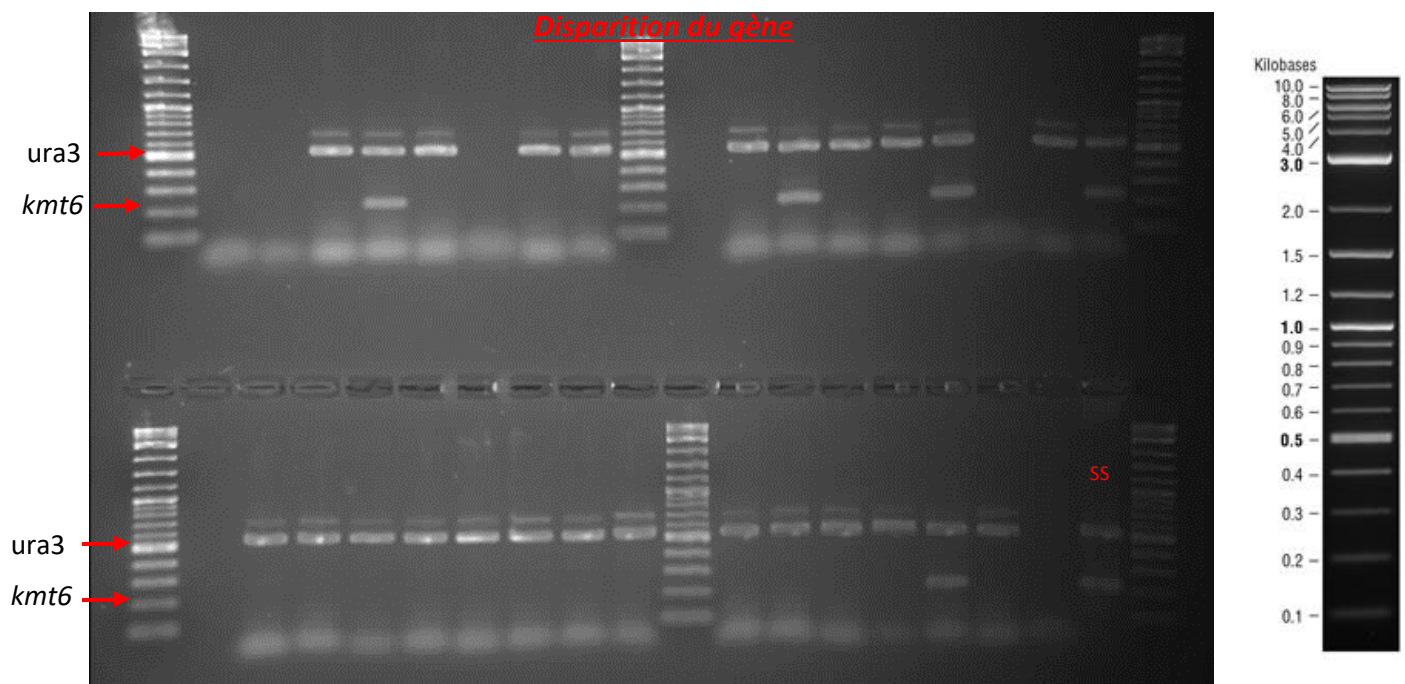
TrA1199W	Putative protein of unknown function
TrA1233C	Putative protein of unknown function
TrA1195W	Putative protein of unknown function
TrF1115C	Putative protein of unknown function
TrB0166W	Putative protein of unknown function
TrB0195W	Putative protein of unknown function
TrB0361C	Putative protein of unknown function
TrB0502C	Putative protein of unknown function
TrB0545C	Putative protein of unknown function
TrB0565C	Putative protein of unknown function
TrB0670C	Putative protein of unknown function
TrB0683W	Putative protein of unknown function
TrB0686C	Putative protein of unknown function
TrB0771C	Putative protein of unknown function
TrB0879W	Putative protein of unknown function
TrB0905W	Putative protein of unknown function
TrB0980W	Putative protein of unknown function
TrB1153W	Putative protein of unknown function
TrB1319C	Putative protein of unknown function
TrB1542W	Putative protein of unknown function
TrB1575C	Putative protein of unknown function
TrB1652C	Putative protein of unknown function
TrB1685W	Putative protein of unknown function
TrB1717C	Putative protein of unknown function
TrB1723W	Putative protein of unknown function
TrB1724W	Putative protein of unknown function
TrB1727C	Putative protein of unknown function
TrB1791C	Putative protein of unknown function
TrC0252C	Putative protein of unknown function
TrC0255W	Putative protein of unknown function
TrC0269W	Putative protein of unknown function
TrC0325W	Putative protein of unknown function
TrB1633C	Putative protein of unknown function
TrB1579W	Putative protein of unknown function
TrB1605W	Putative protein of unknown function
TrC0378W	Putative protein of unknown function
TrC0385C	Putative protein of unknown function
TrC0413C	Putative protein of unknown function
TrC0499C	Putative protein of unknown function
TrC0487C	Putative protein of unknown function
TrC0596C	Putative protein of unknown function
TrC0646W	Putative protein of unknown function
TrC0757W	Putative protein of unknown function
TrC0773W	Putative protein of unknown function
TrC0923W	Putative protein of unknown function
TrA1335W	Putative protein of unknown function

TrA1301W	Putative protein of unknown function
TrA1296C	Putative protein of unknown function
TrA1287C	Putative protein of unknown function
TrA1272W	Putative protein of unknown function
TrA1252C	Putative protein of unknown function
TrA1456W	Putative protein of unknown function
TrA1480W	Putative protein of unknown function
TrA1543C	Putative protein of unknown function
TrA1588W	Putative protein of unknown function
TrC1020C	Putative protein of unknown function
TrC1034C	Putative protein of unknown function
TrC1068C	Putative protein of unknown function
TrC1073C	Putative protein of unknown function
TrC1100C	Putative protein of unknown function
TrC1151W	Putative protein of unknown function
TrC1280C	Putative protein of unknown function
TrG1011C	Putative protein of unknown function
TrG0232W	Putative protein of unknown function

XIX. Gels PCR de vérification des mutants

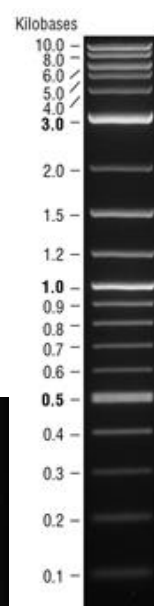
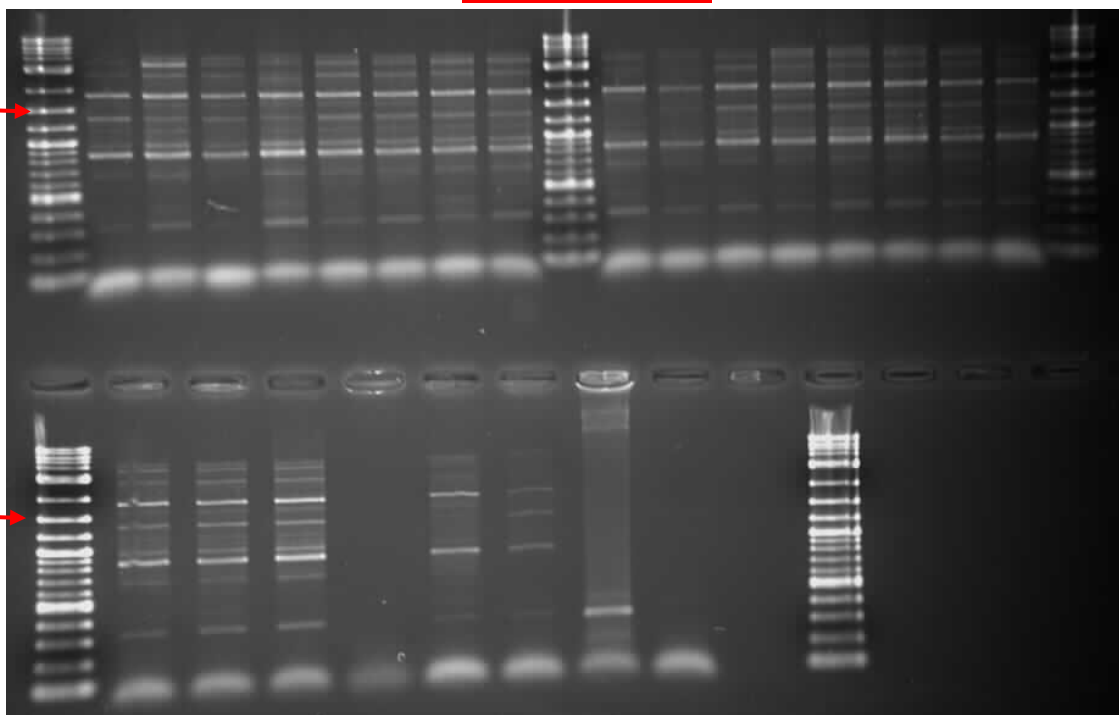
Vérifications par PCR de l'insertion correcte des cassettes d'inactivations pour chacun des gènes visés. La première vérification se fait par PCR en duplex avec l'amplification du gène URA3 pour s'assurer que la réaction a bien eu lieu. Si le gène a bien été substitué, aucune bande n'apparaît aux endroits indiqués par les flèches rouges sur les gels intitulés « Disparition du gène ». Ensuite, les bords des insertions sont ensuite vérifiés pour les mutants ne possédant à priori plus le gène visé.

Mutants $Q\Delta kmt6$



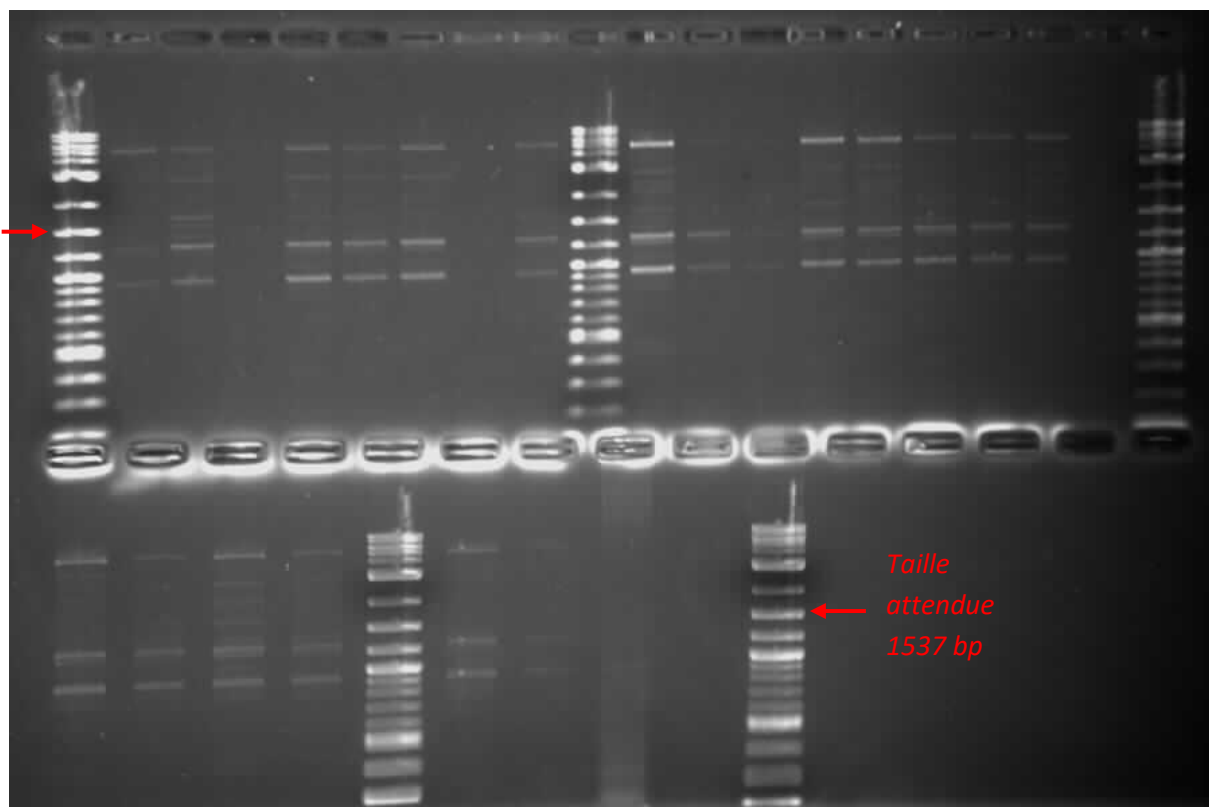
Amplification en 5'

Taille
attendue
1556 bp



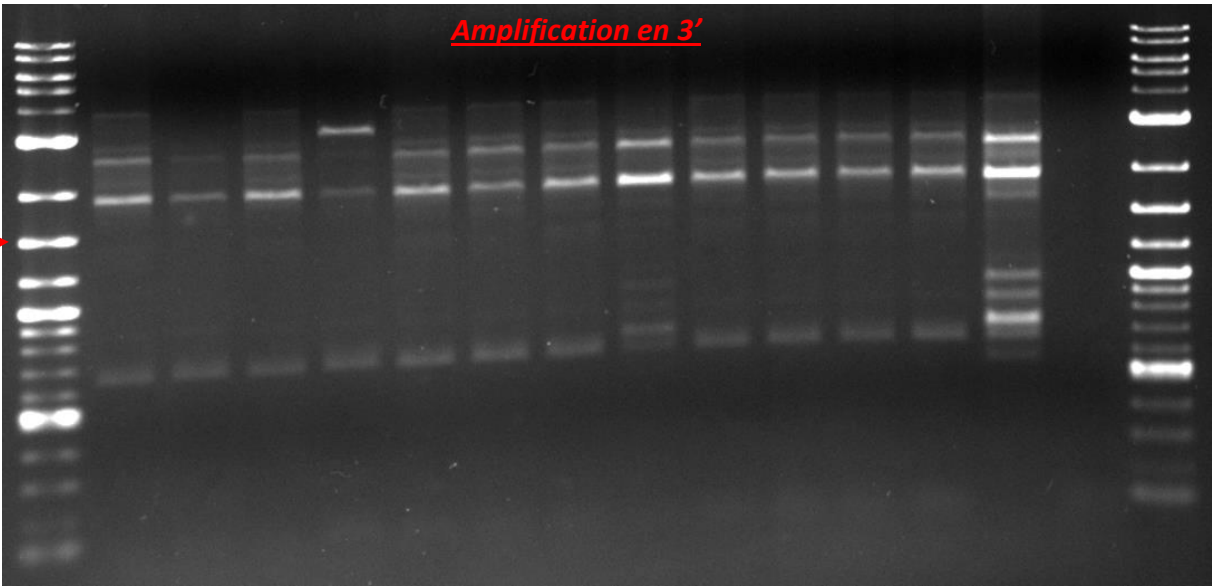
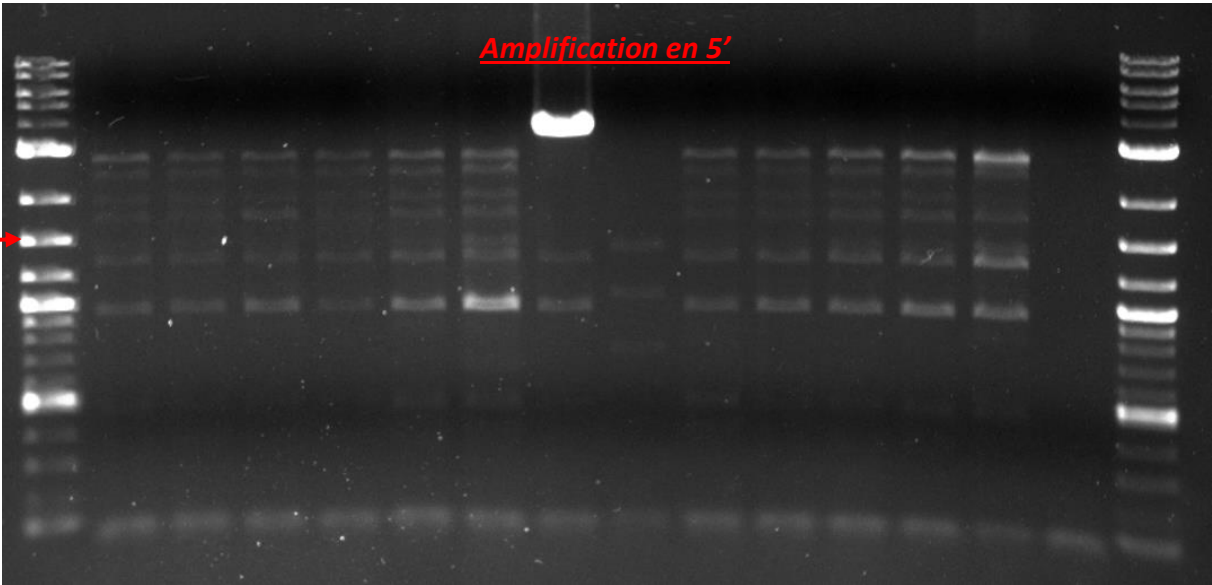
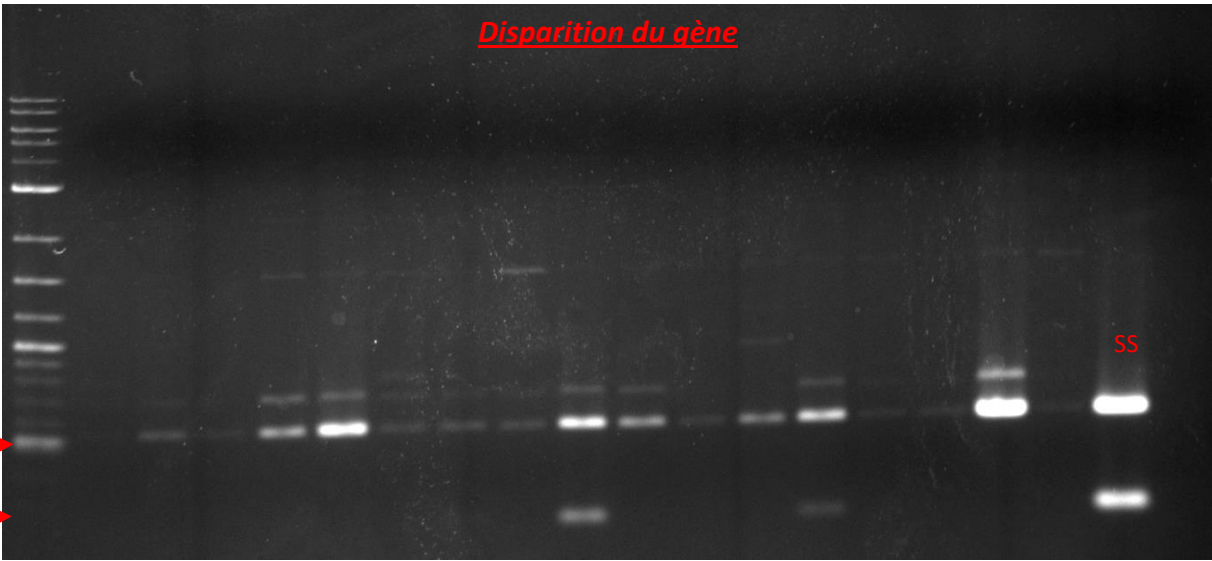
Amplification en 3'

Taille
attendue
1537 bp



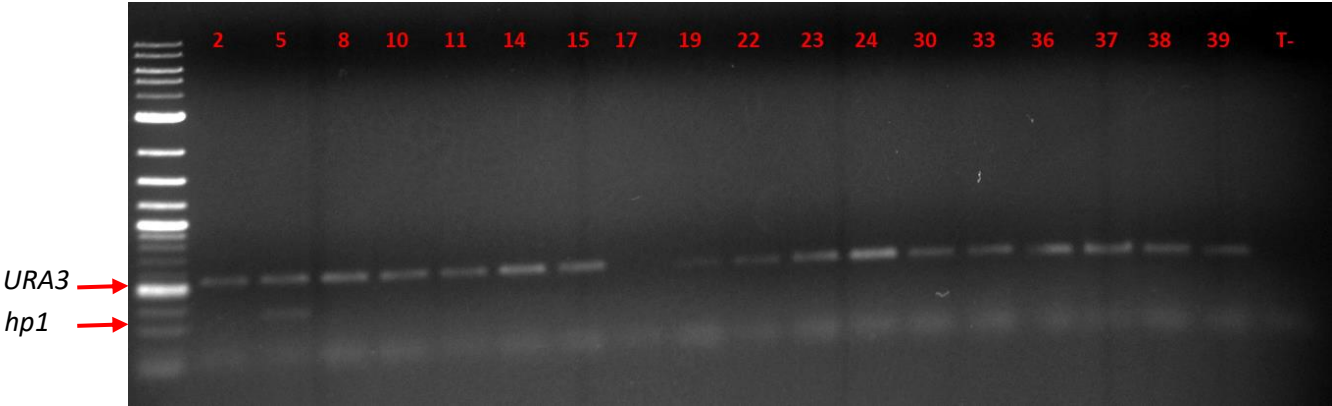
Taille
attendue
1537 bp

Mutants RΔkmt6



Mutants *QΔhp1*

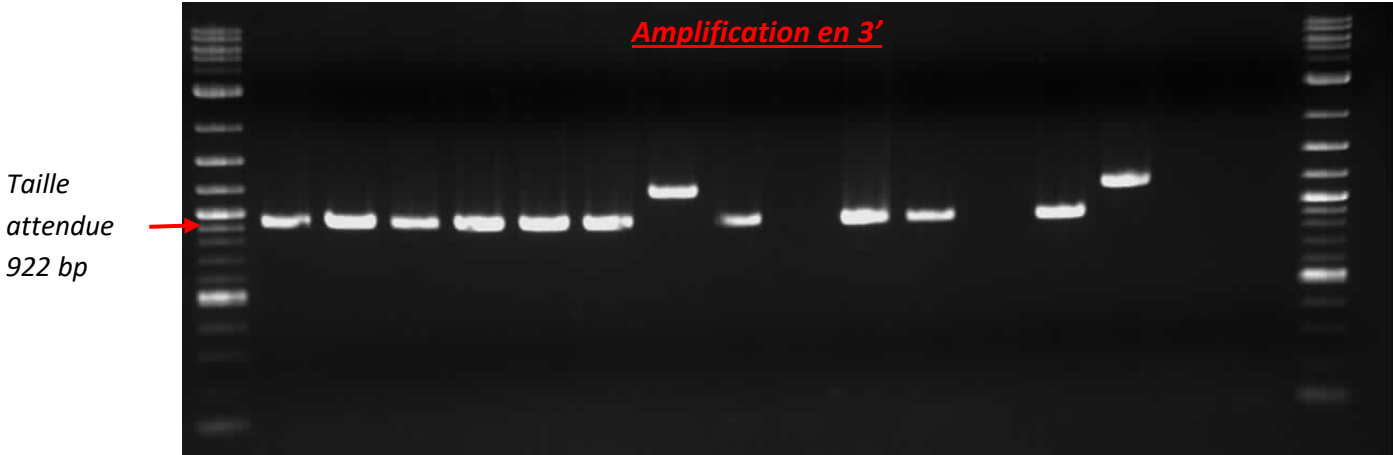
Disparition du gène



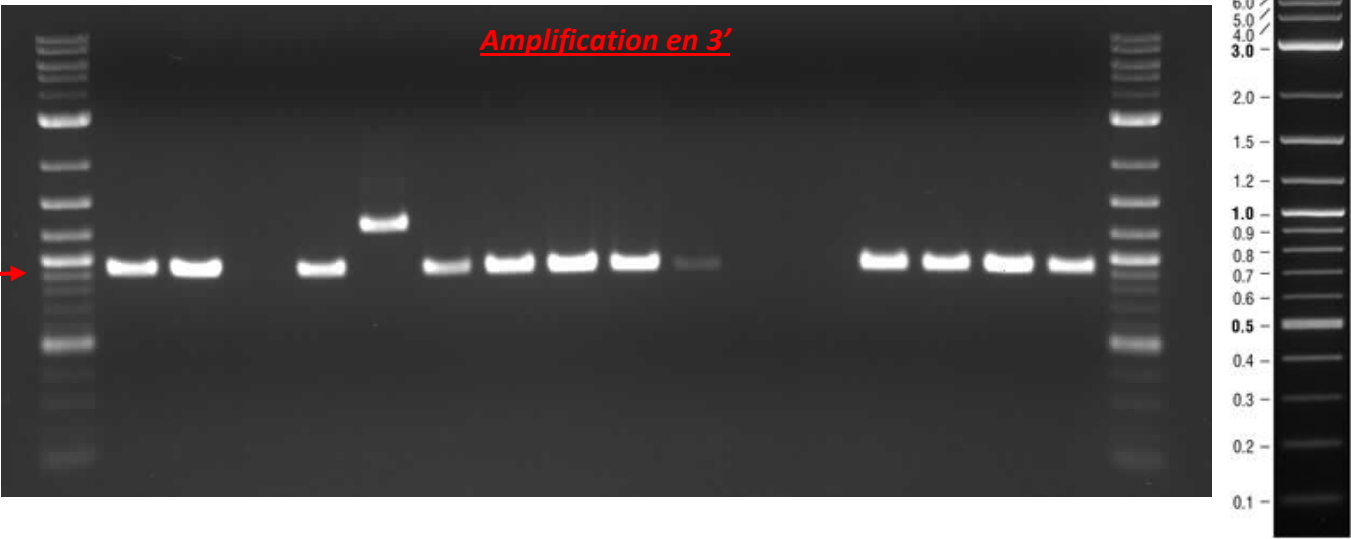
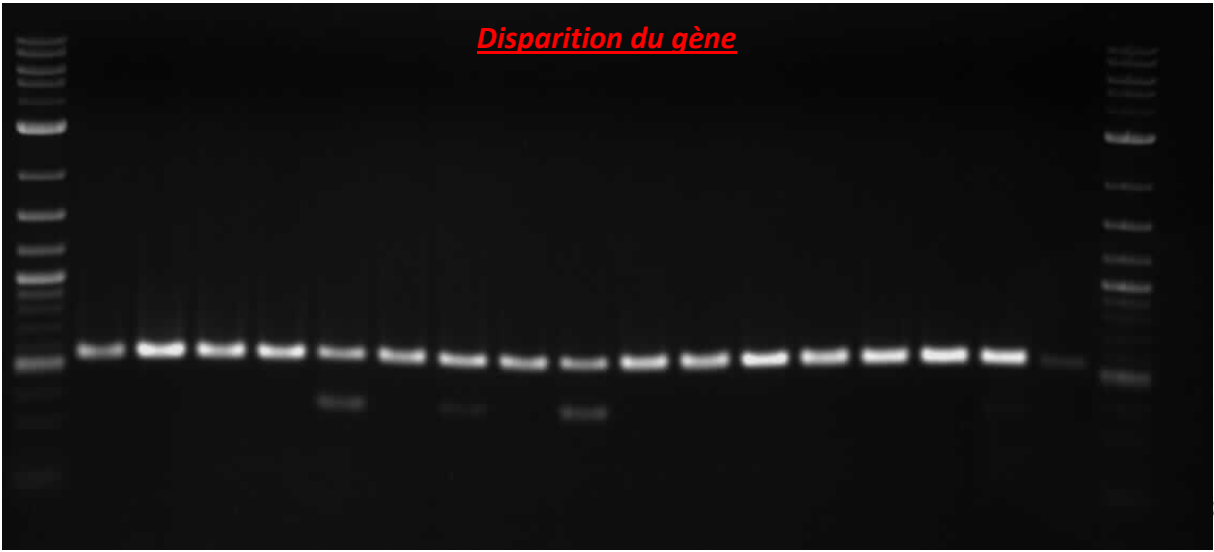
Amplification en 5'



Amplification en 3'

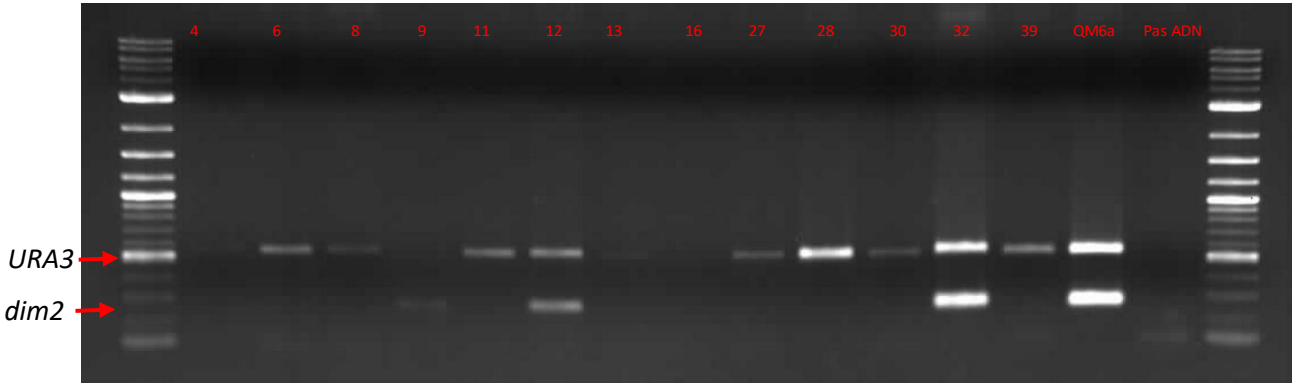


Mutants *RΔhp1*

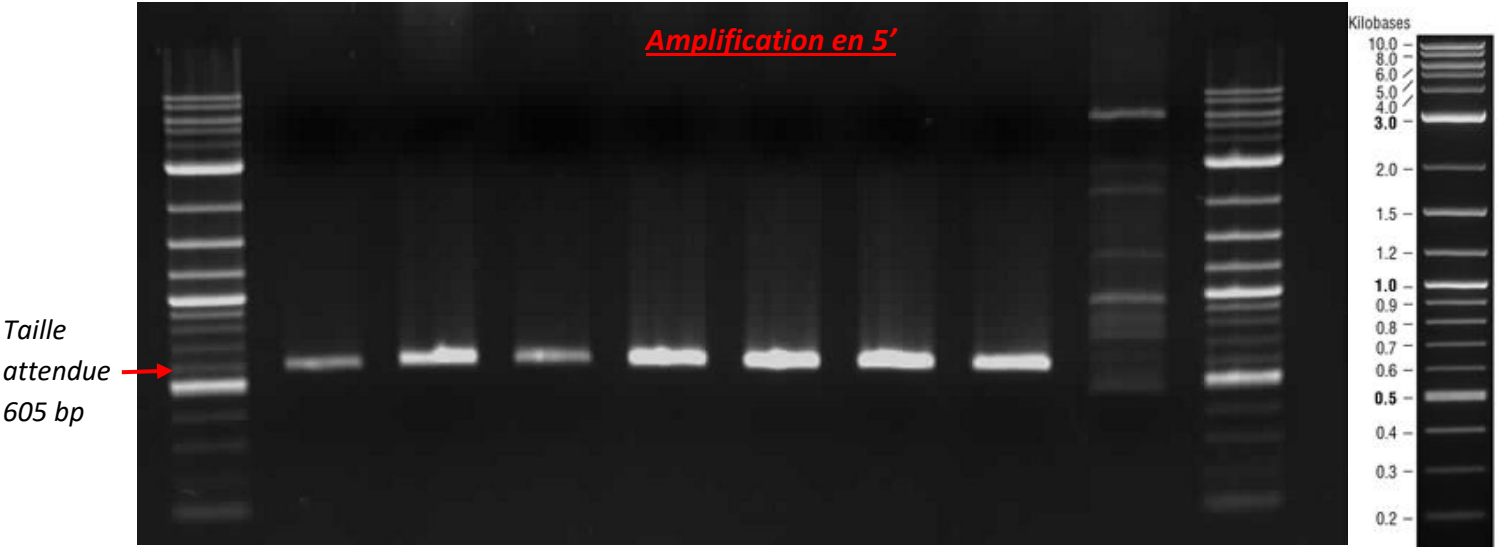


Mutants *QΔdim2*

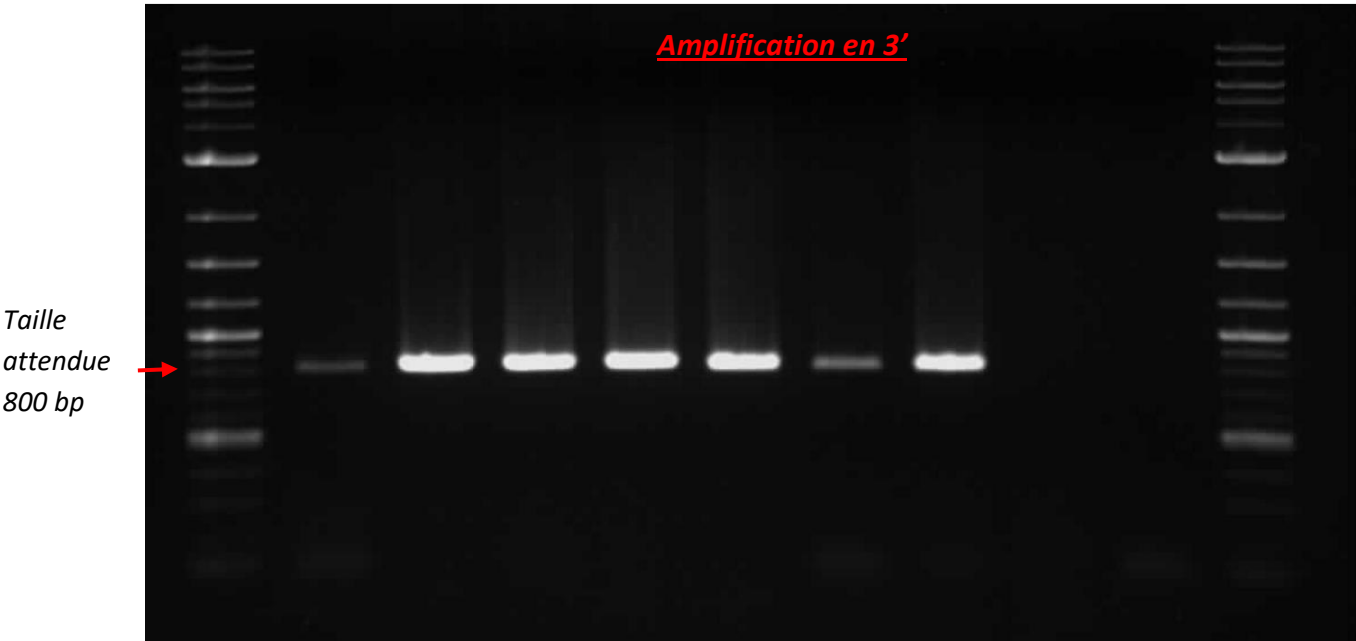
Disparition du gène



Amplification en 5'

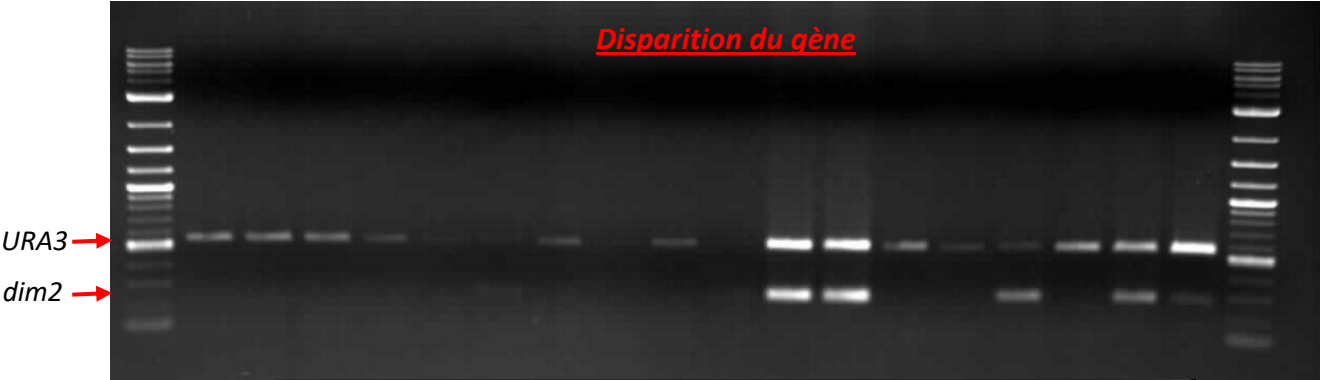


Amplification en 3'

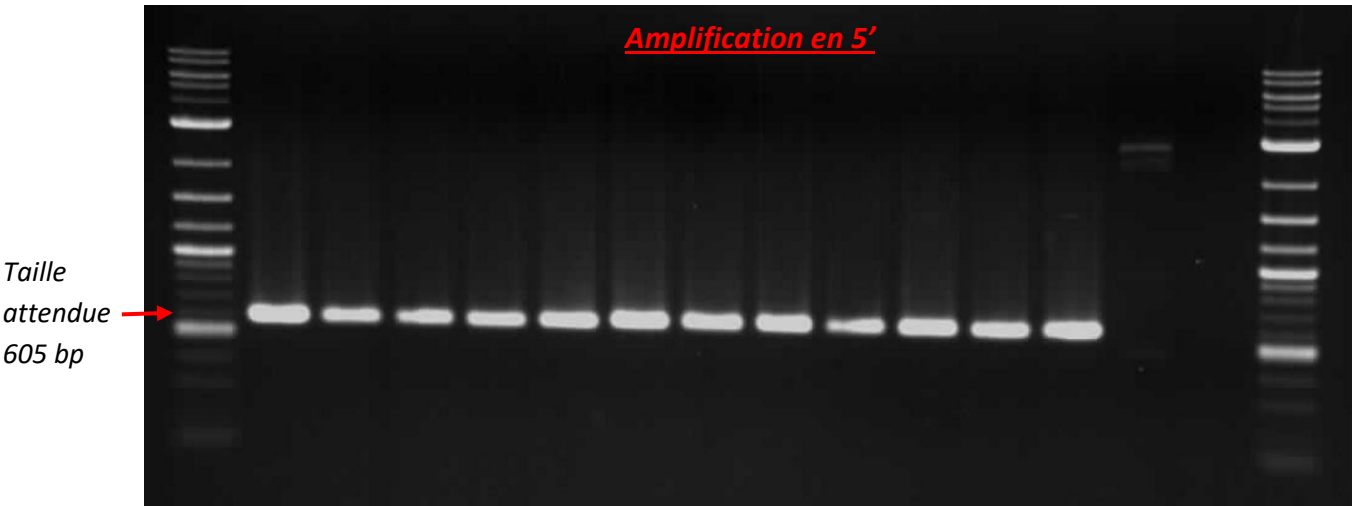


Mutants *RΔdim2*

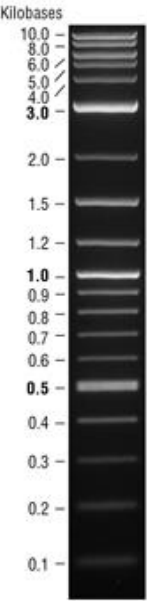
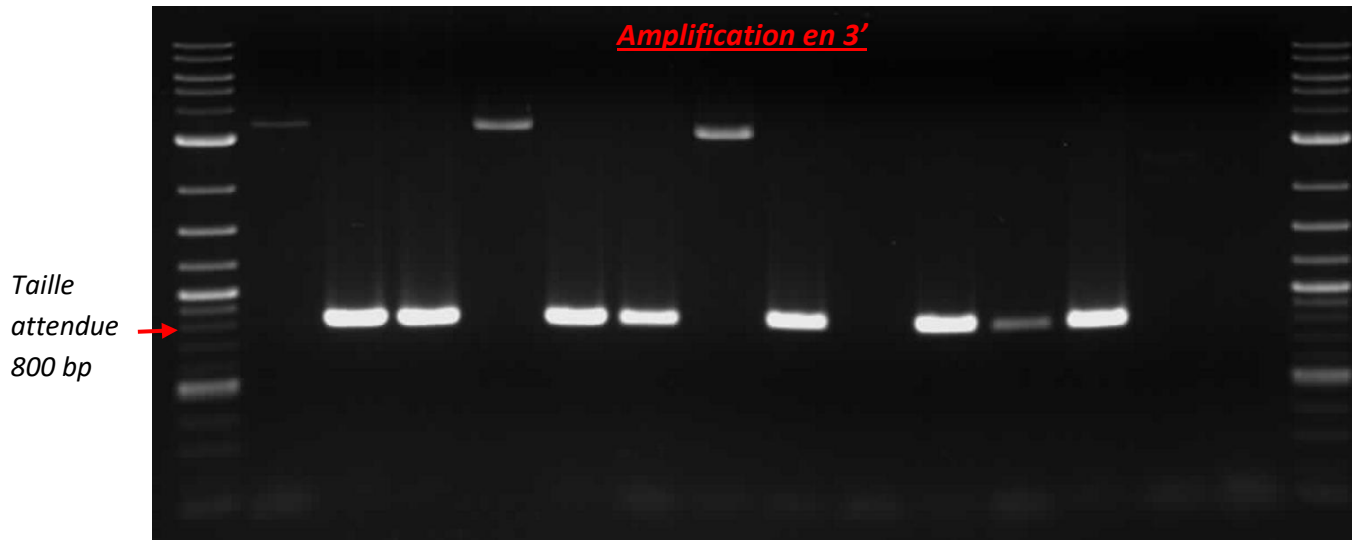
Disparition du gène



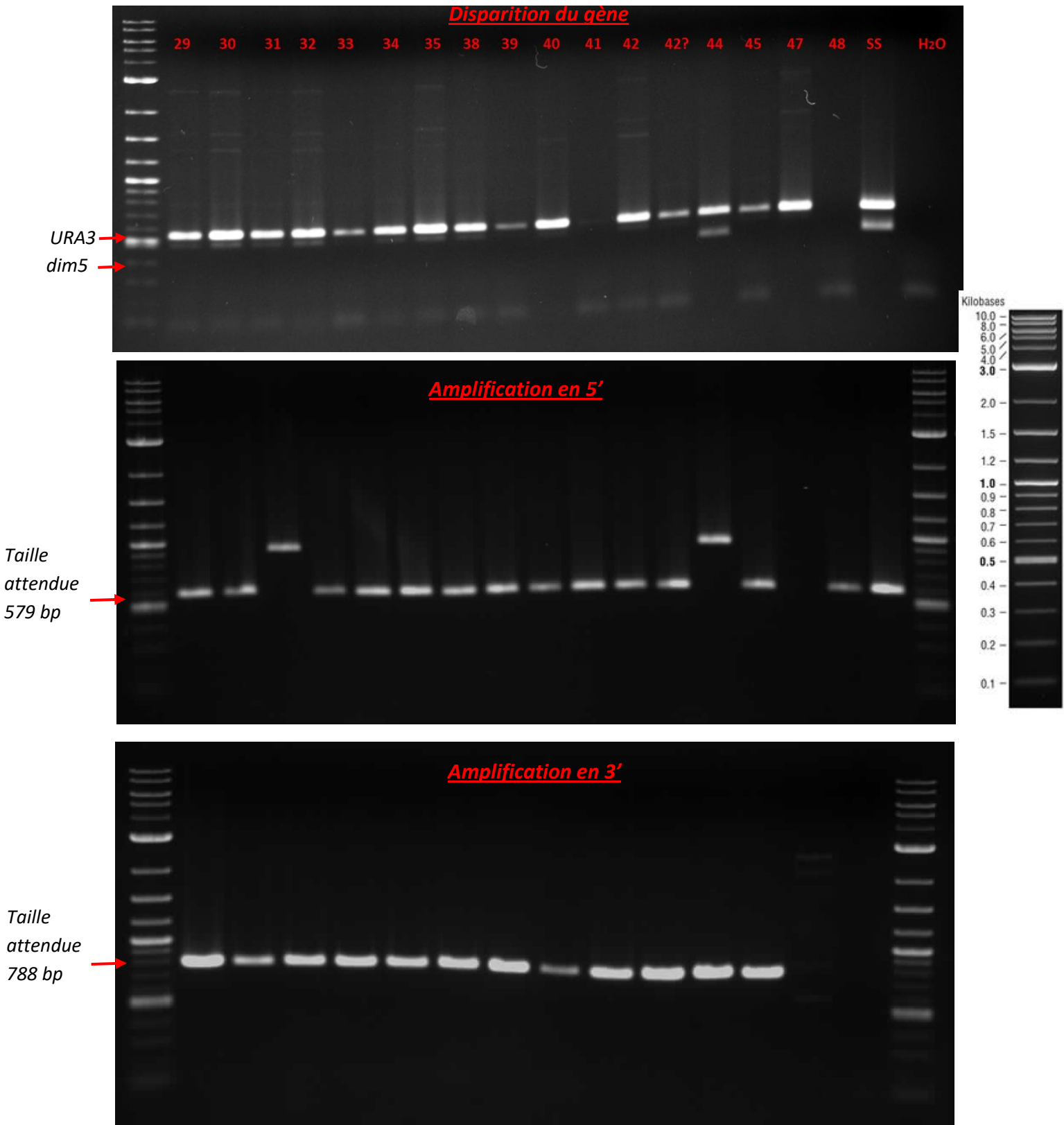
Amplification en 5'



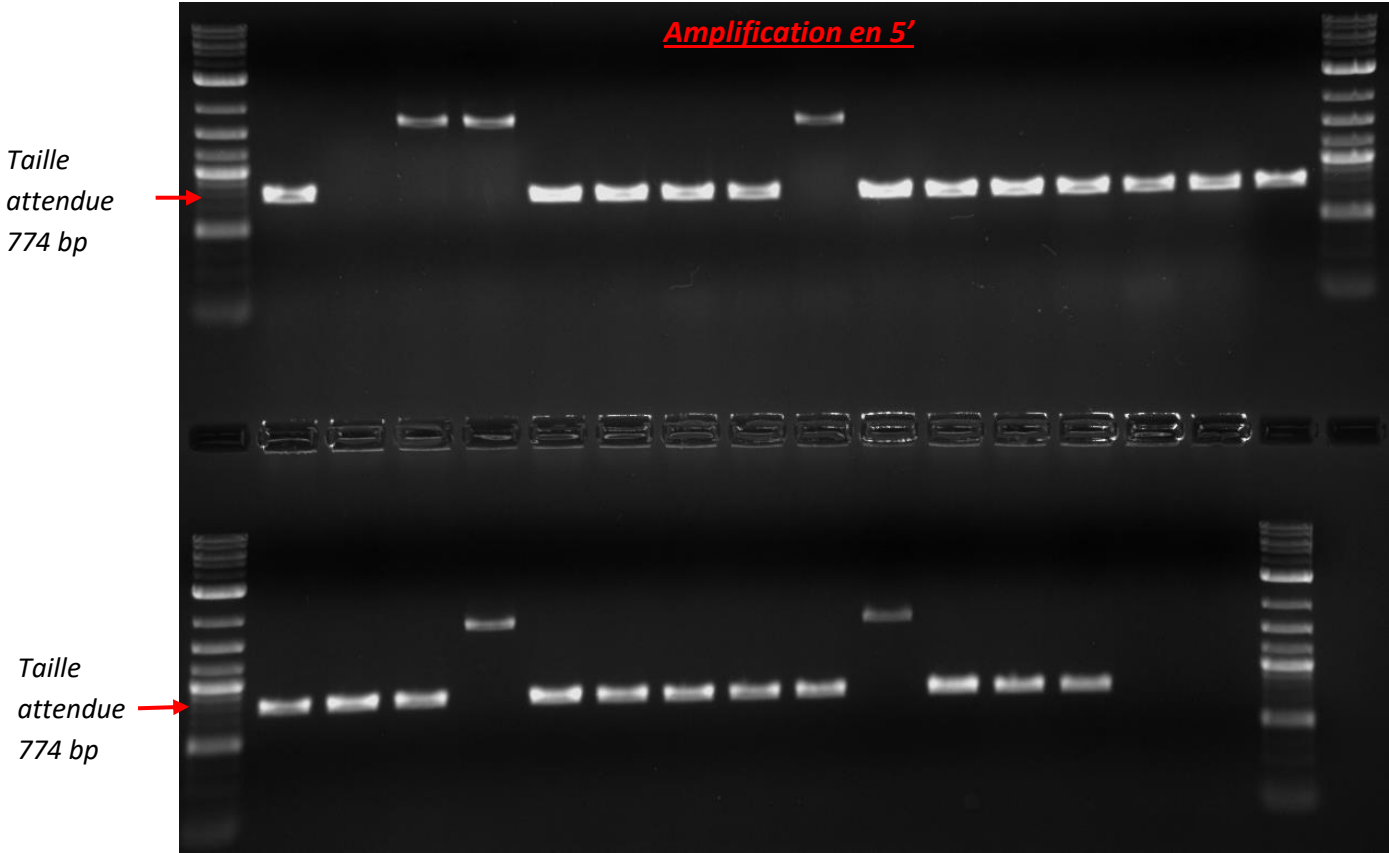
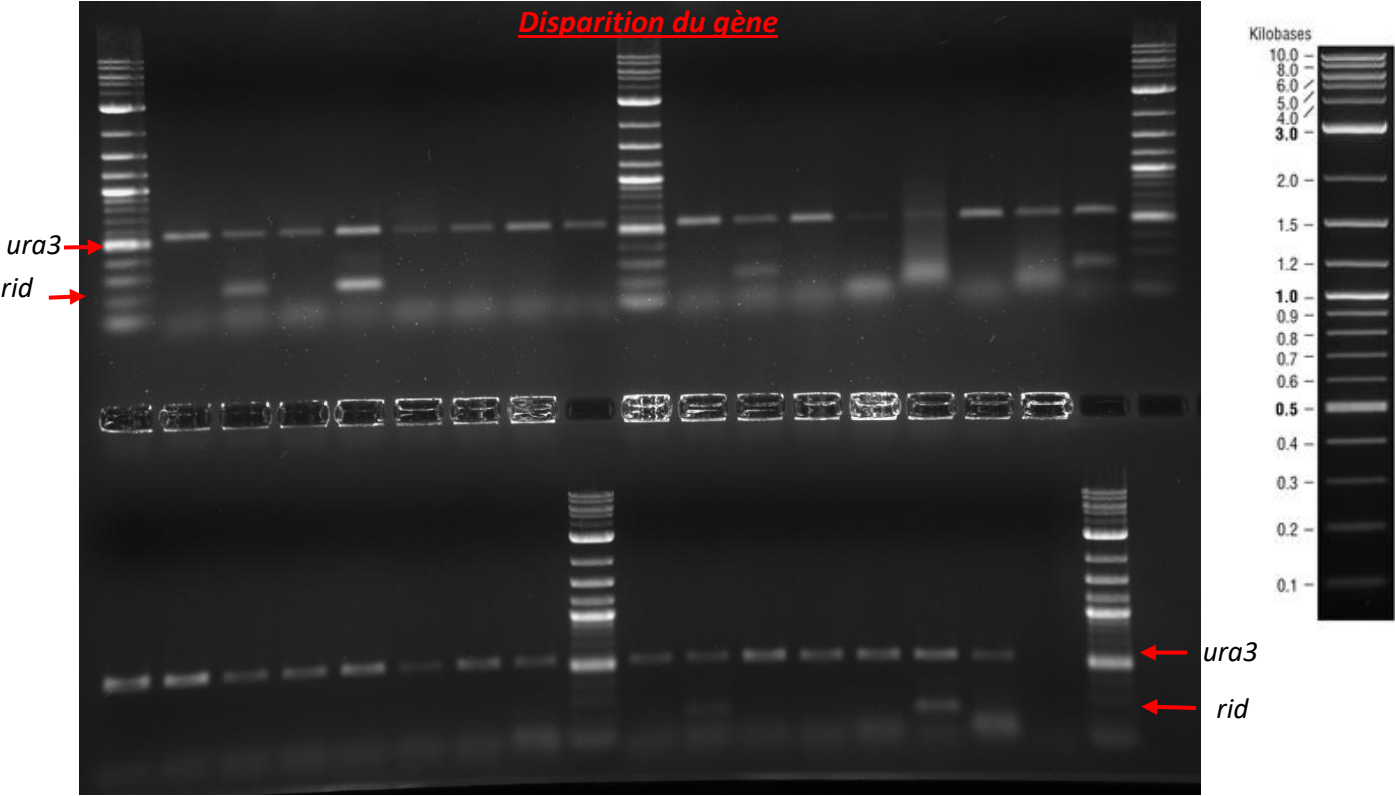
Amplification en 3'

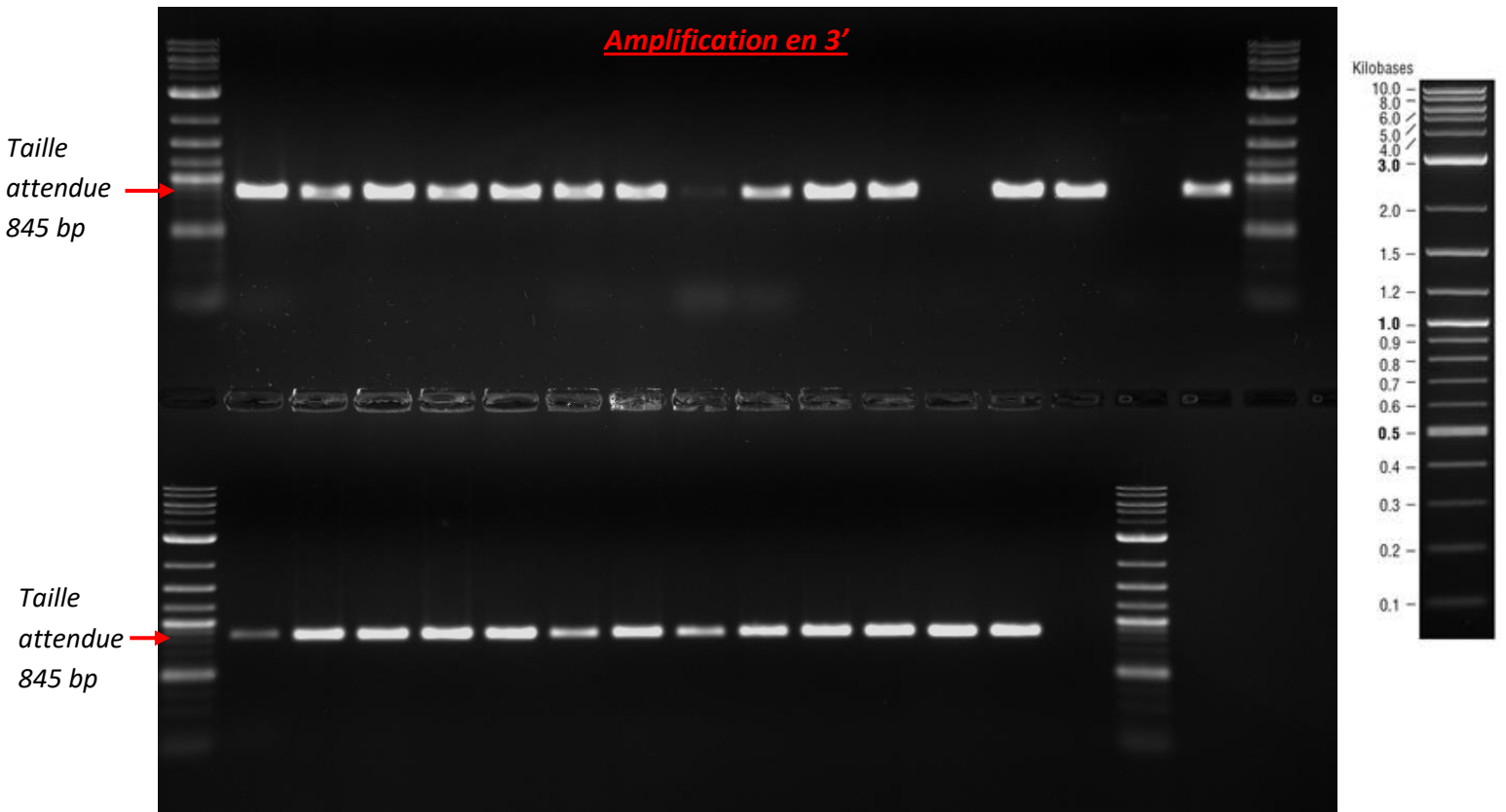


Mutants QΔdim5



Mutants *RΔrid*





Mutants *QΔrid*



XX. Phénotypage sur boîtes, réplicats

Photos des expériences de phénotypages sur boîtes contenant des milieux solides avec différentes sources de carbones ou sources de stress ou ayant été placé dans des environnements aux températures ou luminosités différentes. Le champignon a été étudié après une semaine de croissance. Les deux réplicats effectués pour chaque souche sont présentés.

