

Analyse du transcriptome de *Trichoderma reesei* pour amélioration de la production de cellulases



THESE 2008

Titre de la thèse	Analyse du transcriptome de <i>Trichoderma reesei</i> pour l'amélioration de la production de cellulases
Doctorant	Thomas Portnoy
Université-Ecole doctorale	Logique du Vivant (LDV) Université de Paris 6
Directeur de thèse	Dr. Stéphane Le Crom, Ecole Normale Supérieure, Département de Biologie, INSERM U784, Laboratoire de Génétique Moléculaire du Développement, 46 rue d'Ulm, 75230 Paris
Laboratoires d'accueil	- ENS, Département de Biologie, 46 rue d'Ulm, 75230 Paris - TU Vienna, Institute of Chemical Engineering, Gene Technology and Applied Biochemistry, Getreidemarkt 9/E 1665, A-106 Vienna, Austria - Département de Biotechnologie de IFP Energies nouvelles (IFPEN), 1-4 Avenue de Bois-Préau, 92852 Rueil-Malmaison
Responsable de thèse	Antoine Margeot, Département de Biotechnologie, IFPEN, Rueil-Malmaison
Durée	Trois ans fin 2008 à fin 2011

Résumé

Le présent sujet de thèse se propose d'explorer non pas le cocktail enzymatique produit par le champignon filamenteux, utilisé pour l'hydrolyse enzymatique de la biomasse lignocellulosique, mais les mécanismes au niveau génétique de sa production par fermentation. En effet, une marge de progrès semble envisageable dans ce domaine.

L'idée centrale du projet est de tirer parti du séquençage du génome de *Trichoderma reesei* pour l'étude du transcriptome et du génome de la souche de ce champignon CL 847. A cette fin, des puces à ADN seront utilisées. Cette technologie permet de visualiser l'expression de chacun des gènes du champignon à un instant donné. Une vue d'ensemble de l'état physiologique du microorganisme est obtenue et permet d'identifier des points d'amélioration de la fermentation ou des cibles génétiques pertinentes, afin de diminuer les coûts de production des enzymes cellulolytiques.

A partir d'échantillons obtenus lors de fermentations réalisées dans des conditions proches de celles utilisées à la taille industrielle, l'expression des gènes au cours de l'induction de la production de cellulases par la souche CL 847 sera suivie, et les acteurs génétiques clés de cette induction seront identifiés par analyse bioinformatique. Les acteurs potentiellement les plus influents seront soumis à modifications dans le génome de la souche (surexpression ou délétion) et les effets sur la production de cellulases par le champignon seront évalués.

Les modifications à apporter au génome pouvant être multiples, les travaux porteront également sur le développement d'outils de transformation génétique efficaces et novateurs qui devraient permettre à terme d'introduire différentes constructions dans la même souche productrice.

Résultats

Exploration des mécanismes génétiques sous-jacents aux comportements d'intérêt (production de cellulases) de *Trichoderma reesei* en bioréacteur,
Elargissement des connaissances des réponses cellulaires mises en jeu au cours de l'induction de la production de cellulases,
Description des effets du lactose, un des inducteurs pouvant être utilisé à l'échelle industrielle sur l'expression des facteurs de transcription centraux influençant l'expression des gènes d'enzymes cellulolytiques,
Description globale des effets de l'induction de la production de cellulases en bioréacteur sur l'expression des gènes de souches plus ou moins productrices de *T.reesei*,
Etablissement des premiers scénarios de régulations génétiques mis en jeu dans différentes souches d'intérêt de *T.reesei* dans des conditions de cultures industrielles,
Etablissement de l'implication de certains acteurs de la voie du métabolisme du galactose dans l'induction de la production de cellulases.

Perspectives

Les comparaisons du génome et du transcriptome des souches de *T.reesei* pourront être complétées par des études des flux et des équilibres métaboliques associés,
Des calculs de flux métaboliques apporteront de nombreuses précisions sur l'état physiologique des cellules au cours des différentes étapes du procédé de production de cellulases,
Des études des promoteurs des gènes qui voient leur expression modifiée au cours du processus semblent cruciales pour l'identification d'éventuels motifs de régulation communs et des facteurs de transcription qui leur sont associés,

A terme, ces travaux devraient permettre d'améliorer le procédé de production de cellulases et de ce fait, celui de fabrication d'éthanol à partir de biomasse lignocellulosique.

Livrables

[Page de couverture du manuscrit](#)

[Manuscrit de la thèse](#)

Publications

Jourdren L, Duclos A, Brion C, Portnoy T, Mathis H, Margeot A, Le Crom S (2010). Teolenn : an efficient and customizable workflow to design high-quality probes for microarray experiments. *Nucleic Acids Res.*, 1-9

Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, Sandor E, Hartl L, Karaffa L, Druzhinina IS, Seiboth B, et al (2011). The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei* : a master regulator of carbon assimilation. *BMC Genomics*, 12, 269

Portnoy T, Margeot A, Seidl-Seiboth V, Le Crom S, Ben Chaabane F, Linke R, Seiboth B, Kubicek CP (2011). Differential regulation of the cellulase transcription factors XYR1, ACE2 and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulases. *Eukariotic Cell*, 10, 262-271

Contact

Antoine Margeot

IFP Energies nouvelles, 1-4 Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex
Tel : 01 47 52 71 96

Tel : 01 47 52 71 43 (secrétariat)
antoine.margeot@ifpen.fr

**THÈSE DE DOCTORAT DE
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité Sciences de la Vie

École Doctorale 515 : Complexité du Vivant

Présentée par

M. Thomas Portnoy

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Analyse du transcriptome de *Trichoderma reesei* pour l'amélioration de la production de cellulases

Thèse dirigée par Stéphane Le Crom
et co-encadrée par Antoine Margeot

Soutenue le 22 septembre 2011

Devant le jury composé de :

M. Frédéric Devaux	Président
M. Claude Gaillardin	Rapporteur
M. Michel Werner	Rapporteur
M. Jean-Guy Berrin	Examineur
M. Antoine Margeot	Examineur
M. Stéphane Le Crom	Directeur de thèse

**THÈSE DE DOCTORAT DE
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité Sciences de la Vie

École Doctorale 515 : Complexité du Vivant

Présentée par

M. Thomas Portnoy

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Analyse du transcriptome de *Trichoderma reesei* pour l'amélioration de la production de cellulases

Thèse dirigée par Stéphane Le Crom
et co-encadrée par Antoine Margeot

Soutenue le 22 septembre 2011

Devant le jury composé de :

M. Frédéric Devaux	Président
M. Claude Gaillardin	Rapporteur
M. Michel Werner	Rapporteur
M. Jean-Guy Berrin	Examineur
M. Antoine Margeot	Examineur
M. Stéphane Le Crom	Directeur de thèse

Remerciements

Je veux tout d'abord remercier infiniment Stéphane Le Crom, mon directeur de thèse, et Antoine Margeot, mon co-encadrant, qui m'ont fait l'honneur de me confier ce projet. Attentifs et respectueux, ils ont su ménager tout au long de ces trois années des conditions de travail idéales, tant sur le plan scientifique que sur le plan humain. Grâce à leur expérience et à leur grand sens des responsabilités, ils ont osé de grandes et belles idées pour ce projet, audacieuses et stimulantes, sans jamais perdre de vue la nécessité de se fixer des objectifs réalistes, atteignables et valorisables. Disponibles et attentionnés, ils m'ont également offert une relation de confiance, de respect. Ils ont su m'écouter, me soutenir, m'accompagner quand il le fallait, avec beaucoup de bienveillance et une considération que chacun ne peut qu'espérer recevoir de la part de ses encadrants. C'est une chance immense que j'ai eue de travailler à leurs côtés.

Je remercie Frédéric Monot, responsable du département Biotechnologie d'IFPEN, de m'avoir accordé sa confiance en m'acceptant au sein de son équipe. C'est grâce à sa grande ouverture d'esprit scientifique, à son enthousiasme et à son soutien que ce projet à dominante fondamentale a pu exister et prospérer dans d'excellentes conditions au sein d'IFPEN.

Je remercie Sophie Lemoine et Laurent Jourden, mes collègues et amis de la plateforme transcriptome. Tout d'abord pour leur aide indispensable face aux bases de données, lignes de commandes et autres scripts bioinformatiques qu'ils ont su me faire apprivoiser avec patience et pédagogie, mais surtout pour leur présence stimulante, vivifiante, essentielle. Nos discussions riches et exaltantes, notre complicité et nos délires irremplaçables ont donné à ces trois années toute leur saveur. Leur écoute, leurs attentions et leur soutien sans faille ont également été pour moi source d'un profond réconfort, si précieux et apaisant quand il le fallait. Je les remercie du fond du cœur de m'avoir tout simplement rendu heureux au quotidien.

Merci à Frédérique Bidard-Michelot, dont l'expertise et la patience ont donné à mon travail tout son sens. Mes résultats auraient été bien peu de choses sans ses fichiers d'annotations multicolores qu'elle a eu le courage de constituer cellule par cellule, gène après gène. Nos séances de travail m'ont permis, entre deux fous rires, d'acquérir une véritable vision experte des données qu'il m'aurait été bien difficile de développer si nous n'avions pas « mis les mains dans le cambouis » ensemble.

Je remercie Corinne Blugeon et Fanny Couplier, tout d'abord de m'avoir initié et accompagné avec autant de bienveillance et de disponibilité dans l'art délicat de la manipulation des puces à ADN - qu'elles maîtrisent avec virtuosité -, mais aussi de m'avoir adopté avec autant de gentillesse et de générosité, car c'est aussi grâce à elles que l'atmosphère au sein de la plateforme est si conviviale, heureuse et chaleureuse.

Merci à Sabine Prigent avec qui j'ai fait équipe pendant ces trois années lorsqu'il s'agissait de me frotter aux bioréacteurs, et sans qui il n'aurait tout simplement pas été possible que je réussisse la moindre production de cellulases. Sa maîtrise technique m'a guidé tout au long de ces manipulations lourdes, et sa présence aura jusqu'au bout été pour moi une source de soulagement face à un tel challenge expérimental. Merci également à Bernard Chaussepied, de m'avoir initié à cette technique de la fermentation et d'avoir été si souvent prêt à laisser en plan toutes ses activités pour me prêter main forte, grâce à ses compétences si pointues.

Merci à Fadhel Ben Chaabane pour sa gentillesse, sa patience et son immense sens de la pédagogie, sans lesquels je n'aurais jamais été capable de comprendre la moindre petite équation de modélisation. Je le remercie également de m'avoir permis d'utiliser les bioréacteurs F2P malgré le planning expérimental très chargé de son équipe.

Merci à Maria Bernard pour son énorme travail sur le traitement des données de séquençage à haut débit qui a permis d'agrémenter considérablement mes résultats, mais également pour sa gentillesse, sa gaieté et son sens de l'humour qui nous ont permis de vivre des instants encore plus pétillants lors de son passage à la plateforme.

Merci à Hugues Mathis, pour son aide si précieuse au labo de biologie moléculaire d'IFPEN, sa disponibilité et ses qualités humaines qui contribuent à créer une ambiance de travail si agréable qu'elle me manque dès qu'il n'est plus là.

Je souhaite également remercier les stagiaires que j'ai eu la chance d'encadrer au cours de cette thèse, et dont l'aide m'a été indispensable.

Merci à Hélène Ipas, dont la motivation et la bonne humeur ont accompagné mon été 2009, et qui a fait progresser le projet *xyl1* grâce à son investissement et sa grande aisance au laboratoire. Plein de courage pour ta thèse, je suis certain qu'elle sera de grande qualité.

Merci à Caroline Zemb, qui m'a prêté main forte durant l'été 2010 dans la préparation, la conduite et l'analyse de certaines productions en bioréacteurs, et qui n'a pas démerité même

quand les cellules avaient décidé de s'inviter sur les murs du labo. Bon courage pour la suite de ton parcours.

Merci également à Lydie Vamparys et Valentin Wucher, stagiaires de bioinformatique sous la responsabilité de Stéphane, mais dont l'encadrement ponctuel a également été très formateur pour moi. Merci enfin aux précédents stagiaires, Pauline Sacareau, Sarah Martinez et Christian Brion d'avoir essuyé les plâtres de certaines mises au point expérimentales, et d'avoir débroussaillé le terrain avant mon arrivée.

Je ne peux pas ne pas mentionner nos collaborateurs, Christian Kubicek et Bernhard Seiboth, dont les idées et la culture scientifique colossale ont joué un rôle moteur et décisif dans la définition de nos axes de recherche. Merci à leur équipe et à celle de Levente Karaffa pour leur aide précieuse dans la conduite de certaines expérimentations.

Merci aux membres de mon comité de suivi de thèse Gilles Fischer et Jean-Yves Coppée pour leur regard attentif et les discussions constructives qu'ils ont suscitées, ainsi qu'à Sylvie Hermann-Le Denmat pour son soutien et ses conseils précieux tout au long de ces années, et depuis plus longtemps encore. J'en profite pour lui adresser également un autre immense merci, ainsi qu'à Stéphane, à Antoine, mais aussi à Gérald Peyroche, pour m'avoir permis d'enseigner au cours de cette thèse, chance inouïe et plaisir fantastique.

Merci à l'équipe de Patrick Charnay, Pascale, Michel, Yassine et tous les autres, pour leur accueil, leur écoute, et les mille idées constructives qu'ils m'ont données en réunion de labo.

Merci à Alexandra pour sa joie de vivre et son soutien si amical et rassurant.

Merci à Baptiste, qui m'a accompagné au cours de cette dernière année si intense, pour sa présence et son amitié.

Merci à Marie pour sa bonne humeur et tous ces moments si extra qu'on a partagés.

Merci à Camille et Matthieu pour leur coolitude, et à tous les autres qui ont compté ô combien ! Mathilde, Céline, Anaïs, Paul, Nicolas, Florent, Nassim, Benjamin, Sandra, Juliette, Anne, Étienne, Françoise, Yves, Françoise, Senta, Meriem, Véronique, Brigitte, Martine, pour leur gentillesse et leur aide au quotidien à IFPEN ou à l'ENS.

Merci aussi à ma famille et à mes amis, les plus proches comme les plus lointains, pour tout leur soutien et leurs encouragements essentiels.

Merci enfin à Dante Poggi d'avoir répondu présent pour reprendre le flambeau. Je suis très heureux de te confier ce joli projet, et te souhaite beaucoup de réussite et trois années aussi riches et exaltantes que celles que j'ai eu la chance de vivre.

Résumé

Analyse du transcriptome de *Trichoderma reesei* pour l'amélioration de la production de cellulases

Le champignon filamentueux *Trichoderma reesei* constitue la principale source de cellulases, enzymes employées dans les procédés industriels d'hydrolyse de la biomasse lignocellulosique nécessaire à la production de bioéthanol de deuxième génération. Ce projet de thèse se base sur l'utilisation de techniques de génétique moléculaire et d'outils de génomique fonctionnelle pour décrire les mécanismes génétiques impliqués dans la synthèse et la sécrétion de ces enzymes afin d'envisager une amélioration des capacités de production des souches de *T. reesei* rationnellement et de manière ciblée.

Une première partie de ce travail a consisté à préciser la régulation de l'expression des facteurs de transcription centraux de la production de cellulases lors d'une induction par le lactose. Nous avons notamment démontré que *xyr1*, *ace1* et *ace2* étaient spécifiquement induits par le lactose, et que CRE1, médiateur de la répression catabolique, était nécessaire à une induction complète de *xyr1* et *ace2*. Dans un deuxième temps, des études transcriptomiques nous ont permis (i) d'identifier les gènes cibles de la répression catabolique dépendante ou non de CRE1, et (ii) de décrire à l'échelle du génome total les régulations mises en jeu lors de l'induction de la production de cellulases. Nos travaux ont notamment permis d'établir l'importance de l'expression génétique basale dans l'acquisition des capacités d'hyperproduction des souches, et d'identifier les principales fonctions cellulaires soumises à des régulations au cours du processus d'induction.

L'ensemble de ces résultats permet d'orienter avec précision les stratégies d'amélioration des souches de *T. reesei* par ingénierie génétique.

Mots clés : *Trichoderma reesei*, cellulases, régulations, transcriptome, transcription, *xyr1*, répression catabolique, lactose, QM 9414, RUT C30, hyperproduction, bioéthanol, biocarburants.

Abstract

Analysis of *Trichoderma reesei*'s transcriptome for cellulase production improvement

Trichoderma reesei is the major fungus used for industrial cellulase production, which currently remains the main limiting step in the process of second-generation biofuel production from lignocellulosic biomass. By using molecular genetic methods and functional genomic analyses, this PhD project aimed to describe the genetic mechanisms underlying cellulase synthesis and secretion by *T. reesei*, in order to make way for the improvement of its strain capacities by targeted mutagenesis.

First, we analyzed the regulation of expression of the major transcription factors involved in cellulase production in the course of a lactose induction. We demonstrated that *xyr1*, *ace1* and *ace2* are induced by lactose, and that full induction of *xyr1* and *ace2* requires CRE1, the main mediator of carbon catabolite repression. Subsequently, several transcriptomic analyses allowed us (i) to identify genes that are regulated by carbon catabolite repression and to assess their dependence on CRE1, and (ii) to draw up a global description of the regulations that occur during a cellulase induction process. Our results proved the importance of the basal genetic expression for the acquisition of hyperproducing capacities, and allowed us to identify the main cellular functions that are subject to regulation during the induction process.

Overall, our analyses will help to elaborate strategies for the generation of enhanced *T. reesei* strains by genetic engineering.

Keywords : *Trichoderma reesei*, cellulases, regulations, transcriptome, transcription, *xyr1*, carbon catabolite repression, lactose, QM 9414, RUT C30, hyperproduction, bioethanol, biofuels.

Thèse soutenue et financée par la Fondation Tuck dans le cadre du fond dédié Enerbio



***Ener*bio**

Table des matières

RESUME	3
ABSTRACT	3
TABLE DES MATIERES	5
ABREVIATIONS	9
Liste des figures et des tableaux	10
1 INTRODUCTION	11
2 ETUDE BIBLIOGRAPHIQUE	15
2.1 CELLULOSE ET ENZYMES CELLULOLYTIQUES.....	17
2.1.1 <i>Structure et composition de la biomasse lignocellulosique</i>	17
2.1.2 <i>Nature et fonction des enzymes cellulolytiques</i>	19
2.2 <i>TRICHODERMA REESEI</i> , UN PRODUCTEUR DE CELLULASES	20
2.2.1 <i>Identification de l'espèce</i>	20
2.2.2 <i>Taxonomie, écologie et cycle de développement</i>	20
2.2.3 <i>Le répertoire d'enzymes cellulolytiques de T. reesei</i>	21
2.2.4 <i>Des mutagénèses aléatoires à l'origine de souches hyperproductrices</i>	22
2.2.5 <i>Les apports du séquençage à la connaissance du génome de T. reesei</i>	23
2.2.5.1 <i>Un pool d'enzymes cellulolytiques réduit</i>	23
2.2.5.2 <i>Des regroupements physiques de gènes de CAZymes</i>	24
2.3 REGULATION DE L'EXPRESSION DES CELLULASES	25
2.3.1 <i>Les inducteurs de la production de cellulases</i>	25
2.3.1.1 <i>L'induction des cellulases par la cellulose</i>	25
2.3.1.2 <i>Le lactose, un inducteur inattendu mais efficace</i>	26
2.3.2 <i>La régulation transcriptionnelle des gènes de cellulases</i>	30
2.3.2.1 <i>Les facteurs activateurs de la production de cellulases</i>	31
2.3.2.2 <i>Les facteurs répresseurs de la production de cellulases</i>	32
2.3.2.3 <i>Régulation intégrée des cellulases</i>	34
2.3.2.4 <i>Vers de nouveaux mécanismes régulateurs ?</i>	35
2.4 OBJECTIFS DU PROJET	36
3 MATERIEL ET METHODES	39
3.1 MATERIEL BIOLOGIQUE.....	41
3.2 CONDITIONS DE CULTURES	41
3.2.1 <i>Cultures des souches pour les études cinétiques</i>	41
3.2.1.1 <i>Précultures</i>	41
3.2.1.2 <i>Cultures en bioréacteur</i>	42
3.2.2 <i>Cultures continues et cultures en fioles</i>	42
3.3 METHODES ANALYTIQUES	43
3.3.1 <i>Détermination de la concentration en biomasse</i>	43
3.3.2 <i>Détermination de la concentration en sucres</i>	43
3.3.3 <i>Détermination de la concentration en protéines extracellulaires</i>	44
3.4 PREPARATION DES ECHANTILLONS D'ARN.....	44
3.5 ÉTUDE DES GENES D'INTERET PAR PCR QUANTITATIVE	44
3.5.1 <i>Rétrotranscription</i>	44
3.5.2 <i>PCR quantitative</i>	45
3.5.2.1 <i>Principe</i>	45
3.5.2.2 <i>Protocole</i>	45
3.5.2.3 <i>Analyse</i>	46
3.6 ÉTUDES TRANSCRIPTOMIQUES.....	46
3.6.1 <i>Les puces à ADN : principe</i>	46
3.6.2 <i>Conception de puces à ADN dédiées à l'étude de T. reesei</i>	47

3.6.2.1	Choix du dessin des oligonucléotides	47
3.6.2.2	Conception des sondes.....	47
3.6.2.3	Production des puces à ADN	48
3.6.3	<i>Protocole expérimental</i>	48
3.6.3.1	Marquage des cibles	48
3.6.3.2	Hybridation	49
3.6.3.3	Lecture des lames.....	49
3.6.4	<i>Traitements bioinformatiques et statistiques</i>	50
3.6.5	<i>Analyse des résultats par classification non supervisée</i>	51
3.6.6	<i>Identification des gènes, prédictions fonctionnelles et calculs d'enrichissement</i>	51
3.6.7	<i>Analyses des séquences promotrices</i>	52
3.6.8	<i>Séquençage à haut débit du transcriptome entier</i>	53
4	RESULTATS ET DISCUSSIONS	55
4.1	ÉTUDE DU COMPORTEMENT DES FACTEURS DE TRANSCRIPTION CENTRAUX EN REPONSE A UNE INDUCTION PAR LE LACTOSE.....	57
4.1.1	<i>Validation du système expérimental</i>	57
4.1.1.1	Le lactose induit la sécrétion de cellulases.....	57
4.1.1.2	Le lactose induit la transcription des gènes de CAZymes	59
4.1.2	<i>L'expression de xyr1 est induite par le lactose</i>	59
4.1.3	<i>L'induction de xyr1 par le lactose n'est pas due à une levée de la répression catabolique</i> 60	
4.1.3.1	La souche $\Delta cre1$ est bien dérèprimée.....	60
4.1.3.2	Xyr1 est induit par le lactose dans la souche $\Delta cre1$	61
4.1.4	<i>Xyr1 est régulé positivement par CRE1 au cours de l'induction lactose</i>	61
4.1.5	<i>L'induction de xyr1 par le lactose est indépendante du métabolisme du D-galactose</i>	61
4.1.6	<i>Ace1 et ace2 sont induits par le lactose</i>	62
4.1.7	<i>Vérification du caractère hyperproducteur de la souche CL847 de T. reesei</i>	63
4.1.7.1	CL847 descend de la souche RUT C30, fortement productrice	64
4.1.7.2	CL847 produit bien plus de cellulases que QM 9414.....	64
4.1.8	<i>L'expression de xyr1 et de ace1 est modifiée dans la souche hyperproductrice CL847</i>	65
4.1.9	<i>Discussion</i>	67
4.1.9.1	Induction de l'expression de xyr1 par le lactose	67
4.1.9.2	Quels sont les mécanismes de l'induction de xyr1 par le lactose ?.....	68
4.1.9.3	Les apports d'une comparaison intersouches.....	68
4.1.9.4	Ace2 et ace1 ont un comportement plus difficilement interprétable	69
4.1.9.5	Vers un rôle activateur de CRE1 ?	70
4.2	ÉTUDE DE LA REPRESSION CATABOLIQUE, MECANISME D'INHIBITION DE L'EXPRESSION DES GENES DE CELLULASES	73
4.2.1	<i>Description du dispositif expérimental</i>	73
4.2.1.1	Création de puces à ADN dédiées T. reesei	73
4.2.1.2	Réalisation de cultures continues des souches QM 9414 et $\Delta cre1$	74
4.2.1.3	Hybridations et analyses des résultats.....	74
4.2.2	<i>Caractéristiques générales de la régulation de l'expression des gènes</i>	75
4.2.2.1	Influence de CRE1.....	75
4.2.2.2	Influence du taux de croissance	75
4.2.3	<i>Comportement et fonctions principales des gènes régulés positivement en l'absence de CRE1</i> 77	
4.2.4	<i>Comportement et fonctions principales des gènes régulés négativement en l'absence de CRE1</i> 79	
4.2.5	<i>Comportement et fonctions principales des gènes régulés par le taux de croissance indépendamment de CRE1</i>	79
4.2.6	<i>Un groupe aux profils hétérogènes</i>	80
4.2.7	<i>Présence de sites de liaison CRE1 dans les régions 5' des gènes identifiés</i>	80
4.2.8	<i>Discussion</i>	81
4.2.8.1	CRE1, un régulateur majeur combinant les fonctions d'activateur et de répresseur	82
4.2.8.2	Les protéines de transport : des cibles privilégiées de CRE1.....	82
4.2.8.3	CRE1 agit sur l'expression basale de certaines CAZymes	83
4.2.8.4	Influence de CRE1 sur les régulateurs transcriptionnels.....	83
4.2.8.5	CRE1, RAS1, et le développement de T. reesei.....	84

4.2.8.6	Un nouveau motif de liaison potentiel ?.....	85
4.3	ANALYSE DE L'EXPRESSION GENETIQUE GLOBALE AU COURS DE LA PRODUCTION DE CELLULASES	87
4.3.1	<i>Description du dispositif expérimental</i>	87
4.3.1.1	Cultures	87
4.3.1.2	Hybridations et analyses des résultats.....	89
4.3.2	<i>Caractéristiques générales de la régulation de l'expression des gènes</i>	92
4.3.2.1	Régulations mises en jeu au cours de l'induction de la production de cellulases chez NG14 et RUT C30 92	
4.3.2.2	Régulations de l'expression des gènes hors induction entre QM6a et RUT C30	92
4.3.3	<i>Comportement et fonctions principales des gènes régulés positivement au cours de l'induction</i>	93
4.3.4	<i>Comportement et fonctions principales des gènes régulés négativement au cours de l'induction</i>	95
4.3.5	<i>Un groupe aux profils hétérogènes</i>	96
4.3.6	<i>Les mutations génétiques de NG14 et RUT C30 touchent majoritairement des gènes non différenciellement exprimés</i>	96
4.3.7	<i>Discussion</i>	97
4.3.7.1	Nature et fonction des protéines les plus fortement induites au cours de la production de cellulases..	98
4.3.7.2	Nature des régulations mises en jeu chez l'hyperproductrice	99
4.3.7.3	RUT C30 possède les équipements cellulaires nécessaires à une importante synthèse protéique.....	101
4.3.7.4	Induction du système de sécrétion, répression des protéines de dégradation	102
4.3.7.5	Une régulation réduite des facteurs de transcription.....	104
4.3.7.6	Métabolisme du galactose et production de cellulases	106
4.3.7.7	Comparaison entre le génome et le transcriptome des souches.....	106
5	CONCLUSION ET PERSPECTIVES.....	111
6	REFERENCES.....	117
7	ANNEXES.....	125
7.1	TABLEAUX ANNEXES	127
7.2	FIGURES ANNEXES	159
7.3	PUBLICATIONS	163

Abréviations

ABC	ATP-binding cassette
ADN	Acide desoxyribonucléique
ADNc	Acide desoxyribonucléique complémentaire
AMPc	Adénosine monophosphate cyclique
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messenger
ATP	Adénosine triphosphate
BFB	Avant le fed-batch (before fed-batch)
CAZyme	Enzyme agissant sur les glucides (Carbohydrate-active enzyme)
CBH	Cellobiohydrolase
CT	Cycle seuil (cycle threshold)
Da	Dalton
EC	Enzyme Commission
EG	Endoglucanase
FunCat	Catalogue Fonctionnel
GH	Glycoside hydrolase
GO	Gene Ontology
GTP	Guanosine triphosphate
Kb	Milliers de paires de bases
KDa	Kilodalton
Mb	Millions de paires de bases
MFS	Major facilitator superfamily
Pb	Paires de bases
PCRq	Réactions de polymérisation en chaine quantitatives
pO ₂	Pression en oxygène dissout
RE	Réticulum endoplasmique
RNA-seq	Séquençage à haut débit des transcrits
RT	Rétrotranscription
SNP	Polymorphisme nucléotidique

Liste des Figures et des Tableaux

Figures

FIGURE 1. STRUCTURE DE LA LIGNOCELLULOSE.....	18
FIGURE 2. ACTION DES ENZYMES DEGRADANT LA CELLULOSE ET LES HEMICELLOLOSES	19
FIGURE 3. GENEALOGIE SIMPLIFIEE DES SOUCHES DE <i>T. REESEI</i> UTILISEES AU COURS DE L'ETUDE	23
FIGURE 4. VOIES DE CATABOLISME DU GALACTOSE CHEZ <i>T. REESEI</i>	29
FIGURE 5. LE RESEAU DE REGULATION DE L'EXPRESSION DES GENES DE CELLULASES CHEZ <i>T. REESEI</i>	35
FIGURE 6. LA SOUCHE QM 9414 DE <i>T. REESEI</i> EXPRIME LES GENES <i>CBH1</i> ET <i>BGA1</i> LORS D'UN FED-BATCH LACTOSE.....	58
FIGURE 7. L'EXPRESSION DE <i>XYR1</i> EST INDUITE PAR LE LACTOSE INDEPENDAMMENT D'UNE LEVEE DE LA REPRESSION CATABOLIQUE.	60
FIGURE 8. L'INDUCTION LACTOSE DE <i>XYR1</i> NE DEPEND PAS DU METABOLISME DU GALACTOSE.	62
FIGURE 9. LES GENES <i>ACE2</i> ET <i>ACE1</i> SONT REGULES PAR LE LACTOSE, MAIS SEUL <i>ACE1</i> EST SOUMIS A LA REPRESSION CATABOLIQUE.	63
FIGURE 10. LA SOUCHE HYPERPRODUCTRICE CL847 DE <i>T. REESEI</i> PRODUIT DES CELLULASES LORS D'UN FED-BATCH LACTOSE	64
FIGURE 11. L'EXPRESSION DES GENES REGULATEURS DE LA TRANSCRIPTION DES CELLULASES EST MODIFIEE DANS LA SOUCHE HYPERPRODUCTRICE CL847	66
FIGURE 12. LES NOUVELLES DONNEES DE REGULATION DES GENES DE CELLULASES CHEZ <i>T. REESEI</i>	71
FIGURE 13. PLAN EXPERIMENTAL UTILISE POUR ETUDIER LE ROLE DE <i>CRE1</i> ET DES TAUX DE CROISSANCE SUR L'EXPRESSION DES GENES DE <i>T. REESEI</i>	74
FIGURE 14. HEATMAP DES GENES LES PLUS REGULES.....	76
FIGURE 15. DISTRIBUTION DES GENES PARMI LES DIFFERENTS GROUPES D'EXPRESSION.....	77
FIGURE 16. DISTRIBUTION DES GENES PARMI LES CATEGORIES FONCTIONNELLES PRINCIPALES.....	78
FIGURE 17. PROFILS DES CULTURES DE NG14 ET RUT C30 AU COURS DE L'EXPERIMENTATION	88
FIGURE 18. EVOLUTION DE LA CONCENTRATION EN PROTEINES EXTRACELLULAIRES AU COURS DES CULTURES DE NG14 ET RUT C30.....	88
FIGURE 19. PLAN EXPERIMENTAL EMPLOYE POUR ETUDIER LES REGULATIONS TRANSCRIPTIONNELLES MISES EN JEU CHEZ QM6A, NG14 ET RUT C30.....	89
FIGURE 20. HEATMAP DES GENES SIGNIFICATIVEMENT REGULES.	90
FIGURE 21. DISTRIBUTION DES GENES PARMI LES DIFFERENTS GROUPES D'EXPRESSION	93
FIGURE 22. PROFILS DE REGULATION DES GENES CODANT POUR DES PROTEINES RIBOSOMIQUES.....	102
FIGURE 23. PRINCIPALES REGULATIONS DES FONCTIONS CELLULAIRES AU COURS DE L'INDUCTION DE LA PRODUCTION DE CELLULASES.	109

Tableaux

TABLEAU 1. LISTE DES ENZYMES CELLULOLYTIQUES DE <i>T. REESEI</i>	22
TABLEAU 2. CAZYMES ET CELLULASES CODEES PAR LE GENOME DE <i>T. REESEI</i> ET DE 12 AUTRES ORGANISMES	24
TABLEAU 3. SOUCHES DE <i>TRICHODERMA REESEI</i> UTILISEES AU COURS DE L'ETUDE.....	41
TABLEAU 4. RESULTATS DE L'ANALYSE DES PROMOTEURS DES GENES REGULES	81
TABLEAU 5. INVENTAIRE DES MUTATIONS TOUCHANT DES GENES REGULES AU COURS DE LA PRODUCTION DE CELLULASES.....	97
TABLEAU 6. NIVEAUX D'EXPRESSION CHEZ RUT C30 DE DIFFERENTS GENES LIES AU METABOLISME DES ACIDES AMINES.	101
TABLEAU 7. NIVEAUX D'EXPRESSION CHEZ RUT C30 DE DIFFERENTS GENES CODANT POUR DES PROTEINES ASSOCIEES AU PROTEASOME.	104
TABLEAU 8. NIVEAUX D'EXPRESSION CHEZ RUT C30 DES GENES <i>XYR1</i> , <i>ACE1</i> ET <i>ACE2</i>	105

1 INTRODUCTION

Les phénomènes de réchauffement climatique, de pollution mondiale croissante et de gaz à effet de serre ont contribué ces dernières années à s'interroger sur l'avenir écologique de notre planète. Cette prise de conscience, renforcée par la réalité de la diminution des ressources énergétiques fossiles, a initié un revirement de la politique économique et environnementale internationale pour envisager une nouvelle vision du futur énergétique. C'est dans ce contexte économique et politique que les biocarburants, dont le développement a été initié il y a plus d'une trentaine d'années, connaissent un regain d'intérêt.

Le bioéthanol est le biocarburant dont l'usage est le plus répandu à l'échelle mondiale. Il est aujourd'hui produit à partir de deux grands types de cultures, les plantes sucrières (cannes à sucres, betteraves) et les plantes amylacées (blé, maïs). Vivement controversé pour sa concurrence potentielle aux ressources alimentaires planétaires et son bilan peu prometteur en terme d'émissions de CO₂, ce biocarburant de première génération devrait être progressivement complété et remplacé par du bioéthanol de deuxième génération, dérivant des ressources lignocellulosiques (Ballerini and Alazard-Toux, 2006). Le procédé classique de production de bioéthanol à partir de la biomasse lignocellulosique suit un processus en trois étapes. La première étape consiste à prétraiter le substrat pour rendre la cellulose, polysaccharide principal des parois végétales, accessible à l'hydrolyse ; l'action de cocktails d'enzymes cellulolytiques sur la cellulose libère alors les monosaccharides de glucose qui la constituent ; ces derniers sont ensuite fermentés en éthanol qui sera récupéré par distillation (Margeot *et al.*, 2009). Les phases de production d'enzymes et d'hydrolyse représentent à elles seules 30% du coût de l'éthanol produit à partir de matériaux lignocellulosiques. L'amélioration de ces phases est donc indispensable pour rendre ce procédé viable économiquement.

Dans le but de produire des quantités d'enzymes plus importantes, les acteurs du domaine industriel des biocarburants utilisent des souches du champignon filamentueux *Trichoderma reesei* obtenues par des approches de mutagenèse aléatoire au cours des années 1980 et sélectionnées par criblage pour leurs capacités de sécrétion de cellulases améliorées (Durand *et al.*, 1988a; Durand *et al.*, 1988b). Ces approches « à l'aveugle » ont abouti à des résultats prometteurs (la production des protéines a ainsi été augmentée d'un facteur 10), mais ont désormais atteint leurs limites. Aujourd'hui, il semble donc essentiel d'avoir recours à des stratégies de modifications génétiques plus ciblées, dirigées spécifiquement contre les acteurs moléculaires impliqués dans les processus d'induction de la production de cellulases. La mise en œuvre de ces stratégies nécessite une connaissance précise de ces processus.

Mon projet de thèse se base sur l'utilisation de techniques de génétique moléculaire et d'outils de génomique fonctionnelle pour décrire les mécanismes de régulation de la synthèse et de la sécrétion d'enzymes cellulolytiques chez *T. reesei*. Une première partie de mon travail a consisté à

étudier les gènes déjà connus comme étant impliqués dans la production de cellulases, et à préciser leur comportement transcriptionnel lors d'une induction par le lactose, seul inducteur soluble de cellulases pouvant être utilisé à l'échelle industrielle. Dans un deuxième temps, des études transcriptomiques globales m'ont permis (i) d'identifier les gènes cibles de la répression catabolique – processus impliqué dans l'inhibition de la synthèse des cellulases –, et (ii) de décrire à l'échelle du génome total les régulations mises en jeu lors de cinétiques d'induction de la production de ces enzymes.

Ces travaux résultent d'une collaboration entre le département Biotechnologie d'IFP Énergies nouvelles, l'Institut de Biologie de l'ENS (IBENS), et le laboratoire de Technologie Génétique et de Biochimie appliquée de l'Université Technologique de Vienne (Autriche).

2 ÉTUDE BIBLIOGRAPHIQUE

2.1 Cellulose et enzymes cellulolytiques

La production de biocarburants de deuxième génération se base sur l'exploitation des ressources énergétiques stockées dans les parois des cellules végétales. Cette énergie est accessible moyennant une dégradation des polysaccharides constitutifs de ces parois par les enzymes adaptées.

2.1.1 Structure et composition de la biomasse lignocellulosique

La structure et la composition des parois végétales varient suivant les espèces, les tissus, les types cellulaires, et même les différentes couches moléculaires au sein d'une même paroi. Cette complexité chimique et structurale leur permet d'assurer des fonctions de support mécanique, mais aussi de protection contre les pathogènes ou de transport régulé de molécules (eau, solutés...). Elles sont principalement composées de polysaccharides, cellulose et hémicellulose, et de lignine.

Le composant principal des parois végétales, la cellulose, est formé de longs homopolymères d'environ 8 000 à 12 000 unités de glucose liées par des liaisons glycosidiques β -1-4. Ces chaînes de cellodextrines ne portent pas de branchements, et sont étroitement assemblées entre elles en une structure organisée appelée microfibrille. La cohérence de cette structure est permise par la conformation « en chaise » des résidus de glucose au sein des cellulodextrines, qui force les groupements hydroxyles à adopter une orientation radiale et les atomes d'hydrogènes aliphatiques une position axiale. Ainsi, des liaisons hydrogènes entre chaînes de celluloses adjacentes sont favorisées, conférant aux microfibrilles une résistance forte à l'hydrolyse enzymatique. Cette structure cristalline, insoluble, est acquise par auto-assemblage immédiat des molécules individuelles au site de biosynthèse (Himmel *et al.*, 2007; Martinez *et al.*, 2009; Wei *et al.*, 2009). Dans quelques rares cas, par exemple chez le cotonnier, la cellulose peut être présente sous une forme presque pure. Néanmoins, les fibres de cellulose sont la plupart du temps entourées d'une matrice d'autres biopolymères structuraux, les hémicelluloses et la lignine, qui constituent respectivement de 20 à 35% et de 5 à 30% du poids sec des plantes. Une couche de cellulose amorphe peut également entourer la cellulose cristalline, ce qui en restreint encore d'avantage l'accès et la protège d'éventuelles enzymes cellulolytiques (Wei *et al.*, 2009).

L'hémicellulose n'est pas un composé chimiquement bien défini mais plutôt une famille d'hétéropolysaccharides ramifiés composés de différentes unités monosaccharidiques à 5 ou 6 carbones, comme le D-xylose, le D-mannose, le D-glucose, le L-arabinose et le D-galactose. Certains de ces monosaccharides sont parfois partiellement estérifiés par les acides acétique, férulique et p-coumarique (Rubin, 2008). La lignine est un polymère tridimensionnel d'unités

phénylpropanoïdes, principalement de l'alcool p-coumarique, de l'alcool coniférylique et de l'alcool sinapylique, dont les proportions varient en fonction des espèces et des tissus. Elle joue le rôle de « colle cellulaire » car elle permet de comprimer les fibres polysaccharidiques, conférant ainsi aux parois une forte rigidité et des propriétés de résistance aux insectes et aux pathogènes (Del Rio *et al.*, 2007). Un schéma synoptique de la structure de la lignocellulose est présenté en Figure 1.

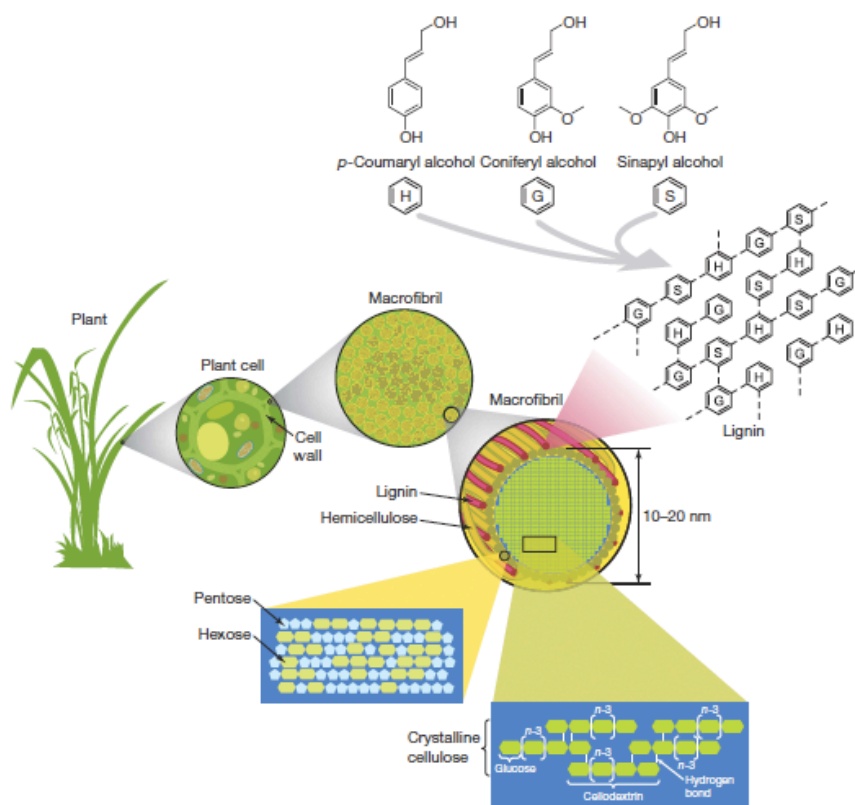


Figure 1. Structure de la lignocellulose. La cellulose (chaînes de résidus glucose assemblées par des liaisons hydrogènes), les hémicelluloses (constituées de différents glucides à 5 ou 6 carbones) et la lignine (composée d'alcools p-coumariques, coniféryliques et sinapyliques), forment des structures appelées microfibrilles qui s'organisent en macrofibrilles assurant la stabilité structurale de la paroi cellulaire végétale. D'après Rubin, 2008.

Les interactions étroites des différents composés biochimiques des parois végétales les protègent donc fortement d'une hydrolyse destinée à libérer leurs monosaccharides constitutifs. Trois catégories d'enzymes sont nécessaires pour assurer cette hydrolyse : les cellulases, les hémicellulases et des enzymes accessoires modifiant et dégradant la lignine. Différents procédés industriels étant capables de solubiliser les hémicelluloses et de se substituer aux effets des enzymes accessoires (Margeot *et al.*, 2009), seule la maîtrise de la production des cellulases représente un enjeu majeur dans la mise en place de procédés industriels de synthèse de bioéthanol.

2.1.2 Nature et fonction des enzymes cellulolytiques

La complexité de la cellulose représente un véritable défi en terme d'hydrolyse. Une dégradation efficace de cette structure n'est possible que par une action synergique de différents types d'enzymes cellulolytiques. Toutes les cellulases ont les mêmes spécificités chimiques vis-à-vis des liaisons glycosidiques β -1-4, mais elles diffèrent en terme de site d'attaque du substrat solide. Traditionnellement, les termes d'exoglucanases et d'endoglucanases sont retenus pour qualifier leurs différences de mode d'action. Les exoglucanases, ou cellobiohydrolases (CBH) (1,4- β -D-glucan cellobiohydrolase, EC 3.2.1.91), clivent les unités de cellulose depuis les extrémités des chaînes polysaccharidiques, et présentent une activité élevée sur la cellulose cristalline. Certaines cellobiohydrolases attaquent les chaînes de cellulose depuis leur extrémité réductrice, d'autres depuis leur extrémité non-réductrice. Les endoglucanases (EG) (1,4- β -D-glucan-4-glucanohydrolase, EC 3.2.1.4) sont capables d'hydrolyser les chaînes polysaccharidiques en leur milieu. Leur action, préférentiellement efficace sur les régions amorphes de la cellulose, génère ainsi de nouvelles extrémités permettant de nouvelles hydrolyses par les cellobiohydrolases. Enfin, les β -glucosidases (EC 3.2.1.21) clivent le cellobiose et les autres oligosaccharides solubles en glucose (Figure 2) (Aro *et al.*, 2005; de Vries and Visser, 2001; Harman and Kubicek, 1998).

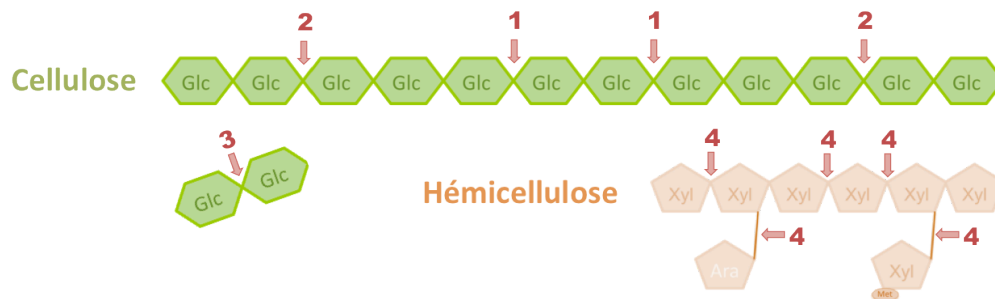


Figure 2. Action des enzymes dégradant la cellulose et les hémicelluloses. Les endoglucanases (1) hydrolysent les cellodextrines en leur milieu tandis que les cellobiohydrolases (2) attaquent les chaînes polysaccharidiques depuis leurs extrémités. Les diholosides résultant de l'action de ces deux enzymes sont clivés par les β -glucosidases (3). Les hémicellulases (4) attaquent les chaînes d'hémicelluloses. Glc : résidu glucose ; Xyl : résidu xylose ; Ara : résidu arabinose.

Une telle distinction entre enzymes ne prend cependant pas en compte leurs similarités de séquences et leurs mécanismes catalytiques ; pour cette raison, le système de classification des « enzymes agissant sur les glucides et leurs dérivés » (Carbohydrate-Active enzymes, CAZymes) mis en place par Couinho et Henrissat (Cantarel *et al.*, 2009) est aujourd'hui communément adopté et utilisé. D'après ce classement, les cellulases appartiennent à la famille des Glycoside Hydrolases (GHs). Cette famille biochimique comprend également les hémicellulases, qui, comme les cellulases,

hydrolysent les hémicelluloses par l'action d'endo-enzymes et d'exo-enzymes libérant les sucres monomères.

2.2 *Trichoderma reesei*, un producteur de cellulases

Différents microorganismes de différents clades du vivant ont la capacité de dégrader la biomasse suivant diverses stratégies. Ainsi, des bactéries aérobies (e.g. *Bacillus*), anaérobies (e.g. *Clostridium*), des actinomycètes (e.g. *Cellulomonas*) et des champignons filamenteux (e.g. *Trichoderma*) sont capables de produire des cellulases. Les caractéristiques génétiques de ces différents organismes étant plus ou moins connues, de nombreuses études sont en cours afin d'accroître notre connaissance de leurs mécanismes de dégradation et de leur répertoire d'enzymes cellulolytiques. À l'échelle industrielle, *Trichoderma reesei* est l'organisme le plus utilisé pour produire des cellulases en raison de ses fortes capacités de sécrétion qui rendent ce modèle intéressant économiquement.

2.2.1 Identification de l'espèce

L'intérêt scientifique pour le genre *Trichoderma*, introduit dans la classification il y a plus de 200 ans, a émergé durant la seconde Guerre Mondiale lorsque l'armée américaine s'est inquiétée de la vitesse à laquelle ses matériaux militaires étaient dégradés dans les régions tropicales, particulièrement dans le Pacifique sud. Face à ce phénomène, le corps des intendants militaires a mis en place un vaste programme d'étude afin de déterminer la nature des agents responsables de ce pourrissement (Reese, 1976). C'est ainsi que la souche aux capacités cellulolytiques les plus développées de la région a été identifiée. Cette souche, que l'on nomma QM6a (pour « quartermaster », grade militaire, Reese *et al.*, 1950), a initialement été rattachée de manière inappropriée à l'espèce *Trichoderma viride*, avant d'être considérée comme une nouvelle espèce à part entière. Cette espèce fut baptisée *Trichoderma reesei*, en l'honneur de son investigateur, Elwyn T. Reese (Simmons, 1977). C'est grâce aux travaux de ce dernier, qui ont largement mis en évidence le potentiel cellulolytique du champignon, que *Trichoderma reesei* a suscité l'intérêt de nombreux laboratoires américains, européens et asiatiques lors du premier choc pétrolier de la fin des années 1960.

2.2.2 Taxonomie, écologie et cycle de développement

Trichoderma est un champignon filamenteux de l'embranchement des ascomycètes. Ce groupe frère des basidiomycètes regroupe l'ensemble des organismes dont les spores sont formées, au cours

du cycle cellulaire, à l'intérieur d'asques. Les eurotiomycètes (regroupant notamment les espèces cellulolytiques du genre *Aspergillus* dont la machinerie de dégradation de la biomasse est souvent étudiée), les saccharomycètes (comprenant les levures ascomycètes non cellulolytiques des genres *Saccharomyces* et *Candida*) et les sordariomycètes constituent les principales classes de cet embranchement. La classe des sordariomycètes, dont *Trichoderma* fait partie, comprend plus de 3 000 espèces identifiées dont un grand nombre possède des capacités de dégradation des polysaccharides végétaux.

Longtemps basée sur une simple identification morphologique des espèces, la taxonomie du genre *Trichoderma* a connu de nombreuses avancées ces dernières décennies, grâce aux analyses de séquences génétiques. Des études pionnières, basées sur les analyses des séquences de gènes d'ARN ribosomiques de différentes espèces, ont notamment permis de prouver que *Trichoderma reesei* était l'anamorphe (forme reproductrice asexuée) de *Hypocrea jecorina*, un ascomycète initialement considéré comme appartenant à un clade plus éloigné de la phylogénie (Kuhls *et al.*, 1996). Des comparaisons des activités cellulolytiques des deux espèces ont confirmé cette relation (Kubicek *et al.*, 1996).

Les espèces du genre *Trichoderma* se trouvent sur les végétaux et les bois en décomposition, et représentent souvent la composante majoritaire de la microflore d'un grand nombre de sols. Elles sont particulièrement répandues dans la litière des forêts humides de feuillus. Ce sont des organismes saprophytes aux besoins nutritionnels assez faibles. Ils ont la capacité d'utiliser une grande variété de matière organique d'origine naturelle ou chimique, et peuvent également attaquer d'autres champignons (Danielson and Davey, 1973; Roiger *et al.*, 1991).

Les souches de *Trichoderma* sont identifiables par leurs aspects morphologiques, notamment grâce à leurs pigments conidiens vert clair ou blancs et à leur structure de conidophore embranchée. La plupart ne présente pas de cycle sexuel ; leur cycle de développement ne se compose que de phases de multiplication végétative par mitoses. Ces sporulations, souvent abondantes, leur assurent un développement rapide. Ce mode de vie est d'autant plus efficace que les organes de dissémination de *Trichoderma* sont capables de rester à l'état quiescent et de se développer uniquement lorsqu'une source d'énergie appropriée est disponible (Papavizas, 1985; Samuels, 2006).

2.2.3 Le répertoire d'enzymes cellulolytiques de *T. reesei*

Différents gènes codants pour des enzymes cellulolytiques ont été isolés chez *T. reesei* avant que le séquençage ne précise leurs caractéristiques et leur répartition au sein du génome (voir partie 2.2.5). Ainsi, deux gènes de cellobiohydrolases (*cbh1* et *cbh2*, codant respectivement pour des exoglucanases attaquant les extrémités réductrices et non réductrices des chaînes de cellulose), huit gènes d'endoglucanases (*egl1* à *egl5*, *cel74a*, *cel61b* et *cel5b*) et sept gènes de β -glucosidases (*bgl1* et *bgl2*

notamment) ont été identifiés (Aro *et al.*, 2005). Le Tableau 1 répertorie l'ensemble de ces gènes, avec leur ancienne et leur nouvelle désignation. À ceux-ci s'ajoutent au moins quatorze gènes d'hémicellulases, dont les xylanases *xyn1* à *xyn4* et dix autres gènes codant par exemple pour une β -xylosidase (*bxl1*), une mannanase (*man1*), une acetyl xylane estérase (*axe1*), ou des arabinofuranosidases (*abf1-3*) (Aro *et al.*, 2005; Foreman *et al.*, 2003). Une grande majorité des produits de ces gènes est effectivement retrouvée dans les sécrétomes de souches de *T. reesei* lorsque le champignon produit ses enzymes. Une étude protéomique révèle que 80 à 90% des protéines sécrétées par le champignon dans ces conditions sont bel et bien des enzymes liées à la dégradation de la biomasse, et que les cellobiohydrolases CBH1 et CBH2 sont les protéines les plus abondamment produites (70 à 80% du total des cellulases) (Herpoel-Gimbert *et al.*, 2008).

Cellobiohydrolases			Endo- β -1,4-glucanases		
Gène	Protéine	Famille de GH	Gène	Protéine	Famille de GH
cbh1/cel7a	CBH1/CEL7A	GH7	egl1/cel7b	EG1/CEL7B	GH7
cbh2/cel6	CBH2/CEL6	GH6	egl2/cel5a	EG2/CEL5A	GH5
β -glucosidases			egl3/cel12a	EG3/CEL12A	GH12
Gène	Protéine	Famille de GH	egl4/cel61a	EG4/CEL61A	GH61
bgl1/cel3a	BGL1/CEL3A	GH3	egl5/cel45a	EG45/CEL45A	GH45
bgl2/cel1a	BGL2/CEL1A	GH1	cel74a	CEL74A	GH74
cel3b	CEL3B	GH3	cel61b	CEL61B	GH61
cel3c	CEL3C	GH3	cel5b	CEL5B	GH5
cel1b	CEL1B	GH1			
cel3d	CEL3D	GH3			
cel3e	CEL3E	GH3			

Tableau 1. Liste des enzymes cellulolytiques de *T. reesei*. Les anciennes et nouvelles nomenclatures des gènes et de leur protéine associée sont indiquées. La répartition des protéines en familles de Glycosides Hydrolases (GH) révèle que les différences de fonctions ne résultent pas nécessairement de différences majeures de structure.

2.2.4 Des mutagénèses aléatoires à l'origine de souches hyperproductrices

En dépit de ses capacités cellulolytiques indiscutables, la souche QM6a n'est pas exploitable en l'état pour assurer une production industrielle rentable. De nombreux programmes d'amélioration génétique, principalement basés sur des mutagénèses aléatoires, ont donc été entrepris dans les années 1970. Ces approches, qui ont combiné l'action de différents agents mutagènes aussi bien chimiques (nitrosoguanidine, éthylméthylsulfonate) que physiques (rayonnements ultra-violets, irradiations par accélérateur linéaire), ont ainsi généré une grande diversité de souches, dont les plus performantes ont été isolées au moyen de différents cribles de sélection (Durand *et al.*, 1988b). Parmi elles, on trouve notamment la souche QM 9414, encore faiblement productrice et largement étudié dans la littérature, le mutant NG14 et son descendant

RUT C30, fortement producteur et très étudié également, et la souche CL847, que l'IFP a choisi d'exploiter en raison de ses excellentes capacités de production et de sa résistance à la répression catabolique par le glucose (voir partie 2.3.2.2.2). La Figure 3 présente un arbre simplifié de la généalogie des principales souches d'intérêt mentionnées dans la suite de ce document, telle qu'elle se présente d'après les données publiées à ce jour.

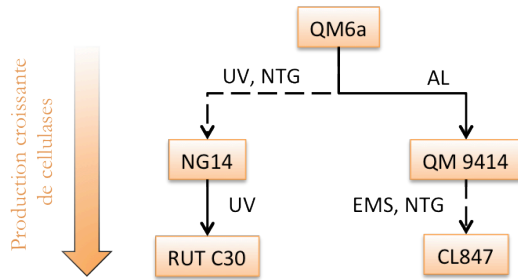


Figure 3. Généalogie simplifiée des souches de *T. reesei* utilisées au cours de l'étude. Les agents mutagènes à l'origine des mutants sont précisés. UV : rayonnement ultra-violet ; NTG : nitrosoguanidine ; AL : accélérateur linéaire ; EMS : éthylméthylsulfonate. Les flèches en pointillés symbolisent des étapes de mutagenèse multiples. Les capacités de production des souches augmentent avec les étapes de mutagenèse. D'après Durand *et al.*, 1988a.

Les techniques d'ingénierie génétique classique employées au cours des dernières décennies ont donc permis d'améliorer remarquablement la souche mère de *T. reesei* en terme de production d'enzymes. Néanmoins, la nécessité de comprendre plus précisément ses caractéristiques moléculaires afin d'accroître encore davantage son potentiel industriel a conduit à entreprendre un séquençage complet de son génome. L'analyse des données issues de ce séquençage révèle une organisation inattendue de ses gènes de CAZymes et un pool de séquences codant pour des enzymes cellulolytiques étonnamment réduit.

2.2.5 Les apports du séquençage à la connaissance du génome de *T. reesei*

Le séquençage aléatoire a permis de reconstituer les 33,4 millions de paires de bases (Mb) du génome de *T. reesei*. La combinaison de différentes méthodes d'annotation a aboutit à la modélisation de 9 129 gènes, d'une longueur moyenne de 1 793 paires de bases (Martinez *et al.*, 2008).

2.2.5.1 Un pool d'enzymes cellulolytiques réduit

Le répertoire des gènes de CAZymes de *T. reesei* a été comparé à celui de treize autres espèces de champignons pour lesquelles les génomes ont été séquencés (le détail de ces espèces est présenté dans le Tableau 2). Alors que l'on s'attendrait à ce qu'un dégradeur efficace des parois des cellules végétales contienne de nombreux gènes codants pour des protéines impliquées dans la digestion de la lignocellulose, seules 200 séquences de GHs ont été répertoriées chez *T. reesei*, contre 211 en moyenne chez les différents sordariomycètes étudiés et 265 chez les eurotiomycètes (voir

Tableau 2). De la même manière, le génome de *T. reesei* contient étonnamment moins de protéines possédant un domaine de liaison aux glucides (domaine pourtant largement rencontré chez les enzymes impliquées dans la dépolymérisation des polysaccharides végétaux) que tous les autres sordariomycètes considérés dans l'étude (36 « carbohydrate-binding modules » chez *T. reesei* contre 49 en moyenne) (Martinez *et al.*, 2008).

Taxons	Espèces	Familles de CAZymes				Cellulases principales					Somme
		GH	Moy GH	CBM	Moy CBM	CBH1 (CEL7A)	CBH2 (CEL6)	EG1 (CEL7)	EG2 (CEL5)	EG4 (CEL61)	
Eurotio- mycètes	<i>A. nid.</i>	247		36		2	2	1	2	9	16
	<i>A. fum.</i>	263	265	55	40	2	1	2	3	7	15
	<i>A. ory.</i>	285		30		2	1	1	2	8	14
Asco- mycètes	<i>M. gris.</i>	231		58		3	2	2	2	17	26
	<i>N. cra.</i>	171	211	39	49	2	2	3	1	14	22
	<i>T. ree.</i>	200		36		1	1	1	2	3	8
	<i>F. gra.</i>	243		61		1	0	1	2	13	17
Saccharo- mycètes	<i>C. alb.</i>	58		4		0	0	0	0	0	0
	<i>S. cer.</i>	45	47	12	9	0	0	0	0	0	0
	<i>C. gla.</i>	38		12		0	0	0	0	0	0
Archiasco- mycètes	<i>S. pom.</i>	46	46	5	5	0	0	0	0	0	0
Basidiomycètes	<i>C. neo.</i>	75		10		0	0	0	0	1	1
	<i>P. chr.</i>					7	1	2	2	14	26

Tableau 2. CAZymes et cellulases codées par le génome de *T. reesei* et de 12 autres organismes. *T. reesei* apparaît en gras. Les moyennes (moy) sont données par groupes taxonomiques. Les plus grands et plus petits nombres d'enzymes de chaque classe sont respectivement indiqués en rouge et en bleu. CAZymes : GH, glycoside hydrolase ; CBM, carbohydrate-binding module. Organismes : *A. nid.*, *Aspergillus nidulans* ; *A. fum.*, *Aspergillus fumigatus* ; *A. ory.*, *Aspergillus oryzae* ; *M. gris.*, *Magnaporthe grisea* ; *N. cra.*, *Neurospora crassa* ; *T. ree.*, *Trichoderma reesei* ; *F. gra.*, *Fusarium graminearum* ; *C. alb.*, *Candida albicans* ; *S. cer.*, *Saccharomyces cerevisiae* ; *C. gla.*, *Candida glabrata* ; *S. pom.*, *Schizosaccharomyces pombe* ; *C. neo.*, *Cryptococcus neoformans* ; *P. chr.*, *Phanerochaete chrysosporium*. Adapté de Martinez *et al.*, 2008.

Une inspection plus précise des gènes de cellulases révèle que les séquences codant pour les endoglucanases ou les cellobiohydrolases les plus connues sont presque toutes présentes en une seule version dans le génome de *T. reesei*, ce qui fait de ce dernier le champignon ayant le moins de gènes de cellulases de tous les organismes cellulolytiques considérés dans l'étude (voir Tableau 2). *T. reesei* a également le plus petit pool d'hémicellulases des champignons analysés avec seulement 16 gènes contre 32 en moyenne chez les autres ascomycètes (Martinez *et al.*, 2008).

2.2.5.2 Des regroupements physiques de gènes de CAZymes

Une analyse détaillée de la distribution des gènes de CAZymes le long du génome de *T. reesei* révèle que plus de 40% de ces gènes sont regroupés dans 25 régions individualisées ayant une longueur variant de 14 à 275 kb. Ces régions contiennent entre 2 et 10 gènes de CAZymes, soit une

densité cinq fois supérieure à la densité attendue si la distribution de ces gènes sur les chromosomes était aléatoire. De plus, 73% des gènes de ces clusters sont dans des régions non synthéniques (où la colinéarité des gènes entre génomes d'espèces apparentées n'est pas respectée) alors que 72% de ces derniers ont des orthologues avec le sordariomycète *Fusarium graminearum*. Des mouvements de gènes au sein du génome de *T. reesei* auraient donc conduit à ce réarrangement des séquences en clusters (Martinez *et al.*, 2008). Une analyse par famille de gènes révèle que 70% des séquences de CAZymes trouvées dans les clusters codent pour des GHs. Des études visant à localiser dans le génome des gènes décrits comme étant activés par les mêmes inducteurs (sophorose et cellulose, voir partie 2.3.1.1) ont par ailleurs répertorié plusieurs exemples pour lesquels des gènes proches ou adjacents étaient coexprimés (Foreman *et al.*, 2003; Martinez *et al.*, 2008). Le regroupement aurait donc une réelle importance biologique, probablement en terme de contrôle de l'expression génétique.

Une telle organisation, véritable couplage fonctionnel des gènes dont les produits sont impliqués dans les mêmes processus de dégradation, pourrait expliquer pourquoi un organisme ayant un nombre de copies aussi réduit de gènes d'enzymes cellulolytiques constitue un des dégradeurs de débris végétaux des plus efficaces. Les mécanismes de régulation de l'expression des gènes de cellulases influençant de manière majeure la production de ces enzymes, une part importante des travaux menés sur *T. reesei* au cours des dernières décennies s'attache à étudier et à décrire précisément ces mécanismes.

2.3 Régulation de l'expression des cellulases

Plusieurs niveaux de régulation interviennent dans le contrôle transcriptionnel de la synthèse d'enzymes cellulolytiques. L'expression des gènes de cellulases nécessite en effet la présence d'un inducteur, et est régulée par différents facteurs de transcription.

2.3.1 Les inducteurs de la production de cellulases

2.3.1.1 L'induction des cellulases par la cellulose

Les cellulases ayant pour fonction de fournir à la cellule des oligomères solubles issus de la dégradation de la cellulose, il est attendu que la biosynthèse de ces enzymes ne soit assurée qu'en présence de leur substrat. Cependant, la cellulose étant un polymère insoluble, elle ne peut traverser la membrane des cellules pour initier le processus. Différentes études ont donc été menées pour déterminer la nature et les mécanismes de formation du véritable inducteur des cellulases. La théorie

la plus communément admise est qu'un niveau basal de cellulases (principalement CBH1 et CBH2) est exprimé en continu par les cellules, ce qui permet d'initier la dégradation de la cellulose lorsque cette dernière est présente dans le milieu et ainsi de libérer des petites quantités d'oligosaccharides, qui déclenchent à leur tour une biosynthèse accrue de cellulases (Carle-Urioste *et al.*, 1997; el-Gogary *et al.*, 1989). Il est également possible que les conidies de *T. reesei* présentent une activité cellulolytique membranaire. Il a en effet été décrit que l'enzyme CBH2 existait sous forme membranaire à la surface de ces structures cellulaires (Kubicek *et al.*, 1988). De plus, des mutants de délétions pour cette enzyme présentent une expression très réduite des autres gènes de cellulases, tandis que ces transcrits sont normalement détectables en cas de délétion de *cbh1* ou de *egl1* (codant pour des enzymes absentes des conidies) (Seiboth *et al.*, 1992).

Une activité constitutive de certaines cellulases aboutirait donc à la formation d'un oligosaccharide inducteur. Les résultats de différentes études semblent pencher pour un rôle du sophorose (2-O- β -D-Glucopyranosyl-D-glucose) dans cette induction. La formation de cette molécule a en effet été détectée dans des cultures de cellules en présence de cellobiose (D-Glucosyl- β -1,4-D-glucose) ou de cellulose. Ainsi, le sophorose pourrait résulter de la transglycosylation du cellobiose issu de la dégradation de la cellulose par les cellulases exprimées constitutivement. Le sophorose serait ensuite intégré par les cellules *via* une perméase ayant une forte affinité pour cet oligosaccharide à faibles concentrations (Kubicek *et al.*, 1993). Différentes données expérimentales prouvent qu'une β -glucosidase extracellulaire ou liée à la membrane pourrait catalyser la transglycosylation du cellobiose en sophorose. L'enzyme BGL1 pourrait par exemple être impliquée dans une telle fonction puisque qu'une perturbation de l'expression du gène *bgl1* entraîne un retard d'expression des autres cellulases en présence de cellulose mais pas en présence de sophorose (Mach *et al.*, 1995).

2.3.1.2 Le lactose, un inducteur inattendu mais efficace

Le lactose (D-galactosyl- β -1,4-D-glucoside) est également capable d'induire efficacement l'expression des gènes de cellulases (Mach and Zeilinger, 2003). Son avantage principal est d'être une molécule soluble et peu coûteuse (par rapport au cellobiose ou au sophorose), ce qui permet de l'utiliser également en tant que source de carbone lors de procédés industriels de production.

2.3.1.2.1 La β -galactosidase et l'hydrolyse du lactose

Chez *T. reesei*, le catabolisme du lactose nécessite une activité β -galactosidase extracellulaire pour que le composé soit hydrolysé en D-galactose et en D-glucose et ainsi assimilé par le champignon *via* des perméases adaptées. En effet, si d'autres espèces fongiques, comme *Kluyveromyces*

lactis ou *Aspergillus nidulans*, sont capables d'absorber le lactose par des lactose perméases spécifiques et de l'hydrolyser ensuite intracellulairement (Chang and Dickson, 1988; Dickson and Barr, 1983; Fekete *et al.*, 2002), ce n'est pas le cas de *T. reesei* qui ne possède pas de β -galactosidase intracellulaire (Seiboth *et al.*, 2005).

La présence dans le génome du champignon du gène *bga1*, codant pour la β -galactosidase extracellulaire BGA1 (famille GH35), est elle-même inattendue. En effet, le lactose n'est *a priori* pas une source de carbone trouvée habituellement dans l'environnement de champignons mycoparasites et saprophytes, puisqu'il n'est naturellement présent que dans le lait des mammifères. Ainsi, il est probable que cette β -galactosidase impliquée dans l'hydrolyse initiale du lactose ait une autre fonction dans le métabolisme habituel du champignon. Des études récentes ont cherché à déterminer ce rôle. Le gène *bga1* semble être induit par le D-galactose et le L-arabinose (Seiboth *et al.*, 2005), indiquant un rôle possible de l'enzyme dans la dégradation d'oligo- ou de polysaccharides présents dans la rhizosphère, l'habitat naturel de *T. reesei*. Cette hypothèse a été confirmée par une étude cinétique de la protéine BGA1, qui a prouvé que le lactose n'était pas le substrat naturel de l'enzyme, et que cette dernière était capable d'agir sur différents polymères de galactose végétaux (β -1,4-galactane et β -1,3- / β -1,6-arabinogalactanes notamment) (Gamauf *et al.*, 2007). *Trichoderma reesei* a donc probablement développé des compétences à utiliser les arabinogalactanes provenant des polysaccharides produits par les racines de plantes de la rhizosphère.

2.3.1.2.2 Le galactose est-il un inducteur en soi ?

Le lactose étant hydrolysé dans le milieu extracellulaire en D-galactose et en D-glucose, il serait attendu que ces monomères soient eux-mêmes capables d'induire la production de cellulases. Différentes données de la littérature prouvent que ce n'est en fait pas le cas. Dans des conditions normales de croissance, ni le D-glucose ni le D-galactose ni un mélange des deux monosaccharides n'induisent la transcription des gènes de cellulases (Seiboth *et al.*, 2004). Ce n'est que lorsqu'un faible taux de croissance est imposé aux cellules (par des cultures continues en présence d'une source de carbone limitée) qu'une induction de l'expression des gènes de cellulases est observée en présence de D-galactose ou d'un mélange de D-galactose et de D-glucose (Karaffa *et al.*, 2006). Néanmoins, l'expression des cellulases dans ces conditions est bien plus faible que lors d'une culture en lactose au même taux de croissance ; le lactose reste donc un meilleur inducteur que le D-galactose ou qu'un mélange de D-galactose et de D-glucose.

2.3.1.2.3 L'importance de la stéréospécificité du galactose dans l'induction

La différence entre un mélange commercial de D-galactose et de D-glucose et le D-glucose et le D-galactose issus de l'action de la β -galactosidase réside dans les propriétés de stéréospécificité des sucres. En effet, alors que le galactose synthétique est composé d'un mélange d'anomères α et β , le D-galactopyranose libéré par BGA1 à partir du lactose est sous forme β (à l'inverse du résidu D-glucose qui peut, lui, être un anomère α ou β) (Kubicek *et al.*, 2009). Ce β -D-galactose peut être converti en anomère α par mutarotation chimique, mais ce procédé étant relativement long, de nombreux organismes ont développé des enzymes mutarotases (encore appelées aldose-1-épipiméras, EC 5.1.3.2) qui augmentent le taux de mutarotation.

Le génome de *T. reesei* contient trois gènes d'aldose-1-épipiméras (*aep1-3*) codant pour deux protéines intracellulaires (AEP1-2) et une protéine extracellulaire (AEP3). Cependant, aucun de ces gènes n'est exprimé au cours d'une croissance en lactose et aucune activité mutarotase n'est détectée dans ces conditions (Fekete *et al.*, 2008). De plus, une étude a montré que lorsque le domaine C-terminal de l'aldose-1-épipimérase Gal10p, responsable de l'activité de mutarotation du D-galactopyranose chez *S. cerevisiae*, était surexprimé dans des cellules de *T. reesei* cultivées en lactose, la transcription des gènes de cellulases était diminuée (Fekete *et al.*, 2008). L'ensemble de ces résultats suggère donc que l'anomère β du D-galactose constitue un intermédiaire important dans l'induction des cellulases par le lactose.

Ce constat était inattendu puisque l'assimilation du D-galactose s'effectue chez la plupart des microorganismes presque exclusivement *via* la voie de Leloir (voir ci-après), initiée par une enzyme qui n'a d'affinité que pour la forme α du monomère (Seiboth *et al.*, 2007b).

2.3.1.2.4 La voie de Leloir et la voie alternative

La voie de Leloir assure la conversion du D-galactose en D-glucose-6-phosphate, intermédiaire métabolique pouvant entrer dans la glycolyse. Le D-galactose est initialement converti en D-galactose-1-phosphate par une réaction ATP-dépendante catalysée par la galactokinase (GAL1; EC 2.7.1.6). L'enzyme suivante, la D-galactose-1-phosphate uridylyltransferase (GAL7; EC 2.7.7.12) réalise le transfert d'un groupe uridine-monophosphate (UMP) d'un UDP-glucose au D-galactose-1-phosphate, générant ainsi un UDP-galactose et un D-glucose-1-phosphate. Ce dernier sera converti en D-glucose-6-phosphate par l'action d'une phosphoglucomutase (PGM1, EC 2.7.5.1), tandis que l'UDP-galactose est recyclé en UDP-glucose nécessaire à la réaction suivante grâce à une UDP-galactose-4-épipimérase (GAL10; EC 5.1.3.2) (Holden *et al.*, 2003) (Figure 4). La galactokinase GAL1 n'est capable de phosphoryler que le carbone C1 d'un α -D-galactose. Les D-

galactose mutarotases jouent donc un rôle clé, puisqu'elles épimérisent habituellement le β -D-galactose en début de voie, ce qui permet le déroulement des réactions suivantes. Ainsi, chez *Escherichia coli*, la délétion d'un gène codant pour une mutarotase entraîne une chute du taux de croissance observé en lactose (Bouffard *et al.*, 1994).

L'absence d'activité mutarotase en lactose chez *T. reesei* suscite donc de nombreuses interrogations, puisqu'il semble peu probable que le champignon puisse se contenter uniquement des α -D-galactoses issus d'une mutarotation chimique spontanée pour maintenir sa croissance. Différentes études cherchant à déterminer par quelle voie biochimique *T. reesei* est capable de métaboliser le β -D-galactose ont donc permis de décrire une « voie alternative ». Cette voie est initiée par la réduction du D-galactose en galactitol par la D-xylose réductase (ou « aldose réductase ») XYL1 (EC 1.1.1.21), dépendante du nicotinamide adénine dinucléotide phosphate (NADP) (Seiboth *et al.*, 2007b). Deux autres enzymes, habituellement impliquées – comme *xyl1* – dans le métabolisme du D-xylose et du L-arabinose, participent également à cette voie alternative de dégradation du D-galactose : la L-arabinitol déshydrogénase LAD1 (EC 1.1.1.12) et la xylitol déshydrogénase XDH1 (EC 1.1.1.9) (Seiboth *et al.*, 2004) (Figure 4). Certaines étapes de cette voie restent cependant encore mal décrites.

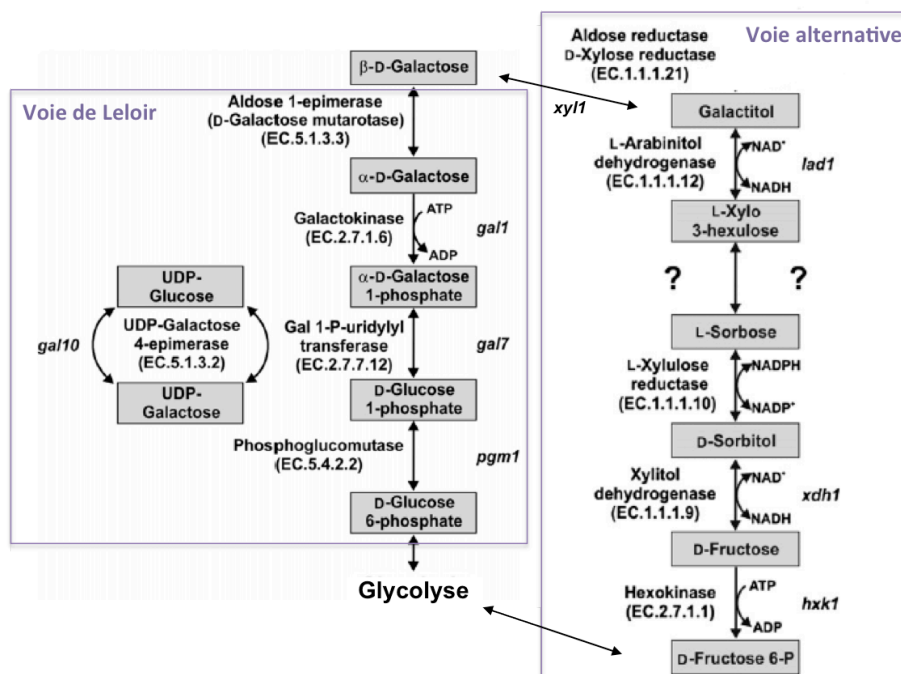


Figure 4. Voies de catabolisme du galactose chez *T. reesei*. Le β -D-galactose est transformé en produit capable de rejoindre la glycolyse *via* la voie de Leloir (à gauche) ou la voie alternative (à droite). Le nom des différentes enzymes impliquées est indiqué s'il est connu. Adapté de Seiboth *et al.*, 2007b.

2.3.1.2.5 Vers un décryptage plus précis des mécanismes d'induction

L'anomère β du D-galactose joue un rôle majeur dans l'induction des cellulases par le lactose. Cependant, le mécanisme exact de cette régulation est probablement complexe. En effet, si le D-galactose issu du lactose doit nécessairement être catabolisé par la voie alternative mentionnée ci-dessus, il a été montré que la délétion du gène *gal1*, codant pour la galactokinase de la voie de Leloir, induisait une forte diminution de la formation de cellulases lors d'une croissance en lactose (Seiboth *et al.*, 2004). Cette observation, *a priori* contradictoire avec les déductions précédentes, a été confirmée par une étude effectuée sur des souches $\Delta gal1$ retransformées par un gène d'*E.coli* codant pour une galactokinase active et ne possédant aucune fonction de régulation transcriptionnelle. Dans ces souches, la formation des cellulases est restaurée, ce qui n'est pas le cas de souches retransformées avec une version inactive de la galactokinase (Hartl *et al.*, 2007). L'activité galactokinase elle-même semble donc importante pour l'induction des cellulases. Étonnamment, l'inactivation de l'étape biochimique suivante de la voie de Leloir, au cours de laquelle GAL7 transforme le D-galactose-1-phosphate en UDP-galactose semble, elle, ne pas avoir d'effet sur l'induction (Seiboth *et al.*, 2002).

Une inactivation de *xy11* (codant pour l'adose réductase initiant la voie alternative) entraîne également une diminution de l'expression des gènes de cellulases en lactose, mais cette diminution est moins forte que dans le cas de l'inactivation de *gal1* (Seiboth *et al.*, 2007a). Il a également été montré qu'une double délétion *xy11/gal1* n'engendre pas d'effet cumulatif mais donne lieu aux mêmes niveaux d'expression que l'inactivation *gal1* (Seiboth *et al.*, 2007a).

Si le lactose constitue, au même titre que les disaccharides issus de la dégradation de la cellulose, un inducteur efficace de la production de cellulases, les mécanismes par lesquels il engendre cette production ne sont donc à ce jour que partiellement caractérisés et nécessitent de nouvelles investigations.

2.3.2 La régulation transcriptionnelle des gènes de cellulases

La présence obligatoire d'un inducteur pour initier l'expression des gènes de cellulases implique une régulation fine des promoteurs de ces gènes. À ce jour, différents activateurs et répresseurs transcriptionnels ont été décrits comme étant impliqués dans cette régulation.

2.3.2.1 Les facteurs activateurs de la production de cellulases

2.3.2.1.1 XYR1 est un activateur majeur

Le gène *xyr1* (xylanase regulator 1) du génome de *T. reesei* code pour un facteur de transcription à doigts de zinc de 102 kDa (934 acides aminés) capable de se fixer sur des motifs GGCTAA disposés en répétitions inversées (Rauscher *et al.*, 2006). Ce facteur de transcription est considéré comme l'activateur principal de l'expression des gènes de cellulases et d'hémicellulases. Le rôle essentiel de XYR1 dans le processus d'induction a été établi par une étude démontrant qu'une délétion du gène *xyr1* éliminait l'expression des gènes de cellulases (*cbh1*, *cbh2* et *egl1*) dans des souches cultivées en présence de cellulose ou de sophorose (Stricker *et al.*, 2006). *Xyr1* est également essentiel pour la formation des transcrits des hémicellulases principales *xyn1* et *xyn2* en présence de xylobiose ou de xylanes et de la β -glucosidase *bgl1* en présence de sophorose. Il est enfin indispensable à l'expression de l'aldose réductase *xyl1* en présence de D-xylose (Stricker *et al.*, 2006).

Lorsque le lactose est utilisé en guise d'inducteur, *xyr1* joue un rôle tout aussi fondamental. Aucun transcrit *cbh1*, *cbh2* ou *xyn1* n'est en effet détecté dans des souches Δ *xyr1* cultivées en présence du disaccharide (Stricker *et al.*, 2007). L'expression de *xyl1* est également affectée dans ces conditions, ce qui explique une croissance ralentie des souches, qui ne peuvent plus cataboliser le β -D-galactose *via* la voie alternative et doivent donc se contenter du métabolisme des α -D-galactoses issus d'une mutarotation chimique spontanée (Stricker *et al.*, 2007). L'absence d'activité mutarotase engendre également une diminution de l'expression du gène *bga1*, car ce dernier est induit par le galactitol (Fekete *et al.*, 2007). L'ensemble de ces résultats est cohérent avec les données *in silico* prouvant que les promoteurs *cbh1*, *cbh2*, *egl1* et *xyn1* portent respectivement 6, 3, 1 et 1 séquences GGCTAA (Aro *et al.*, 2001), contrairement à celui de *bga1*, qui n'en présente pas (Fekete *et al.*, 2007).

Certaines expériences menées sur des souches Δ *xyr1* retransformées par le gène *xyr1* sous le contrôle d'un promoteur *nag1* de la N-acétyl- β -D-glucosaminidase de *Hypocrea atroviridis* semblent suggérer qu'une augmentation de l'expression constitutive de *xyr1* pourrait accroître la formation des cellulases et hémicellulases (Mach-Aigner *et al.*, 2008). Des analyses complémentaires, mettant en jeu des promoteurs exprimés plus fortement que *nag1*, restent néanmoins indispensables pour clarifier la question.

2.3.2.1.2 ACE2 joue le rôle d'activateur secondaire

La protéine à doigts de zinc ACE2 (activator of cellulases 2) a été également caractérisée comme activatrice des gènes de cellulases. En effet, lors d'une croissance sur un milieu enrichi en

cellulose, la délétion du gène *ace2* conduit à une diminution de la cinétique d'induction des ARNm de cellulases et à une activité enzymatique de 30 à 70% moins grande (Aro *et al.*, 2001; Stricker *et al.*, 2008a). Étonnamment, l'induction des cellulases par le sophorose n'est en revanche pas affectée par une telle délétion (Aro *et al.*, 2001). Des expériences *in vitro* ont révélé que le domaine de liaison à l'ADN d'ACE2 était capable, de se fixer à la séquence 5'-GGCTAATAA présente dans les promoteurs de *cbh1* et *cbh2* (Aro *et al.*, 2001). Une étude initialement menée sur l'implication de la protéine dans la régulation de *xyn2* suggère que la liaison de ACE2 à son promoteur cible nécessite une phosphorylation et une dimérisation (Stricker *et al.*, 2008a).

2.3.2.2 Les facteurs répresseurs de la production de cellulases

2.3.2.2.1 ACE1 est un régulateur négatif

Le gène *ace1* code pour un facteur de transcription à doigts de zinc de type Cys₂His₂ capable de se fixer sur huit sites contenant la séquence 5'-AGGCA du promoteur de *cbh1* (Saloheimo *et al.*, 2000). Contrairement au rôle initialement attribué à la protéine ACE1 (activator of cellulases 1), Aro *et al.* ont montré que la délétion du gène *ace1* provoquait une augmentation de l'expression des gènes de cellulases (*cbh1*, *cbh2*, *egl1* et *egl2*) et d'hémicellulases (*xyn1* et *xyn2*) dans des cellules cultivées en présence de sophorose ou de cellulose (Aro *et al.*, 2003). ACE1 est donc un répresseur des gènes de cellulases et de xylanases. La même étude montre par ailleurs que les niveaux d'expression des cellulases et des xylanases d'une souche portant une double délétion des gènes *ace1* et *ace2* étaient similaires à ceux d'une souche $\Delta ace1$.

2.3.2.2.2 La répression catabolique

Chez *T. reesei* et chez les autres espèces de champignons cellulolytiques, une grande majorité de gènes de cellulases n'est pas exprimée lorsque les souches sont cultivées en présence de glucose. Cela s'explique d'une part parce que le glucose peut interférer dans l'absorption de l'inducteur par les cellules ; la présence de glucose inhibe par exemple le fonctionnement de la perméase spécifique du sophorose ou du cellobiose (Kubicek *et al.*, 1993). D'autre part, parce que sa présence dans l'environnement de *T. reesei* déclenche un phénomène de répression catabolique.

La répression catabolique est un processus permettant à de nombreux microorganismes d'inhiber la synthèse d'enzymes impliquées dans l'utilisation de sources de carbone complexes lorsque des sources de carbone plus facilement assimilables sont présentes dans le milieu. Ainsi, chez les organismes cellulolytiques, la présence de glucose entraîne une inhibition de la synthèse des cellulases, puisqu'un catabolisme à fort rendement en ATP est possible sans leur intervention (Gancedo, 1998; Gorke and Stulke, 2008; Ronne, 1995).

Chez *T. reesei*, l'acteur principal de cette répression par le glucose est le facteur de transcription à doigts de zinc de type Cys₂His₂ nommé CRE1. Il s'agit de l'homologue de la protéine CreA d'*Aspergillus nidulans* et d'*Aspergillus niger*, caractérisée comme étant le médiateur de la répression catabolique chez ces deux espèces (Arst *et al.*, 1990; Ruijter and Visser, 1997). Des études pionnières menées sur le mutant fortement producteur RUT C30 de *T. reesei* prouvent que la majeure partie de la séquence du gène *cre1* a été perdue par son génome au cours de la mutagenèse, et que la réintroduction d'un gène *cre1* sauvage dans cette souche naturellement dérégulée permettait de rétablir une répression de *cbh1* en présence de glucose (Ilmen *et al.*, 1996). De plus, les régions à doigts de zinc de CRE1/CreA ont de grandes similarités avec les doigts de zinc des protéines Mig1/Mig2/Mig3, responsables de la répression catabolique chez *Saccharomyces cerevisiae* (Dowzer and Kelly, 1991; Westholm *et al.*, 2008).

CRE1/CreA se fixe sur les promoteurs de ses cibles *via* un motif consensus 5'-SYGGRG-3', dont la fonctionnalité a notamment été montrée *in vivo* dans les promoteurs de *cbh1* et *xyn1* (Panozzo *et al.*, 1998; Strauss *et al.*, 1995; Takashima *et al.*, 1996). En effet, des mutations dans les séquences consensus de fixation de CRE1 donnent lieu à une expression constitutive de *cbh1* et de *xyn1* en présence de D-glucose. Le plus souvent, les sites de liaison CRE1/CreA fonctionnels consistent en fait en deux motifs 5'-SYGGRG rapprochés, et certaines études suggèrent que cette répétition est indispensable pour que le régulateur accomplisse ses fonctions de répression (Cubero and Scazzocchio, 1994; Mach *et al.*, 1996).

La fixation de CRE1 sur ses sites de liaison requiert la phosphorylation d'un résidu sérine du facteur de transcription par une caséine kinase de classe 2 encore non identifiée (Cziferszky *et al.*, 2002). Le principal candidat à une telle fonction était la kinase SNF1 ; cette protéine est en effet présente chez *T. reesei* et a pour fonction de phosphoryler l'équivalent de CRE1 (Mig1p) chez *S. cerevisiae*. Une étude montre cependant qu'elle n'est pas impliquée dans la phosphorylation du répresseur chez *T. reesei* (Cziferszky *et al.*, 2003).

La manière dont la présence de glucose induit la répression catabolique reste encore peu décrite chez les champignons filamenteux. Chez *S. cerevisiae*, deux hexokinases et une glucokinase sont impliquées dans la détection du glucose. L'hexokinase Hxk2p est l'actrice principale de la répression médiée par Mig1p (Rolland *et al.*, 2002; Trumbly, 1992). Le mécanisme par lequel Hxk2p contribue à la répression n'a pas encore été élucidé, mais il ne semble pas impliquer l'activité catalytique de l'enzyme. La transmission du signal dépendrait donc de changements conformationnels de Hxk2p suite à sa liaison au substrat, ou même d'un rôle régulateur de l'enzyme, qui pourrait directement intervenir dans le noyau. *T. reesei*, comme *A. nidulans*, ne possède qu'une hexokinase (HXK1) et une glucokinase (GLK1) et la répression catabolique par le D-glucose est

maintenue si l'une des deux enzymes est mutée. Les souches doubles mutantes pour les gènes codant pour ces kinases sont en revanche dérèprimées (Hartl L, 2009).

Il convient de noter que les travaux mentionnés dans ce chapitre montrent systématiquement que, chez *T. reesei*, la répression catabolique agit surtout sur l'expression constitutive des gènes de cellulases et de xylanases. Une levée de la répression augmente donc l'expression de ces gènes mais n'est pas suffisante pour induire une hyperproduction. Ainsi, la présence d'un inducteur reste indispensable à une synthèse et une sécrétion élevée d'enzymes, même chez des souches dérèprimées (par mutation du gène *cre1* par exemple).

2.3.2.3 Régulation intégrée des cellulases

L'expression des gènes de cellulases dépend donc de différents régulateurs, qui doivent agir en étroite coopération. Parmi eux, *xyr1* semble jouer un rôle majeur, puisqu'il est impliqué dans l'activation des gènes de cellulases quelque soit le type d'inducteur. Il est donc fondamental de déterminer quels sont les signaux influençant la transcription de *xyr1* lui-même. Lors d'une croissance en xylose ou xylobiose, les niveaux de transcription de *xyr1* sont plus élevés dans une souche $\Delta ace1$ que dans une souche sauvage (Mach-Aigner *et al.*, 2008). L'effet inhibiteur d'ACE1 sur la transcription des gènes de cellulases pourrait donc être la conséquence de la répression de *xyr1* par ACE1. Plus étonnamment, une élévation des niveaux de transcription de *xyr1* semble également être détectée chez une souche $\Delta ace2$ cultivée en xylobiose ou en sophorose (Stricker *et al.*, 2008b). La protéine ACE2 étant décrite comme un activateur des cellulases, la raison pour laquelle elle exercerait un effet répresseur sur un autre activateur reste difficilement explicable.

En 2008, Mach-Aigner *et al.* ont mené une étude destinée à déterminer si l'expression de *xyr1* était également dépendante de l'inducteur ou de la source de carbone utilisée. Les auteurs ont mesuré les niveaux d'expression de *xyr1* dans des souches mises en présence de glucose, de D-xylose ou de xylobiose après une préculture en glycérol. Ces niveaux étaient identiques en D-xylose et en xylobiose, mais très diminués en glucose. De plus, les niveaux en D-xylose et xylobiose étaient équivalents à ceux détectés dans des cellules transférées dans un milieu sans source de carbone après la préculture (Mach-Aigner *et al.*, 2008). Ainsi, la transcription de *xyr1* semble être influencée par la répression catabolique, mais pas par un mécanisme d'induction. Stricker *et al.* ont également constaté que *xyr1* était plus exprimé en lactose qu'en glucose. Les auteurs expliquent eux aussi ce phénomène par un effet de la levée de la répression catabolique et non d'une induction (Stricker *et al.*, 2007). L'identification de dix sites de liaison CRE1 dans une région de 1 400 pb en amont de la séquence de *xyr1* confirme que ce dernier est probablement une cible du médiateur de la répression catabolique (Mach-Aigner *et al.*, 2008).

La Figure 5 propose une synthèse de l'ensemble des données de régulation présentées dans ce chapitre.

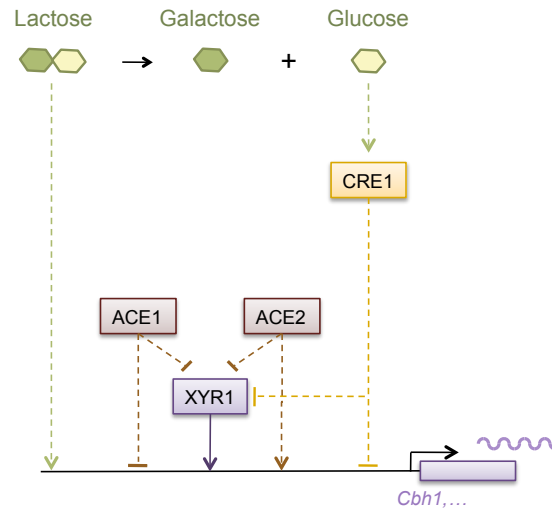


Figure 5. Le réseau de régulation de l'expression des gènes de cellulases chez *T. reesei*. En présence de lactose, l'expression des gènes de cellulases est activée. Celle-ci est régulée par le facteur de transcription XYR1, activateur majeur, et les protéines ACE1 (régulateur négatif) et ACE2 (régulateur positif). L'influence de ACE1 et ACE2 sur la transcription de *xyl1* est également représentée d'après les données disponibles dans la littérature. La présence de glucose initie le processus de répression catabolique qui inhibe l'expression de *xyl1* et des gènes de cellulases. Ce processus est orchestré par la protéine CRE1. Les lignes pointillées indiquent une action des régulateurs sur leurs cibles nécessitant d'éventuels intermédiaires non caractérisés.

2.3.2.4 Vers de nouveaux mécanismes régulateurs ?

Certaines études suggèrent que d'autres processus cellulaires pourraient également influencer l'expression des gènes de cellulases. *Ace1* a par exemple récemment été identifié comme étant un orthologue du gène *stzA* d'*A. nidulans* codant pour un facteur de réponse au stress (Chilton *et al.*, 2008). Or les promoteurs des gènes *cpcA* d'*A. nidulans* et d'*Aspergillus fumigatus*, codant pour un facteur de transcription régulant le métabolisme des acides aminés (cross pathway control regulator of amino acid biosynthesis), présentent sept sites potentiels de liaison à ACE1/StzA, dont six ont une position très conservée au sein des génomes. De plus, il existe réciproquement des sites potentiels de liaison à CPCA dans les promoteurs de *stzA* et *ace1* de nombreuses espèces (Chilton *et al.*, 2008). Il pourrait donc exister un lien entre la disponibilité intracellulaire d'acides aminés et l'expression des gènes de cellulases. Cette supposition a été récemment appuyée par une étude révélant qu'une addition de méthionine pouvait augmenter les niveaux de transcription de certaines cellulases (Gremel *et al.*, 2008).

La répression catabolique pourrait également être couplée à d'autres processus cellulaires fondamentaux. En plus de *creA* (l'orthologue de *cre1*), *A nidulans* possède en effet trois autres gènes participant à la répression catabolique : *creB*, *creC* et *creD*. *CreB* code pour une enzyme de désubiquitination (Lockington and Kelly, 2001) et *creC* pour une protéine contenant cinq motifs WD40 et une région riche en proline (domaines habituellement mis en jeu dans des interactions protéines-protéines) (Todd *et al.*, 2000). *CreB* et *CreC* formeraient un complexe permettant d'éliminer des ubiquitines de *CreA* ou d'autres protéines, stabilisant ainsi ces substrats en fonction de l'existence ou non d'une répression (Lockington and Kelly, 2002). Des mutations de ces gènes engendrent une utilisation perturbée des sources de carbone, et une mutation de *creD* supprime les effets de ces mutations (Boase and Kelly, 2004). Ainsi, il semble exister un lien entre des mécanismes d'ubiquitination et la régulation des gènes de cellulases. Des orthologues de *creB*, *creC* et *creD* sont présents dans le génome de *T. reesei* (*cre2*, *cre3* et *cre4*, respectivement). Leurs effets ne sont pas caractérisés à ce jour, mais de tels mécanismes pourraient donc également s'appliquer à la répression catabolique de *T. reesei*.

Des comparaisons récentes des données issues du reséquençage de souches plus ou moins productrices confirment l'importance de mécanismes cellulaires tels que le transport nucléocytoplasmique ou la machinerie de dégradation des ARNs dans la production de cellulases par *T. reesei* (Le Crom *et al.*, 2009). Des expériences de mutations ponctuelles ont par ailleurs montré que l'expression de *cbh2* nécessitait la fixation du complexe protéique HAP2/3/5 sur une séquence « CCAAT box » présente dans son promoteur (Zeilinger *et al.*, 2001; Zeilinger *et al.*, 1998). L'analyse des fonctions des orthologues des protéines HAP présents chez d'autres espèces suggère que ce complexe pourrait jouer un rôle dans le remodelage de la chromatine, permettant ainsi la mise en place d'une structure décondensée propice à une activation transcriptionnelle (Li *et al.*, 1998; Linhoff *et al.*, 1997; Zeilinger *et al.*, 2003). Si certains gènes sont effectivement co-régulés du fait de leur proximité physique au sein du génome de *T. reesei* (voir partie 2.2.5.2), un tel complexe pourrait donc avoir une importance majeure dans le contrôle de l'expression génétique.

2.4 Objectifs du projet

L'ensemble des données de la littérature, résumé ici, rend compte de la grande diversité des mécanismes moléculaires de régulation d'un processus aussi complexe que la production de cellulases. Les études menées ces vingt dernières années ont permis des avancées notoires dans la description de ces mécanismes, grâce notamment à la détermination précise des effets des inducteurs saccharidiques ou à l'identification des différentes protéines régulatrices présentées ci-dessus.

Néanmoins, notre connaissance de ces mécanismes reste encore très incomplète sur de nombreux points. L'exploration minutieuse des réseaux de régulations transcriptionnelles apparaît notamment indispensable, puisque ces processus semblent jouer un rôle majeur dans la synthèse et la sécrétion des enzymes. Le séquençage du génome de *T. reesei* permet de surcroît d'envisager de nouvelles approches pour l'élucidation de ces processus.

C'est dans ce cadre que s'inscrit mon projet de thèse, qui s'attache à exploiter les outils de la génétique moléculaire classique et les capacités de techniques de génomique fonctionnelle à haut débit pour améliorer notre connaissance de *T. reesei* et de ses voies de production de cellulases.

Trois approches différentes ont été adoptées. Afin de préciser le comportement de facteurs de transcription connus comme étant impliqués dans la production d'enzymes cellulolytiques, des études d'expression génétique classique ont été menées dans un premier temps. Celles-ci, principalement basées sur des analyses par PCR quantitatives (PCRq), permettent notamment de décrire l'effet du lactose, utilisé dans notre système comme inducteur des gènes de cellulases, sur les régulateurs XYR1, ACE1 et ACE2. Des études transcriptomiques sur puces à ADN nous ont ensuite permis d'étudier en détail le processus de répression catabolique. Ce phénomène assurant une inhibition efficace de la synthèse d'enzymes cellulolytiques en présence de certains types de source de carbone, la description des gènes régulés par CRE1, médiateur majeur du processus, apparaît en effet comme un élément clé pour la connaissance des mécanismes impactant la production de cellulases. Enfin, une analyse de l'expression génétique globale au cours de cinétiques d'induction de différents mutants plus ou moins producteurs de *T. reesei* a été entreprise. Le comportement d'un grand nombre de gènes régulés au cours de la production de cellulases a ainsi été décrit grâce à des études sur puces à ADN et de séquençage à haut débit des transcrits (RNA-seq). Ces résultats ont également été confrontés aux données de reséquençage du génome des mutants d'intérêt et à leurs performances en bioréacteur afin d'identifier des mécanismes liés à l'acquisition des phénotypes d'hyperproduction de certaines souches.

3 MATÉRIEL ET MÉTHODES

Afin de comprendre la réponse génétique de *Trichoderma reesei* lors de la production de cellulases, des cultures de cellules dans des conditions induisant (ou non) la synthèse et la sécrétion de ces protéines sont réalisées. À différents instants des cultures, des échantillons de biomasse sont prélevés. Le matériel mycélien ainsi collecté est immédiatement broyé en une fine poudre, à partir de laquelle les ARN cellulaires sont extraits. Le contenu des échantillons de matériel génétique est ensuite analysé par PCR quantitative au moyen de sondes permettant la détection de différents transcrits d'intérêt, ou par études transcriptomiques, au moyen de puces à ADN dont les sondes couvrent l'ensemble du génome de *T. reesei*.

3.1 Matériel biologique

Les souches de *T. reesei* employées dans le cadre de ce projet sont répertoriées dans le Tableau 3. Des cultures de souches ont été réalisées sur milieu gélosé Potato Dextrose Agar (Difco Laboratories USA) à 30°C. Après sporulation, leurs spores ont été resuspendues dans une solution stérile de NaCl (9 g.L⁻¹) additionnée de 20% de glycérol, et conservées à -80°C.

Nom/Génotype	Référence
QM6a	Reese (1976)
QM 9414	Mandels et Andeotti (1978)
QM 9414 $\Delta gal1$	Seiboth <i>et al.</i> (2004)
QM 9414 $\Delta xy11$	Stricker <i>et al.</i> (2007)
QM 9414 $\Delta gal1\Delta xy11$	Seiboth <i>et al.</i> (2007)
QM 9414 $\Delta cre1$	Portnoy <i>et al.</i> (2011b)
NG14	Eveleigh et Montenecourt (1979)
RUT C30	Eveleigh et Montenecourt (1979)
CL847	Durand <i>et al.</i> (1988)

Tableau 3. Souches de *Trichoderma reesei* utilisées au cours de l'étude.

3.2 Conditions de cultures

3.2.1 Cultures des souches pour les études cinétiques

3.2.1.1 Précultures

Des précultures sont réalisées à partir des spores dans une fiole de Fernbach contenant 250 mL de milieu de culture stérile (glucose 30 g.L⁻¹, cornsteep 2 g.L⁻¹; (NH₄)₂SO₄ 1,4 g.L⁻¹; KOH 0,8 g.L⁻¹; H₃PO₄ 85% 4 mL.L⁻¹; phtalate de potassium 5 g.L⁻¹; MgSO₄.7H₂O 0,3 g.L⁻¹; CaCl₂ 0,3 g.L⁻¹; FeSO₄.7H₂O 5,0 mg.L⁻¹; MnSO₄.H₂O 1,6 mg.L⁻¹; ZnSO₄.7H₂O 1,4 mg.L⁻¹; CoCl₂.6H₂O 2,0 mg.L⁻¹) et placé à 30°C sous agitation orbitale à 150 rpm pendant 72 h.

3.2.1.2 Cultures en bioréacteur

Le protocole de production de cellulases en bioréacteur, établi à IFP Énergies nouvelles, est articulé autour de deux phases principales : la première phase (batch) correspond à une phase de croissance des cellules en présence de glucose ; lorsque la majorité du glucose a été consommée, la souche est alimentée par un apport à débit constant de lactose (fed-batch), induisant la production de cellulases par les cellules. Des cultures contrôles sont réalisées dans les mêmes conditions en remplaçant le lactose par du glucose ; aucune production de cellulases n'est alors enregistrée.

La culture est initiée en inoculant le contenu de la fiole de préculture dans 1,75 L de milieu de culture stérile à 30 g.L⁻¹ de glucose placé dans un bioréacteur de 5 L. Le contenu du bioréacteur est maintenu à un pH de 4,8 par ajout d'ammoniac (NH₃ 5,5 N), et à une température de 28°C. L'aération est de 0,5 vvm (volume de gaz par volume de liquide par minute) et l'agitation est initialement fixée à 500 rpm (rotation par minute) ; cette valeur est augmentée au cours du temps de manière à maintenir la pression en oxygène dissout (pO₂) dans le bioréacteur au dessus de 40%.

Lorsque la concentration en glucose devient inférieure à 20% de la concentration initiale, le fed-batch est initié par l'injection à un débit de 0,98 g.h⁻¹ d'une solution de lactose (ou de glucose) à 250 g.L⁻¹. Le pH du milieu est alors ajusté à 4,0 et la température progressivement ramenée à 26°C. Afin d'éviter la formation de mousse, quelques gouttes de Symparonic PEL 61 stérile (BASF) sont injectées chaque jour. La sortie de gaz est refroidie par un condenseur connecté à une arrivée d'eau froide extérieure à 4°C pour limiter les pertes d'H₂O.

L'évolution du pH, de la température, de la pO₂ et du débit d'aération est suivie tout au long de l'expérimentation grâce au logiciel BioXpert NT (version 2.51.109). Les acquisitions en ligne des teneurs en CO₂ et en O₂ sont réalisées au moyen des analyseurs COSMA Beryl 100 et 102, et la balance Satorius CP4202S permet de suivre l'évolution de l'ajout de la solution de base. Des prélèvements sont effectués régulièrement au cours de la culture afin de suivre la concentration en biomasse, en sucres (glucose et lactose) et en protéines extracellulaires.

3.2.2 Cultures continues et cultures en fioles

Les cultures des mutants de délétion de la voie de métabolisme du galactose ont été réalisées à Vienne (Autriche) suivant le protocole décrit par Seiboth *et al.* (Seiboth *et al.*, 2007a). Les cultures destinées à évaluer la capacité du galactose à induire l'expression de *xyl1* ont été réalisées à Vienne (Autriche) dans des flasques de 1 L contenant 250 mL du milieu décrit par Mandels et Andreotti (1978) (Mandels and Andreotti, 1978) additionné de 1% glycérol en guise de source de carbone, pendant 24 h à 30°C et sous agitation de 250 rpm. Les cellules ont ensuite été collectées, lavées dans

un milieu sans source de carbone et transférées dans un milieu frais contenant 4 mM de D-glucose ou de D-galactose. Les échantillons destinés à l'étude ont été collectés après 4 h.

Les cultures continues des souches QM 9414 et QM 9414 $\Delta cre1$ destinées à identifier les gènes soumis à la répression catabolique dépendante ou non de CRE1 ont été réalisées à Debrecen (Hongrie) suivant le protocole décrit par Karaffa *et al.* (Karaffa *et al.*, 2006).

3.3 Méthodes analytiques

3.3.1 Détermination de la concentration en biomasse

Afin d'évaluer la concentration en biomasse, 10 mL du milieu de culture collecté sont filtrés à l'aide d'une pompe à vide sur une membrane GF/C en fibres de verre 1,2 μm (Whatman) préalablement séchée et pesée. Après filtration, les membranes sont séchées à 105 °C dans une étuve Memmert pendant 48 h, puis pesées. La masse sèche est alors déterminée par différence entre la masse avant et celle après filtration de la suspension cellulaire sur la membrane.

3.3.2 Détermination de la concentration en sucres

Au cours de la phase de batch, la concentration en glucose est déterminée par le glucoSTAT Beckman Glucose Analyser 2. Son principe 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, selon la réaction : $\text{Glucose} + 2\text{H}_2\text{O} + \text{O}_2 \rightarrow \text{acide gluconique} + 2\text{H}_2\text{O}_2$. 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. Les estimations ainsi obtenues permettent d'évaluer le moment approprié pour lancer le fed-batch.

La concentration précise des sucres tout au long des cultures est dosée *a posteriori* par chromatographie en phase liquide à haute performance (HPLC) sur une colonne HPX-87P (Biorad) à 85 °C. De l'eau préalablement dégazée à l'hélium est utilisée comme éluant à un débit de 0,4 mL.min⁻¹. Des standards (Beckman Coulter) sont injectés avant et après les échantillons, et l'intégration des chromatogrammes obtenus est réalisée avec le logiciel Chromeleon (version 6.8, Dionex).

3.3.3 Détermination de la concentration en protéines extracellulaires

Les mesures de concentration de protéines sont effectuées suivant la méthode de Bradford, dosage colorimétrique utilisant les propriétés de changement d'absorbance du bleu de Coomassie. Ce composé absorbe aux alentours de 465-470 nm lorsqu'il est libre sous forme cationique, tandis que sa forme anionique (bleue) issue de sa fixation sur les acides aminés aromatiques présents dans les protéines a un maximum d'absorbance à 595 nm. Ce changement d'absorbance étant proportionnel à la quantité de colorant fixée, les concentrations des échantillons sont déterminées par mesure de leur absorbance à 595 nm et comparaison à une gamme standard d'albumine sérique bovine de 0,125 à 1 g.L⁻¹ (Biorad). Les mesures sont effectuées en microcuvettes de 1 mL contenant 20 µL d'échantillon et 1 mL de réactif (Biorad).

3.4 Préparation des échantillons d'ARN

Les échantillons de biomasse destinés à une extraction d'ARN sont filtrés sur membranes de microfibrilles de verre (GF/B, Whatman). Le gâteau de mycélium obtenu est plongé dans un mortier contenant de l'azote liquide, et les parois cellulaires sont cassées mécaniquement par broyage. L'extraction d'ARN s'effectue à partir de la fine poudre de mycélium ainsi obtenue. Deux cents microgrammes de cette poudre sont soumis à un traitement à la TRI Reagent Solution (Applied Biosystem) qui sépare les acides nucléiques (resuspendus dans la phase aqueuse) des protéines (retrouvées dans la phase organique phénolique). Les ARN sont ensuite isolés et précipités par traitement au bromochloropropane et ajout d'isopropanol (Sigma) suivant les instructions du fournisseur. Après un lavage à l'éthanol, les échantillons sont resuspendus dans de l'eau, purifiés suivant la procédure RNeasy (Qiagen), et soumis à une digestion à la DNase sur colonne (RNase-Free DNase Set, Qiagen). L'élution finale est effectuée dans 50 µL d'H₂O. La qualité des ARN ainsi obtenus est contrôlée au Nanodrop ND1000 (Thermo Fisher Scientific).

3.5 Étude des gènes d'intérêt par PCR quantitative

3.5.1 Rétrotranscription

Afin de pouvoir être analysés par PCR quantitative (PCRq), les ARN sont rétrotranscrits en ADN complémentaires (ADNc) au moyen du kit iScript cDNA Synthesis (Biorad). Cinq cent nanogrammes d'ARN sont mélangés à 4 µL de tampon réactionnel 5X (iScript Reaction Mix)

contenant des amorces aléatoires et des oligo(dT), et à 1 μL de reverse transcriptase, pour un volume final de 19 μL complété avec de l'eau. Afin d'éviter toute dégradation du matériel génétique au cours du protocole, 1 μL de RNAsin (Promega), réactif inhibant l'action d'éventuelles RNAses, est ajouté au mélange. La rétrotranscription (RT) est réalisée dans un thermocycleur (iCycler, Biorad) pendant 30 min à 42°C après une phase d'hybridation des amorces de 5 min à 25°C. L'enzyme est ensuite inactivée et les ARN dégradés au cours d'une phase de 5 min à 85°C. Des contrôles négatifs sont systématiquement réalisés suivant la même procédure en remplaçant l'enzyme par de l'eau.

3.5.2 PCR quantitative

3.5.2.1 Principe

La PCR quantitative (PCRq) est une réaction de polymérisation en chaîne suivie en temps réel et permettant de quantifier l'ADN présent dans les échantillons testés. Son principe repose sur l'utilisation du SYBR green, fluorochrome dont la fluorescence, très faible lorsqu'il est en solution, augmente fortement lorsqu'il s'associe à un double brin d'ADN. Un suivi de l'augmentation de la fluorescence des échantillons au fil des cycles d'amplifications permet donc une estimation *a posteriori* de la quantité initiale de matrice.

3.5.2.2 Protocole

Les PCRq sont réalisées dans un thermocycleur iCycler iQ real-time detection system (Biorad). Les milieux réactionnels d'un volume final de 25 μL contenant 1 μL de matrice ADNc (dilution 1/10 du produit de RT), 12,5 μL de tampon réactionnel SYBR Green Supermix 2X (Biorad), 0,8 pmol. μL^{-1} d'amorces sens et anti-sens (Eurofins MWG) et complétés avec de l'eau, sont disposés dans une plaque 96 puits. Le protocole de PCRq comprend 3 min de dénaturation initiale à 95°C suivie de 40 cycles de 10 secondes à 95°C (dénaturation) et 30 secondes à 60°C (hybridation et amplification). Afin de s'assurer la spécificité des produits, une courbe de fusion est réalisée à l'issue de la réaction par augmentation progressive de la température de 60°C à 95°C par paliers de 0,5 °C toutes les 10 secondes.

Les amorces oligonucléotidiques utilisées sont présentées dans le Tableau annexe 1. Les contrôles négatifs de RT permettent de vérifier l'absence d'ADN génomique dans les échantillons, et des réactions sans matrice sont réalisées pour s'assurer que les réactifs de PCR ne sont pas contaminés. Toutes les PCR sont effectuées en triplicats au sein d'une plaque, et chaque condition expérimentale est testée au moins sur deux plaques différentes. Un échantillon identique

« calibrateur » est présent sur chacune des plaques afin de corriger d'éventuelles variations inter-réactions.

3.5.2.3 Analyse

Les données sont analysées en utilisant le logiciel iQ5 Optical System Software (Biorad), qui définit automatiquement le seuil de détection de fluorescence pour chaque réaction. Lorsque la valeur de fluorescence d'un échantillon dépasse ce seuil, le cycle en cours est désigné comme le « cycle seuil » (cycle threshold, C_T) de l'échantillon. Ce C_T est caractéristique de la quantité initiale d'ADNc de l'échantillon. Toutes les valeurs de C_T d'une plaque sont ajustées de manière à éliminer les différences de valeurs de C_T entre les calibrateurs. Afin de pouvoir comparer différents échantillons, les C_T de tous les gènes testés sont normalisés par les C_T issus de l'amplification systématique d'un gène dont l'expression est connue pour être invariante dans les conditions étudiées, ici le gène *gpd1* codant pour une glycéraldéhyde-3-phosphate déshydrogénase constitutive.

Pour étudier le comportement des transcrits d'intérêt au cours des inductions de cellulases, l'analyse des ratios d'expression a été effectuée sur des échantillons collectés juste avant le début du fed-batch (before fed-batch, BFB) ou 1 h, 6 h et 24 h après le début de l'accumulation de protéines extracellulaires pendant le fed-batch. La souche QM 9414 commençant sa production de protéines 18 h après le début du fed-batch (temps nécessaire à la levée de la répression catabolique dans la souche, voir Figure 6), les niveaux d'expression au cours du fed-batch ont été mesurés 1 h, 6 h et 24 h après le début du fed-batch pour les souches QM 9414 $\Delta cre1$, CL847 et RUT C30, et 19h, 24h et 42 h après le début du fed-batch pour la souche parentale QM 9414. Les moyennes et les écarts types des réplicats ont été calculés pour chaque condition testée. Pour chaque transcrit, le plus faible ratio mesuré sur l'ensemble des souches et des conditions a été ramené à une valeur de 1 et les autres ratios ajustés en conséquence afin d'assurer une meilleure lisibilité des résultats. Des tests *t* de Student ont été réalisés sur l'ensemble des jeux de données. Toutes les différences mentionnées dans le texte sont conformes à une p-value inférieure à 0,05.

3.6 Études transcriptomiques

3.6.1 Les puces à ADN : principe

L'accès au transcriptome d'un organisme est possible entre autre grâce à l'utilisation de puces à ADN, lames de verre sur lesquelles ont été fixées individuellement de courtes séquences nucléiques simple brin nommées « sondes ». Chaque sonde a été dessinée de façon à être spécifique

d'un seul et unique gène de l'organisme étudié. Plusieurs dizaines de milliers de spots sont ainsi fixés dans un ordre bien défini sur la lame de verre de quelques cm². Lorsque cette lame est mise en présence d'un mélange d'ADNc issus de la rétrotranscription des ARN contenus dans une cellule, les ADNc correspondant aux sondes s'hybrident à celles-ci. L'échantillon de matériel génétique à tester ayant été marqué préalablement par des molécules fluorescentes, chaque spot de la lame est associé, après lecture par un scanner à haute résolution, à une intensité de fluorescence directement proportionnelle à la quantité de molécules fixée sur la sonde. Ainsi, si deux molécules fluorescentes émettant dans des longueurs d'ondes différentes sont utilisées (ici, le Cy3 émettant dans le vert et le Cy5 émettant dans le rouge), une comparaison directe du transcriptome de deux populations est possible. Cette méthode nous permet de comparer l'expression des gènes (i) chez différentes souches soumises ou non à une répression catabolique, ou (ii) à différents moments de la cinétique d'induction de la production des cellulases chez une même souche.

3.6.2 Conception de puces à ADN dédiées à l'étude de *T. reesei*

Afin de pouvoir étudier le transcriptome de *T. reesei* dans différentes souches et différentes conditions, il était nécessaire de créer des puces à ADN dédiées à l'analyse du génome du champignon, car aucun fournisseur ne proposait de tels produits.

3.6.2.1 Choix du dessin des oligonucléotides

Une sélection des oligonucléotides selon le protocole tiling a été retenue pour la conception des puces. Cette solution consiste à choisir les oligonucléotides de manière à couvrir l'ensemble du génome sans tenir compte de son annotation et en répartissant des sondes de manière homogène ; elle permet donc aussi bien des analyses transcriptomiques que des expériences d'hybridation génomique comparative (CGH) sur l'ADN génomique total du champignon. Il a été choisi de concevoir dans un premier temps des puces comprenant 244 000 sondes oligonucléotidiques réparties sur les deux brins. La longueur du génome de *T. reesei* étant de 33,4 Mb, la résolution obtenue est d'une sonde toutes les 140 paires de bases (pb).

3.6.2.2 Conception des sondes

La séquence du génome de *T. reesei* a été téléchargée à partir du site du Department of Energy Joint Genome Institute (JGI) : <http://genome.jgi-psf.org/Trire2/Trire2.home.html>. Les séquences des sondes (60 pb) ont été choisies au moyen du logiciel Teolenn, développé sur la Plateforme transcriptome de la Montagne Sainte Geneviève (Jourden *et al.*, 2010). Ce logiciel permet de sélectionner les meilleures sondes parmi l'ensemble des oligonucleotides possibles le long du

génomique. Cette sélection s'effectue au moyen du calcul par le logiciel de scores estimant leur sensibilité, leur spécificité et tenant compte de l'homogénéité de leur répartition. Pour ces calculs, les différents paramètres considérés ont été pondérés par les coefficients suivants : 0,4 pour le T_m , 0,3 pour l'unicité, 0,2 pour le contenu en GC et 0,1 pour la complexité. L'unicité a été calculée par genome tools avec un MAX_PREFIX_LENGTH défini à 15. Pour le calcul des scores finaux, des coefficients de 0,75 et de 0,25 ont été respectivement assignés aux scores de qualité et de position.

3.6.2.3 Production des puces à ADN

Des puces de 1x244k comportant les 238 589 sondes sélectionnées réparties de manière aléatoire sur la lame ont été commandées auprès du fournisseur Agilent. Ces puces ont été employées pour étudier les gènes soumis à la répression catabolique dépendante ou non de CRE1, pour l'analyse des cinétiques d'induction de la souche RUT C30, ainsi que pour la comparaison de l'expression basale de QM6a et de RUT C30

L'évolution des formats de lames proposés par Agilent nous a conduit à dessiner un second jeu de puces uniquement dédié aux études transcriptomiques. Ces lames de 8x60k comportent l'ensemble des oligonucleotides de la première série localisées dans les séquences codantes répertoriées dans le génome de *T. reesei*. Elles ont été employées pour l'analyse des cinétiques d'induction de la souche NG14.

3.6.3 Protocole expérimental

3.6.3.1 Marquage des cibles

Le marquage des ARN est réalisé par une étape de reverse transcription (RT), au cours de laquelle sont incorporés des groupements amino-allyl dUTP (aa-dUTP) aux brins néoformés, suivie d'un couplage chimique des ADNc ainsi synthétisés aux fluorochromes. Les cyanines utilisées (Cy3 et Cy5) portent en effet des groupements N-hydroxysuccinimides (NHS) qui ont une forte affinité pour les fonctions aminoallyl des nucléotides incorporés. Un microgramme d'ARN total purifié est mélangé à 2 µg d'amorces oligo dT et à 5 µg d'amorces aléatoires (Invitrogen) dans un volume final de 23 µL et dénaturé pendant 10 min à 70°C. Dix sept microlitres de mélange réactionnel de RT contenant 8 µL de tampon Superscript III 5X (5X First Strand Buffer, Invitrogen), 4 µL de DTT (0,1 M, Invitrogen), 2 µL de mélange de aadNTPs (25 µl dATP, dCTP, dGTP à 100 mM, 10 µl dTTP et 15 µl d'aa-dUTP, qsp 1 mL d'H₂O RNase free) et 2 µL de Superscript III Reverse Transcriptase (Invitrogen, 200U/µL) sont ensuite ajoutés aux échantillons. Après une incubation pendant 10 min à 25°C, l'élongation est effectuée pendant 1 h à 50°C. La reverse transcriptase est

ensuite dénaturée pendant 15 min à 70°C, puis une hydrolyse chimique des ARN est réalisée en ajoutant 15 µL de NaOH à 0,1 N. Après 10 min à 70°C, le pH des échantillons est neutralisé par addition de 15 µL de HCl à 0,1 N. Les ADNc sont alors précipités pendant 1 h à -20°C en présence de 2,5 volumes d'éthanol à 96% et d'un dixième de volume d'acétate de sodium 3 M à pH 5,3, puis culotés par centrifugation à 4°C pendant 30 min à 12 000 rpm. Après resuspension dans 40 µL d'H₂O, les ADNc sont purifiés sur colonne suivant la procédure QIAquick (Qiagen), en utilisant un volume de départ de 200 µL de tampon PB additionné de 4 µL d'acétate de sodium 3 M, pH 5,2. À l'issue de la procédure, les ADNc sont élués dans 30 µL d'H₂O préchauffée à 37°C.

Le couplage de ces échantillons aminoallylés avec le Cy3 (ou Cy5)-NHS est alors effectué pendant 1 h à l'obscurité. Pour favoriser la réaction d'estérification, le pH du mélange ADNc et cyanines est préalablement ajusté à 9 par ajout de 3 µL de bicarbonate de sodium à 0,5 M. Six microlitres d'hydroxylamine 4 M sont ensuite ajoutés aux échantillons dans le but de bloquer d'éventuels sites aminoallyls non occupés par les fluorochromes et d'éviter ainsi les couplages croisés lors du mélange des deux échantillons. À l'issue de ce mélange, une nouvelle purification sur colonne est réalisée pour éliminer les fluorochromes non fixés. Afin d'éviter des biais de fluorochromes et d'améliorer ainsi la fiabilité des résultats, ce protocole est systématiquement répété en inversant les fluorochromes pour marquer les mêmes échantillons (dye-swap, réplicats techniques) ou des échantillons différents (dye-switch, réplicats biologiques).

3.6.3.2 Hybridation

Les cibles sont récupérées dans 50 µL (lames 1x244k)/28µL (lames 8x60k) d'H₂O préchauffée à 37°C, et mélangées à 50 µL/7µL d'agent bloquant 10X (10 Hybs-Lyophilised, Agilent), 250 µL/35µL de tampon d'hybridation 2X (GEx HI-RPM, Agilent) et 150 µL/0µL supplémentaires d'H₂O. 490 µL/40µL du mélange sont alors déposés sur la lame. Maintenu par un joint dans une chambre d'hybridation, l'ensemble formé de la lame et du mélange d'échantillons marqués est incubé pendant 17 h à 65°C sous une agitation rotative à 10 rpm assurant une répartition homogène des cibles sur la puce pendant l'hybridation. Les lames sont ensuite lavées dans deux bains successifs de 1 min chacun dans les tampons fournis par Agilent (Gene Expression Wash Buffers 1-2) et conservées à l'abri de la lumière jusqu'à la lecture.

3.6.3.3 Lecture des lames

La lecture des lames de 1x244k est effectuée par le scanner GenePix 4000B de Axon (Molecular Dynamics), à une résolution de 5 µm par pixel. La lecture des lames de 8x60k est réalisée par le scanner Agilent G2565CA, à une résolution de 2 µm par pixel. Ces lectures génèrent deux images en niveaux de gris (16 bits pour le scanner Axon, 20 bits pour le scanner Agilent) qui

correspondent aux deux longueurs d'ondes d'émission des fluorochromes : 532 nm (Cy3) et 635 nm (Cy5). Les niveaux de gris de chaque image sont ensuite convertis en fausses couleurs vert (Cy3) ou rouge (Cy5). La superposition des deux images permet de générer une image TIFF où l'intensité de chaque spot révèle le niveau d'expression du transcrite correspondant. Ainsi, les spots verts ou rouges correspondent respectivement à des gènes fortement exprimés dans l'échantillon marqué en Cy3 ou en Cy5, tandis que les spots jaunes correspondent aux gènes dont les niveaux d'expression sont équivalents dans les deux échantillons. Une fois l'image de la puce obtenue, l'étape suivante consiste à placer une grille qui identifie chaque spot par le nom de l'oligonucléotide correspondant. Pour les lames 1x244k, cette étape s'effectue de manière semi-automatique au moyen du logiciel GenePix Pro (version 6.1), tandis que la procédure est entièrement automatisée avec les lames 8x60k scannées sur l'appareil Agilent. À ce stade, les spots artéfactuels sont identifiés et éliminés. Une fois la grille positionnée, le logiciel calcule pour chaque spot détecté sur la lame (et donc pour chaque sonde) une valeur d'intensité de fluorescence ainsi que le ratio de fluorescence Cy5 sur Cy3, et génère un tableau de résultats qui sert de base au traitement informatique des données.

3.6.4 Traitements bioinformatiques et statistiques

Les données brutes d'intensités mesurées sur les puces à ADN sont normalisées par la méthode Lowess globale (locally weighted polynomial regression) au moyen du programme Goulphar (Lemoine *et al.*, 2006). Cette normalisation par régression non-linéaire permet de corriger des variations d'intensités de fluorescence entre Cy3 et Cy5 qui ne sont pas dues à une réelle expression différentielle des gènes (incorporation différente des deux fluorochromes, mesure de leur émission par des lasers différents du scanner).

L'analyse globale de l'expression des gènes est ensuite réalisée au moyen de scripts spécifiquement développés pour ce projet et permettant le traitement des fichiers de données normalisées. Pour chaque condition expérimentale, les spots présentant une intensité de fluorescence inférieure à la moyenne ajoutée de deux écarts types du bruit de fond sur l'ensemble de la lame sont considérés comme non détectés et éliminés de l'analyse. Un ratio final pour chaque transcrite est obtenu en moyennant les valeurs d'hybridation détectées de l'ensemble des sondes positionnées sur le brin codant et à l'intérieur de la séquence codante correspondante. Les transcrits ne possédant aucune ou une seule sonde considérée comme détectée sont éliminés. Les résultats issus des répliquats biologiques et techniques sont traités parallèlement et indépendamment.

À l'issue de la procédure, l'analyse différentielle est effectuée à l'aide d'une modélisation linéaire (lmFit) et de tests bayésiens (eBayes), implémentés dans le package R limma (Smyth, 2004) pour chaque condition. Les p-values sont ensuite ajustées par la méthode de correction des tests multiples

de Benjamini-Hochberg, et considérées comme significatives quand elles sont inférieures à 0,05. Seuls les transcrits les plus régulés ($|\log_2(\text{ratios d'hybridations})| > 1$ pour les analyses cinétiques et ($|\log_2(\text{ratios d'hybridations})| > 2$ pour l'étude de la répression catabolique) sont conservés.

3.6.5 Analyse des résultats par classification non supervisée

Une matrice d'expression regroupant les \log_2 (ratios d'hybridations) de chaque gène pour chacune des conditions expérimentales testées est réalisée. Pour les analyses du transcriptome au cours de la production de cellulases, les gènes présentant plus de 30% de valeurs manquantes sur l'ensemble des hybridations sont filtrés au moyen du pipeline GEPAS (Herrero *et al.*, 2003) et éliminés. Les gènes significativement régulés sont ensuite groupés par des méthodes de nuées dynamiques à l'aide de la distance euclidienne. Le nombre optimal de groupes pour chaque étude est déterminé en minimisant conjointement le nombre de groupes (k) et la somme des variances intragroupes pour chaque k testé. Les groupes d'expression finaux sont déterminés par intersection des listes obtenues à l'issue de 5 procédures de regroupement répétées. Les gènes ne faisant pas partie de cette intersection (i.e. les gènes qui ne se regroupent pas systématiquement ensemble) sont réassociés entre eux ou réaffectés aux autres groupes en répétant la procédure sur les nouveaux sous-ensembles obtenus. Une classification non supervisée hiérarchique est enfin réalisée pour chacun des groupes d'expression en utilisant la distance euclidienne et la méthode de saut moyen entre éléments. La procédure utilisée pour réaliser les groupes de gènes de l'étude de la répression catabolique est présentée en annexe (Portnoy *et al.*, 2011a). L'ensemble de ces analyses a été réalisé au moyen du logiciel MultiExperiment Viewer (Saeed *et al.*, 2003).

3.6.6 Identification des gènes, prédictions fonctionnelles et calculs d'enrichissement

Une annotation manuelle experte, réalisée par Frédérique Bidard-Michelot, a permis d'identifier l'ensemble des séquences codantes de *T. reesei* répertoriées à l'issue du séquençage de son génome (Martinez *et al.*, 2008). Une exploration des annotations Gene Ontology (GO), Enzyme Commission (EC) et des prédictions Signal Peptide (Signalp) disponibles sur le site du JGI a été systématiquement entreprise afin de caractériser fonctionnellement chaque séquence. Dans le cas de gènes peu ou mal annotés, les séquences en acides aminés correspondantes ont été soumises à un BALST protéine contre protéines (blastp) via la plateforme du NCBI, ou confrontées aux bases de données cfog (<http://embg.igmors.u-psud.fr/fungipath/>) (Grossetete *et al.*, 2010) ou superfam (<http://supfam.org/SUPERFAMILY/>) (Gough *et al.*, 2001) dans le but d'identifier les meilleurs

homologues parmi les gènes d'ascomycètes (*Neurospora crassa*, *Fusarium graminearum*, *Saccharomyces cerevisiae*, *Podospora anserina* et *Aspergillus nidulans* principalement), ou, à défaut, d'obtenir des informations fonctionnelles présumées.

Les analyses d'enrichissement ont été réalisées pour chaque groupe de gènes obtenus lors du regroupement ou sur l'ensemble des gènes répertoriés pour chaque étude à partir de l'annotation GO disponible sur le site du JGI (4 977 des 9 129 gènes de *T. reesei* sont annotés avec au moins terme GO). Pour chaque catégorie, la significativité du ratio d'enrichissement a été calculée à l'aide d'un test exact de Fischer basé sur la distribution hypergéométrique. Les fonctions présentant une p-value inférieure à 0,05 ont été considérées comme sur- ou sous-représentées. Une procédure similaire a été menée pour l'étude de la répression catabolique à partir des catégories du Catalogue Fonctionnel (FunCat) implémenté pour *T. reesei* sur le serveur Pendant du Munich Information Center for Protein Sequences (<http://pendant.helmholtz-muenchen.de>) (Portnoy *et al.*, 2011a).

3.6.7 Analyses des séquences promotrices

L'ensemble des analyses a été effectué au moyen de la suite logicielle RSAT (Thomas-Chollier *et al.*, 2008). Les séquences promotrices des gènes répertoriés dans l'étude de la répression catabolique ont été récupérées au moyen de l'algorithme « retrieve sequence », des coordonnées -1000 à -1, et de l'option « noORF » désactivée. La recherche de motifs a été effectuée au moyen de l'algorithme « DNA Pattern Matching » et de l'option « prevent overlapping matches ». Pour chaque jeu de données (listes des gènes de chaque groupe ou liste des gènes de l'ensemble du génome), le nombre total de motifs identifiés a été enregistré, et un nombre moyen de motifs par gène calculé. Ce nombre a été normalisé par rapport au nombre moyen de motifs par gène sur l'ensemble du génome. La significativité des différences entre les valeurs de chaque groupe et les valeurs pour l'ensemble du génome a été testée par un test *t* de Student, et des p-values inférieures à 0,1 et à 0,05 ont été retenues (voir Tableau 4).

La recherche de nouveaux motifs de régulation a été effectuée pour chaque groupe considéré au moyen des algorithmes « oligo analysis » et « dyad analysis ». Pour s'assurer que les motifs identifiés étaient bien pertinents, trois groupes de gènes de même taille et tirés aléatoirement par la fonction « random gene selection » ont été testés en parallèle pour chaque condition (longueur de 5, 6, 7 ou 8, ou dyad). Seuls les motifs présentant des scores supérieurs au plus haut score obtenu pour les jeux de données contrôles additionné de 0,5 ont été retenus. Les tests d'enrichissement et tests statistiques présentés ci-dessus ont été réappliqués à chaque nouveau motif afin de valider sa significativité.

3.6.8 Séquençage à haut débit du transcriptome entier

Afin de bénéficier d'informations sur les niveaux absolus d'expression des gènes de *T. reesei* au cours de la production de cellulases, des expériences de séquençage à haut débit des transcrits (RNA-Seq) ont été effectuées sur les échantillons récoltés à T0, T1h, T3h et T6h du premier réplicat biologique de la cinétique d'induction de RUT C30. La plateforme transcriptome et épigénome de l'Institut Pasteur a fabriqué les banques pour le séquençage à partir de 3 µg d'ARN total, en respectant les instructions du protocole mRNA-Seq d'Illumina. Le séquençage de ces banques a ensuite été réalisé sur un Genome Analyzer pour obtenir autour de 23 millions de lectures de 36 bases de long par échantillon. Ces lectures ont été alignées sur le génome de *T. reesei* à l'aide du logiciel SOAP en autorisant deux mismatches. Les alignements uniques ont alors été conservés et les niveaux d'abondances de chacun des transcrits estimés en comptabilisant le nombre de lectures par phase ouverte de lecture.

4 RÉSULTATS ET DISCUSSIONS

4.1 Étude du comportement des facteurs de transcription centraux en réponse à une induction par le lactose

Différents facteurs de transcription, tels que XYR1, ACE1 ou ACE2 ont été identifiés comme influençant l'expression des gènes codant pour des enzymes cellulolytiques (voir partie 2.3.2). Bien que des données sur leurs caractéristiques et sur leurs effets transcriptionnels soient maintenant disponibles, la description de leur régulation reste imprécise. L'objectif de cette étude est donc de décrire le comportement des gènes codant pour ces protéines régulatrices dans des conditions d'induction de la production de cellulases par le lactose, seul inducteur efficace qui peut également être utilisé comme source de carbone dans le cadre d'une production industrielle (voir partie 2.3.1). La dépendance de l'expression de ces gènes au phénomène de répression catabolique orchestré par CRE1 ou aux voies métaboliques de dégradation du lactose a également été étudiée. Ces analyses ont été effectuées sur une souche référence de la littérature, QM 9414, puis comparées aux mêmes analyses reproduites sur la souche hyperproductrice CL847 (propriété d'IFP Énergies nouvelles).

4.1.1 Validation du système expérimental

4.1.1.1 Le lactose induit la sécrétion de cellulases

Afin d'étudier l'effet d'une induction lactose sur le comportement transcriptionnel des gènes régulateurs de cellulases, une production de ces enzymes a été réalisée en bioréacteur avec la souche QM 9414. Au cours de cette procédure, les cellules sont d'abord cultivées en présence de D-glucose à une concentration initiale de 30 g.L^{-1} (phase de croissance, ou « batch »), puis, lorsque la concentration en glucose a atteint moins de 20% de la concentration initiale (6 g.L^{-1}), les cellules sont alimentées en lactose à un taux de $0,98 \text{ g.h}^{-1}$ (phase d'induction, ou « fed-batch »). L'efficacité d'un tel système à induire la production de cellulases a été démontrée au cours d'études du secrétome de *T. reesei* (Herpoel-Gimbert *et al.*, 2008). En guise de contrôle, une culture alimentée en glucose au cours de la phase de fed-batch est menée en parallèle.

La Figure 6A (i) présente les courbes de croissance, de consommation de sucres, et d'évolution du taux spécifique de production de protéines pour la souche QM 9414. Le fed-batch a été lancé à l'issue d'une phase de croissance d'environ 30 h au cours de laquelle le D-glucose a été consommé par les cellules en cours de multiplication. L'injection de lactose dans le milieu a initié une sécrétion active de cellulases, comme en témoigne l'évolution de la courbe du taux spécifique de production de protéines (traits pleins). Ce taux est calculé sur la base d'un dosage des protéines totales du milieu extracellulaire rapporté à l'évolution de la biomasse ; l'étude du secrétome de

T. reesei prouve en effet que plus de 80% des protéines s'accumulent dans le milieu au cours d'un tel protocole correspondent à des cellulases et des hémicellulases (Herpoel-Gimbert *et al.*, 2008). À l'inverse, en dépit d'une évolution similaire de la biomasse et d'une consommation identique du glucose au cours de la phase de croissance, pratiquement aucune accumulation de protéine n'a été détectée dans le cas d'un fed-batch glucose (Figure 6B, i). Ces résultats démontrent la pertinence de la procédure de culture des cellules en fed-batch lactose pour induire efficacement la sécrétion de cellulases.

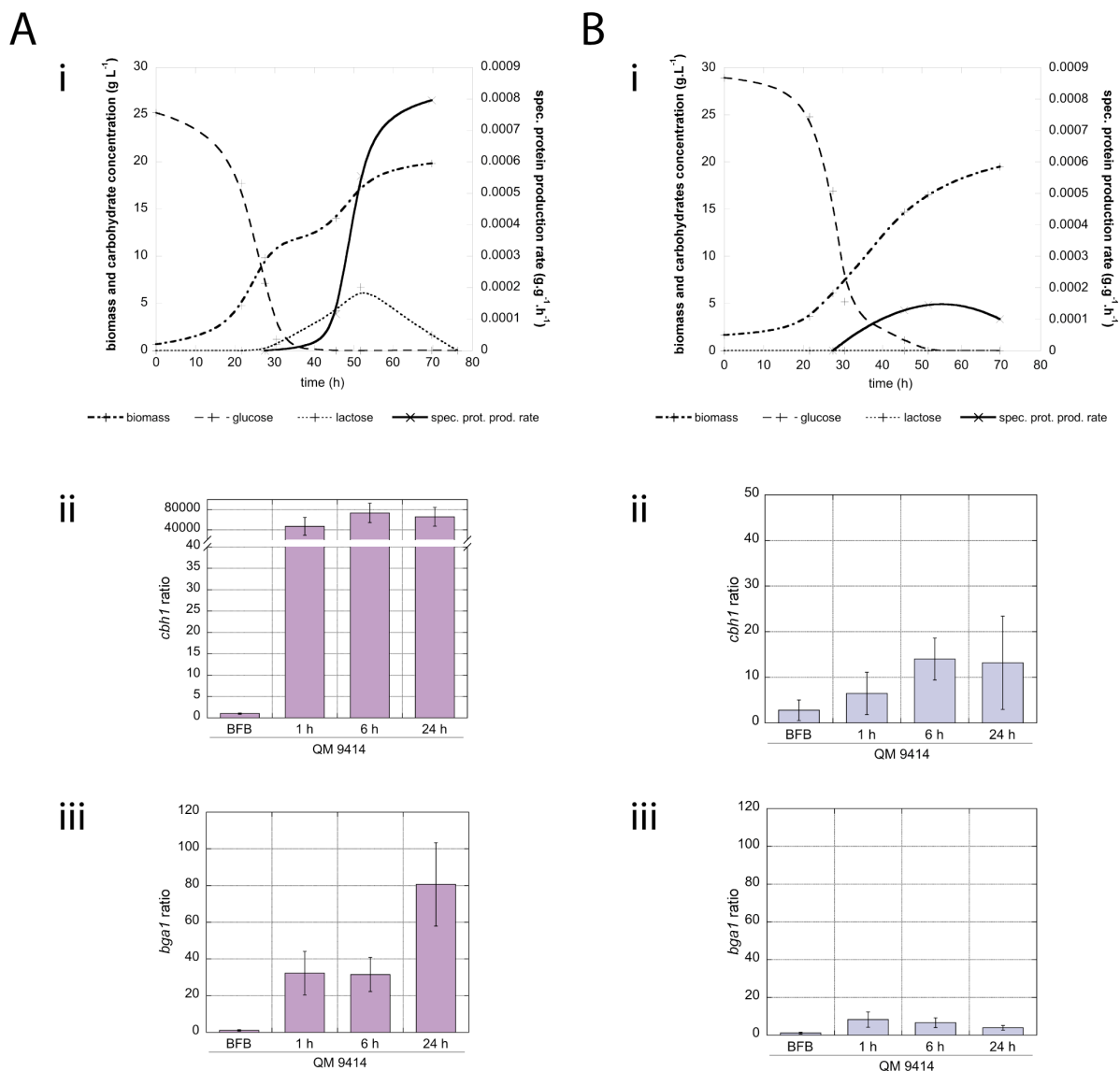


Figure 6. La souche QM 9414 de *T. reesei* exprime les gènes *cbh1* et *bga1* lors d'un fed-batch lactose. Après une phase de croissance en glucose, la souche QM 9414 de *T. reesei* a été cultivée en fed-batch lactose (A) ou glucose (B). La concentration en biomasse sèche, la consommation de sucres et le taux spécifique de production de protéines ont été enregistrés au cours des cultures (i). Les ratios d'expression des transcrits *cbh1* (ii) et *bga1* (iii) ont été calculés par rapport au gène de référence *gpd1* juste avant (BFB) et 1h, 6h et 24h après le début du fed-batch.

4.1.1.2 Le lactose induit la transcription des gènes de CAZymes

L'accumulation de protéines dans le milieu extracellulaire observée au cours du fed-batch lactose témoigne d'une expression accrue des gènes de cellulases par les cellules. Néanmoins, afin de confirmer l'action du lactose sur les niveaux d'expression des gènes de glycoside hydrolases d'intérêt, la quantité de transcrits *cbh1* (codant pour la cellulase principale CBH1) a été contrôlée avant et pendant les fed-batch glucose et lactose (Figure 6A et B, ii). La β -galactosidase BGA1 étant une enzyme indispensable à l'initiation du processus d'induction (voir partie 2.3.1.2.1), les niveaux de transcrits du gène *bga1* ont également été mesurés (Figure 6A et B, iii). Le détail de l'ensemble des niveaux mesurés dans cette étude et mentionnés dans la suite du texte est présenté dans le Tableau annexe 2.

La Figure 6A (ii et iii) montre que le fed-batch lactose induit l'accumulation des transcrits *bga1* et *cbh1*. En effet, l'expression des deux gènes est relativement basse avant l'injection de lactose dans le milieu (barre « before fed-batch », indiquée « BFB » sur les histogrammes), mais augmente considérablement pendant le fed-batch (à 1h, 6h et 24h). À l'inverse, un fed-batch glucose n'entraîne pas ou peu de changements de l'expression des GH. Les niveaux enregistrés au cours d'un tel fed-batch restent en effet très bas et proches de ceux mesurés avant son lancement (Figure 6B, ii et iii).

L'ensemble de ces données confirme donc que notre système expérimental est efficace pour induire l'expression des cellulases, et donc approprié pour étudier la régulation transcriptionnelle des protéines régulatrices de ces gènes.

4.1.2 L'expression de *xyr1* est induite par le lactose

Xyr1 étant décrit comme l'activateur majeur des gènes de cellulases et d'hémicellulases, l'influence du lactose sur son expression a été testée. Comme le montre la Figure 7, l'abondance du transcrit *xyr1* dans la souche QM 9414 était faible à l'issue de la phase de batch, mais a augmenté de près de 250 fois au cours du fed-batch lactose. Afin de vérifier que cette augmentation était effectivement due à une induction par le lactose, l'expression de *xyr1* au cours de la culture alimentée en glucose a également été mesurée. La Figure 7 montre que le fed-batch glucose n'entraîne qu'une augmentation modérée de l'expression de *xyr1* par rapport à la phase de batch. Le niveau final du transcrit *xyr1* en glucose ne correspond en effet qu'à 5% de celui enregistré au cours d'un fed-batch lactose. Ces résultats démontrent donc un rôle spécifique du lactose dans l'induction de l'expression de *xyr1*.

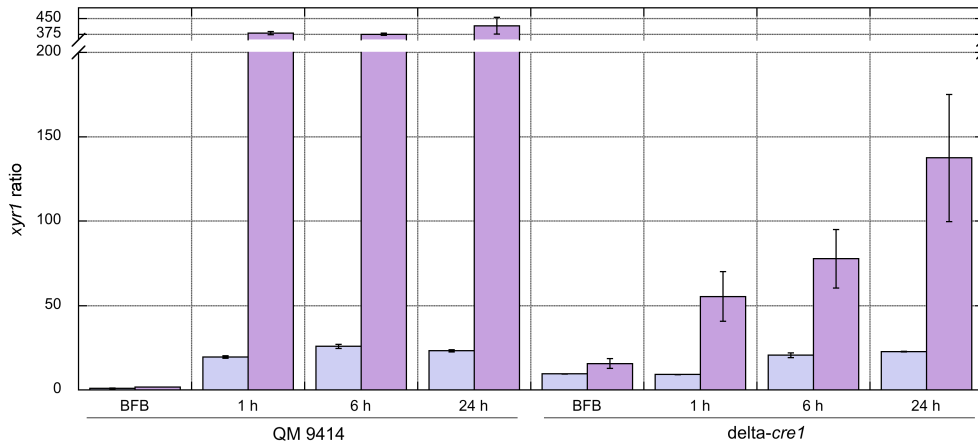


Figure 7. L'expression de *xyr1* est induite par le lactose indépendamment d'une levée de la répression catabolique. Les ratios d'expression des transcrits *xyr1* ont été calculés par rapport au gène de référence *gpd1* dans une souche sauvage ou $\Delta cre1$ de *T. reesei* QM 9414 alimentée en glucose (barres bleues) ou en lactose (barres violettes) après une croissance en glucose de 25 h. Les ratios avant (BFB) et à différents temps pendant le fed-batch sont présentés.

4.1.3 L'induction de *xyr1* par le lactose n'est pas due à une levée de la répression catabolique

En 2007, Stricker *et al.* ont réalisé une comparaison des niveaux d'expression du transcrit *xyr1* dans la souche *cre1*-négative RUT C30 retransformée ou non par *cre1*. Cette étude laissait supposer que, dans le cas d'une culture en présence de lactose, l'accumulation de transcrits *xyr1* était due à une levée de la répression catabolique et non à une induction (Stricker *et al.*, 2007). Afin de confirmer que les niveaux d'expression enregistrés dans nos cellules étaient bien la conséquence d'une induction par le lactose et non uniquement d'une dérèpression, des cultures d'une souche QM 9414 délétée pour le gène *cre1* ont été réalisées dans les mêmes conditions et analysées.

4.1.3.1 La souche $\Delta cre1$ est bien dérèprimée

Le caractère dérèprimé du mutant $\Delta cre1$ de la souche QM 9414, construit dans le cadre de cette étude (voir Portnoy *et al.* 2011b, en annexe), a été vérifié par comparaison des niveaux basaux de transcrit *cbh1* avec une souche QM 9414 sauvage. Une inactivation efficace de *cre1* doit en effet nécessairement donner lieu, en conditions de répression catabolique, à une augmentation de la transcription constitutive de ce gène cible de CRE1. Comme mentionné dans le Tableau annexe 2, ce niveau est de 3 à 10 fois plus élevé dans la souche délétée pour *cre1* au cours des phases de batch. L'expression de *xyr1* est également augmentée de près de 10 fois dans cette souche (voir Figure 7), confirmant l'influence de la répression catabolique sur les niveaux basaux de *xyr1* et son absence dans la souche $\Delta cre1$. La souche étant dérèprimée, le passage à un fed-batch glucose n'a induit aucun effet majeur sur le niveau d'expression de *xyr1*.

Au vu de ces résultats, les observations réalisées à partir de la souche QM 9414 $\Delta cre1$ ne pourront, dès lors, être attribuées à des mécanismes liés à la répression catabolique.

4.1.3.2 Xyr1 est induit par le lactose dans la souche $\Delta cre1$

Comme montré Figure 7, le passage du batch à un fed-batch glucose n'induit qu'une faible augmentation du niveau d'expression de *xyr1* chez le mutant $\Delta cre1$. Ceci est cohérent puisque la souche est déjà déréprimée avant le lancement du fed-batch. En revanche, l'injection de lactose augmente de 10 fois le niveau d'expression de *xyr1* avant fed-batch, qui atteint ainsi une valeur 6 fois supérieure à celle enregistrée dans le cas du fed-batch glucose.

Xyr1 est donc régulé à la fois par le phénomène de répression catabolique et par une induction provoquée par le lactose, indépendamment.

4.1.4 Xyr1 est régulé positivement par CRE1 au cours de l'induction lactose.

Si dans la souche $\Delta cre1$ l'expression du gène *xyr1* au cours du fed-batch glucose atteint les mêmes niveaux que ceux de la souche QM 9414 sauvage, l'abondance des transcrits *xyr1* au cours du fed-batch lactose est plus faible chez le mutant $\Delta cre1$. En effet, bien que 10 fois plus élevés qu'à la fin du batch, les niveaux de transcrits *xyr1* après 24h de fed-batch lactose n'atteignent dans la souche $\Delta cre1$ qu'un tiers des niveaux enregistrés dans la souche QM 9414 (Figure 7). L'induction de *xyr1* par le lactose semble donc être accrue en présence de la protéine CRE1.

4.1.5 L'induction de *xyr1* par le lactose est indépendante du métabolisme du D-galactose

Comme décrit dans la partie 2.3.1.2.4, l'induction des cellulases par le lactose est dépendante du métabolisme du lactose *via* la voie de Leloir et *via* la voie alternative de catabolisme du D-galactose (Seiboth *et al.*, 2007a; Seiboth *et al.*, 2007b). Nous avons donc testé si l'induction de *xyr1* par le lactose dans la souche QM 9414 dépendait également de ce métabolisme. Pour ce faire, des souches QM 9414 possédant des délétions dans les gènes *gal1* (galactokinase, bloquant la voie de Leloir, (Seiboth *et al.*, 2004)), *xyl1* (aldose réductase, bloquant la voie alternative, (Seiboth *et al.*, 2007a)) ou la double mutation *gal1/xyl1* (Seiboth *et al.*, 2007a), ont été cultivées en présence de lactose. La Figure 8A montre que l'expression de *xyr1* n'était affectée par aucune de ces mutations, et ne semble donc pas dépendante du métabolisme du galactose.

Le métabolisme du lactose chez *T. reesei* étant initié par une hydrolyse extracellulaire du lactose en D-glucose et D-galactose (Seiboth *et al.*, 2005; Seiboth *et al.*, 2007b), ce résultat suggère

que le D-galactose pourrait lui-même jouer le rôle d'inducteur de *xyr1*. Afin de tester cette hypothèse, l'expression de *xyr1* a été mesurée chez des souches QM 9414 cultivées en glycérol, en D-glucose ou en D-galactose (ces cultures ont été réalisées à des taux de croissance faibles, auxquels le D-galactose est capable d'induire une production de cellulases (Karaffa *et al.*, 2006)). Comme le montre la Figure 8B, les niveaux de transcrits *xyr1* en présence de D-galactose sont 5 à 10 fois supérieurs à ceux enregistrés en présence de glucose et de glycérol, confirmant ainsi la capacité du monosaccharide à induire l'expression du facteur de transcription.

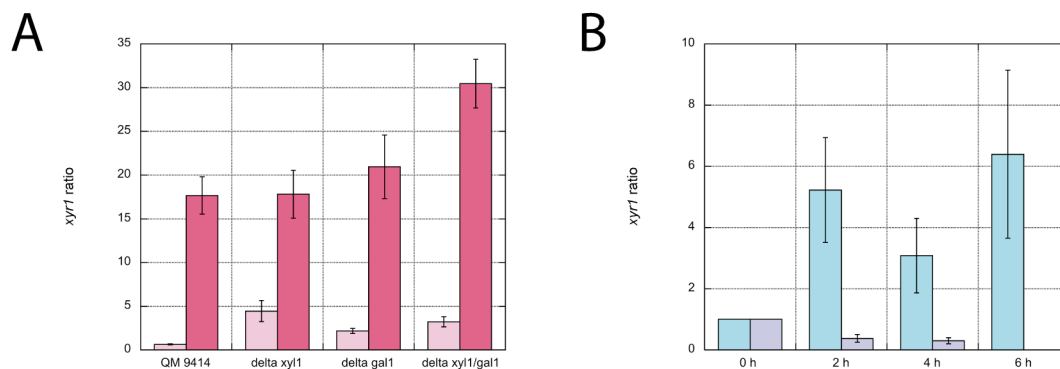


Figure 8. L'induction lactose de *xyr1* ne dépend pas du métabolisme du galactose. Les ratios d'expression des transcrits *xyr1* ont été calculés par rapport au gène de référence *gpd1* dans des mutants de délétions pour les voies métaboliques du galactose (A) ou dans la souche sauvage QM 9414 (B). A : les souches $\Delta gal1$, $\Delta xyl1$ et $\Delta gal1 \Delta xyl1$ de QM 9414 ont été cultivées en fioles en présence de lactose et les ratios d'expressions calculés après 20 h (barres roses) ou 30h (barres rouges). B : La souche QM 9414 a été pré-cultivée en présence de glycérol pendant 24h puis collectée, lavée, et transférée sur un milieu frais contenant du D-galactose (barres turquoise) ou du D-glucose (barres bleues) pendant 2h, 4h ou 6h, comme indiqué. Seul le ratio pour le galactose a été mesuré à 6h. Le ratio obtenu à l'issue de la pré-culture en glycérol a été fixé à 1,0 et les autres ratios ajustés en conséquence.

4.1.6 *Ace1* et *ace2* sont induits par le lactose

L'influence du lactose sur l'expression des régulateurs secondaires ACE1 et ACE2 a également été testée. La Figure 9A montre que *ace2* est constitutivement exprimé à des niveaux faibles pendant les phases de croissance en D-glucose, et qu'un apport de lactose augmente cette expression de plus d'un facteur 3. Bien que cet écart soit moins spectaculaire que dans le cas de *xyr1*, une induction d'*ace2* par le lactose est donc bien détectée. L'expression d'*ace2* ne semble pas soumise à la répression catabolique puisque les niveaux transcriptionnels ne sont pas plus élevés chez la souche $\Delta cre1$; au contraire, les niveaux atteints dans cette souche après l'induction au lactose sont plus faibles que ceux observés dans la souche sauvage. Ainsi, ces résultats suggèrent que, comme pour *xyr1*, CRE1 pourrait favoriser, directement ou indirectement, l'expression d'*ace2* dans des conditions de dérégulation.

Le gène *ace1*, qui code pour un répresseur de la transcription des gènes de cellulases, est faiblement exprimé au cours des phases de batch glucose. Alors que cette expression est restée invariante lorsqu'un fed-batch glucose a été initié, le passage à un fed-batch lactose a entraîné une augmentation du niveau de transcrits *ace1* (Figure 9B). Aussi bien l'expression basale que l'expression induite d'*ace1* semblent être soumises à la répression catabolique car leur niveaux étaient bien plus élevés dans la souche $\Delta cre1$.

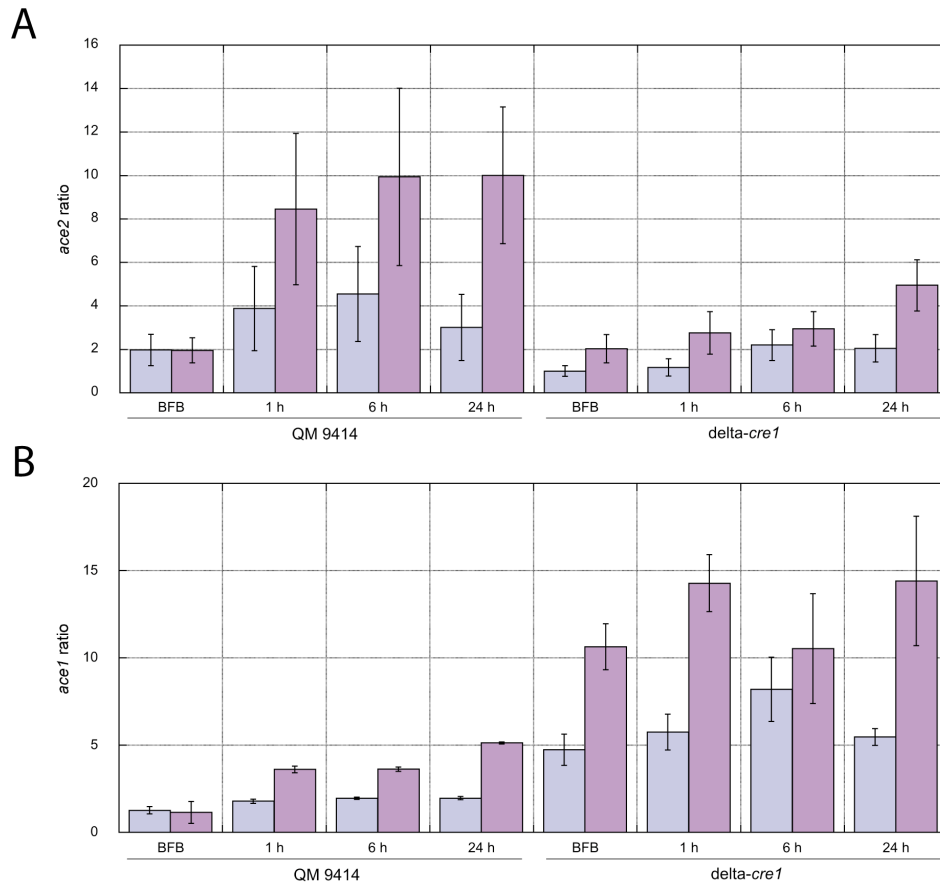


Figure 9. Les gènes *ace2* et *ace1* sont régulés par le lactose, mais seul *ace1* est soumis à la répression catabolique. Les ratios d'expression des transcrits *ace2* (A) et *ace1* (B) ont été calculé par rapport au gène de référence *gpd1* dans une souche sauvage et une souche $\Delta cre1$ de *T. reesei* QM 9414 alimentées en glucose (barres bleues) ou lactose (barres violettes) après une croissance de 25 h en glucose. Les ratios avant (BFB) et à différents temps pendant le fed-batch sont présentés.

4.1.7 Vérification du caractère hyperproducteur de la souche CL847 de *T. reesei*.

Afin de déterminer comment l'expression des régulateurs transcriptionnels XYR1, ACE2 et ACE1 était modifiée en réponse à une induction lactose dans une souche hyperproductrice de cellulases, les expériences décrites ci-dessus ont été reproduites avec la souche CL847 de *T. reesei*.

4.1.7.1 CL847 descend de la souche RUT C30, fortement productrice

Le mutant CL847 de *T. reesei* est la souche la plus productrice de cellulases décrite à ce jour (Durand *et al.*, 1988b). D'après les étapes de mutagenèse à l'origine des mutants hyperproducteurs décrites par Durand *et al.* (voir partie 2.2.4), cette souche a pour origine la lignée de QM 9414. Nous avons donc testé si CL847, censée être dérèprimée, contenait un locus *cre1* intact, comme c'est le cas chez sa parente supposée. À notre grande surprise, nos données de reséquençage prouvent que CL847 présente un locus *cre1* tronqué exactement identique à celui de la souche RUT C30, autre souche référence de la littérature, dérèprimée et fortement productrice (Eveleigh and Montenecourt, 1979) (voir Figure annexe 3). Ces données indiquent donc que l'hyperproductrice CL847 n'est autre qu'un descendant de la souche RUT C30, déjà fortement productrice.

4.1.7.2 CL847 produit bien plus de cellulases que QM 9414

Des cultures de la souche CL847 ont été réalisées dans les mêmes conditions que celles utilisées pour QM 9414. La Figure 10 montre que la consommation de D-glucose et l'évolution de la quantité de biomasse de CL847 au cours de la phase de batch étaient similaires à QM 9414. Comme pour la souche $\Delta cre1$, la production de protéines a été enregistrée immédiatement après le lancement du fed-batch lactose, indiquant que cette souche n'est pas réprimée. La concentration en protéines sécrétées dans le milieu de culture était de 5 à 15 fois plus élevée pour CL847 que pour QM 9414 à des temps comparables, confirmant le caractère hyperproducteur de la souche industrielle (e.g. 0,08 g.L⁻¹ pour QM 9414 contre 1,33 g.L⁻¹ pour CL847 après 24 h d'induction, ou 0,58 g.L⁻¹ pour QM 9414 contre 4,23 g.L⁻¹ pour CL847 après 50 h d'induction).

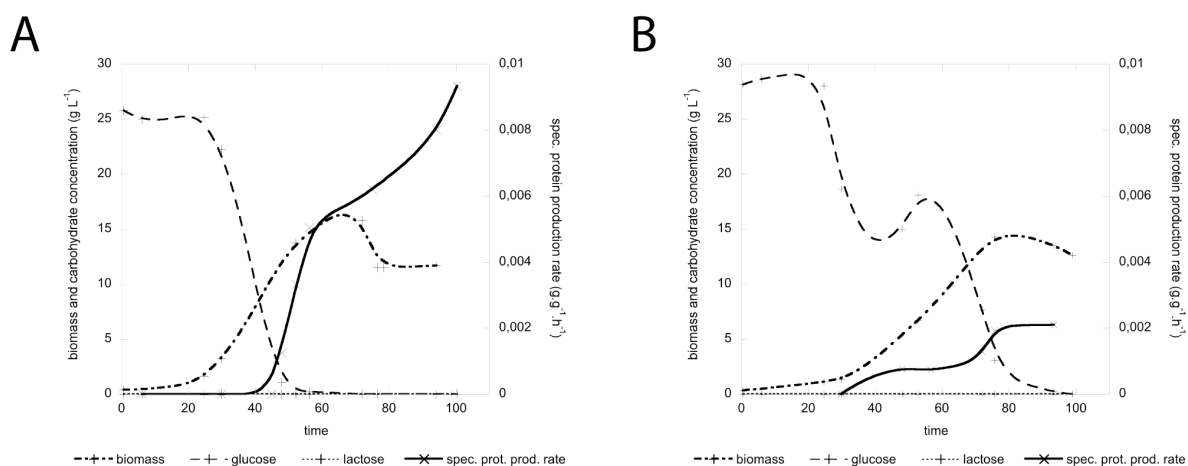


Figure 10. La souche hyperproductrice CL847 de *T. reesei* produit des cellulases lors d'un fed-batch lactose. Après une phase de croissance en glucose, la souche CL847 de *T. reesei* a été cultivée en fed-batch lactose (A) ou glucose (B). La concentration en biomasse sèche, la consommation de sucres et le taux spécifique de production de protéines ont été enregistrés au cours des cultures.

4.1.8 L'expression de *xyr1* et de *ace1* est modifiée dans la souche hyperproductrice CL847

Comme effectué pour QM 9414 et son mutant $\Delta cre1$, l'accumulation des transcrits *cbh1*, *xyr1*, *ace2* et *ace1* a été mesurée tout au long des cultures de CL847. L'expression du gène *cbh1* est déjà très élevée chez CL847 avant le lancement du fed-batch (Figure 11A); le niveau de transcrits *cbh1* atteint par CL847 avant l'induction lactose correspond en effet à plus de la moitié du niveau atteint par QM 9414 après induction. Après 24 h de fed-batch, l'abondance des transcrits *cbh1* est 10 fois supérieure chez CL847 par rapport à QM 9414.

L'expression de *xyr1* est également significativement plus forte chez CL847 avant le début du fed-batch, et les niveaux atteints 24 h après le début de l'induction sont 6 à 7 fois plus élevés que chez QM 9414 (Figure 11B). Chez CL847, le niveau basal d'expression de *xyr1* dans la culture contrôle, 8 fois plus élevé que dans la souche $\Delta cre1$, n'a pratiquement pas été modifié par le lancement du fed-batch glucose. En revanche, l'abondance du transcrit *xyr1* après 24 h de fed-batch lactose est presque 15 fois supérieure à son niveau enregistré avant le lancement de l'induction. Le mutant CL847 ayant été identifié comme descendant de la souche RUT C30, des expériences complémentaires ont été menées afin d'établir si la transcription de *xyr1* était également augmentée chez RUT C30, dans les mêmes conditions de culture et d'induction. Comme montré dans la Figure 11B, l'abondance des transcrits *xyr1* atteint des niveaux similaires à ceux enregistrés pour QM 9414, aussi bien pour l'expression basale que pour l'expression induite du gène. La forte surexpression de *xyr1* est donc spécifique de l'hyperproductrice CL847.

Les différences d'expression du gène *ace2* sont moins prononcées que celles de *xyr1* (Figure 11C). Les niveaux basaux d'expression enregistrés chez CL847 sont proches de ceux de la souche $\Delta cre1$, et s'expliquent donc probablement par l'absence de protéine CRE1 fonctionnelle chez ces mutants. L'induction lactose a cependant engendré des niveaux similaires à ceux enregistrés chez QM 9414, ce qui semble indiquer que, chez CL847, cette perte de fonction de CRE1 n'affecte pas l'expression d'*ace2* en réponse à l'induction.

L'expression basale d'*ace1* chez CL847 est également proche de celle enregistrée pour la souche $\Delta cre1$ (Figure 11D), et peut s'expliquer par la levée de la répression catabolique chez ces deux souches. Étonnamment, cette expression décroît au cours du fed-batch glucose. L'expression d'*ace1* en réponse au lactose, bien qu'inférieure à celle enregistrée pour la souche $\Delta cre1$, est tout de même plus élevée chez CL847 que chez QM 9414. Cette expression ne semble pas induite au cours du fed-batch.

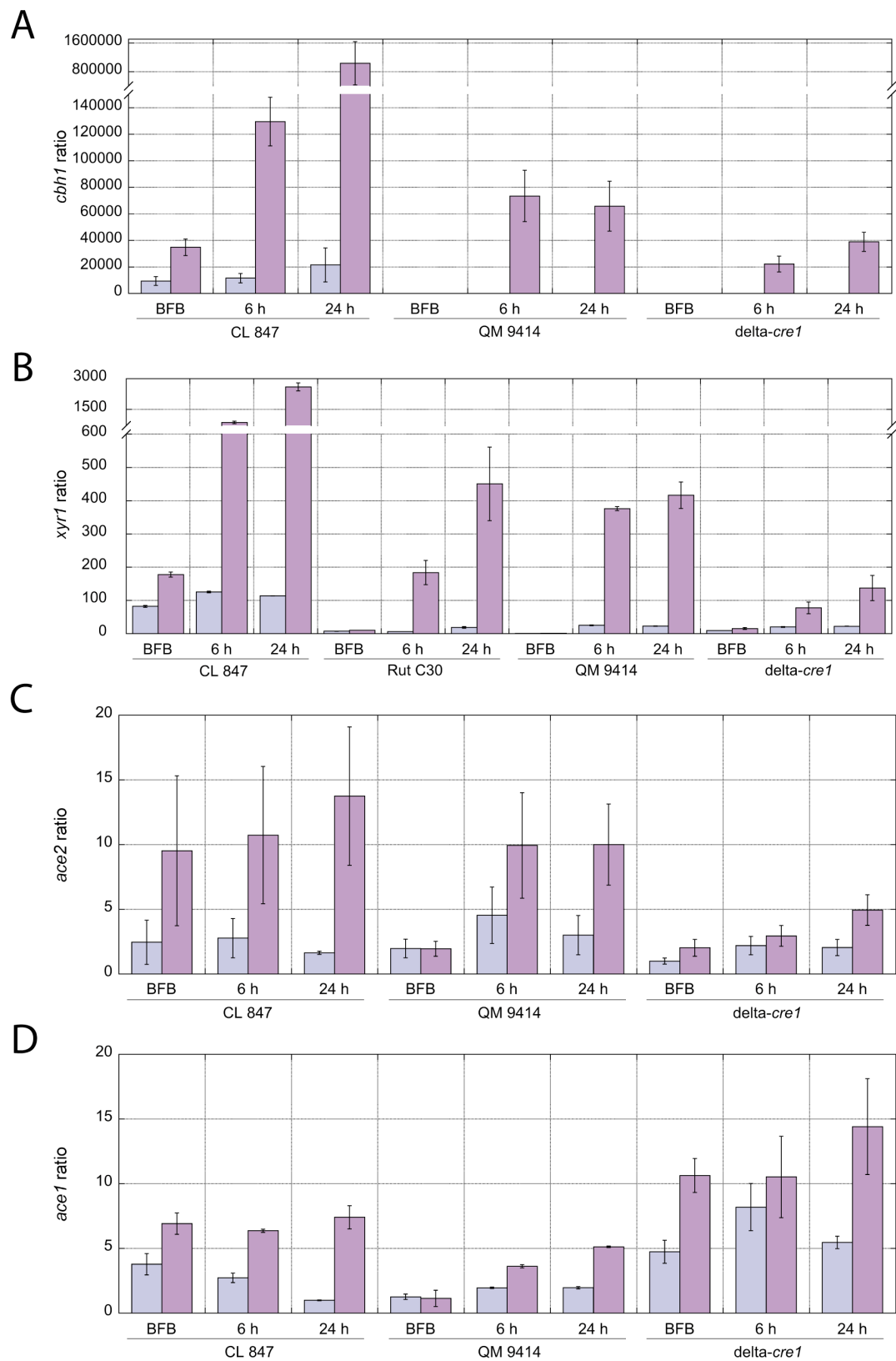


Figure 11. L'expression des gènes régulateurs de la transcription des cellulases est modifiée dans la souche hyperproductrice CL847. Les ratios d'expression des transcrits *cbh1* (A), *xyr1* (B), *ace2* (C) et *ace1* (D) ont été calculés par rapport au gène de référence *gpd1* dans la souche CL847 (et RUT C30 (B)) de *T. reesei* alimentée(s) en glucose (barres bleues) ou en lactose (barres violettes) après une croissance en glucose de 25 h. Les ratios avant (BFB) et à différents temps pendant le fed-batch sont présentés. Les ratios correspondants des souches sauvages et $\Delta cre1$ de QM 9414 sont reproduits pour faciliter la comparaison.

4.1.9 Discussion

4.1.9.1 Induction de l'expression de *xyr1* par le lactose

XYR1 est le seul facteur de transcription identifié à ce jour dont l'inactivation entraîne l'absence de formation de toutes les cellulases, xylanases et de certaines des enzymes impliquées dans le métabolisme des monomères d'hémicellulose, tels que le D-xylose, le L-arabinose et le D-galactose (Seiboth *et al.*, 2007a; Stricker *et al.*, 2006). La régulation de la formation de ce facteur a donc nécessairement une importance majeure dans la capacité de *T. reesei* à dégrader les polysaccharides végétaux. Les analyses présentées ici confirment, en accord avec les données de la littérature, qu'un faible niveau de transcrite *xyr1* est formé en présence de D-glucose. Cependant, contrairement à l'idée évoquée dans certaines publications que la transcription du gène n'est pas régulée par une induction (Mach-Aigner *et al.*, 2008; Stricker *et al.*, 2007), nos résultats prouvent que l'expression de *xyr1* est bien induite par le lactose, ainsi que par le produit issu de son hydrolyse, le D-galactose.

Une telle induction n'a jamais été décrite précédemment ; deux études publiées ces dernières années s'accordent en effet à dire que les augmentations de l'expression de transcrits *xyr1* détectées jusque là n'étaient dues qu'à des levées de la répression catabolique dépendante de *cre1*. L'une de ces études n'a cependant été réalisée que sur des cellules cultivées en présence de D-xylose ou de xylobiose, et non en présence de lactose (Mach-Aigner *et al.*, 2008). Ainsi, si ces molécules ne sont pas capables d'induire l'expression de *xyr1* au delà du niveau atteint par dérépression, les observations de Mach-Aigner *et al.* n'excluent pas la possibilité que d'autres inducteurs en soient capables.

Dans une autre étude, Stricker *et al.* observent des niveaux plus élevés de transcrits *xyr1* chez une souche QM 9414 cultivée en présence de lactose que chez cette même souche cultivée en milieu glucose (Stricker *et al.*, 2007). Cependant, ces travaux ne présentent pas de mesures de l'expression du gène dans des conditions de dérépression des cellules en l'absence de lactose. Ainsi, les niveaux observés en lactose pourraient correspondre non pas au seul effet d'une dérépression, mais à l'addition des effets de la dérépression et d'une induction. Les auteurs cherchent à infirmer cette hypothèse, mais réalisent pour cela des expériences chez la souche RUT C30, naturellement déréprimée. L'expression de *xyr1* est alors peu différente en glucose et en lactose, bien que visiblement plus élevée en lactose. À l'inverse, au cours de notre étude, l'ensemble des expériences a été effectué chez QM 9414, ce qui nous conforte dans la robustesse de nos observations. Nos données obtenues sur RUT C30 mettent par ailleurs en évidence des différences plus nettes entre les niveaux d'expression de *xyr1* en glucose et en lactose, allant également dans le sens d'une induction lactose. Ces différences peuvent s'expliquer par le fait que les analyses de Stricker *et al.* ont été

réalisées à partir de cultures en fioles, au cours desquelles la surexpression de *xyr1* peut être transitoire et non détectable. Notre système expérimental en bioréacteur permet au contraire de maintenir et donc de détecter cette induction au cours du temps sans ambiguïté.

4.1.9.2 Quels sont les mécanismes de l'induction de *xyr1* par le lactose ?

Comme précisé dans la partie 2.3.1.2.5, l'induction des cellulases par le lactose semble dépendante du métabolisme du disaccharide *via* les deux voies de catabolisme du galactose (Seiboth *et al.*, 2007a; Seiboth *et al.*, 2004). À l'inverse, d'après nos résultats, l'induction de *xyr1* par le lactose semble complètement indépendante de ce métabolisme et peut également être assurée par le D-galactose. Le mécanisme par lequel *xyr1* répond à ces deux molécules reste inconnu. Néanmoins, le lactose n'est pas censé être intégré par les cellules (Seiboth *et al.*, 2005); il est donc possible que son effet en absence d'un métabolisme actif du D-galactose soit dû à une voie de transduction. Brunner *et al.* ont récemment identifié un gène codant pour un récepteur couplé à une protéine G AMP cyclique-dépendante chez *Trichoderma arthrovidie* (Brunner *et al.*, 2008), dont l'inactivation est une des causes affectant la capacité du champignon à pousser en présence de galactooligosaccharides. Un orthologue de ce gène est présent dans le génome de *T. reesei* et est également exprimé dans des conditions de synthèse de cellulases (EST id. 17696735, (Foreman *et al.*, 2003)). Le rôle possible de ce récepteur dans l'induction lactose de *xyr1* est actuellement étudié par l'équipe de C.P. Kubicek.

Si les mécanismes moléculaires par lesquels le lactose engendre une augmentation de l'expression de *xyr1* restent encore flous, nos données confirment que, biologiquement parlant, la présence de β -galactosides dérivés des polymères constitutifs des parois végétales est un signal physiologique majeur conduisant *T. reesei* à orienter sa réponse transcriptionnelle en faveur d'une dégradation efficace des composé cellulotiques.

4.1.9.3 Les apports d'une comparaison intersouches

Les régulations transcriptionnelles semblant jouer un rôle majeur dans le contrôle de la biosynthèse des cellulases, les profils des transcrits de *xyr1* dans la souche hyperproductrice CL847 et dans sa souche mère RUT C30 (identifiée ici pour la première fois), ont été comparés à nos résultats chez QM 9414. Dans les conditions de l'étude, CL847 a produit jusqu'à 15 fois plus d'enzymes cellulolytiques que QM 9414. Cette abondance est corrélée à une augmentation de l'expression du gène *cbh1* de plus de 15 fois également, ce qui semble confirmer que les mécanismes transcriptionnels constituent bien un facteur limitant primordial dans la biosynthèse des enzymes cellulolytiques. Les niveaux de transcrits *xyr1* chez CL847 confirment son rôle d'activateur transcriptionnel majeur des cellulases puisque l'expression basale du gène était très augmentée chez

l'hyperproductrice. Cette différence d'expression en présence de glucose est cohérente, puisque le gène *cre1* est inactif chez CL847 et que la souche est donc déjà déréprimée lors du batch glucose. Le passage à un fed-batch lactose a encore accru l'expression du gène, confirmant le rôle inducteur du lactose sur *xyl1*. Le niveau final d'expression de *xyl1* était en effet 6 fois plus élevé dans CL847 que dans QM 9414. Nos données de reséquençage prouvant que le locus *xyl1* de CL847 ne possède aucune mutation ni dans son promoteur ni dans sa séquence codante (voir Figure annexe 3), cette différence s'explique probablement par une efficacité accrue de la machinerie transcriptionnelle de la souche industrielle. De plus, le fait que le niveau d'expression de *xyl1* en glucose soit de 5 à 8 fois plus élevé chez CL847 que dans la souche QM 9414 $\Delta cre1$ montre que la levée de la répression catabolique exercée par *cre1* n'est pas le seul facteur à l'origine de l'expression accrue du régulateur chez CL847. Enfin, RUT C30, pourtant fortement productrice, semblait présenter des niveaux d'expression de *xyl1* similaires à ceux de QM 9414. Ces dernières observations nous confortent donc dans l'idée qu'un examen détaillé de l'ensemble des transcrits exprimés lors d'une induction chez des souches plus ou moins productrices peut s'avérer déterminant dans l'identification de facteurs clés intervenant dans le processus.

4.1.9.4 *Ace2* et *ace1* ont un comportement plus difficilement interprétable

Les effets des facteurs ACE2 et ACE1 restent à ce jour mal identifiés, et les données de la littérature relatives à leurs profils d'expression sont parfois contradictoires (voir partie 2.3.2). LA protéine ACE2 a par exemple été décrite comme activatrice des cellulases lors de cultures en cellulose, mais ne semble pas avoir d'influence lors de cultures en sophorose (Aro *et al.*, 2001), et peut aussi parfois réprimer l'expression de *xyl1* (Stricker *et al.*, 2008b). Concernant *ace1*, Aro *et al.* semblent suggérer, sur la base de données non présentées, que l'expression du gène n'est pas dépendante de la répression catabolique médiée par CRE1 (Aro *et al.*, 2003) ; pourtant, 11 séquences consensus de fixation CRE1 ont été décrites dans son promoteur.

Si l'on considère qu'ACE2 a bel et bien un rôle activateur des cellulases, il est cohérent que son expression soit induite par le lactose. Cela n'est pas le cas pour ACE1, dont l'expression est également induite par le lactose, alors qu'il n'est *a priori* pas attendu qu'un répresseur des cellulases soit activé dans des conditions de production de ces enzymes. L'abondance des transcrits d'*ace1* est par ailleurs augmentée dans la souche hyperproductrice par rapport à QM 9414, ce qui semble également contradictoire avec son rôle de répresseur supposé – à noter que le locus *ace1*, comme celui d'*ace2*, ne présente aucune mutation chez CL847, voir Figure annexe 3. Néanmoins, le comportement du gène *ace1* dans nos expériences semble en accord avec les analyses *in silico* de son promoteur puisque aussi bien son expression basale que son expression induite sont réprimées en

présence de CRE1. *Ace1* étant soumis à la répression catabolique, ses niveaux d'expression dans CL847 sont à comparer avec ceux de la souche $\Delta cre1$; une telle analyse révèle que l'expression basale du gène est plus faible dans l'hyperproductrice que dans le mutant dérèprimé, et décroît au cours du fed-batch glucose. Ainsi, la dérèpression d'ACE1 semble restreinte chez CL847, ce qui irait dans le sens d'un rôle de régulateur négatif. On retrouve cependant chez l'hyperproductrice une augmentation d'un facteur deux des niveaux basaux d'*ace1* en présence de lactose, comme chez les autres souches.

4.1.9.5 Vers un rôle activateur de CRE1 ?

Si CRE1 a été largement décrit comme étant l'acteur principal de la répression catabolique exercée lorsque les cellules de *T. reesei* sont en présence de glucose (voir partie 2.3.2.2.2), cette protéine pourrait ne pas se limiter à des fonctions de régulateur négatif. Nos résultats montrent en effet sans équivoque que l'induction par le lactose de l'expression de *xyr1* est moins prononcée en l'absence de CRE1 ; l'abondance des transcrits *xyr1* au cours du fed-batch lactose, bien que très supérieure aux niveaux enregistrés avant l'induction, reste trois fois moins forte dans la souche $\Delta cre1$ que chez son parent exprimant la protéine. De même, les niveaux diminués de transcrit *ace2* dans la souche $\Delta cre1$ prouvent que l'induction d'*ace2* est elle aussi accrue en présence de la protéine CRE1 fonctionnelle. Comme pour *xyr1*, l'expression correcte d'*ace2* chez CL847 est probablement assurée grâce à des mécanismes de compensation acquis par la souche hyperproductrice.

Un tel effet positif de CRE1 sur l'expression des gènes n'avait à notre connaissance jamais été décrit auparavant. La plupart des études portant sur ce facteur de transcription (ou sur son homologue CreA chez *Aspergillus nidulans* ou *Aspergillus niger*) ont été menées dans l'idée que la répression catabolique allait nécessairement contre l'expression des gènes, et s'attachent donc à identifier les transcrits dont l'expression est réprimée par CRE1. Pourtant, il est vraisemblable que cette protéine régulatrice puisse également activer la transcription de certains gènes. Des observations de Morgensen *et al.* semblent aller dans le sens de cette hypothèse. Lors d'une analyse transcriptionnelle menée sur une partie du génome d'*Aspergillus nidulans*, l'équipe de Morgensen a en effet identifié quelques gènes dont l'expression était diminuée dans une souche délétée pour le gène *creA* (Morgensen *et al.*, 2006). Si les auteurs « n'excluent pas » que cette observation pourrait s'expliquer par un effet positif de CreA sur ces gènes, l'hypothèse n'a pas été testée plus avant.

En résumé, cette étude a permis de décrire le comportement en réponse à l'induction lactose des facteurs de transcription XYR1, ACE2 et ACE1, qui avaient précédemment été identifiés comme étant impliqués dans la production de cellulases de *T. reesei*. Nos résultats confirment l'importance des régulations transcriptionnelles dans le contrôle de cette production, précisent l'influence de CRE1 sur ces régulations, et mettent pour la première fois en évidence l'existence d'une induction lactose et non pas d'une simple dérégulation de l'expression du facteur central *xyr1*. Ils indiquent également que l'acquisition du phénotype hyperproducteur de la souche CL847 s'accompagne de dérégulations majeures de l'expression de ces protéines. La Figure 12 présente le réseau de régulation des cellulases décrit dans la Figure 5 et complété par les données issues de nos analyses. L'ensemble de ces travaux a fait l'objet d'une publication présentée en annexe (Portnoy *et al.*, 2011b).

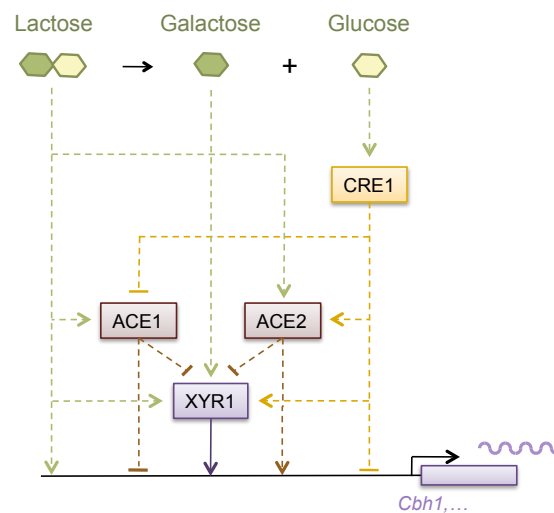


Figure 12. Les nouvelles données de régulation des gènes de cellulases chez *T. reesei*. Nos analyses démontrent que le lactose induit l'expression de *xyr1*, *ace1* et *ace2*, que *ace1* est soumis à la répression catabolique et que CRE1 accentue l'induction de *xyr1* et *ace2*. Le D-galactose est également un inducteur de *xyr1*. Les lignes pointillées indiquent une action des régulateurs sur leurs cibles nécessitant d'éventuels intermédiaires non caractérisés.

4.2 Étude de la répression catabolique, mécanisme d'inhibition de l'expression des gènes de cellulases

La répression catabolique est un mécanisme majeur influençant la production de cellulases (voir partie 2.3.2.2.2). Chez *T. reesei*, elle est orchestrée par le facteur de transcription CRE1, qui, on l'a vu, a un impact significatif sur le fonctionnement du réseau de régulation de l'expression de ces enzymes. Chez *A. nidulans* et *A. niger*, un certain nombre de gènes dont l'expression est contrôlée par la répression catabolique a été identifié à ce jour, mais notre connaissance des effets de CRE1/CreA sur l'expression génétique globale reste très partielle. Comme mentionné précédemment, la plupart des études ont jusqu'ici été menées dans l'idée que la répression catabolique allait nécessairement contre l'expression des gènes ; elles s'attachent donc le plus souvent à identifier les transcrits dont l'expression est réprimée par CRE1/CreA. Pourtant, nos résultats suggèrent que cette protéine régulatrice pourrait également favoriser la transcription de certains gènes (voir partie 4.1.9.5). De plus, s'il est acquis que CRE1 joue un rôle central dans la mise en place de la répression catabolique, il n'est pas exclu que d'autres voies indépendantes de CRE1 prennent part dans la réalisation du processus. L'objectif de cette deuxième étude, menée à l'échelle du génome total, est donc d'identifier quels gènes chez *T. reesei* sont contrôlés par CRE1, et d'établir si certains gènes dont l'expression est modifiée en conditions de répression catabolique sont régulés indépendamment de ce facteur de transcription.

4.2.1 Description du dispositif expérimental

4.2.1.1 Création de puces à ADN dédiées *T. reesei*

Notre objectif étant d'examiner les effets transcriptionnels de CRE1 et de la répression catabolique sur l'ensemble des gènes de *T. reesei*, des puces à ADN dédiées à l'exploration du génome de cet organisme ont été conçues. Grâce au logiciel Teolenn, développé sur la Plate-Forme Transcriptome de l'IBENS pour l'occasion (manuscrit publié dans *Nucleic Acids Res*, présenté en annexe (Jourden *et al.*, 2010)), des sondes oligonucleotidiques réparties sur l'ensemble des 33,4 millions de bases du génome de *T. reesei* ont été dessinées à la résolution d'un oligonucleotide de 60 mers toutes les 140 paires de bases. Dans le cadre d'études transcriptomiques, ce design permet d'atteindre une couverture moyenne de 9,7 sondes par transcrit (voir les méthodes partie 3.6.2).

4.2.1.2 Réalisation de cultures continues des souches QM 9414 et $\Delta cre1$

Afin d'identifier l'ensemble des gènes dont l'expression est modifiée dans des conditions de répression catabolique et de préciser leur dépendance ou non au facteur de transcription CRE1, les transcriptomes de la souche QM 9414 et de son mutant $\Delta cre1$ issus de cultures reproduisant ou non des conditions de répression catabolique ont été comparés. La souche $\Delta cre1$ présentant –comme précédemment référencé (Nakari-Setälä *et al.*, 2009)– une croissance ralentie (voir Figure annexe 1 et Tableau annexe 3), des cultures continues en chémostat ont été préférées à des cultures classiques en fioles. Ce dispositif permet en effet de contrôler parfaitement le taux de croissance des cellules et donc de s'assurer que les différences d'expression génétiques observées entre les souches sont bien la conséquence des phénomènes testés et non des répercussions indirectes de la délétion. Les cellules de QM 9414 et celles de son mutant $\Delta cre1$ ont donc été cultivées et maintenues à l'état d'équilibre (biomasse constante) en présence de glucose à deux taux de dilution (ou taux de croissance), l'un induisant une répression catabolique (0.07 h^{-1}), l'autre une dérèpression (0.025 h^{-1}) (Ilyes *et al.*, 2004).

4.2.1.3 Hybridations et analyses des résultats

Les échantillons de matériel génétique récoltés ont été hybridés deux à deux sur puces à ADN suivant le plan expérimental présenté Figure 13. Chaque expérience a été réalisée en double (dye-switch) à partir de réplicats biologiques issus de cultures indépendantes.

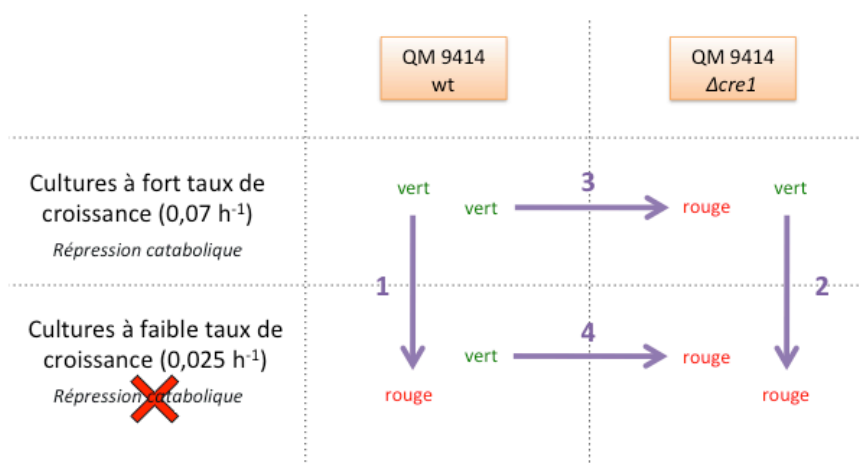


Figure 13. Plan expérimental utilisé pour étudier le rôle de CRE1 et des taux de croissance sur l'expression des gènes de *T. reesei*. Deux souches de *T. reesei* QM 9414 ont été comparées, l'une sauvage (wt), l'autre délétée pour le gène *cre1* ($\Delta cre1$). Deux taux de croissance ont été comparés, l'un induisant une répression catabolique ($0,07 \text{ h}^{-1}$), l'autre une dérèpression ($0,025 \text{ h}^{-1}$). Deux hybridations « dye-switch » ont été réalisées par expérience. Les nombres au-dessus des flèches indiquent les numéros d'expériences mentionnés dans le **Tableau annexe 4**. Les couleurs représentant les niveaux d'expression dans les groupes de la **Figure 15** sont également indiquées pour chaque condition.

Après prétraitement et normalisation, les données issues des hybridations ont été soumises à différents tests statistiques (voir méthodes partie 3.6.4). À l'issue de ces tests, 251 gènes dont l'expression a été significativement régulée dans au moins une des conditions expérimentales ont été identifiés. Le gène *cre1* lui-même en faisait partie, et a été exclu de la suite des analyses.

Suivant une méthode de classification hiérarchique non supervisée, l'ensemble des gènes a été réparti en neuf groupes d'après leur profil de régulation (Figure 14). La liste de l'ensemble des gènes, de leurs ratios d'expression dans chaque condition expérimentale et de leur groupe associé est présentée dans le Tableau annexe 4. Chaque gène a été annoté et s'est vu attribuer une ou plusieurs catégories fonctionnelles (voir méthodes parties 3.6.5 et 3.6.6). L'enrichissement des catégories fonctionnelles majeures a été testé pour chaque groupe, et une analyse d'enrichissement Gene Ontology (GO) a été effectuée sur l'ensemble des données en guise de contrôle supplémentaire (voir Tableau annexe 5 et Tableau annexe 6).

4.2.2 Caractéristiques générales de la régulation de l'expression des gènes

Les effets de CRE1 sur l'expression des gènes, en relation avec l'influence des taux de croissances, sont résumés dans la Figure 15.

4.2.2.1 Influence de CRE1

Alors qu'il est classiquement admis que CRE1 est un régulateur transcriptionnel négatif, nos résultats indiquent que seuls 47,3% des gènes identifiés comme étant régulés au cours des expériences sont réprimés par CRE1 (groupes C, E, G et H ; 118 gènes), tandis que 29% sont induits par le facteur de transcription (groupes D et F ; 72 gènes). Par ailleurs, 17,2% des gènes ont vu leur expression régulée indépendamment de CRE1 (groupes A et B ; 43 gènes), ce qui confirme que la répression catabolique met également en jeu des voies de régulation n'impliquant pas le facteur de transcription.

4.2.2.2 Influence du taux de croissance

62,2% des gènes ont vu leur expression régulée dans des conditions de répression catabolique (fort taux de croissance ; 155 gènes), mais seuls 44,5% d'entre eux sont réprimés (groupes A et E ; 69 gènes) tandis que les autres (55,5%) sont induits (groupes B, F et G ; 86 gènes). Dans des conditions de dérèpression (faible taux de croissance), 24,9% des gènes ont vu leur expression augmenter (groupes D et H ; 62 gènes), tandis que 6,4% des gènes sont régulés identiquement par CRE1 quel que soit le taux de croissance (groupe C ; 16 gènes).

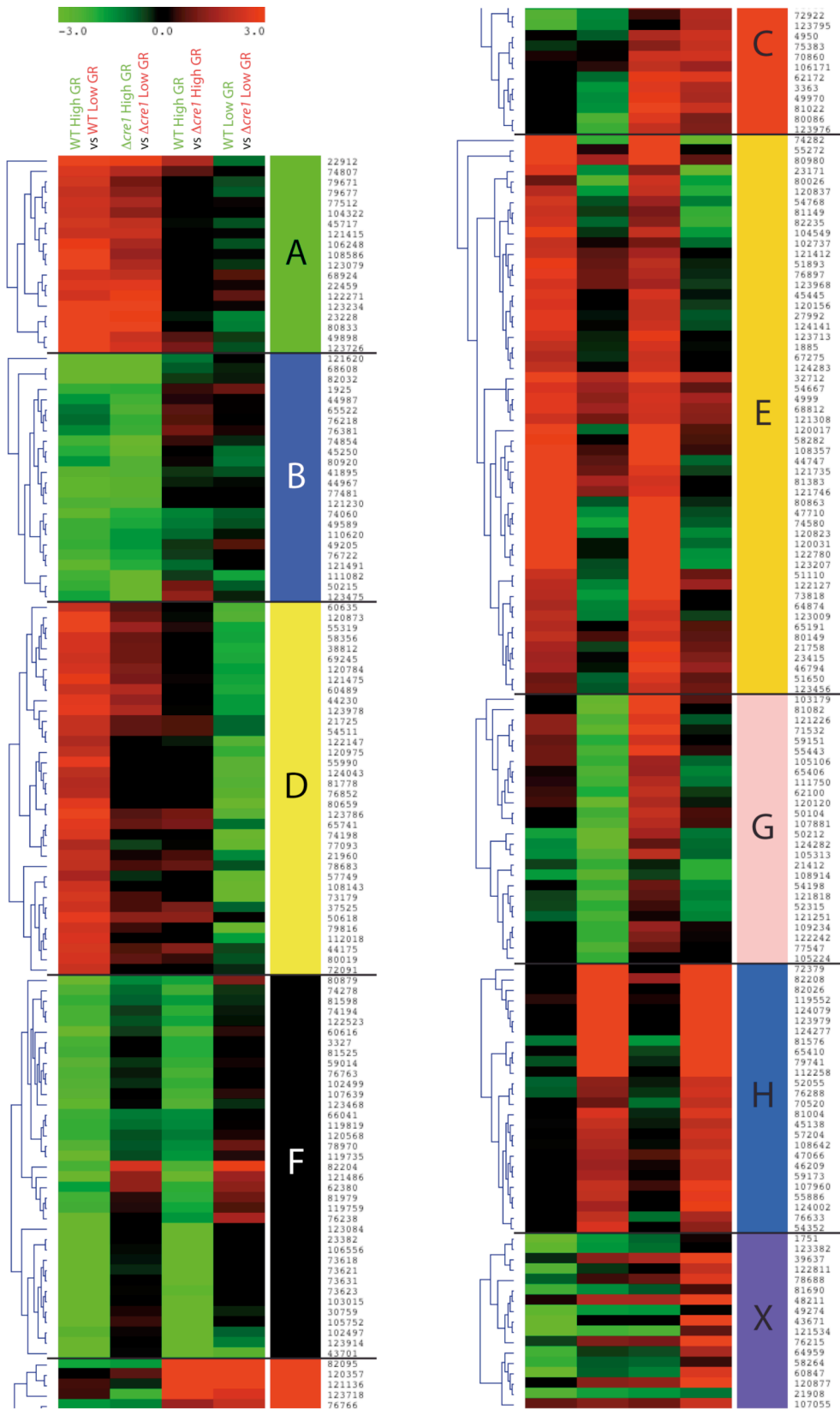


Figure 14. Heatmap des gènes les plus régulés. Le résultat de la classification hiérarchique non supervisée est présentée pour les 250 gènes identifiés comme étant les plus régulés (leur identifiant est indiqué au bout de chaque ligne). Les barres verticales de couleur indiquent les lettres identifiant le groupe correspondant. GR : Growth rate (taux de croissance) ; wt : QM 9414 souche sauvage

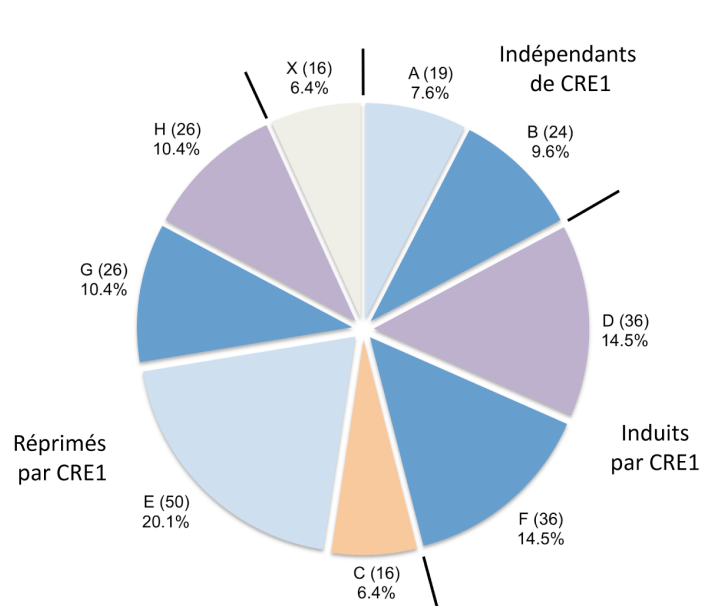


Figure 15. Distribution des gènes parmi les différents groupes d'expression. Les neuf groupes (de A à H et X) identifiés parmi les profils des gènes les plus fortement régulés ont été rassemblés en fonction de l'influence de CRE1 sur leurs gènes : répression, induction ou absence d'effet. Le code couleur indique l'influence du taux de croissance. Bleu : influence du fort taux de croissance (bleu foncé : régulation positive, bleu clair : régulation négative). Violet : influence du faible taux de croissance (régulation positive). Orange : pas d'influence du taux de croissance. Pour chaque groupe, le nombre de gènes (entre parenthèses) et le pourcentage du nombre total sont indiqués.

4.2.3 Comportement et fonctions principales des gènes régulés positivement en l'absence de CRE1

Comme présenté Figure 15, les gènes ayant vu leur expression augmenter dans la souche $\Delta cre1$ par rapport à la souche sauvage peuvent être classés en quatre catégories :

- (i) les gènes régulés positivement chez le mutant $\Delta cre1$ indépendamment du taux de croissance (groupe C, 16 gènes) ;
- (ii) les gènes régulés positivement chez le mutant $\Delta cre1$ à fort taux de croissance uniquement (groupe E, 50 gènes) ;
- (iii) les gènes dont l'induction à fort taux de croissance – détectée chez le mutant $\Delta cre1$ – est empêchée par CRE1 dans la souche sauvage (groupe G ; 26 gènes) ;
- (iv) les gènes dont l'induction à faible taux de croissance – détectée chez le mutant $\Delta cre1$ – est empêchée par CRE1 dans la souche sauvage (groupe H ; 26 gènes).

La Figure 16a présente la distribution, parmi ces quatre groupes, des catégories fonctionnelles principales des gènes identifiés. Les gènes associés au métabolisme des sucres sont retrouvés en majorité dans le groupe E, qui rassemble les gènes réprimés par CRE1 dans des conditions de répression catabolique – ce qui était attendu au vu des données de la littérature. Parmi

eux, on retrouve notamment des glucosidases, une aldose épimérase, ou encore une amylase. La répression catabolique dépendante de CRE1 ayant pour effet d'inhiber la sécrétion des CAZymes de *T. reesei* lorsque du glucose extracellulaire est détecté, ce groupe présente également en toute logique un enrichissement significatif en gènes impliqués dans le transport cellulaire. Un enrichissement pour cette même catégorie est détecté pour le groupe C, qui rassemble les gènes réprimés par CRE1 indépendamment du taux de croissance. Les protéines de transport identifiées dans le groupe E ont des natures diverses (transport de peptides, de sucres, mais aussi d'acides aminés notamment), tandis que le groupe C compte notamment des transporteurs de composés azotés.

Le groupe E (gènes réprimés par CRE1 dans des conditions de répression catabolique) présente également des gènes liés à la régulation de la transcription (par exemple l'hélicase SNF2 impliquée dans le remodelage de la chromatine (Dechassa *et al.*, 2008), ou le régulateur transcriptionnel MedA impliqué dans la sporulation (Ohara *et al.*, 2004)), ainsi que différents gènes participant au développement du champignon (RAS1, PhiA ou PAG1 par exemple). Cette fonction est aussi représentée par le gène *bys1* dans le groupe G, qui rassemble les gènes dont CRE1 empêche l'induction à fort taux de croissance. Le groupe G comprend également des gènes liés au métabolisme des sucres (glycosides hydrolases et glycosides transférases), ainsi que la protéine membranaire Hsp30, connue pour être induite par différents stress cellulaires, dont la limitation en glucose (Piper *et al.*, 1997). Comme le groupe E, le groupe rassemblant les gènes dont CRE1 empêche l'induction à faible taux de croissance (groupe H) comporte une majorité de gènes impliqués dans le transport cellulaire. Une protéine membranaire liée à la sporulation, CMP1, est également retrouvée dans ce groupe (Puyesky *et al.*, 1999).

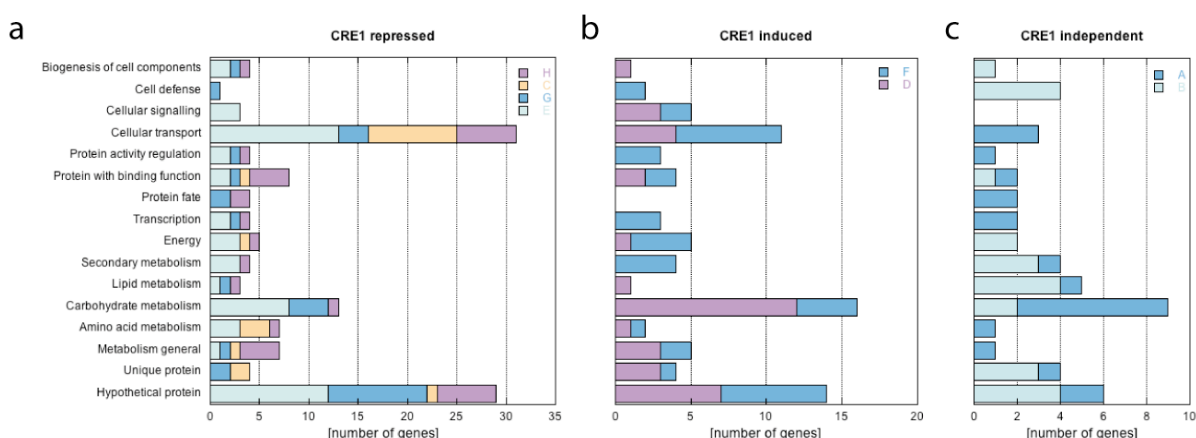


Figure 16. Distribution des gènes parmi les catégories fonctionnelles principales. Le nombre de gènes appartenant à chaque catégorie est présenté pour chaque groupe, suivant le même code couleur que celui de la **Figure 15**. Les sous-figures distinguent les gènes suivant la même classification que celle employée dans le texte : (a) gènes régulés positivement en l'absence de CRE1, (b) gènes régulés négativement en l'absence de CRE1, (c) gènes régulés par le taux de croissance indépendamment de CRE1. Unique protein : protéines n'ayant pas d'homologue ayant pu être identifié chez les ascomycètes.

4.2.4 Comportement et fonctions principales des gènes régulés négativement en l'absence de CRE1

Sur les 190 gènes identifiés comme étant significativement régulés en fonction de la présence ou de l'absence de CRE1 (groupes C à H), 72 ont vu leur expression diminuée dans le mutant *Δcre1* par rapport à la souche sauvage, ce qui suggère que CRE1 agirait positivement sur près de 38% de ses cibles directes ou indirectes. Trente-six gènes semblent induits par CRE1 à fort taux de croissance (groupe F), tandis que 36 autres sont induits par CRE1 à faible taux de croissance (groupe D). Les gènes qui voient leur expression régulée positivement par CRE1 dans des conditions de répression catabolique (groupe F) codent en majorité pour des protéines impliquées dans le transport cellulaire ou pour des protéines pour lesquelles la fonction n'a pas été caractérisée (Figure 16b). Ce groupe comprend également 4 gènes participant au métabolisme des sucres (dont une énoïase et une glyceraldéhyde deshydrogénase impliquées dans la glycolyse), 2 synthases et 2 oxidoréductases impliquées dans la formation de métabolites secondaires, ainsi que 3 facteurs de transcription à domaines Zn₂Cys₆ ou bZIP de liaison à l'ADN. Le groupe D, qui rassemble les gènes induits par CRE1 dans des conditions de dérpression, présente un enrichissement significatif en gènes de CAZymes (dont 9 glycoside hydrolases, glucanases et mannosidases notamment). Sa composition est par ailleurs assez hétéroclite, avec un grand nombre de fonctions différentes représentées – les protéines impliquées dans le transport cellulaire et les protéines non caractérisées étant les plus abondantes.

4.2.5 Comportement et fonctions principales des gènes régulés par le taux de croissance indépendamment de CRE1.

Parmi les 250 gènes identifiés comme étant significativement régulés dans au moins une des conditions expérimentales, 43 présentent un profil d'expression indépendant de la présence ou de l'absence de CRE1 (17,2%). Pour ces gènes, les variations de taux de croissance au sein d'une même souche engendrent en effet une modification notable de leur expression, mais aucune différence n'est détectée lors des comparaisons inter-souches à taux de croissance identique. Dans des conditions de répression catabolique (fort taux de croissance), 19 d'entre eux sont réprimés (groupe A) tandis que 24 sont régulés positivement (groupe B). Parmi les gènes réprimés, on note un enrichissement significatif de gènes impliqués dans le métabolisme des sucres, liés pour plus de la moitié d'entre eux au catabolisme de la N-acetylglucosamine en fructose-6-phosphate. Les gènes induits présentent des fonctions variées, le métabolisme des acides gras et la défense cellulaire étant les plus représentées (Figure 16c).

4.2.6 Un groupe aux profils hétérogènes

Un dernier groupe, nommé « X » a été identifié dans notre étude. Les gènes de ce groupe présentent des profils d'expression très hétérogènes, mais chez la souche sauvage, leur expression est systématiquement plus élevée à fort taux de croissance. À faible taux de croissance, leur expression est également systématiquement plus élevée dans le mutant *Δcre1*. Les gènes auxquels des fonctions ont été attribuées ne semblent pas présenter de caractéristiques communes. Un orthologue du gène *creD* d'*A. nidulans* y est cependant représenté. Une mutation de ce gène, qui code pour une protéine d'interaction avec les ubiquitines ligases (voir partie 2.3.2.4), semble engendrer une levée partielle de la répression catabolique chez *A. nidulans* (Boase and Kelly, 2004).

4.2.7 Présence de sites de liaison CRE1 dans les régions 5' des gènes identifiés

Dans le but de préciser si les effets régulateurs de CRE1 identifiés dans cette étude sont directs ou indirects, une recherche systématique du motif de liaison CRE1 (5'-SYGGRG-3') a été réalisée grâce à la suite logicielle RSAT (Thomas-Chollier *et al.*, 2008) parmi les séquences situées en amont des régions codantes des gènes identifiés, dans une fenêtre de 1kb. Pour chaque groupe, un nombre moyen de sites par gène a été établi et normalisé par rapport au nombre moyen de sites par gène enregistré dans l'ensemble du génome de *T. reesei*. La fixation de CRE1/CreA ayant été décrite comme nécessitant le plus souvent deux sites adjacents (Cubero and Scazzocchio, 1994; Mach *et al.*, 1996), un comptage des motifs présents par paires (répétitions directes ou inversées) a également été effectué pour chaque gène, puis moyenné pour chaque groupe et normalisé par rapport à la moyenne de ces « motifs doubles » par gène sur l'ensemble du génome. Deux distances, de 20 pb ou de 50 pb maximum entre les deux motifs, ont été testées. Pour une distance de 20 pb, le nombre moyen de motifs doubles (tous types confondus) par gène sur l'ensemble du génome est de 0,5 tandis qu'il est de 0,9 pour une distance de 50 pb. La significativité des différences enregistrées entre la moyenne par groupe et la moyenne sur l'ensemble du génome a été testée au moyen de *t* tests de Student bidirectionnels, mais les écart-types élevés – qui s'expliquent par une fréquence relativement forte des motifs, combinée à la présence possible dans les groupes de cibles aussi bien directes qu'indirectes de CRE1 – rendent les résultats de cette analyse difficiles à valider statistiquement. Néanmoins, comme présenté dans le Tableau 4, les motifs de liaison de CRE1 sont significativement plus abondants en amont des gènes du groupe E (gènes réprimés par CRE1 en conditions de répression catabolique) et du groupe D (gènes induits par CRE1 en conditions de dérèglement). Un enrichissement semble également être détecté pour les gènes du groupe C (que CRE1 réprime indépendamment du taux de croissance), mais n'a pu être validé statistiquement.

Aucun enrichissement n'est détecté pour les groupes F et G (gènes induits par CRE1 en conditions de répression catabolique et gènes pour lesquels CRE1 s'oppose à l'induction à fort taux de croissance, respectivement), tandis que les gènes du groupe H (gènes pour lesquels CRE1 s'oppose à l'induction à faible taux de croissance) semblent présenter un nombre légèrement plus important de motifs doubles espacés de 20 pb que la moyenne du génome. Ces résultats suggèrent que les groupes D, E, et peut-être C sont enrichis en cibles directes de CRE1, tandis que les groupes F, G et peut-être H, seraient soumis à une régulation indirecte de CRE1. Les valeurs obtenues pour les gènes des groupes A et B (régulés indépendamment de CRE1) sont situées dans ou en-dessous de la moyenne du génome, ce qui confirme la cohérence de nos résultats.

Nous avons également cherché si d'autres motifs nucléotidiques étaient partagés par les gènes identifiés dans notre étude. L'un d'entre eux, 5'-SAGGGRG-3', habituellement retrouvé 0,57 fois par gène en moyenne chez *T. reesei*, est significativement enrichi dans les groupes D et E présentant déjà un enrichissement en motifs CRE1.

Motifs	SYGGRG	SYGGRG ... CYCCRS CYCCRS ... SYGGRG SYGGRG ... SYGGRG		SAGGGRGR	
		distance de 20bp	distance de 50bp		
Nombre moyen par gène dans le génome entier	4.13	0.47	0.9	0.57	
Enrichissements dans les groupes	A	-8%	-66%**	-24%	-17%
	B	-3%	-2%	-7%	+17%
	C	+26%	+33%	+11%	+34%
	D	+23%**	+36%	+48%**	+105%**
	E	+16%**	+58%*	+26%	+72%**
	F	-3%	+9%	+11%	-15%
	G	+1%	-43%*	-6%	-26%
	H	+1%	+23%*	+3%	+8%
	X	-16%	-50%**	-28%	+24%

Tableau 4. Résultats de l'analyse des promoteurs des gènes régulés. Les enrichissements par groupes, comparés aux valeurs moyennes du génome, sont présentés. Les enrichissements significatifs, principalement discutés dans le texte, sont présentés en gras (*t*-test de Student : * 0,05<p-value<0,1 ; ** p-value<0,05).

4.2.8 Discussion

Le facteur de transcription CRE1/CreA est un des régulateurs majeurs ayant été identifié chez les champignons filamenteux. Acteur central de la répression catabolique, son rôle de régulateur négatif a notamment été prouvé chez *T. reesei* sur des gènes impliqués dans le métabolisme des polysaccharides (sur les gènes *cbh1* et *xyn1* notamment)(Mach *et al.*, 1996; Takashima *et al.*, 1996), ou chez *A. nidulans* et *A. niger* sur l'utilisation de l'éthanol ou d'acides aminés comme sources de

carbone (Felenbok *et al.*, 2001; Panozzo *et al.*, 1998). Si l'impact de la répression catabolique sur l'expression de certains transcrits a été largement décrit, les travaux publiés à ce jour ne permettent pas d'établir une vision plus globale du processus. De plus, nos données suggèrent que notre connaissance des effets de CRE1/CreA est incomplète (voir partie 4.1.9.5). L'objectif de ce travail n'était donc pas d'expliquer au niveau moléculaire comment certains gènes sont régulés directement ou indirectement par CRE1, mais de donner un aperçu général de la réponse transcriptionnelle engendrée par la répression catabolique, et de préciser quels gènes sont affectés par une perte de fonction du régulateur majeur de ce processus.

4.2.8.1 CRE1, un régulateur majeur combinant les fonctions d'activateur et de répresseur

Notre approche expérimentale, menée sur la totalité du transcriptome de *T. reesei* et dans des conditions de culture minutieusement contrôlées, démontre clairement que la protéine CRE1 est un facteur de transcription agissant de manière complexe sur un grand nombre de gènes et à de nombreux niveaux. Nos données prouvent notamment que le taux de croissance des cellules influe sur le rôle de CRE1 ; l'état physiologique du champignon semble ainsi pouvoir modifier le comportement de ce facteur de différentes manières en fonction des cibles considérées. Au total, neuf groupes rassemblant 250 gènes de *T. reesei* ont ainsi pu être définis à partir des différents profils d'expressions rencontrés. Deux d'entre eux présentent des gènes ayant vu leur expression diminuée en l'absence du facteur de transcription. Alors qu'il était jusque là communément considéré comme un répresseur, CRE1 pourrait donc, comme le suggéraient nos résultats précédents, activer la transcription de près de 38% de ses cibles. Ce chiffre est supérieur à celui enregistré en 2006 chez *A. nidulans* par Morgensen *et al.* (voir partie 4.1.9.5), mais l'étude n'avait été menée à l'époque que sur une partie du génome. Au-delà de la répression catabolique, CRE1 semble donc jouer un rôle majeur dans la régulation de l'assimilation des composés carbonés. La répression catabolique semble à l'inverse ne pas se résumer exclusivement aux effets de CRE1 puisque 17,2% des gènes identifiés dans notre étude sont régulés *via* des mécanismes indépendants du facteur de transcription.

4.2.8.2 Les protéines de transport : des cibles privilégiées de CRE1

Sur les 190 gènes identifiés comme étant régulés par CRE1, 118 sont réprimés par le facteur de transcription. Hormis les protéines dont la fonction n'est pas caractérisée, la majorité de ces gènes codent pour des protéines impliquées dans le transport cellulaire, notamment des perméases, ce qui semble indiquer que la répression catabolique dépendante de CRE1 agit très largement sur l'entrée des substrats dans la cellule ainsi que sur les voies impliquées dans la sécrétion. Il avait déjà été démontré en ce sens qu'une hexose perméase d'*A. nidulans* à forte affinité – permettant aux

cellules de détecter efficacement d'infimes quantités de glucose – était soumise à la répression catabolique dépendante de CreA (MacCabe *et al.*, 2003). Étonnamment, quelques unes de ces protéines de transport sont réprimées par CRE1 indépendamment du taux de croissance. C'est notamment le cas de perméases impliquées dans le transport de composés azotés, ce qui suggère que les peptides et leurs produits de dégradation pourraient constituer des substituts aux sucres facilement métabolisables lorsque ces derniers sont indisponibles. Une augmentation de la consommation d'acides aminés en conditions de répression catabolique a déjà été détectée chez *S. cerevisiae*, mais le phénomène était indépendant de Mig1p (homologue de CRE1) dans ce cas (Peter *et al.*, 2006). L'habitat naturel de *T. reesei* étant pauvre aussi bien en azote qu'en sources de carbones induisant une répression catabolique, ce mécanisme pourrait favoriser l'assimilation des composés azotés disponibles. Quelques protéines de transport, enfin, sont induites par CRE1 dans des conditions de répression. Il s'agit notamment de transporteurs du métabolisme secondaire impliqués dans la résistance et la survie cellulaire.

4.2.8.3 CRE1 agit sur l'expression basale de certaines CAZymes

Il est généralement admis que la transcription des gènes de CAZymes est réprimée par CRE1. Nos résultats le confirment pour certains de ces gènes, mais seule une petite proportion du CAZome de *T. reesei* est représentée dans notre étude. Des glycosyltransférases, des carbohydate estérases et des glycoside hydrolases (dont un bon nombre de glucosidases) sont régulées négativement dans des conditions de répression catabolique, mais les enzymes cellulolytiques ou hémicellulolytiques, dont la régulation par CRE1 a été démontrée précédemment (Mach *et al.*, 1996; Nakari-Setälä *et al.*, 2009), ne sont pas régulées positivement dans le mutant $\Delta cre1$. Nos résultats confirment donc de précédentes observations suggérant que la répression catabolique dépendante de CRE1 affecte l'induction de ces enzymes, mais pas leur expression basale (Karaffa *et al.*, 2006). Ainsi, si notre étude identifie 250 gènes dont l'expression a été fortement régulée par le taux de croissance et/ou le facteur CRE1 (ce qui représente 2,8% de l'ensemble des gènes prédits de *T. reesei*), le nombre de gènes contrôlé par ces facteurs est donc probablement encore supérieur, et doit dépendre dans de nombreux cas de la source de carbone présente dans l'environnement cellulaire.

4.2.8.4 Influence de CRE1 sur les régulateurs transcriptionnels

Malgré l'impact très large de CRE1 sur l'expression génétique, il est intéressant de noter que peu de facteurs de transcription font partie des cibles du régulateur. Dans des conditions de répression catabolique, on observe même une réduction significative du nombre de facteurs de transcription parmi les cibles de CRE1 par rapport à l'ensemble du génome. Quelques gènes clés

impliqués dans des mécanismes transcriptionnels voient tout de même leur expression régulée par la protéine. C'est notamment le cas de Med2, composant du « complexe médiateur ». Cet édifice moléculaire, partiellement conservé chez l'ensemble des eucaryotes, assure un dialogue entre facteurs de transcriptions et machinerie transcriptionnelle (Malik and Roeder, 2000), et il a été montré que Med2 était responsable, avec la protéine Gal11, des activations les plus fortes chez *S. cerevisiae* (Balciunas *et al.*, 2003), ce qui en fait une cible de choix dans des mécanismes de régulation. D'après nos résultats, CRE1 s'oppose à son induction à faible taux de croissance, ce qui lui permet probablement de contrôler du même coup l'expression d'un grand nombre de gènes dont l'induction serait peu économe en terme d'énergie dans un tel contexte cellulaire. Un composant du complexe Swi/Snf de remodelage de la chromatine, le gène *snf2*, est également réprimé par CRE1, dans des conditions de répression catabolique. Ce complexe, capable de modifier la structure chromatinienne en agissant sur les interactions ADN-histones, est impliqué dans la régulation transcriptionnelle d'environ 5% des gènes de *S. cerevisiae* (Becker and Horz, 2002; Sudarsanam *et al.*, 2000). La protéine Snf2 y joue un rôle clé, puisqu'elle est associée à l'enroulement de l'ADN autour du nucléosome (Dechassa *et al.*, 2008). De précédentes observations avaient établi qu'une perte de fonction de CRE1/CreA entraînait une altération dans le positionnement du nucléosome en présence de glucose (Garcia *et al.*, 2004; Zeilinger *et al.*, 2003). Notre identification de *snf2* comme cible de CRE1 explique donc ces observations. CRE1 affecte enfin l'expression de quelques autres facteurs de transcriptions potentiels (protéines à domaine Zn₂Cys₆ ou bZIP), ainsi que celle de MedA, régulateur impliqué dans des procédés développementaux.

4.2.8.5 CRE1, RAS1, et le développement de *T. reesei*

Une étude récente a démontré qu'une perte de fonction de CRE1/CreA induisait des modifications dans la morphologie des cellules et une altération de la sporulation. Nos observations de la souche *Δcre1* confirment ces modifications (voir Figure annexe 1) que certains des gènes identifiés dans nos groupes pourraient expliquer. La protéine RAS1, petite GTPase réprimée par CRE1 à fort taux de croissance, pourrait en effet jouer un rôle central dans ces altérations phénotypiques. Chez *Schizosaccharomyces pombe*, Ras1 régule deux voies moléculaires distinctes : l'une contrôle la différenciation sexuelle nécessaire à la conjugaison *via* la cascade Byr2/MAPK (Van Aelst *et al.*, 1993), l'autre permet le maintien d'une morphologie cellulaire allongée par l'intermédiaire de la voie Scd1/Cdc42 (Chang *et al.*, 1994). Chez *Neurospora crassa* et *A. fumigatus*, les orthologues de RAS1 régulent le développement asexué, la morphologie, et la synthèse et l'intégrité de la membrane cellulaire (Fortwendel *et al.*, 2008; Kana-uchi *et al.*, 1997). Enfin, la protéine RAS1 de *S. cerevisiae* est régulée par le glucose (Breviario *et al.*, 1988). Il est donc tout à fait possible que RAS1 soit impliquée,

chez *T. reesei*, dans les conséquences phénotypiques de l'absence de CRE1, voire dans la régulation d'autres gènes identifiés dans nos groupes. Trois autres gènes liés à la morphologie ou à la sporulation sont en effet également réprimés par CRE1 à fort taux de croissance : ils codent pour les protéines PAG1, PhiA, et pour le facteur de transcription MedA. Chez *S. cerevisiae* Pag1 est associée à la protéine kinase Cbk1p nécessaire à la morphogénèse et à la prolifération cellulaire (Du and Novick, 2002). *MedA* code pour un orthologue des régulateurs transcriptionnels REN1 de *Fusarium oxysporum*, ACR1 de *Magnaporthe grisea* et MedA d'*A. nidulans*, impliqués dans le bon déroulement de la conidiogénèse (Ohara *et al.*, 2004), processus mettant en jeu la protéine membranaire PhiA (Melin *et al.*, 2003). À fort taux de croissance, CRE1 s'oppose enfin à l'induction d'un orthologue de BYS1, une protéine extracellulaire de *Blastomyces dermatitis* pouvant être impliquée dans la morphologie cellulaire (Bono *et al.*, 2001). L'ensemble de ces données suggère donc que CRE1, et son interaction probable avec RAS1, pourrait constituer un facteur régulateur important de procédés liés au développement de *T. reesei*.

4.2.8.6 Un nouveau motif de liaison potentiel ?

Si notre étude ne permet pas d'établir précisément si les régulations de CRE1 enregistrées pour chaque gène sont directes ou indirectes, notre recherche des occurrences de la séquence cible de CRE1 précédemment définie (5'-SYGGRG-3') confirme son implication dans les effets du facteurs de transcription, et révèle que deux des six groupes régulés par CRE1 sont enrichis en ce motif. Un enrichissement encore plus important est détecté lorsque l'on considère des « motifs doubles » de cette séquence, ce qui confirme la fonctionnalité précédemment décrite de tels motifs (Cubero and Scazzocchio, 1994; Mach *et al.*, 1996). Nos recherches ont également permis d'identifier une nouvelle séquence consensus (5'-SAGGGRGR-3') surreprésentée dans les promoteurs des gènes régulés par CRE1. Ce motif, riche en GC, n'a jamais été décrit chez les champignons filamenteux ni même chez d'autres organismes ; il nous est donc impossible d'établir quelles protéines pourraient s'y fixer. Cependant, une étude sur les motifs présents dans les promoteurs de *S. cerevisiae* régulés par Mig1/Mig2/Mig3 a précédemment identifié le motif 5'-GGGAGG-3', qui est pleinement inclus dans la séquence 5'-SAGGGRGR-3' (Westholm *et al.*, 2008). Les auteurs de cette étude suggèrent qu'un facteur de transcription régulant des gènes impliqués dans le métabolisme du phosphate pourrait se fixer à ce motif, mais aucun des gènes de nos groupes possédant une séquence 5'-SAGGGRGR-3' dans leur promoteur n'est impliqué dans une telle fonction. Étonnamment, la présence ou l'absence des séquences 5'-SYGGRG-3' et 5'-SAGGGRGR-3' dans les promoteurs des gènes régulés n'est pas corrélée à la nature de la régulation exercée par CRE1, puisque ces motifs sont retrouvés chez des gènes réprimés comme chez des gènes induits.

En résumé, cette deuxième étude a permis de confirmer nos premiers résultats (voir partie 4.1.9.5) suggérant que la protéine CRE1 pouvait, en plus de ses fonctions reconnues de répresseur, jouer un rôle d'activateur transcriptionnel pour certains gènes de *T. reesei*. Plus généralement, nos résultats ont permis d'identifier un grand nombre de cibles de ce régulateur majeur, jusque là non caractérisées. Nos observations montrent notamment l'influence de CRE1 sur l'expression de différents transporteurs cellulaires ou de protéines impliquées dans le développement de *T. reesei*. Des gènes influencés par la répression catabolique indépendamment de ce facteur ont également été identifiés. L'ensemble de ces travaux a fait l'objet d'une publication présentée en annexe (Portnoy *et al.*, 2011a).

4.3 Analyse de l'expression génétique globale au cours de la production de cellulases

Après avoir décrit le comportement des régulateurs majeurs des gènes de cellulases et identifié l'ensemble des cibles de la répression catabolique dépendante ou non du facteur de transcription CRE1, nous avons souhaité élargir notre connaissance des mécanismes impliqués dans la production d'enzymes cellulolytiques, en examinant sans *a priori* l'ensemble des régulations de l'expression génétique mises en jeu au cours du processus d'induction par le lactose. Afin de dresser un tableau général du comportement des gènes de *T. reesei* à différents temps de la cinétique et d'identifier d'éventuelles voies dont la régulation influencerait les capacités de production des souches, cette approche globale a été menée sur les mutants plus ou moins producteurs de la lignée QM6a-NG14-RUT C30 (voir partie 2.2.4), cultivés en bioréacteur.

4.3.1 Description du dispositif expérimental

4.3.1.1 Cultures

Des productions de cellulases des souches NG14 (moyennement productrice) et RUT C30 (hyperproductrice) ont été réalisées suivant le protocole industriel en bioréacteur présenté précédemment (voir partie 4.1.1). Les cellules ont été cultivées en présence de D-glucose à une concentration initiale de 30 g.L⁻¹ durant une phase de croissance (ou « batch ») d'une trentaine d'heures. Puis, lorsque la concentration en glucose a atteint moins de 20% de la concentration initiale (6 g.L⁻¹), la production de cellulases a été induite par l'injection régulière de lactose dans le milieu de culture à un taux de 0,98 g.h⁻¹ (phase de production ou « fed-batch »). Les Figure 17A et B présentent les courbes de croissance, de consommation de sucres, et d'évolution du taux spécifique de production de protéines pour les souches NG14 et RUT C30, respectivement. Une seconde culture indépendante de la souche RUT C30, réalisée suivant les mêmes conditions, présente un comportement proche de celui de la Figure 17B (voir Figure annexe 2). Une comparaison de l'évolution de la concentration en protéines extracellulaires au cours des cultures est présentée Figure 18. Durant le batch, les cellules de NG14 et de RUT C30 se sont multipliées en consommant le glucose disponible dans le milieu. Une fois l'intégralité de ce sucre consommé, le lancement du fed-batch a vu une stabilisation de la concentration en biomasse de RUT C30, comme attendu. NG14 a quant à elle légèrement poursuivi sa croissance au cours de l'injection de lactose. Ce comportement indique que la souche utilise une fraction de ce sucre inducteur à des fins de multiplication, tandis que l'utilisation du lactose est probablement uniquement dédiée chez RUT C30 à l'induction des cellulases. En outre, la concentration en protéines extracellulaires

augmente immédiatement après le début du fed-batch chez RUT C30, tandis qu'un temps de latence d'une dizaine d'heures est détecté pour NG14. Aucune accumulation de lactose n'est cependant détectée dans aucune des cultures. Après 24h de fed-batch, on enregistre un niveau de protéines déjà 2 fois plus élevé dans la culture RUT C30 (1 g.L^{-1}) que dans la culture NG14 ($0,5 \text{ g.L}^{-1}$).

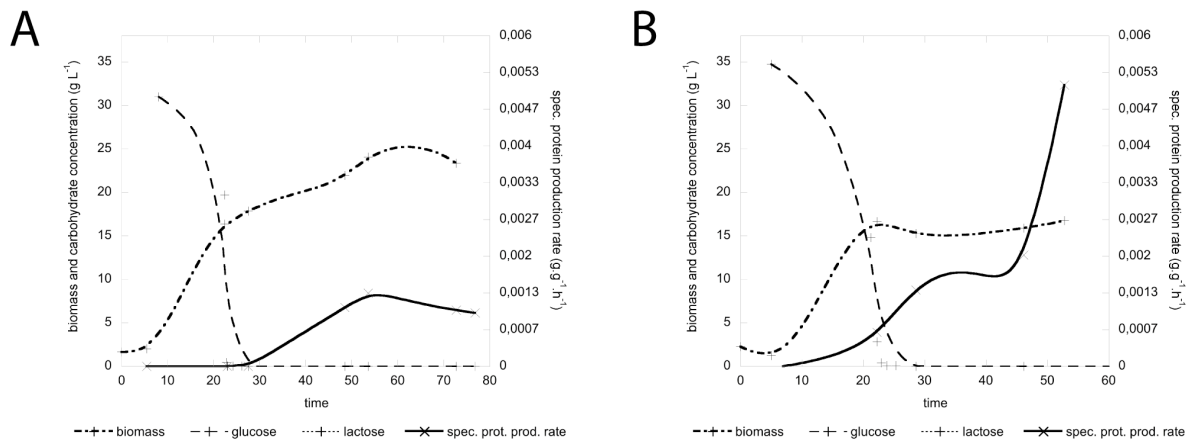


Figure 17. Profils des cultures de NG14 et RUT C30 au cours de l'expérimentation. Après une phase de croissance en glucose, les souches NG14 (A) et RUT C30 (B) ont été cultivées en fed-batch lactosé induisant une production de cellulases. La concentration en biomasse sèche, la consommation de sucres et les taux spécifiques de production de protéines ont été enregistrés au cours des cultures.

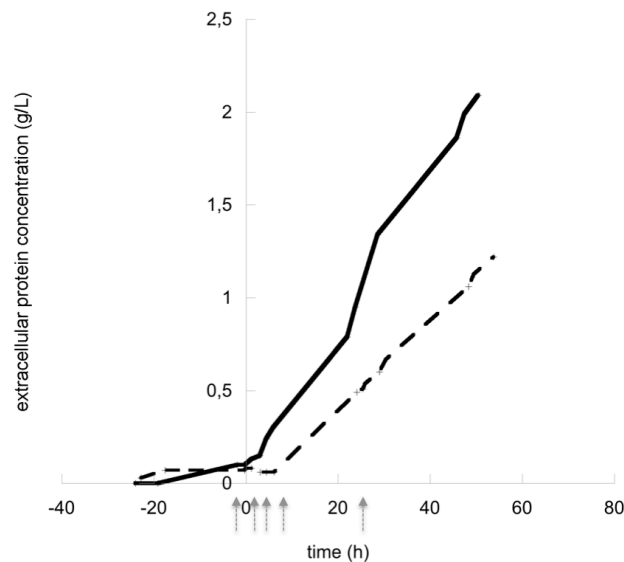


Figure 18. Evolution de la concentration en protéines extracellulaires au cours des cultures de NG14 et RUT C30. La concentration en protéines présentes dans le milieu de culture de NG14 (ligne pointillée) et de RUT C30 (ligne pleine) est présentée au cours du temps par rapport au lancement du fed-batch (temps 0). Cette valeur est représentative de la quantité de cellulases sécrétées par les cellules (voir partie 4.1.1). Les flèches représentent les instants auxquels sont effectués les prélèvements destinés à l'analyse transcriptomique.

4.3.1.2 Hybridations et analyses des résultats

Afin de comparer l'expression génétique globale des souches au cours de la production des cellulases, des extractions d'ARN ont été effectuées pour chaque culture à l'issue de la phase de croissance (référence T0 avant l'injection de lactose), puis 1h, 3h, 6h et 24h après le début de l'induction au lactose (voir Figure 18). Les échantillons ainsi récoltés ont été hybridés sur des puces à ADN dédiées à l'exploration du transcriptome de *T. reesei* (voir méthodes partie 3.6.2) suivant le plan expérimental présenté Figure 19. Chaque échantillon correspondant à un des temps de la cinétique d'induction a été hybridé contre la référence T0 récoltée avant le début de cette induction. Un réplicat technique a été réalisé pour chacune des conditions étudiées (dye-swap), et les réplicats biologiques de la seconde culture de RUT C30 ont été soumis au même protocole.

Afin d'établir d'éventuelles différences dans l'expression génétique basale (i.e. hors induction) entre la souche hyperproductrice RUT C30 et son ancêtre faiblement producteur QM6a, l'échantillon de référence (T0) de RUT C30 a été également hybridé avec un échantillon du matériel génétique de QM6a récolté dans les mêmes conditions à l'issue d'un batch glucose (T0 QM6a).

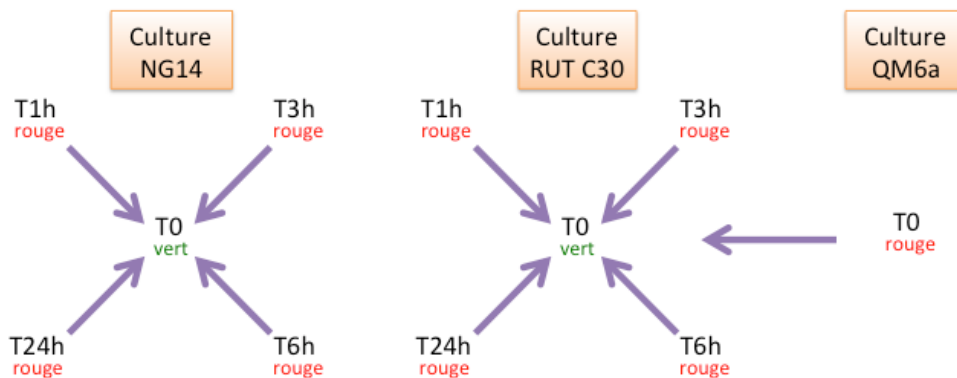


Figure 19. Plan expérimental employé pour étudier les régulations transcriptionnelles mises en jeu chez QM6a, NG14 et RUT C30. Après une phase de croissance en glucose, les souches NG14 et RUT C30 de *T. reesei* ont été cultivées en fed-batch lactose. Pour chaque souche, les échantillons de matériel génétique récoltés 1h, 3h, 6h et 24h après le début de l'induction ont été hybridés contre un échantillon prélevé juste avant le lancement du fed-batch (T0). L'échantillon T0 de la culture de RUT C30 a également été hybridé contre un échantillon T0 d'une culture de QM6a réalisée dans les mêmes conditions. Un réplicat technique de chaque hybridation (dye-swap) a été réalisé par inversion des fluorochromes. Les hybridations RUT C30 ont chacune été réalisées pour les deux réplicats biologiques.

Après prétraitement et normalisation, différents filtrages et tests statistiques ont été appliqués aux données issues des hybridations (voir méthodes parties 3.6.4 et 3.6.5). À l'issue de cette procédure, 592 gènes ayant vu leur expression significativement régulée dans au moins une des neuf conditions expérimentales testées ont été identifiés. Des méthodes de classification hiérarchique non supervisée et de nuées dynamiques ont permis de répartir l'ensemble de ces gènes en dix groupes en fonction de leur profil de régulation (Figure 20).

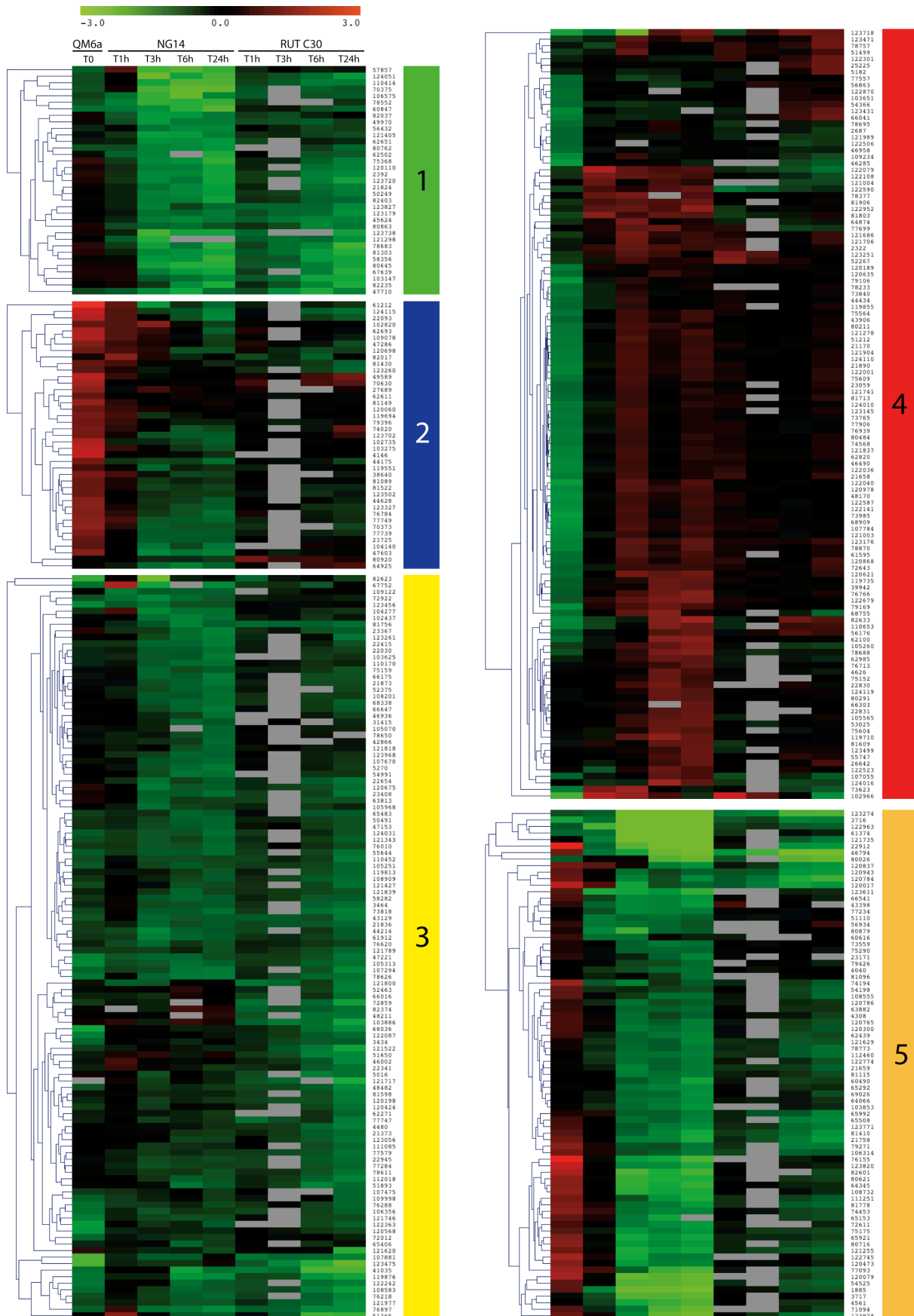


Figure 20. Heatmap des gènes significativement régulés. Les 592 gènes identifiés comme étant significativement régulés dans les conditions expérimentales testées ont été rassemblés en 10 groupes numérotés de 1 à 10. L'identifiant de chaque gène est donné en bout de ligne. Les cellules grises correspondent aux valeurs manquantes. Suite page suivante.

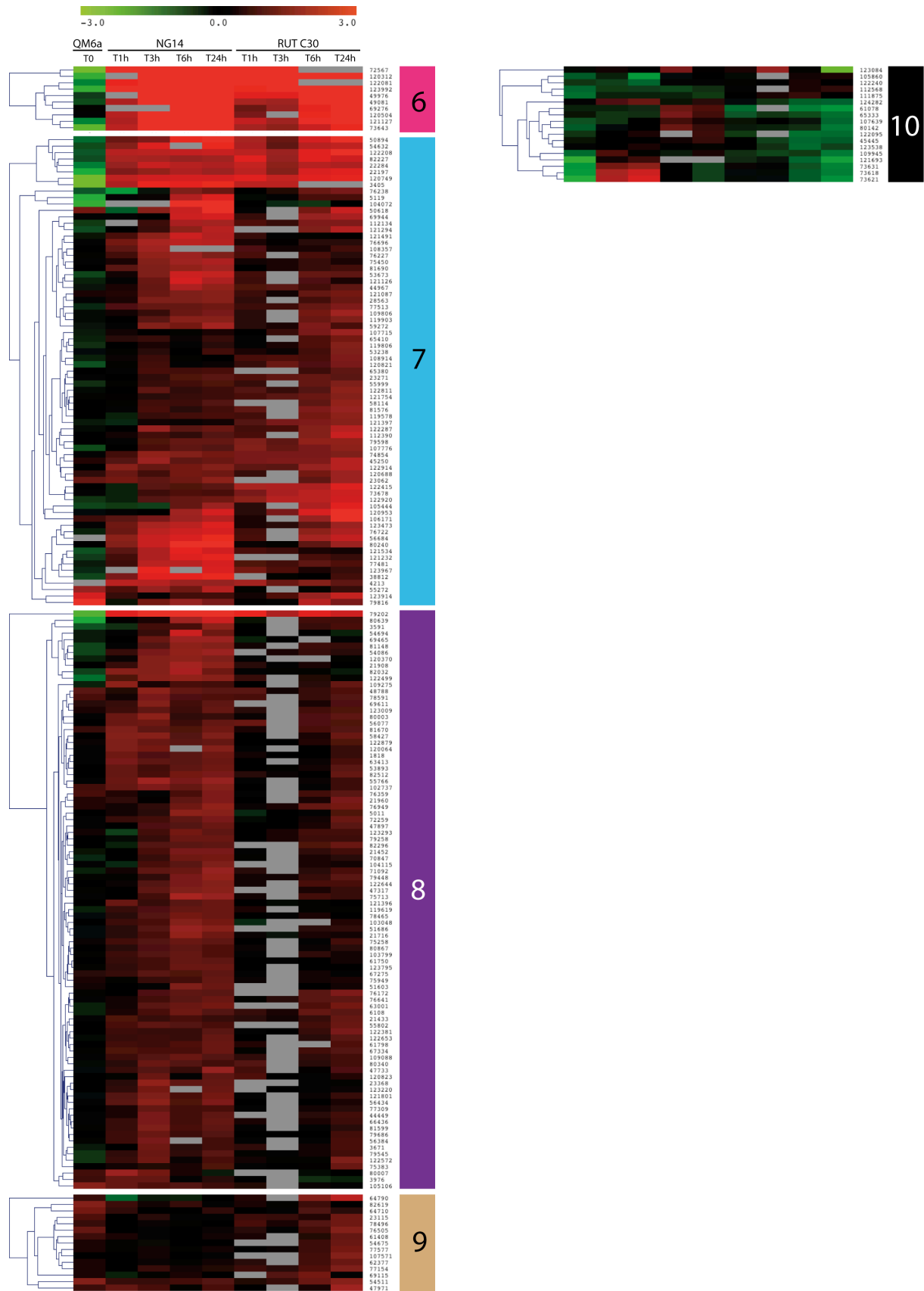


Figure 20. Heatmap des gènes significativement régulés. Suite.

Une annotation manuelle a permis d'identifier la fonction de chaque gène, et des analyses d'enrichissement GO ont été effectuées pour chacun des groupes (voir Tableau annexe 8). Une analyse d'enrichissement global a été également effectuée sur l'ensemble des données (voir Tableau annexe 9). La liste de l'ensemble des gènes, de leur fonction, de leurs ratios d'expression dans chaque condition expérimentale et de leur groupe associé est présentée dans le Tableau annexe 7. Par souci de simplification, les noms de gènes employés dans la suite du manuscrit se réfèrent à leurs meilleurs homologues identifiés dans les génomes des autres ascomycètes (voir méthodes partie 3.6.6 et les annotations complètes Tableau annexe 7).

Afin de bénéficier d'informations sur les niveaux d'expression absolus des transcrits d'intérêt, des analyses par séquençage haut débit des ADNc (RNA-seq) ont également été effectuées sur les échantillons les plus précoces (T0 à T6h) d'une des deux cinétiques d'induction de RUT C30. Différentes données issues de ces analyses seront mentionnées au cours de la discussion.

4.3.2 Caractéristiques générales de la régulation de l'expression des gènes

La nature des régulations mises en jeu dans chaque groupe de gènes au cours de la production de cellulases et les différences d'expression basale entre RUT C30 et QM6a, sont résumées dans la Figure 21.

4.3.2.1 Régulations mises en jeu au cours de l'induction de la production de cellulases chez NG14 et RUT C30

Sur les 592 gènes identifiés comme étant significativement régulés dans l'une de nos conditions expérimentales, 51,7% ont vu leur expression induite au cours de la production de cellulases (groupes 6 à 9 et 4 ; 306 gènes), tandis que 45,3% ont été réprimés (groupes 1 à 3 et 5 ; 268 gènes). Parmi les gènes induits, 61,4% l'étaient aussi bien chez NG14 que chez RUT C30 (groupes 6 à 9 ; 188 gènes soit 31,8% du total), mais 38,6% ne l'étaient que chez NG14 (groupe 4 ; 118 gènes soit 19,9% du total). 84,7% des gènes réprimés au cours de la production de cellulases étaient régulés chez les deux souches (groupes 1, 3 et 5 ; 227 gènes soit 38,3 % du total), tandis que 15,3% (groupe 2 ; 41 gènes soit 6,9% du total) ne l'étaient que chez NG14.

4.3.2.2 Régulations de l'expression des gènes hors induction entre QM6a et RUT C30

40,9% des gènes répertoriés présentaient une expression basale plus forte chez RUT C30 que chez QM6a (groupes 3, 4 et 6 ; 242 gènes), dont 41,8% des gènes induits au cours de la production (groupes 4 et 6 ; 128 gènes) et 42,5% des gènes réprimés (groupe 3 ; 114 gènes). 22,6% des gènes étaient à l'inverse moins exprimés hors induction chez l'hyperproductrice que chez la

souche ancestrale (groupes 2, 5 et 9 ; 134 gènes), dont 4,9% des gènes induits (groupe 9 ; 15 gènes) et 44,4% des gènes réprimés (groupes 2 et 5 ; 119 gènes). 33,4% ne présentaient pas de différence significative de leur niveau d'expression basal entre les deux souches (groupes 1, 7 et 8 ; 198 gènes), soit 53,5 % des induits (groupes 7 et 8 ; 163 gènes) et 13,1 % des réprimés (groupe 1 ; 35 gènes).

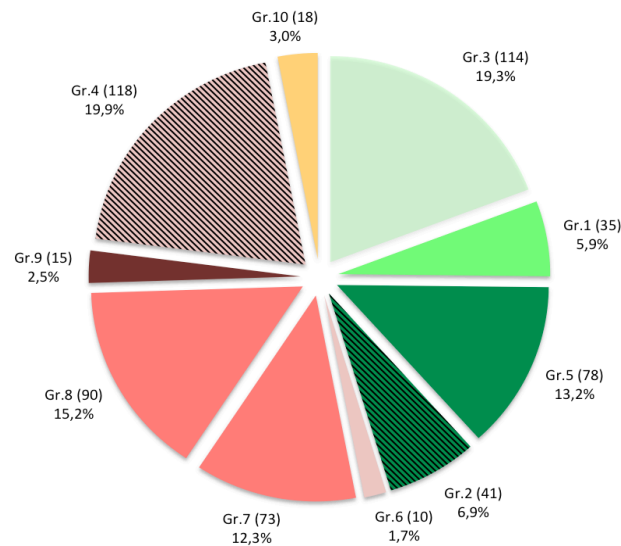


Figure 21. Distribution des gènes parmi les différents groupes d'expression. Les 10 groupes (Gr.1-Gr.10) identifiés parmi les profils des gènes significativement régulés sont représentés en fonction de la nature des régulations mises en jeu au cours de la production de cellulases chez NG14 et RUT C30. Vert : gènes réprimés au cours de la production de cellulases ; Rouge : gènes induits au cours de la production de cellulases. Les nuances des couleurs indiquent les différences d'expression basale (hors induction) entre RUT C30 et QM6a. Couleur claire : expression basale plus forte chez RUT C30 que chez QM6a ; Couleur foncée : expression basale moins forte chez RUT C30 que chez QM6a ; Couleur médiane : pas de différence. Les hachures indiquent une régulation détectée chez NG14 uniquement (absence de régulation chez RUT C30). Le groupe 10 (orangé) regroupe des profils de gènes hétérogènes. Pour chaque groupe, le nombre de gènes (entre parenthèses) et le pourcentage du nombre total sont indiqués.

4.3.3 Comportement et fonctions principales des gènes régulés positivement au cours de l'induction

Les gènes voyant leur expression augmentée au cours de l'induction de la production de cellulases sont rassemblés dans les groupes 4, 6, 7, 8 et 9.

Le groupe 6 (10 gènes) rassemble les gènes présentant les régulations positives les plus fortes au cours de l'induction, aussi bien chez NG14 que chez RUT C30. Ce groupe est constitué presque exclusivement de gènes impliqués dans la dégradation des sucres (9 gènes ; 90%), dont la plupart codent pour des cellulases majeures (le dernier gène code pour une protéine dont la fonction n'a pas été identifiée). On y retrouve notamment les gènes des endoglucanases *egl1*, *egl2*, *egl4*, et *cel74a*, de la cellobiohydrolase *ccb2*, et de la swollenine *swo1* ; ce dernier gène code pour une protéine à domaine de liaison pour la cellulose présentant de grandes similarités avec les expansines végétales qui sont

impliquées dans la désorganisation des polysaccharides des parois (Saloheimo *et al.*, 2002). En plus de leur très forte induction, ces gènes présentent également une expression basale bien plus élevée chez l'hyperproductrice que chez son ancêtre QM6a.

Les groupes 7 et 8 rassemblent une majorité des autres gènes dont l'expression est augmentée au cours de l'induction de la production de cellulases dans les souches NG14 et RUT C30. Les profils de ces deux groupes sont semblables, mais les gènes du groupe 7 présentent des régulations plus prononcées que celles du groupe 8. La plupart des gènes de ces deux groupes présentent une expression basale inchangée entre QM6a et RUT C30.

Le groupe 7 (73 gènes) présente un enrichissement significatif en gènes impliqués dans le métabolisme des sucres. On y retrouve ainsi 10 gènes de glycoside hydrolases codant pour 5 GH caractérisées – la β -galactosidase BGA1 et les β -glucosidases BGL2, CEL1B, CEL3C et CEL3E – et 5 autres GH présumées. Un enrichissement significatif en gènes codant pour des protéines membranaires de transport, notamment des transporteurs présumés de sucres, est également détecté. On retrouve par exemple dans ce groupe 4 protéines à domaines MFS (major facilitator superfamily), une xylose perméase, ainsi qu'une protéine ayant des similitudes avec les lactose perméases de différentes espèces d'*Aspergillus* (voir partie 2.3.1.2.1). Ce groupe présente enfin un enrichissement significatif en gènes codant pour les protéines du réticulum endoplasmique impliquées dans la sécrétion. *Xyr1* et trois autres facteurs de transcription présumés (dont l'orthologue de *creD* d'*A. nidulans*, et de GAL80 de *S. cerevisiae*, le troisième n'ayant pas d'orthologue) sont présents dans ce groupe, attestant de la régulation positive de leur expression au cours de l'induction au lactose.

Le groupe 8 (90 gènes) rassemble des transcrits de natures plus hétéroclites et codant pour des protéines impliquées dans des processus biologiques variés. On y détecte un enrichissement significatif de gènes codant pour des protéines de liaison à l'ARN – deux d'entre elles semblent notamment impliquées dans la maturation des ARN ribosomiques –, pour des protéines impliquées dans la biosynthèse des acides gras – la protéine FAS2 par exemple –, ou des protéines impliquées dans la signalisation dépendante des protéines G. Ce groupe comporte également onze transporteurs membranaires présumés à domaines MFS ou ABC (ATP-binding cassette).

Le groupe 9 (15 gènes) rassemble les transcrits sous-exprimés à l'état basal chez RUT C30 par rapport à QM6a, et dont l'expression est augmentée au cours de l'induction. L'expression différentielle de ces gènes est tardive chez RUT C30 (à partir de 6h après le début de l'injection de lactose) et très faible chez NG14. Ces gènes codent également pour des protéines à fonctions diverses, parmi lesquelles on distingue PAH1 – l'une des seules protéines de développement

répertoriées dans notre étude –, la protéine expansin-like EEL2, ou encore SAR1, protéine de liaison au GTP impliquée dans la sécrétion.

Le groupe 4, enfin, rassemble les gènes qui présentent une augmentation modérée de leur expression au cours de l'induction dans NG14, pas ou peu d'expression différentielle dans RUT C30, et une expression basale hors induction supérieure à celle de QM6a chez l'hyperproductrice. Les calculs d'enrichissements fonctionnels de ce groupe (118 gènes) sont sans équivoque, puisque près de 40% des gènes représentés codent pour des protéines ribosomiques présumées. Des protéines associées aussi bien à la sous-unité 60S qu'à la sous-unité 40S sont représentées. Onze gènes codant pour des transporteurs MFS et ABC et 6 gènes de glycoside hydrolases sont également présents dans ce groupe.

4.3.4 Comportement et fonctions principales des gènes régulés négativement au cours de l'induction

Les groupes 1, 3 et 5 rassemblent les gènes qui voient leur expression régulée négativement chez NG14 et RUT C30 au cours de l'induction de la production de cellulases.

Les gènes réunis dans le groupe 5 (78 gènes) sont les gènes présentant les niveaux de répression les plus importants de ces trois groupes. Ils ont également une expression basale plus élevée dans QM6a que dans RUT C30. On y distingue 8 gènes de la voie de dégradation des acides gras, dont 3 codant pour les enzymes spécifiques du cycle du glyoxylate : la malate synthase MLS1 et les isocitrate lyases ICL1 et ICL2. Ce groupe présente également un enrichissement significatif en gènes codant pour des transporteurs de sucre, comme le transporteur de glucose présumé HXT1. Les gènes codant pour l'amylase SGA1, la glucosidase CEL3B, ainsi que pour les enzymes du métabolisme du galactose AEP3 et LXR1 appartiennent également à ce groupe et sont donc réprimés au cours de la production de cellulases.

Les gènes du groupe 3 (114 gènes) présentent des niveaux de répression un peu moins élevés que ceux du groupe 5 au cours de la production, et sont surexprimés chez RUT C30 à l'état basal. Les enrichissements significatifs enregistrés dans ce groupe concernent des gènes liés à la dégradation protéique, codant aussi bien pour des peptidases (7 gènes) que pour des protéines du système ubiquitine-protéasome (5 gènes). On retrouve également deux enzymes du catabolisme du galactose dans ce groupe : l'épimérase AEP1 et la phosphoglucomutase PGM2.

Les gènes rassemblés dans le groupe 1, également réprimés au cours de la production de cellulases, ne présentent pas de différence dans leur expression basale entre QM6a et RUT C30. Ce groupe (35 gènes) enregistre un enrichissement significatif en transporteurs présumés de sucres (à domaines MFS) et en enzymes impliquées dans le métabolisme des acides aminés (aminotransférases

et décarboxylases notamment). On y trouve également 5 protéines participants à la dégradation des acides gras, dont 4 enzymes peroxisomales impliquées dans la β -oxydation.

Le groupe 2 (41 gènes) rassemble les gènes dont la répression au cours de la production de cellulases n'est détectée que chez NG14, mais pas chez l'hyperproductrice RUT C30, où il ne sont pas ou très peu différenciellement exprimés. L'expression basale de ces gènes est plus élevée chez QM6a que chez RUT C30. Un enrichissement significatif en gènes codant pour des protéines membranaires de transport (à domaines MFS ou ABC) est détecté. On trouve également dans ce groupe une dizaine de gènes liés au métabolisme des acides aminés, notamment des acides aminés polaires non chargés et aromatiques (sérine, cystéine, tyrosine ou tryptophane), et 2 facteurs de transcription : AZF1, qui contrôle la transcription de la cycline CLN3 en réponse au glucose, et PacC, répresseur impliqué dans les réponses cellulaires au pH.

4.3.5 Un groupe aux profils hétérogènes

Le dernier groupe identifié dans notre étude (groupe 10 ; 18 gènes) rassemble des gènes présentant des profils d'expression hétérogènes, et difficilement interprétables. On y trouve deux glycoside hydrolases, la peptidase TYR2, une chaperonne de la famille des HSP104 ainsi que différentes enzymes présumées et 6 protéines dont la fonction est inconnue.

4.3.6 Les mutations génétiques de NG14 et RUT C30 touchent majoritairement des gènes non différenciellement exprimés

Afin d'établir si l'expression différentielle de certains gènes au cours de la production de cellulases était liée à la présence de mutations apparues dans ces gènes au cours de la mise en place de la lignée QM6a-NG14-RUT C30, nous avons réalisé une comparaison des profils d'expression des gènes de notre étude avec la liste des mutations des souches NG14 et RUT C30 disponible. Cette liste résulte d'une analyse par séquençage à haut débit du génome des souches NG14 et RUT C30, entreprise en 2008 par Stéphane Le Crom et Antoine Margeot (Le Crom *et al.*, 2009), et complétée depuis par une autre étude effectuée à l'aide de puces à ADN à haute résolution (Vitikainen *et al.*, 2010). Une recherche systématique des gènes présents dans une fenêtre de 800 pb autour de chaque mutation a été effectuée. Sur l'ensemble des 280 mutations considérées, seules 20 touchaient des gènes présentant une expression différentielle significative dans au moins une des neuf conditions de notre étude (7,1%). 112 mutations répertoriées sont spécifiques de la souche RUT C30 ; seules 8 d'entre elles (7,1%) touchent un gène significativement régulé au moins une fois parmi les différentes conditions considérées (Tableau 5). Les gènes concernés présentent des

fonctions diverses ; leur détail est présenté dans le Tableau annexe 10. On trouve notamment 4 transporteurs présumés, une ATPase membranaire, un récepteur à l'AMPc et deux facteurs de transcription.

Souches		Total (NG14, RUT C30 ou les deux)	NG14 & RUT C30	RUT C30
Nombre total de mutations		280	167	112
Mutations touchant des gènes différenciellement exprimés	toutes	20 mutations 7,1%	12 mutations 7,2%	8 mutations 7,1%
	snp	16	9	7
	délétions	3	2	1 (<i>cre1</i>)
	transloc.	1	1	0

Tableau 5. Inventaire des mutations touchant des gènes régulés au cours de la production de cellulases. Les mutations communes à NG14 et RUT C30 et les mutations spécifiques de RUT C30 sont distinguées en fonction de leur nature : polymorphismes nucléotidiques (snp), délétions ou translocation. Le nombre de gènes concernés est présenté pour chaque catégorie.

4.3.7 Discussion

Les études ayant pour but de décrypter les régulations génétiques mises en jeu au cours de la production de cellulases chez *T. reesei* s'attachent le plus souvent à décrire l'état transcriptionnel des quelques gènes connus comme étant impliqués dans une fonction physiologique précise, reliée de près ou de loin au processus d'intérêt. De telles études, indispensables pour la connaissance des mécanismes en question, ne fournissent néanmoins qu'une vision partielle de la réponse cellulaire mise en jeu au cours de la production de cellulases.

Les travaux présentés ici ont été entrepris dans l'idée que les très fortes capacités de certaines souches de *T. reesei* en terme de synthèse et de sécrétion d'enzymes cellulolytiques ne pouvaient résulter que de la mise en place, au cours de la production, de modifications physiologiques majeures intervenant à l'échelle de la cellule toute entière. En menant une étude globale des régulations transcriptionnelles mises en jeu chez *T. reesei* au cours de la production de cellulases, nous avons souhaité dresser un tableau général du comportement des gènes de cet ascomycète sans *a priori* fonctionnels ou mécanistiques. Une comparaison des « scénarios physiologiques » mis en place par NG14, souche moyennement productrice, et par RUT C30, sa descendante hyperproductrice, a également été entreprise dans le but d'identifier d'éventuels mécanismes responsables de l'acquisition du phénotype d'hyperproduction.

4.3.7.1 Nature et fonction des protéines les plus fortement induites au cours de la production de cellulases

Les gènes présentant les régulations positives les plus fortes au cours de l'induction de la production de cellulases sont réunis dans le groupe 6. Ce groupe présente le plus petit effectif de l'ensemble des groupes de cette étude (10 gènes). On y retrouve les gènes codant pour les endoglucanases majeures (*egl1,2* et *4*) et pour la cellobiohydrolase CBH2. Parmi ces enzymes clés de la dégradation de la cellulose, seul manque le gène *cbh1*. Le profil de ce gène a en fait été supprimé de la liste finale des gènes retenus car son expression a tellement augmenté dès la première heure de l'induction que les sondes *cbh1* présentaient un signal saturant sur l'ensemble des temps testés pour RUT C30. Sur les 9 129 transcrits annotés dans le génome de *T. reesei*, ce comportement est un cas unique. Quelques autres gènes, tels que *cbh2* ou *egl1* présentent en effet un signal saturant chez RUT C30 au cours de l'expérimentation, mais uniquement dans des temps tardifs (i.e. 6h et 24h post-induction) ; leurs profils sont donc suffisamment renseignés pour ne pas avoir été éliminés au cours de l'analyse. *Cbh1*, seul gène éliminé pour cause d'expression trop élevée, peut donc être considéré comme faisant partie de ce groupe de cellulases majeures.

Outre les glycosides hydrolases habituellement associées à la dégradation de la cellulose par *T. reesei*, ce groupe 6 contient également deux hemicellulases : la β -xylosidase BXL1 (Akel *et al.*, 2009) et une protéine de la famille des GH30 à probable activité xylanase, ainsi qu'une endoglucanase de la famille GH45. Au vu de la régulation très élevée de leurs gènes, ces protéines pourraient donc jouer un rôle important dans la faculté de *T. reesei* à dégrader la biomasse lignocellulosique. BXL1 avait déjà été détectée dans le sécrétome de RUT C30 en présence de lactose, mais était étonnamment absente de celui de CL847 dans les mêmes conditions (Herpoel-Gimbert *et al.*, 2008).

La swollenine semble également être un composant majeur du cocktail d'enzymes cellulolytiques de *T. reesei*, puisque l'expression du gène *swo1* est également très fortement augmentée au cours de l'induction. Cette protéine à domaine de liaison pour la cellulose présente de grandes similarités avec les expansines végétales, qui désorganisent les polysaccharides des parois lors de la croissance cellulaire pour permettre leur expansion. De précédentes études ont démontré que le gène *swo1* était sensible aux mêmes inducteurs et répresseurs que les gènes de cellulases (Saloheimo *et al.*, 2002), et que la protéine SWO1 était sécrétée par les souches hyperproductrices RUT C30 et CL847 au cours d'une induction lactose (Herpoel-Gimbert *et al.*, 2008). Nos résultats confirment donc ces observations, et établissent l'importance de la régulation transcriptionnelle de *swo1* en présence d'inducteur des cellulases. Cette régulation, qui atteint un niveau similaire à celui des cellulases principales, témoigne d'un rôle probablement clé de la swollenine dans la dégradation de la

cellulose par *T. reesei*. Comme les expansines végétales, SWO1 présente une activité de perturbation des liaisons hydrogènes des microfibrilles de cellulose sans hydrolyse. La combinaison des effets de la swollenine à l'activité hydrolytique des cellulases contribue ainsi probablement à optimiser l'efficacité de dégradation de *T. reesei*. En désorganisant en amont les parois végétales, SWO1 assurerait l'accessibilité des exo- et endoglucanases hydrolytiques à leurs substrats. Un autre gène codant pour une expansin-like, *eel2*, est retrouvé dans le groupe 9 de notre étude, ce qui confirme de précédents travaux décrivant l'induction de son expression au cours de la production de cellulases (Verbeke *et al.*, 2009).

Certains gènes du groupe 7 présentent également de fortes régulations. Parmi eux, notre annotation experte a permis de mettre en évidence l'existence d'un gène codant pour une lactose perméase potentielle. La régulation positive d'une telle protéine au cours d'une induction par le lactose pourrait remettre en cause l'idée selon laquelle la dégradation du disaccharide est nécessairement extracellulaire (Seiboth *et al.*, 2005). De plus, si le lactose peut être intégré par les cellules, il est également possible que *T. reesei* possède des β -galactosidases intracellulaires, bien qu'aucune n'ait pu être identifiée à ce jour dans son génome. La régulation de telles protéines capables de prendre en charge le lactose pourrait jouer un rôle déterminant dans l'induction de la production de cellulases.

4.3.7.2 Nature des régulations mises en jeu chez l'hyperproductrice

Si certains gènes présentent une expression différentielle dans la souche NG14 et pas chez RUT C30 (groupes 2 et 4), tous les gènes qui voient leur expression modifiée au cours de l'induction dans la souche hyperproductrice étaient régulés de manière similaire dans la souche NG14. Il n'existe donc pas à proprement parler de groupe de gènes significatifs présentant des régulations propres à la souche RUT C30. Ces observations suggèrent que l'acquisition du caractère hyperproducteur ne passerait pas par la mise en place de mécanismes de régulation spécifiques au cours de l'induction de la production de cellulases. Les potentialités des souches ancestrales seraient en fait optimisées chez RUT C30 car la mutagenèse aurait engendré « un tri » entre les régulations nécessaires et celles qui ne le sont pas (ou celles qui sont contre-productives). À titre d'exemple, on enregistre chez NG14 la répression d'un gène impliqué dans la formation du complexe de remodelage de la chromatine SWI-SNF (groupe 2, voir partie 4.2.8.4). Cette répression pourrait s'avérer défavorable pour la cellule dans des conditions nécessitant de nombreuses régulations transcriptionnelles, or on constate que ce gène n'est quasiment plus réprimé chez RUT C30. En favorisant le contrôle transcriptionnel de certains mécanismes précis, la souche améliorerait donc son rendement de production.

L'hyperproductrice semble en revanche bénéficier d'une expression génétique basale modifiée par rapport à ses ancêtres. En effet, hors induction, sept des dix groupes de gènes identifiés dans cette étude présentent une expression chez QM6a significativement différente de celle de RUT C30. 193 gènes de l'étude enregistrent ainsi une expression modifiée de plus d'un facteur 2 entre l'hyperproductrice et son ancêtre. Cette modification de l'état transcriptionnel « constitutif » des cellules a nécessairement un impact important leurs capacités physiologiques. Ainsi, les gènes des cellulases majeures (réunies dans le groupe 6) semblaient déjà être très surexprimés chez RUT C30 avant même le début de l'induction ; cette dernière n'a donc pour conséquence que d'augmenter encore davantage leur transcription.

La modification des niveaux d'expression basaux pourrait également expliquer l'absence de certaines régulations chez RUT C30. Par exemple, certains gènes liés au métabolisme des acides aminés voient leur expression réprimée au cours du fed-batch chez NG14, mais pas chez l'hyperproductrice (groupe 2). Ces gènes présentent un niveau de transcription basal moins élevé chez RUT C30 que chez QM6a. D'autres gènes du métabolisme des acides aminés, qui eux ne présentent pas de différence à l'état basal entre l'hyperproductrice et son ancêtre, subissent une répression aussi bien chez NG14 que chez RUT C30 au cours du fed-batch (groupe 1). Une comparaison des niveaux d'expression chez RUT C30 de l'ensemble de ces gènes révèle qu'à T0, les gènes du groupe 1 ont une expression de 2 à 8 fois supérieure à ceux du groupe 2. Après le lancement du fed-batch, leur expression est réprimée jusqu'à atteindre des niveaux proches de ceux des gènes du groupe 2, qui sont restés stables au cours du temps (Tableau 6). Ainsi, les niveaux basaux des gènes du groupe 2 étant élevés chez la souche ancestrale, une répression a lieu chez NG14. On pourrait imaginer que cette régulation négative permette d'assurer une synthèse protéique correcte tout en limitant les pertes d'énergie à nécessaire l'hypersécrétion. La transcription de ces gènes étant déjà contrôlée hors induction chez RUT C30, cette dernière n'aurait en revanche pas besoin de mettre en place une répression, et les cellules seraient donc pleinement mobilisées pour assurer immédiatement une production efficace d'enzymes. L'expression basale des gènes du groupe 1 n'étant pas modifiée chez RUT C30, leur répression est enregistrée aussi bien chez NG14 que chez l'hyperproductrice. Il n'est bien sûr pas exclu que cette dérégulation acquise par RUT C30 soit au contraire contre-productive, et qu'une augmentation de la transcription des gènes d'acides aminés augmenterait encore davantage les capacités de production de la souche.

Ces observations suggèrent que l'état de transcription basale des gènes pourrait donc influencer la mise en place de régulations au cours de la production de cellulases. Plus généralement, la modification des niveaux d'expression des cellules de *T. reesei* hors induction conditionnerait de manière cruciale l'acquisition des phénotypes d'hyperproduction.

Groupe	Identifiant	Fonction	niveau moyen à T0	niveau moyen à 6h
1	67639	Putative Acetolactate synthase	5416	1259
1	110414	Putative 4-aminobutyrate aminotransferase	4389	1385
1	123738	Putative Arginase ARG1	3119	459
2	44628	Putative Kynureninase BNA5	633	375
2	70630	Putative Homoserine acetyltransferase family protein	1139	2274
2	81089	Putative Cystathionine beta-synthase CYS4	1394	1168
2	81522	Putative Selenocysteine lyase-like protein	2628	1076

Tableau 6. Niveaux d'expression chez RUT C30 de différents gènes liés au métabolisme des acides aminés. Les valeurs sont données en nombre de lectures RNA-seq sur l'ensemble du gène.

4.3.7.3 RUT C30 possède les équipements cellulaires nécessaires à une importante synthèse protéique

L'une des observations les plus marquantes de notre étude concerne la présence parmi les gènes sélectionnés d'un grand nombre de transcrits codant pour des protéines ribosomiques (46 gènes, soit près de 8% de l'effectif total des gènes de notre étude). Quarante-cinq de ces gènes (soit près de la moitié des gènes de protéines ribosomiques de l'ensemble du génome) se réunissent au sein du groupe 4. Le quarante-sixième, retrouvé dans le groupe 8, correspond au seul gène de notre étude codant pour une protéine associée aux ribosomes mitochondriaux. Une exploration détaillée des profils d'expression du groupe 4 prouve que tous les gènes associés aux ribosomes cytoplasmiques présentent sans exception une expression basale au moins doublée dans RUT C30 par rapport à QM6a (voir Figure 22). Leur expression n'est modifiée dans NG14 qu'après les premières heures de l'induction, et l'augmentation détectée reste modérée. Chez RUT C30, leur régulation est quasi-nulle jusqu'à 24h après le début de l'induction, où certains gènes enregistrent une expression très légèrement augmentée, dans la limite d'un facteur 1,5. Ainsi, la surexpression basale des gènes associés aux ribosomes chez RUT C30 semble suffire à garantir une synthèse protéique efficace et immédiate, ce qui n'est pas le cas chez NG14 où une induction est nécessaire afin de répondre aux besoins engendrés par la présence de lactose. Les régulations transcriptionnelles propres aux mutants hyperproducteurs de *T. reesei* (voir les parties 4.1.8 et 4.1.9.3, et ci-dessus) pourraient donc s'accompagner d'une augmentation de la capacité de production de leur machinerie de synthèse protéique par rapport aux souches précoces de la généalogie. Des observations de microscopie électronique semblent aller dans le sens de cette hypothèse. Il a en effet été décrit que le réticulum endoplasmique (RE) de RUT C30 était bien plus développé que celui de QM6a. Le RE de l'hyperproductrice semble notamment présenter de nombreux saccules aplatis formant de véritables nappes parallèles superposées occupant un très gros volume du cytoplasme

cellulaire, tandis que le réticulum de QM6a est bien moins dense et complexe (Ghosh *et al.*, 1984). L'étude montre cependant que cette morphologie n'est pas permanente chez l'hyperproductrice, mais acquise au cours d'une croissance en présence de lactose ; la surface membranaire du RE de RUT C30 semble en effet du même ordre de grandeur que celle de QM6a au début de l'expérimentation, mais augmente considérablement au cours des premières heures de culture en présence de l'inducteur. Ainsi, c'est la machinerie traductionnelle elle-même qui serait davantage développée chez les cellules hyperproductrices, ce qui leur permettrait dès le début de l'induction d'assurer efficacement la synthèse des premières enzymes cellulolytiques, mais aussi des protéines membranaires du RE afin de permettre son développement rapide et d'accroître les capacités de sécrétion. Nos résultats montrent notamment que les gènes *sec61* et *sec63*, qui codent pour les éléments principaux du complexe de translocation des protéines néosynthétisées à travers la membrane du RE (Zimmermann *et al.*, 2011), voient leur expression augmentée dès les premières heures du fed-batch (groupe 7).

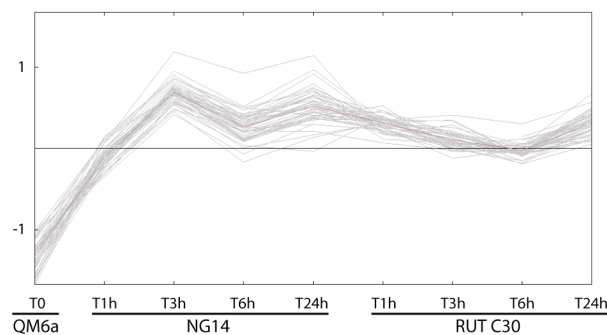


Figure 22. Profils de régulation des gènes codant pour des protéines ribosomiques. Les \log_2 (ratios d'expression) des 45 gènes du groupe 4 codant pour les protéines ribosomiques sont présentés pour chaque condition expérimentale testée.

4.3.7.4 Induction du système de sécrétion, répression des protéines de dégradation

Nos observations prouvent que les fortes capacités de synthèse protéique de RUT C30 s'accompagnent, dès le début de l'induction, d'une surexpression des gènes des acteurs du système de sécrétion. Les groupes 7 et 8 rassemblent notamment plusieurs gènes codant pour des protéines résidentes du RE. Parmi elles, on trouve BiP1, composant associé au complexe SEC61 qui se fixe aux protéines néosynthétisées après leur translocation et qui les maintient dans l'état nécessaire à leur repliement (Gething, 1999), PDI1, qui catalyse l'isomérisation des ponts disulfures des protéines afin d'assurer leur repliement (LaMantia and Lennarz, 1993), ou CNE1, qui permet aux protéines correctement glycosylées de sortir du RE (Hauri *et al.*, 2000). Différentes chaperonnes impliquées dans l'acquisition de la conformation tridimensionnelle des protéines transitant par le RE sont

également retrouvées parmi ces groupes 7 et 8, notamment des protéines des familles HSP23, HSP30, HSP70 et HSP78. Enfin, plusieurs protéines impliquées dans le transport vésiculaire sont également induites. On retrouve notamment SAR1, protéine de liaison au GTP qui constitue l'un des principaux composants du complexe COPII (coat protein complex-II) associé aux vésicules (Aridor *et al.*, 2001), ainsi que d'autres récepteurs conditionnant la biogénèse des vésicules par leurs interactions avec les protéines cargo, tels que ERV29, protéine vésiculaire permettant la reconnaissance d'un signal d'export porté par des protéines solubles (Otte and Barlowe, 2004), ou ERV41 et ERV46, qui portent des motifs d'export du RE et pourraient également jouer le rôle de récepteur aux cargo (Welsh *et al.*, 2006). L'induction de l'ensemble de ces acteurs au cours du fed-batch permet donc aux cellules d'assurer l'efficacité des voies de sécrétion ; cette efficacité est nécessaire à la prise en charge des peptides résultant de la traduction accrue des gènes de glycoside hydrolases notamment.

Face à ce flux important de protéines transloquées dans le RE, la cellule doit également mettre en place un contrôle renforcé de la qualité des protéines qu'elle produit. L'augmentation de la synthèse protéique s'accompagne donc habituellement de l'induction de réponses au stress, notamment de la voie ERAD (ER-associated protein degradation), au cours de laquelle les protéines présentant un défaut conformationnel sont rétrotransloquées dans le cytoplasme et dégradées par le protéasome (Brodsky and Wojcikiewicz, 2009). Pourtant, les gènes liés à la dégradation des protéines détectés dans notre étude sont quasiment tous réprimés pendant la production de cellulases (21 gènes sur les 24 répertoriés). Étonnamment, la plupart d'entre eux (12 gènes) voient leur expression basale augmentée dans la souche hyperproductrice par rapport à QM6a. Une analyse plus précise des fonctions de ces 21 gènes révèle néanmoins que seuls 6 d'entre eux sont associés au système ubiquitine/protéasome. Ces gènes codent pour des protéines du cœur catalytique 20S, et les analyses de RNA-seq indiquent que leurs niveaux d'expression restent importants au cours du fed-batch malgré la répression (Tableau 7). De plus, aucun gène du complexe régulateur 19S (gènes RPN) n'est sélectionné dans notre étude, et leurs niveaux d'expression sont également élevés. En dépit de la répression de certains d'entre eux, les différents constituants du protéasome sont donc bien exprimés au cours de la production.

Les autres gènes de dégradation de notre étude codent majoritairement pour des peptidases, dont plusieurs d'entre elles présentent un signal de sécrétion. Cette observation pourrait confirmer que les peptides et leurs produits de dégradation constitueraient des substituts aux polysaccharides dans certaines conditions (voir partie 4.2.8.2). En présence de sucres inducteurs, la cellule réprimerait donc ces voies alternatives d'assimilation pour se consacrer à la traduction des enzymes nécessaires au métabolisme de ces sucres.

Groupe	Identif.	Fonction	niveau moyen à T0	niveau moyen à 6h
3	55644	Putative 20S proteasome alpha 5 subunit PUP2	1370	900
3	76010	20S proteasome alpha subunit 7 PSA7 / PRE6	1691	1106
3	121343	Putative 20S Proteasome alpha subunit PRE5	1637	1145
3	124031	Putative 20S proteasome alpha 3 subunit PRE9	1339	967
non sélectionné	54454	Putative 26S proteasome regulatory subunit RPN6	1491	1175
non sélectionné	65591	Putative 26S proteasome regulatory subunit RPN10	1166	785
non sélectionné	77591	Putative 26S proteasome regulatory subunit RPN3	2458	1931
non sélectionné	78423	Putative 26S proteasome regulatory subunit RPN2	2845	2710
non sélectionné	120153	Putative 26S proteasome regulatory subunit RPN1	3481	2987

Tableau 7. Niveaux d'expression chez RUT C30 de différents gènes codant pour des protéines associées au protéasome. Les valeurs sont données en nombre de lectures RNA-seq sur l'ensemble du gène.

4.3.7.5 Une régulation réduite des facteurs de transcription

Parmi les différents groupes de gènes étudiés, on ne répertorie qu'un nombre restreint de gènes codant pour des facteurs de transcription (16 gènes sur les 592 de notre étude). Un calcul de l'enrichissement fonctionnel global toutes souches confondues confirme cette tendance : sur l'ensemble de l'étude, un nombre significativement réduit de facteurs de transcription est détecté comme présentant une expression différentielle (voir Tableau annexe 9). Le contrôle transcriptionnel des souches serait donc optimisé au point qu'il ne semble pas nécessaire aux cellules de mettre en place des variations majeures de l'expression de régulateurs au cours de la production de cellulases. Néanmoins, si nos précédents résultats montrent que la mise en place du processus de production et l'acquisition des capacités d'hyperproduction sont associées à des modifications transcriptionnelles majeures (voir l'étude 4.1), des régulations post-transcriptionnelles pourraient également tout à fait influencer sur ces mécanismes. L'activation ou l'inactivation des facteurs de transcription peut également être assurée à l'échelle cellulaire par un contrôle de leur entrée/sortie dans le noyau par exemple. De tels mécanismes ne sont pas identifiables sur puces à ADN.

Parmi les gènes répertoriés, on note sans surprise la présence de *xyr1*, qui est fortement induit dès la première heure du fed-batch. Les gènes *ace1* et *ace2* ne sont en revanche pas retenus dans notre liste. Des variations minimales des niveaux d'expression des facteurs de transcription – qui peuvent avoir des conséquences importantes sur la régulation de leurs gènes cibles – sont en effet difficilement détectables par une analyse globale comme celle menée ici. Les analyses de RNAseq confirment cependant l'induction – modérée – d'*ace1* et *ace2* par le lactose, et révèlent que *ace1* est environ 20 fois plus exprimé que *ace2* (Tableau 8).

Groupe	Identifiant	Fonction	niveau moyen à T0	niveau moyen à 6h
7	122208	Xylanase regulator 1 Xyr1	3557	34453
non sélectionné	75418	Transcription factor ACE1	5246	7893
non sélectionné	78445	Transcription factor ACE2	209	412

Tableau 8. Niveaux d'expression chez RUT C30 des gènes *xyr1*, *ace1* et *ace2*. Les valeurs sont données en nombre de lectures RNA-seq sur l'ensemble du gène.

Parmi les autres facteurs de transcription, on détecte la répression d'AZF1, qui influence la prolifération cellulaire en présence de glucose en régulant positivement la transcription de la cycline CLN3 (Newcomb *et al.*, 2002), et de PacC, orthologue de RIM101 de *S. cerevisiae*, qui gouverne les réponses cellulaires au pH. Chez *A. nidulans* comme chez *S. cerevisiae*, PacC/RIM101 active les gènes nécessaires à la survie cellulaire à pH basique et réprime les gènes exprimés en conditions acides (Lamb *et al.*, 2001; Penalva *et al.*, 2008). Le milieu de culture de nos cellules étant maintenu à pH 4,8 pendant la phase de croissance, puis régulé à pH 4 pendant le fed-batch, il n'est pas surprenant de voir l'expression de ce régulateur diminuer.

L'activateur RPN4, seul facteur de transcription identifié à ce jour chez les ascomycètes comme influençant la transcription des gènes du protéasome (Xie, 2010; Xie and Varshavsky, 2001), fait partie des facteurs de transcription induits au cours de la production. Son induction n'est pourtant pas couplée à celle de ses principaux gènes cibles, qui ne présentent pas de modifications d'expression significatives au cours des premières 24h de notre expérimentation (voir les niveaux des gènes RPN, Tableau 8). Le gène codant pour la protéine GAL80 est également induit pendant le fed-batch. GAL80 est un répresseur majeur des gènes du métabolisme du galactose, largement étudié chez *S. cerevisiae* notamment. En interagissant avec GAL4 et GAL3, GAL80 régule principalement l'expression des gènes *gal1*, *gal7* et *gal10* impliqués dans la voie de Leloir (Lohr *et al.*, 1995). En l'absence de galactose, GAL80 se lie au facteur de transcription GAL4, ce qui empêche ce dernier de recruter les facteurs nécessaires à la transcription dépendante de l'ARN polymérase POLII. Au contraire, en présence de galactose, GAL3 séquestre GAL80 dans le cytoplasme des cellules, ce qui lève l'inhibition de GAL4 et permet l'expression des gènes du système GAL (Pilauri *et al.*, 2005). À ce jour, l'orthologue de GAL3 n'a pas été identifié chez *T. reesei*, tandis que l'homologue présumé de GAL4 n'est pas répertorié parmi les gènes de notre étude. L'augmentation de l'expression de GAL80 au cours du fed-batch pourrait donc assurer une inhibition de l'expression des enzymes de la voie de Leloir au cours de la production de cellulases.

4.3.7.6 Métabolisme du galactose et production de cellulases

Le galactose libéré par la β -galactosidase BGA1 à partir du lactose étant sous la forme β , le métabolisme de ce monosaccharide doit principalement avoir lieu par la voie alternative, et non celle de Leloir (voir partie 2.3.1.2.4). Une étude a en effet précédemment démontré que les aldoses-épimérasas de *T. reesei*, enzymes catalysant la mutarotation du β -galactose indispensable à sa prise en charge par la voie de Leloir, n'étaient pas exprimées au cours d'une croissance en lactose (Fekete *et al.*, 2008). Nos observations confirment l'inhibition de l'expression des gènes d'aldoses-épimérasas *aep1* (groupe 3) et *aep3* (groupe 5) au cours du fed-batch. Les gènes *gal1*, *gal7* et *gal10*, sous le contrôle de GAL80, ne sont pas détectés comme significativement régulés au cours de notre expérimentation – il est possible que la surexpression de GAL80 nécessite plus de 24h pour se répercuter sur l'expression de ses cibles –, mais le gène de la phosphoglucomutase *pgm1* est également réprimé (groupe 3), ce qui renforce l'hypothèse d'une inactivation de la voie de Leloir.

Comme mentionné précédemment (voir partie 2.3.1.2.5), l'aldolase XYL1 initiant la voie alternative joue donc un rôle crucial dans le métabolisme du galactose et par conséquent dans les mécanismes d'induction. Nos résultats indiquent d'ailleurs que l'expression du gène *xy11* est fortement induite dès les premières heures du fed-batch (groupe 7). Le gène de la xylulose réductase LXR1, qui catalyse la quatrième des six étapes de la voie alternative (conversion du L-sorbose en D-sorbitol, voir Figure 4), est en revanche réprimé au cours de la production de cellulases (groupe 5). Cette observation suggère que ce sont les métabolites synthétisés au cours des trois premières étapes de cette voie alternative (le galactitol, le L-xylo-3-hexulose, ou le L-sorbose) qui pourraient constituer des acteurs majeurs de la transmission du signal d'induction de la production des cellulases. Des études de Northern blot ayant précédemment démontré qu'une croissance des cellules de *T. reesei* en L-sorbose pouvait induire la transcription de *cbh1*, *cbh2*, *egl1*, *egl2*, *egl3* et *egl5* (Nogawa *et al.*, 2001), ce métabolite pourrait donc effectivement assurer la transmission du signal lactose.

4.3.7.7 Comparaison entre le génome et le transcriptome des souches

Stéphane Le Crom, Antoine Margeot et leurs collaborateurs ont effectué en 2008 une analyse du génome des souches NG14 et RUT C30 par séquençage à haut débit. Alors que la génération des mutants de *T. reesei* avait jusqu'alors été effectuée de façon aléatoire et qu'il n'existait aucun inventaire des modifications génétiques acquises par les souches, ces travaux ont permis d'établir la liste des mutations retrouvées dans les génomes de NG14 et RUT C30. 233 polymorphismes nucléotidiques, 14 petites insertions ou délétions, et 18 délétions plus larges ont été ainsi identifiés (Le Crom *et al.*, 2009). Une autre étude a ensuite complété cette liste avec 15 mutations

supplémentaires (Vitikainen *et al.*, 2010). Notre analyse systématique des profils d'expression des gènes situés à proximité de chacune de ces mutations révèle que seules 20 de ces 280 mutations touchent des gènes différentiellement exprimés au cours de l'induction de la production de cellulases. Sur les 112 mutations spécifiques de RUT C30 (mutations résultant de la mutagenèse appliquée sur NG14, et donc en principe responsables de l'acquisition du caractère hyperproducteur de RUT C30), huit seulement concernent des gènes significativement régulés au cours de la production. Ainsi, les gènes qui voient leur expression modifiée au cours du processus ne sont majoritairement pas ceux qui sont touchés par des modifications génétiques. Cette observation n'est pas nécessairement étonnante, puisque les mutations peuvent modifier les propriétés des gènes sans lien avec leur expression. Certaines d'entre elles peuvent ainsi entraîner une perte de fonction des protéines associées, ou modifier l'activité de certaines enzymes en impactant leur conformation tridimensionnelle par exemple, sans que l'expression de leur gène ne soit modifiée. À l'inverse, les mutations touchant des facteurs de transcription (au nombre de cinq d'après l'étude de Le Crom *et al.*, 2009) peuvent entraîner une modification de l'expression de plusieurs gènes sans que ceux-ci ne soient eux-mêmes mutés. Cependant, on aurait pu imaginer que les polymorphismes nucléotidiques spécifiques de la souche RUT C30 se traduiraient par une perte de régulation dans cette souche des gènes concernés, or aucun gène des groupes 2 et 4 (présentant des régulations chez NG14 mais pas chez RUT C30) n'est touché par une mutation. De plus, le reséquençage des souches avait identifié plusieurs mutations dans des gènes associés aux systèmes de production de cellulases, notamment à la sécrétion ; il est donc surprenant qu'aucune d'entre elles ne touche des gènes différentiellement régulés malgré le grand nombre de gènes de ces catégories dans notre étude. Cette absence de lien entre les caractéristiques génétiques des souches et leurs profils transcriptomiques pourrait s'expliquer par le fait que les cribles permettant de sélectionner les souches issues de la mutagenèse (voir partie 2.2.4) sont réalisés sur des milieux solides et des durées longues. La procédure a donc probablement sélectionné des mutants présentant des performances améliorées à l'état basal, stable, et non au cours d'une induction telle qu'elle est mise en œuvre en bioréacteur. Les informations apportées par l'analyse du transcriptome des souches de *T. reesei* viennent donc compléter l'étude de leur génome, et ouvrent la voie à de nouvelles investigations sur les systèmes nouvellement identifiés.

Comme indiqué dans le Tableau 5, l'une des mutations majeures apparues entre NG14 et RUT C30 est la perte d'une partie du locus *cre1* chez l'hyperproductrice (voir également la partie 4.1.7.1 et la Figure annexe 3). Étonnamment, l'ensemble des cibles de *cre1* identifiées dans la partie 4.2 n'est pas retrouvé dans notre étude, et les liens entre les groupes de régulation de l'une et l'autre des études ne sont pas systématiques. On s'attendait notamment à ce qu'une majorité des

gènes régulés par le facteur de transcription soient retrouvés en grand nombre dans les groupes 2 et 4, mais seuls 5 et 11 gènes sont concernés, respectivement. Les souches employées et les conditions expérimentales propres à chaque étude expliquent probablement en grande partie ces différences, mais de telles observations suggèrent néanmoins que les effets de *cre1* au cours de l'induction lactose pourraient être moins larges et déterminants que le modèle communément admis. De plus, nos données indiquent que l'expression de *cre1* serait elle-même augmentée au cours du fed-batch. Le gène présente en effet chez NG14 un \log_2 (ratio d'expression) de 1,04 à 24h par rapport au temps 0 ; son expression a donc doublé entre ces deux instants. Le nombre de lectures RNA-seq correspondant à la séquence résiduelle de *cre1* chez RUT C30 (promoteur + région 5' du gène) augmente également de plus de 5 fois entre T0 et T6h (2110 et 11898 lectures, respectivement). Ces résultats vont dans le sens d'un rôle de CRE1 dans l'induction de *xyl1*, activateur majeur des gènes de cellulases (voir partie 4.1.9.5), mais remettent en question l'idée largement acquise que l'expression du facteur de transcription va nécessairement de pair avec une répression de la production des enzymes cellulolytiques (Nakari-Setälä *et al.*, 2009). L'équilibre entre les rôles répresseur et activateur de CRE1 semble donc complexe et dépend probablement largement des conditions de culture des cellules.

En résumé, cette troisième étude a permis de fournir un aperçu global des effets de l'induction de la production de cellulases sur l'expression des gènes de *T. reesei*. Les principales fonctions cellulaires soumises à ces régulations sont résumées dans la Figure 23. Parmi elles, on distingue notamment les voies de synthèse et de sécrétion des protéines, mais aussi le transport cellulaire, le métabolisme des acides aminés, des acides gras ou du galactose. Nos résultats prouvent que la majorité des mutations acquises par NG14 et RUT C30 touchent des gènes qui ne sont pas régulés au cours de la production de cellulases, et suggèrent que l'acquisition des phénotypes d'hyperproduction pourrait être associée à des dérégulations majeures de l'expression génétique basale des souches. Nos observations indiquent également que l'absence de la protéine CRE1 n'engendre pas de modifications transcriptionnelles majeures chez la souche hyperproductrice. Un manuscrit reprenant l'ensemble de ces résultats est en cours de préparation pour publication.

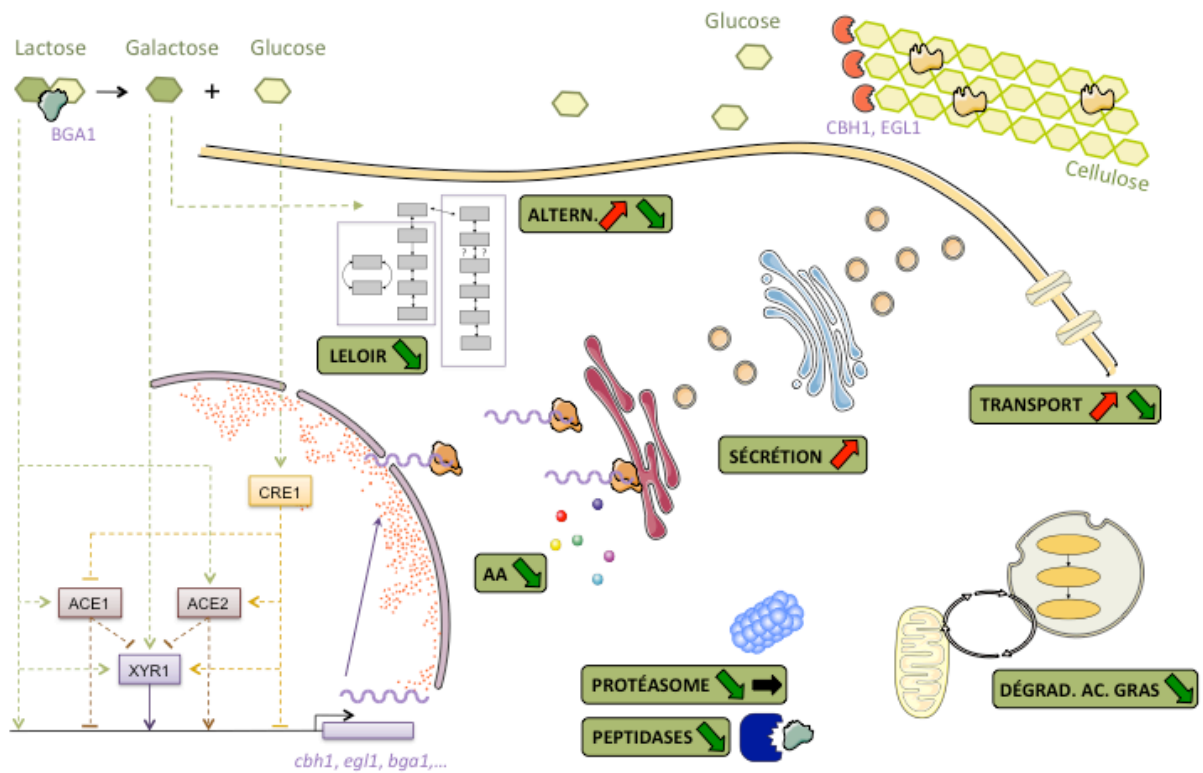


Figure 23. Principales régulations des fonctions cellulaires au cours de l'induction de la production de cellulases. Les fonctions et leurs régulations enregistrées chez NG14 et RUT C30 sont représentées. Rouge : régulation positive ; Vert : régulation négative ; Noir : pas de régulation. Le comportement des principaux facteurs de transcription en réponse au lactose et à ses produits de dégradation est également rappelé.

5 CONCLUSION ET PERSPECTIVES

L'amélioration de la production d'enzymes cellulolytiques par *T. reesei* constitue une étape majeure dans le développement de procédés de production de biocarburants de deuxième génération. De nombreuses souches aux capacités de production accrues ont été générées ces dernières décennies par mutagénèse aléatoire, mais il semble désormais indispensable de comprendre leurs mécanismes d'induction de la synthèse et de la sécrétion de cellulases afin d'accroître davantage leurs performances.

À ce jour, la plupart des études de l'expression génétique des souches de *T. reesei* ont été réalisées à partir de cultures en fioles, qui permettent une expérimentation aisée dans un laboratoire. Dans ces conditions, la densité cellulaire est souvent assez faible, et le milieu de culture est en permanence soumis à des modifications, notamment de pH et de concentration en oxygène dissout. Ainsi, si ces études ont permis la description des profils d'expression de certains gènes de *T. reesei* dans différents contextes environnementaux, ces résultats ne sont pas nécessairement valables dans des conditions de production industrielles, qui permettent d'optimiser les performances des cellules par un contrôle minutieux de l'ensemble des paramètres de culture.

L'un des atouts principaux de ce projet de thèse est de combiner les apports de techniques de recherche fondamentales à la robustesse de ce procédé industriel appliqué. Différentes analyses du transcriptome de *T. reesei* nous ont en effet permis d'amorcer une exploration des mécanismes génétiques sous-jacents aux comportements d'intérêt de cet ascomycète en bioréacteur. Notre volonté d'étudier ces mécanismes au moyen d'approches globales a par ailleurs contribué à élargir notre connaissance des réponses cellulaires mises en jeu au cours de l'induction de la production de cellulases.

Nos travaux ont permis de décrire les effets du lactose, un des seuls inducteurs pouvant être utilisé à l'échelle industrielle, sur l'expression des facteurs de transcription centraux influençant l'expression des gènes d'enzymes cellulolytiques. L'existence d'une régulation positive exercée par le lactose sur les gènes *xyr1*, *ace1* et *ace2* a ainsi été démontrée de manière inédite. Nos résultats ont également prouvé sans ambiguïté que la protéine CRE1, qui orchestre le processus de répression catabolique responsable de l'inhibition de la production de cellulases, pouvait dans certains cas exercer des fonctions d'inducteur, contrairement au modèle généralement admis. L'ensemble des cibles de ce facteur a pu être identifié aussi bien dans un contexte de répression catabolique que de dérépression par le glucose. Nos investigations ont enfin permis d'établir une description globale des effets de l'induction de la production de cellulases en bioréacteur sur l'expression des gènes de souches plus ou moins productrices de *T. reesei*. Une telle approche a révélé l'importance de l'expression génétique basale dans l'acquisition des capacités d'hyperproduction des mutants et a

permis d'identifier les principales fonctions cellulaires soumises à des régulations au cours du processus d'induction.

Dans un contexte énergétique où l'optimisation des procédés de conversion de la biomasse lignocellulosique en bioéthanol est devenue une nécessité, nos travaux ont donc permis d'établir les premiers scénarios de régulations génétiques mis en jeu dans différentes souches d'intérêt de *T. reesei* dans des conditions de cultures industrielles. Ces apports permettent d'orienter les futurs champs d'investigation qui permettront d'établir de nouvelles stratégies d'amélioration des capacités de ces souches en terme de production d'enzymes cellulolytiques.

La description des mécanismes et l'identification des différents acteurs responsables de cette production pourront notamment être poursuivis à la lumière de nos résultats. Ainsi, si CRE1 semble nécessaire à une induction accrue de *xyl1* et d'*ace2* en présence de lactose, et que ses cibles sont nombreuses dans des conditions de répression ou de dérèglement par le glucose, nous n'avons pas observé de signature précise de la disparition de ce facteur de transcription dans la régulation des gènes de la souche RUT C30 au cours d'une induction lactose. Il est donc indispensable d'étudier plus précisément le comportement et les effets de CRE1 au cours d'une telle induction. L'étude du transcriptome d'une souche NG14 délétée pour le gène *cre1* permettra notamment d'établir si certaines régulations que nous avons observées dans la souche NG14 étaient la conséquence d'une action de CRE1, et de confirmer que les mécanismes mis en place chez RUT C30 ne résultent pas uniquement de la délétion de CRE1, comme généralement admis.

De même, si le galactose semble lui-même pouvoir induire la transcription de *xyl1*, il semble désormais clair que certains acteurs de la voie du métabolisme de ce monosaccharide sont impliqués dans l'induction de la production des cellulases. Notre description des régulations des gènes impliqués dans les voies alternative et de Leloir fournit notamment des précisions sur les métabolites susceptibles d'assurer la transmission du signal d'induction lactose. Des études du comportement de souches mutées pour les différents gènes de ces voies permettront d'identifier les acteurs enzymatiques et les métabolites impliqués.

Plus généralement, nos comparaisons du génome et du transcriptome des souches de *T. reesei* pourront être complétées par des études des flux et des équilibres métaboliques associés. Des analyses du métabolome des souches au cours de la production de cellulases en bioréacteur contribueront à l'identification des acteurs moléculaires jouant un rôle dans la mise en place des différents mécanismes associés à cette production. De telles analyses permettront également d'établir quelles sont les régulations pertinentes pour la cellule et celles qui sont contre-productives – et qui constituent donc des points limitants à améliorer par ingénierie génétique. Une analyse des niveaux

d'acides aminés dans les cellules indiquerait par exemple l'influence de la répression de leur métabolisme sur la synthèse massive de protéines.

Des calculs de flux métaboliques apporteront également de nombreuses précisions quant à l'état physiologique des cellules au cours des différentes étapes du procédé. De telles données permettraient par exemple d'expliquer pourquoi les gènes du catabolisme des acides gras sont réprimés au cours de la production de cellulases. La perception par les cellules d'une diminution trop importante des niveaux de glucose au cours de la phase de croissance pourrait en effet entraîner une surexpression des gènes associés à la β -oxydation et au cycle du glyoxylate en fin de batch (T0), dans le but d'assurer une néoglucogenèse efficace. Les cellules étant ensuite alimentées en lactose, cette surexpression serait arrêtée, ce qui expliquerait la diminution des niveaux de transcrits détectée. Au contraire, si les conditions physiologiques des cellules ne leur permettent pas d'assurer une telle détection des niveaux de glucose, la répression des gènes du catabolisme des acides gras pourrait faire partie intégrante de la mise en place d'un processus de production de cellulases efficace.

Des études des promoteurs des gènes qui voient leur expression modifiée au cours du processus semblent enfin cruciales pour l'identification d'éventuels motifs de régulation communs, et des facteurs de transcription qui leur sont associés.

Les résultats présentés dans ce manuscrit ont donc permis une avancée significative dans la description des mécanismes impliqués dans la production de cellulases. Complétés par les différentes études suggérées ci-dessus, ils permettront l'identification de cibles pertinentes pour entreprendre une ingénierie génétique des souches d'intérêt et tenter d'accroître leurs performances en terme de synthèse et de sécrétion d'enzymes cellulolytiques. Les effets d'une surexpression de *xyr1* peuvent par exemple dorénavant et déjà être testés. Des délétions de *ace1*, *ace2* ou des nouveaux facteurs de transcription identifiés par nos travaux comme étant régulés au cours de la production de cellulases peuvent également dès maintenant être envisagées afin de confirmer l'importance de ces facteurs dans le processus.

À terme, ces travaux devraient permettre d'améliorer le procédé de production de bioéthanol de deuxième génération à partir de biomasse lignocellulosique, et ainsi favoriser le développement de ce carburant renouvelable aux bilans positifs pour l'environnement.

6 RÉFÉRENCES

- Akel, E., Metz, B., Seiboth, B., and Kubicek, C.P. (2009). Molecular regulation of arabinan and L-arabinose metabolism in *Hypocrea jecorina* (*Trichoderma reesei*). *Eukaryot Cell*.
- Aridor, M., Fish, K.N., Bannykh, S., Weissman, J., Roberts, T.H., Lippincott-Schwartz, J., and Balch, W.E. (2001). The Sar1 GTPase coordinates biosynthetic cargo selection with endoplasmic reticulum export site assembly. *J Cell Biol* **152**, 213-229.
- Aro, N., Ilmen, M., Saloheimo, A., and Penttila, M. (2003). ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Appl Environ Microbiol* **69**, 56-65.
- Aro, N., Pakula, T., and Penttila, M. (2005). Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiol Rev* **29**, 719-739.
- Aro, N., Saloheimo, A., Ilmen, M., and Penttila, M. (2001). ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. *J Biol Chem* **276**, 24309-24314.
- Arst, H.N., Jr., Tollervey, D., Dowzer, C.E., and Kelly, J.M. (1990). An inversion truncating the *creA* gene of *Aspergillus nidulans* results in carbon catabolite derepression. *Mol Microbiol* **4**, 851-854.
- Balciunas, D., Hallberg, M., Bjorklund, S., and Ronne, H. (2003). Functional interactions within yeast mediator and evidence of differential subunit modifications. *The Journal of biological chemistry* **278**, 3831-3839.
- Ballerini, D., and Alazard-Toux, N. (2006). Les biocarburants, état des lieux, perspectives et enjeux du développement (Technip).
- Becker, P.B., and Horz, W. (2002). ATP-dependent nucleosome remodeling. *Annu Rev Biochem* **71**, 247-273.
- Boase, N.A., and Kelly, J.M. (2004). A role for *creD*, a carbon catabolite repression gene from *Aspergillus nidulans*, in ubiquitination. *Mol Microbiol* **53**, 929-940.
- Bono, J.L., Jaber, B., Fisher, M.A., Abuodeh, R.O., O'Leary-Jepson, E., Scalarone, G.M., and Smith, L.H., Jr. (2001). Genetic diversity and transcriptional analysis of the *bys1* gene from *Blastomyces dermatitidis*. *Mycopathologia* **152**, 113-123.
- Bouffard, G.G., Rudd, K.E., and Adhya, S.L. (1994). Dependence of lactose metabolism upon mutarotase encoded in the gal operon in *Escherichia coli*. *J Mol Biol* **244**, 269-278.
- Breviario, D., Hinnebusch, A.G., and Dhar, R. (1988). Multiple regulatory mechanisms control the expression of the *RAS1* and *RAS2* genes of *Saccharomyces cerevisiae*. *The EMBO journal* **7**, 1805-1813.
- Brodsky, J.L., and Wojcikiewicz, R.J. (2009). Substrate-specific mediators of ER associated degradation (ERAD). *Curr Opin Cell Biol* **21**, 516-521.
- Brunner, K., Omann, M., Pucher, M.E., Delic, M., Lehner, S.M., Domnanich, P., Kratochwill, K., Druzhinina, I., Denk, D., and Zeilinger, S. (2008). *Trichoderma* G protein-coupled receptors: functional characterisation of a cAMP receptor-like protein from *Trichoderma atroviride*. *Curr Genet* **54**, 283-299.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233-238.
- Carle-Urioste, J.C., Escobar-Vera, J., El-Gogary, S., Henrique-Silva, F., Torigoi, E., Crivellaro, O., Herrera-Estrella, A., and El-Dorry, H. (1997). Cellulase induction in *Trichoderma reesei* by cellulose requires its own basal expression. *J Biol Chem* **272**, 10169-10174.
- Chang, E.C., Barr, M., Wang, Y., Jung, V., Xu, H.P., and Wigler, M.H. (1994). Cooperative interaction of *S. pombe* proteins required for mating and morphogenesis. *Cell* **79**, 131-141.
- Chang, Y.D., and Dickson, R.C. (1988). Primary structure of the lactose permease gene from the yeast *Kluyveromyces lactis*. Presence of an unusual transcript structure. *J Biol Chem* **263**, 16696-16703.
- Chilton, I.J., Delaney, C.E., Barham-Morris, J., Fincham, D.A., Hooley, P., and Whitehead, M.P. (2008). The *Aspergillus nidulans* stress response transcription factor StzA is ascomycete-specific and shows species-specific polymorphisms in the C-terminal region. *Mycol Res* **112**, 1435-1446.
- Cubero, B., and Scazzocchio, C. (1994). Two different, adjacent and divergent zinc finger binding sites are necessary for CREA-mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans*. *The EMBO journal* **13**, 407-415.
- Cziferszky, A., Mach, R.L., and Kubicek, C.P. (2002). Phosphorylation positively regulates DNA binding of the carbon catabolite repressor Cre1 of *Hypocrea jecorina* (*Trichoderma reesei*). *J Biol Chem* **277**, 14688-14694.
- Cziferszky, A., Seiboth, B., and Kubicek, C.P. (2003). The Snf1 kinase of the filamentous fungus *Hypocrea jecorina* phosphorylates regulation-relevant serine residues in the yeast carbon catabolite repressor Mig1 but not in the filamentous fungal counterpart Cre1. *Fungal Genet Biol* **40**, 166-175.
- Danielson, R.M., and Davey, C.B. (1973). The abundance of *Trichoderma* propagules and the distribution of species in forest soils. *Soil Biology and Biochemistry* **5**, 485-494.
- de Vries, R.P., and Visser, J. (2001). *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol Mol Biol Rev* **65**, 497-522, table of contents.

- Dechassa, M.L., Zhang, B., Horowitz-Scherer, R., Persinger, J., Woodcock, C.L., Peterson, C.L., and Bartholomew, B. (2008). Architecture of the SWI/SNF-nucleosome complex. *Molecular and cellular biology* **28**, 6010-6021.
- Del Rio, J.C., Marques, G., Rencoret, J., Martinez, A.T., and Gutierrez, A. (2007). Occurrence of naturally acetylated lignin units. *J Agric Food Chem* **55**, 5461-5468.
- Dickson, R.C., and Barr, K. (1983). Characterization of lactose transport in *Kluyveromyces lactis*. *J Bacteriol* **154**, 1245-1251.
- Dowzer, C.E., and Kelly, J.M. (1991). Analysis of the *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. *Mol Cell Biol* **11**, 5701-5709.
- Du, L.L., and Novick, P. (2002). Pag1p, a novel protein associated with protein kinase Cbk1p, is required for cell morphogenesis and proliferation in *Saccharomyces cerevisiae*. *Molecular biology of the cell* **13**, 503-514.
- Durand, H., Baron, M., Calmels, T., and Tiraby, G. (1988a). Classical and molecular genetics applied to *Trichoderma reesei* for the selection of improved cellulolytic industrial strains. *FEMS Symp* **43**, 135-152.
- Durand, H., Clanet, M., and Tiraby, G. (1988b). Genetic improvement of *Trichoderma reesei* for large scale cellulase production. *Enzyme and Microbial Technology* **10**, 341-346.
- el-Gogary, S., Leite, A., Crivellaro, O., Eveleigh, D.E., and el-Dorry, H. (1989). Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. *Proc Natl Acad Sci U S A* **86**, 6138-6141.
- Eveleigh, D.E., and Montencourt, B.S. (1979). Increasing yields of extracellular enzymes. *Adv Appl Microbiol* **25**, 57-74.
- Fekete, E., Karaffa, L., Kubicek, C.P., Szentirmai, A., and Seiboth, B. (2007). Induction of extracellular beta-galactosidase (Bga1) formation by D-galactose in *Hypocrea jecorina* is mediated by galactitol. *Microbiology* **153**, 507-512.
- Fekete, E., Karaffa, L., Sandor, E., Seiboth, B., Biro, S., Szentirmai, A., and Kubicek, C.P. (2002). Regulation of formation of the intracellular beta-galactosidase activity of *Aspergillus nidulans*. *Arch Microbiol* **179**, 7-14.
- Fekete, E., Seiboth, B., Kubicek, C.P., Szentirmai, A., and Karaffa, L. (2008). Lack of aldose 1-epimerase in *Hypocrea jecorina* (anamorph *Trichoderma reesei*): a key to cellulase gene expression on lactose. *Proc Natl Acad Sci U S A* **105**, 7141-7146.
- Felenbok, B., Flippin, M., and Nikolaev, I. (2001). Ethanol catabolism in *Aspergillus nidulans*: a model system for studying gene regulation. *Prog Nucleic Acid Res Mol Biol* **69**, 149-204.
- Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., et al. (2003). Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *J Biol Chem* **278**, 31988-31997.
- Fortwendel, J.R., Fuller, K.K., Stephens, T.J., Bacon, W.C., Askew, D.S., and Rhodes, J.C. (2008). *Aspergillus fumigatus* RasA regulates asexual development and cell wall integrity. *Eukaryotic Cell* **7**, 1530-1539.
- Gamauf, C., Marchetti, M., Kallio, J., Puranen, T., Vehmaanpera, J., Allmaier, G., Kubicek, C.P., and Seiboth, B. (2007). Characterization of the *bga1*-encoded glycoside hydrolase family 35 beta-galactosidase of *Hypocrea jecorina* with galacto-beta-D-galactanase activity. *FEBS J* **274**, 1691-1700.
- Gancedo, J.M. (1998). Yeast carbon catabolite repression. *Microbiol Mol Biol Rev* **62**, 334-361.
- Garcia, I., Gonzalez, R., Gomez, D., and Scazzocchio, C. (2004). Chromatin rearrangements in the *prnD-prnB* bidirectional promoter: dependence on transcription factors. *Eukaryotic Cell* **3**, 144-156.
- Gething, M.-J. (1999). Role and regulation of the ER chaperone BiP. *Seminars in Cell and Developmental Biology* **10**, 465-472.
- Ghosh, A., Ghosh, B.K., Trimino-Vazquez, H., Eveleigh, D.E., and Montencourt, B.S. (1984). Cellulase secretion from a hyper-cellulolytic mutant of *Trichoderma reesei* Rut-C30. *Archives of Microbiology* **140**, 126-133-133.
- Gorke, B., and Stulke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**, 613-624.
- Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of molecular biology* **313**, 903-919.
- Gremel, G., Dorrer, M., and Schmoll, M. (2008). Sulphur metabolism and cellulase gene expression are connected processes in the filamentous fungus *Hypocrea jecorina* (anamorph *Trichoderma reesei*). *BMC Microbiol* **8**, 174.
- Grossetete, S., Labedan, B., and Lespinet, O. (2010). FUNGIpath: a tool to assess fungal metabolic pathways predicted by orthology. *BMC Genomics* **11**, 81.
- Harman, G., and Kubicek, C. (1998). *Trichoderma* and *Gliocladium* (Taylor & Francis).
- Hartl L, K.C., Seiboth B (2009). Carbon signaling by hexose phosphorylating enzymes in *Hypocrea jecorina* (soumis). *Eukaryot Cell*
- Hartl, L., Kubicek, C.P., and Seiboth, B. (2007). Induction of the gal pathway and cellulase genes involves no transcriptional inducer function of the galactokinase in *Hypocrea jecorina*. *J Biol Chem* **282**, 18654-18659.

- Hauri, H., Appenzeller, C., Kuhn, F., and Nufer, O. (2000). Lectins and traffic in the secretory pathway. *FEBS letters* **476**, 32-37.
- Herpoel-Gimbert, I., Margeot, A., Dolla, A., Jan, G., Molle, D., Lignon, S., Mathis, H., Sigoillot, J.C., Monot, F., and Asther, M. (2008). Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnol Biofuels* **1**, 18.
- Herrero, J., Al-Shahrour, F., Diaz-Uriarte, R., Mateos, A., Vaquerizas, J.M., Santoyo, J., and Dopazo, J. (2003). GEPAS: A web-based resource for microarray gene expression data analysis. *Nucleic acids research* **31**, 3461-3467.
- Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., and Foust, T.D. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**, 804-807.
- Holden, H.M., Rayment, I., and Thoden, J.B. (2003). Structure and function of enzymes of the Leloir pathway for galactose metabolism. *J Biol Chem* **278**, 43885-43888.
- Ilmen, M., Thrane, C., and Penttila, M. (1996). The glucose repressor gene *cre1* of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form. *Mol Gen Genet* **251**, 451-460.
- Ilyes, H., Fekete, E., Karaffa, L., Sandor, E., Szentirmai, A., and Kubicek, C.P. (2004). CreA-mediated carbon catabolite repression of beta-galactosidase formation in *Aspergillus nidulans* is growth rate dependent. *FEMS microbiology letters* **235**, 147-151.
- Jourdren, L., Duclos, A., Brion, C., Portnoy, T., Mathis, H., Margeot, A., and Le Crom, S. (2010). Teolenn: an efficient and customizable workflow to design high-quality probes for microarray experiments. *Nucleic Acids Res.*
- Kana-uchi, A., Yamashiro, C.T., Tanabe, S., and Murayama, T. (1997). A ras homologue of *Neurospora crassa* regulates morphology. *Molecular & general genetics : MGG* **254**, 427-432.
- Karaffa, L., Fekete, E., Gamauf, C., Szentirmai, A., Kubicek, C.P., and Seiboth, B. (2006). D-Galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates. *Microbiology* **152**, 1507-1514.
- Kubicek, C.P., Bolzlbauer, U.M., Kovacs, W., Mach, R.L., Kuhls, K., Lieckfeldt, E., Borner, T., and Samuels, G.J. (1996). Cellulase formation by species of *Trichoderma* sect. *Longibrachiatum* and of *Hypocrea* spp. with anamorphs referable to *Trichoderma* sect. *Longibrachiatum*. *Fungal Genet Biol* **20**, 105-114.
- Kubicek, C.P., Messner, R., Gruber, F., Mandels, M., and Kubicek-Pranz, E.M. (1993). Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei*. Involvement of a constitutive, sophorose-inducible, glucose-inhibited beta-diglycoside permease. *J Biol Chem* **268**, 19364-19368.
- Kubicek, C.P., Mikus, M., Schuster, A., Schmoll, M., and Seiboth, B. (2009). Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnol Biofuels* **2**, 19.
- Kubicek, C.P., Muhlbauer, G., Klotz, M., John, E., and Kubicek-Pranz, E.M. (1988). Properties of a Conidial-Bound Cellulase Enzyme-System from *Trichoderma reesei*. *J Gen Microbiol* **134**, 1215-1222.
- Kuhls, K., Lieckfeldt, E., Samuels, G.J., Kovacs, W., Meyer, W., Petrini, O., Gams, W., Borner, T., and Kubicek, C.P. (1996). Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. *Proc Natl Acad Sci U S A* **93**, 7755-7760.
- LaMantia, M., and Lennarz, W.J. (1993). The essential function of yeast protein disulfide isomerase does not reside in its isomerase activity. *Cell* **74**, 899-908.
- Lamb, T.M., Xu, W., Diamond, A., and Mitchell, A.P. (2001). Alkaline response genes of *Saccharomyces cerevisiae* and their relationship to the RIM101 pathway. *The Journal of biological chemistry* **276**, 1850-1856.
- Le Crom, S., Schackwitz, W., Pennacchio, L., Magnuson, J.K., Culley, D.E., Collett, J.R., Martin, J., Druzhinina, I.S., Mathis, H., Monot, F., et al. (2009). Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* **106**, 16151-16156.
- Lemoine, S., Combes, F., Servant, N., and Le Crom, S. (2006). Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics* **7**, 467.
- Li, Q., Herrler, M., Landsberger, N., Kaludov, N., Ogryzko, V.V., Nakatani, Y., and Wolffe, A.P. (1998). *Xenopus* NF-Y pre-sets chromatin to potentiate p300 and acetylation-responsive transcription from the *Xenopus hsp70* promoter in vivo. *EMBO J* **17**, 6300-6315.
- Linhoff, M.W., Wright, K.L., and Ting, J.P. (1997). CCAAT-binding factor NF-Y and RFX are required for in vivo assembly of a nucleoprotein complex that spans 250 base pairs: the invariant chain promoter as a model. *Mol Cell Biol* **17**, 4589-4596.
- Lockington, R.A., and Kelly, J.M. (2001). Carbon catabolite repression in *Aspergillus nidulans* involves deubiquitination. *Mol Microbiol* **40**, 1311-1321.
- Lockington, R.A., and Kelly, J.M. (2002). The WD40-repeat protein CreC interacts with and stabilizes the deubiquitinating enzyme CreB in vivo in *Aspergillus nidulans*. *Mol Microbiol* **43**, 1173-1182.
- Lohr, D., Venkov, P., and Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *Faseb J* **9**, 777-787.
- MacCabe, A.P., Miro, P., Ventura, L., and Ramon, D. (2003). Glucose uptake in germinating *Aspergillus*

- nidulans conidia: involvement of the creA and sorA genes. *Microbiology* **149**, 2129-2136.
- Mach, R.L., Seiboth, B., Myasnikov, A., Gonzalez, R., Strauss, J., Harkki, A.M., and Kubicek, C.P. (1995). The bgl1 gene of Trichoderma reesei QM 9414 encodes an extracellular, cellulose-inducible beta-glucosidase involved in cellulase induction by sophorose. *Mol Microbiol* **16**, 687-697.
- Mach, R.L., Strauss, J., Zeilinger, S., Schindler, M., and Kubicek, C.P. (1996). Carbon catabolite repression of xylanase I (xyn1) gene expression in Trichoderma reesei. *Mol Microbiol* **21**, 1273-1281.
- Mach, R.L., and Zeilinger, S. (2003). Regulation of gene expression in industrial fungi: Trichoderma. *Appl Microbiol Biotechnol* **60**, 515-522.
- Mach-Aigner, A.R., Pucher, M.E., Steiger, M.G., Bauer, G.E., Preis, S.J., and Mach, R.L. (2008). Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in Hypocrea jecorina. *Appl Environ Microbiol* **74**, 6554-6562.
- Malik, S., and Roeder, R.G. (2000). Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem Sci* **25**, 277-283.
- Mandels, M., and Andreotti, R.E. (1978). Problems and Challenges in Cellulose to Cellulase Fermentation. *Process Biochemistry* **13**, 6-8.
- Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R., and Monot, F. (2009). New improvements for lignocellulosic ethanol. *Curr Opin Biotechnol* **20**, 372-380.
- Martinez, A.T., Ruiz-Duenas, F.J., Martinez, M.J., Del Rio, J.C., and Gutierrez, A. (2009). Enzymatic delignification of plant cell wall: from nature to mill. *Curr Opin Biotechnol* **20**, 348-357.
- Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D., et al. (2008). Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). *Nat Biotechnol* **26**, 553-560.
- Melin, P., Schnurer, J., and Wagner, E.G. (2003). Characterization of phiA, a gene essential for phialide development in Aspergillus nidulans. *Fungal genetics and biology : FG & B* **40**, 234-241.
- Mogensen, J., Nielsen, H.B., Hofmann, G., and Nielsen, J. (2006). Transcription analysis using high-density micro-arrays of Aspergillus nidulans wild-type and creA mutant during growth on glucose or ethanol. *Fungal Genetics and Biology* **43**, 593-603.
- Nakari-Setälä, T., Paloheimo, M., Kallio, J., Vehmaanpera, J., Penttilä, M., and Saloheimo, M. (2009). Genetic modification of carbon catabolite repression in Trichoderma reesei for improved protein production. *Appl Environ Microbiol* **75**, 4853-4860.
- Newcomb, L.L., Hall, D.D., and Heideman, W. (2002). AZF1 is a glucose-dependent positive regulator of CLN3 transcription in Saccharomyces cerevisiae. *Molecular and cellular biology* **22**, 1607-1614.
- Nogawa, M., Goto, M., Okada, H., and Morikawa, Y. (2001). L-Sorbose induces cellulase gene transcription in the cellulolytic fungus Trichoderma reesei. *Current genetics* **38**, 329-334.
- Ohara, T., Inoue, I., Namiki, F., Kunoh, H., and Tsuge, T. (2004). REN1 is required for development of microconidia and macroconidia, but not of chlamydospores, in the plant pathogenic fungus Fusarium oxysporum. *Genetics* **166**, 113-124.
- Otte, S., and Barlowe, C. (2004). Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. *Nat Cell Biol* **6**, 1189-1194.
- Panozzo, C., Cornillot, E., and Felenbok, B. (1998). The CreA repressor is the sole DNA-binding protein responsible for carbon catabolite repression of the alcA gene in Aspergillus nidulans via its binding to a couple of specific sites. *The Journal of biological chemistry* **273**, 6367-6372.
- Papavizas, G.C. (1985). Trichoderma and Gliocladium: Biology, Ecology, and Potential for Biocontrol. *Annual Review of Phytopathology* **23**, 23-54.
- Penalva, M.A., Tilburn, J., Bignell, E., and Arst, H.N., Jr. (2008). Ambient pH gene regulation in fungi: making connections. *Trends Microbiol* **16**, 291-300.
- Peter, G.J., During, L., and Ahmed, A. (2006). Carbon catabolite repression regulates amino acid permeases in Saccharomyces cerevisiae via the TOR signaling pathway. *The Journal of biological chemistry* **281**, 5546-5552.
- Pilauri, V., Bewley, M., Diep, C., and Hopper, J. (2005). Gal80 dimerization and the yeast GAL gene switch. *Genetics* **169**, 1903-1914.
- Piper, P.W., Ortiz-Calderon, C., Holyoak, C., Coote, P., and Cole, M. (1997). Hsp30, the integral plasma membrane heat shock protein of Saccharomyces cerevisiae, is a stress-inducible regulator of plasma membrane H(+)-ATPase. *Cell Stress Chaperones* **2**, 12-24.
- Portnoy, T., Margeot, A., Linke, R., Atanasova, L., Fekete, E., Sandor, E., Hartl, L., Karaffa, L., Druzhinina, I.S., Seiboth, B., et al. (2011a). The CRE1 carbon catabolite repressor of the fungus Trichoderma reesei: a master regulator of carbon assimilation. *BMC Genomics* **12**, 269.
- Portnoy, T., Margeot, A., Seidl-Seiboth, V., Le Crom, S., Ben Chaabane, F., Linke, R., Seiboth, B., and Kubicek, C.P. (2011b). Differential Regulation of the Cellulase Transcription Factors XYR1, ACE2, and ACE1 in Trichoderma reesei Strains Producing High and Low Levels of Cellulase. *Eukaryotic Cell* **10**, 262-271.
- Puyesky, M., Benhamou, N., Noyola, P.P., Bauw, G., Ziv, T., Van Montagu, M., Herrera-Estrella, A., and Horwitz, B.A. (1999). Developmental regulation of cmp1, a gene encoding a multidomain conidiospore surface protein of Trichoderma. *Fungal genetics and biology : FG & B* **27**, 88-99.

- Rauscher, R., Wurleitner, E., Wacenovskiy, C., Aro, N., Stricker, A.R., Zeilinger, S., Kubicek, C.P., Penttila, M., and Mach, R.L. (2006). Transcriptional regulation of *xyn1*, encoding xylanase I, in *Hypocrea jecorina*. *Eukaryot Cell* **5**, 447-456.
- Reese, E. (1976). History of the cellulase program at the US Army Natick Development Center. *Biotech Bioeng Symp* **6**, 9-20.
- Reese, E., Levinsons, H., and Downing, M. (1950). Quartermaster culture collection. *Farlowia* **4**, 45-86.
- Roiger, D.J., Jeffers, S.N., and Caldwell, R.W. (1991). Occurrence of *Trichoderma* species in apple orchard and woodland soils. *Soil Biology and Biochemistry* **23**, 353-359.
- Rolland, F., Winderickx, J., and Thevelein, J.M. (2002). Glucose-sensing and -signalling mechanisms in yeast. *FEMS Yeast Res* **2**, 183-201.
- Ronne, H. (1995). Glucose repression in fungi. *Trends in Genetics* **11**, 12-17.
- Rubin, E.M. (2008). Genomics of cellulosic biofuels. *Nature* **454**, 841-845.
- Ruijter, G.J., and Visser, J. (1997). Carbon repression in *Aspergilli*. *FEMS Microbiol Lett* **151**, 103-114.
- Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374-378.
- Saloheimo, A., Aro, N., Ilmen, M., and Penttila, M. (2000). Isolation of the *ace1* gene encoding a Cys(2)-His(2) transcription factor involved in regulation of activity of the cellulase promoter *cbh1* of *Trichoderma reesei*. *J Biol Chem* **275**, 5817-5825.
- Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyyssonen, E., Bhatia, A., Ward, M., and Penttila, M. (2002). Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *European journal of biochemistry / FEBS* **269**, 4202-4211.
- Samuels, G.J. (2006). *Trichoderma*: Systematics, the Sexual State, and Ecology. *Phytopathology* **96**, 195-206.
- Seiboth, B., Gamauf, C., Pail, M., Hartl, L., and Kubicek, C.P. (2007a). 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. *Mol Microbiol* **66**, 890-900.
- Seiboth, B., Hartl, L., Pail, M., Fekete, E., Karaffa, L., and Kubicek, C.P. (2004). The galactokinase of *Hypocrea jecorina* is essential for cellulase induction by lactose but dispensable for growth on d-galactose. *Mol Microbiol* **51**, 1015-1025.
- Seiboth, B., Hartl, L., Salovuori, N., Lanthaler, K., Robson, G.D., Vehmaanpera, J., Penttila, M.E., and Kubicek, C.P. (2005). Role of the *bga1*-encoded extracellular {beta}-galactosidase of *Hypocrea jecorina* in cellulase induction by lactose. *Appl Environ Microbiol* **71**, 851-857.
- Seiboth, B., Hofmann, G., and Kubicek, C.P. (2002). Lactose metabolism and cellulase production in *Hypocrea jecorina*: the *gal7* gene, encoding galactose-1-phosphate uridylyltransferase, is essential for growth on galactose but not for cellulase induction. *Mol Genet Genomics* **267**, 124-132.
- Seiboth, B., Messner, R., Gruber, F., and Kubicek, C.P. (1992). Disruption of the *Trichoderma-Reesei-Cbh2* Gene Coding for Cellobiohydrolase-Ii Leads to a Delay in the Triggering of Cellulase Formation by Cellulose. *J Gen Microbiol* **138**, 1259-1264.
- Seiboth, B., Pakdaman, B.S., Hartl, L., and Kubicek, C.P. (2007b). Lactose metabolism in filamentous fungi: how to deal with an unknown substrate. *Fungal Biol Reviews* **21**, 42-48.
- Simmons, E. (1977). Classification of some cellulase-producing *Trichoderma* species. In *Biology of Trichoderma symposium, 2nd International Mycological Congress Abstracts* (Tampa, FL, USA, University of South Florida), p. 618.
- Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3.
- Strauss, J., Mach, R.L., Zeilinger, S., Hartler, G., Stoffler, G., Wolschek, M., and Kubicek, C.P. (1995). Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*. *FEBS Lett* **376**, 103-107.
- Stricker, A.R., Grosstessner-Hain, K., Wurleitner, E., and Mach, R.L. (2006). Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryot Cell* **5**, 2128-2137.
- Stricker, A.R., Mach, R.L., and de Graaff, L.H. (2008a). Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). *Appl Microbiol Biotechnol* **78**, 211-220.
- Stricker, A.R., Steiger, M.G., and Mach, R.L. (2007). Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. *FEBS Lett* **581**, 3915-3920.
- Stricker, A.R., Trefflinger, P., Aro, N., Penttila, M., and Mach, R.L. (2008b). Role of Ace2 (Activator of Cellulases 2) within the *xyn2* transcriptosome of *Hypocrea jecorina*. *Fungal Genet Biol* **45**, 436-445.
- Sudarsanam, P., Iyer, V.R., Brown, P.O., and Winston, F. (2000). Whole-genome expression analysis of *snf/swi* mutants of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3364-3369.

- Takashima, S., Ikura, H., Nakamura, A., Masaki, H., and Uozumi, T. (1996). Analysis of Cre1 binding sites in the *Trichoderma reesei* cbh1 upstream region. *FEMS microbiology letters* **145**, 361-366.
- Thomas-Chollier, M., Sand, O., Turatsinze, J.V., Janky, R., Defrance, M., Vervisch, E., Brohee, S., and van Helden, J. (2008). RSAT: regulatory sequence analysis tools. *Nucleic acids research* **36**, W119-127.
- Todd, R.B., Lockington, R.A., and Kelly, J.M. (2000). The *Aspergillus nidulans* creC gene involved in carbon catabolite repression encodes a WD40 repeat protein. *Mol Gen Genet* **263**, 561-570.
- Trumbly, R.J. (1992). Glucose repression in the yeast *Saccharomyces cerevisiae*. *Mol Microbiol* **6**, 15-21.
- Van Aelst, L., Barr, M., Marcus, S., Polverino, A., and Wigler, M. (1993). Complex formation between RAS and RAF and other protein kinases. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 6213-6217.
- Verbeke, J., Coutinho, P., Mathis, H., Quenot, A., Record, E., Asther, M., and Heiss-Blanquet, S. (2009). Transcriptional profiling of cellulase and expansin-related genes in a hypercellulolytic *Trichoderma reesei*. *Biotechnology Letters* **31**, 1399-1405.
- Vitikainen, M., Arvas, M., Pakula, T., Oja, M., Penttila, M., and Saloheimo, M. (2010). Array comparative genomic hybridization analysis of *Trichoderma reesei* strains with enhanced cellulase production properties. *BMC Genomics* **11**, 441.
- Wei, H., Xu, Q., Taylor, L.E., 2nd, Baker, J.O., Tucker, M.P., and Ding, S.Y. (2009). Natural paradigms of plant cell wall degradation. *Curr Opin Biotechnol* **20**, 330-338.
- Welsh, L.M., Tong, A.H., Boone, C., Jensen, O.N., and Otte, S. (2006). Genetic and molecular interactions of the Erv41p-Erv46p complex involved in transport between the endoplasmic reticulum and Golgi complex. *J Cell Sci* **119**, 4730-4740.
- Westholm, J.O., Nordberg, N., Muren, E., Ameer, A., Komorowski, J., and Ronne, H. (2008). Combinatorial control of gene expression by the three yeast repressors Mig1, Mig2 and Mig3. *BMC Genomics* **9**, 601.
- Xie, Y. (2010). Feedback regulation of proteasome gene expression and its implications in cancer therapy. *Cancer Metastasis Rev* **29**, 687-693.
- Xie, Y., and Varshavsky, A. (2001). RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3056-3061.
- Zeilinger, S., Ebner, A., Marosits, T., Mach, R., and 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. *Mol Genet Genomics* **266**, 56-63.
- Zeilinger, S., Mach, R.L., and Kubicek, C.P. (1998). Two adjacent protein binding motifs in the cbh2 (cellobiohydrolase II-encoding) promoter of the fungus *Hypocrea jecorina* (*Trichoderma reesei*) cooperate in the induction by cellulose. *J Biol Chem* **273**, 34463-34471.
- Zeilinger, S., Schmoll, M., Pail, M., Mach, R.L., and Kubicek, C.P. (2003). Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter cbh2 associated with cellulase induction. *Mol Genet Genomics* **270**, 46-55.
- Zimmermann, R., Eyrisch, S., Ahmad, M., and Helms, V. (2011). Protein translocation across the ER membrane. *Biochim Biophys Acta* **1808**, 912-924.

7 ANNEXES

7.1 Tableaux annexes

Tableau annexe 1. Oligonucléotides de PCRq utilisés dans cette étude

Transcrit étudié	Sens de l'amorce	Séquence
<i>gpd1</i>	Sens	ggagctctttgaagagga
<i>gpd1</i>	Anti-sens	ggcaggtacttgacgttttc
<i>xyr1</i>	Sens	ccatcaaccttctagacgac
<i>xyr1</i>	Anti-sens	aacctgcaggagatagac
<i>cbh1</i>	Sens	ccgagcttggtagtactctg
<i>cbh1</i>	Anti-sens	ggtagccttcttgactgagt
<i>bga1</i>	Sens	gtacaaggttgactcccttc
<i>bga1</i>	Anti-sens	ccatagtcagagccgtagag
<i>ace1</i>	Sens	ggacgaggaggagattatg
<i>ace1</i>	Anti-sens	gtgagtcttctcgtgctt
<i>ace2</i>	Sens	gacaagaagctcaggtgtc
<i>ace2</i>	Anti-sens	actgtgttcattggctgtg

Tableau annexe 2. Ratios d'expression des transcrits mesurés par PCR quantitative

transcrit	souche	QM 9414	QM 9414	$\Delta cre1$	$\Delta cre1$	CL847	CL847
	condition	Glucose	Lactose	Glucose	Lactose	Glucose	Lactose
<i>xyr1</i>	BFB	1,00	1,68	9,62	15,70	82,24	178,20
<i>xyr1</i>	1 h	19,56	381,34	9,17	55,35		
<i>xyr1</i>	6 h	25,85	376,13	20,60	77,83	125,30	855,63
<i>xyr1</i>	24 h	23,24	416,36	22,75	137,49	113,94	2603,39
<i>cbh1</i>	BFB	2,78	1,00	7,75	11,72	9415,90	34801,59
<i>cbh1</i>	1 h	6,47	46614,66	10,66	31,83		
<i>cbh1</i>	6 h	14,01	73488,23	15,56	22173,82	11561,49	129577,15
<i>cbh1</i>	24 h	13,17	65806,45	19,47	38912,24	21607,00	1034672,23
<i>ace1</i>	BFB	1,27	1,15	4,74	10,64	3,79	6,92
<i>ace1</i>	1 h	1,79	3,61	5,75	14,28		
<i>ace1</i>	6 h	1,96	3,63	8,20	10,53	2,73	6,37
<i>ace1</i>	24 h	1,97	5,13	5,47	14,42	1,00	7,41
<i>ace2</i>	BFB	1,98	1,96	1,00	2,03	2,46	9,53
<i>ace2</i>	1 h	3,88	8,46	1,17	2,76		
<i>ace2</i>	6 h	4,55	9,94	2,20	2,95	2,78	10,74
<i>ace2</i>	24 h	3,01	10,01	2,05	4,95	1,65	13,75

transcrit	souche	RUT C30	RUT C30
	condition	Glucose	Lactose
<i>xyr1</i>	BFB	7,82	10,54
<i>xyr1</i>	6 h	6,55	183,76
<i>xyr1</i>	24 h	19,20	450,60

Tableau annexe 3. Largeur moyenne des hyphes et des parois de la souche QM 9414 de *T. reesei* et de son mutant $\Delta cre1$. Les données sont issues de la moyenne de 50 mesures.

Souche	Largeur des hyphes (μm)	Largeur des parois (μm)
QM 9414	2,277 \pm 0,2628	0,3981 \pm 0,0449
$\Delta cre1$	2,816 \pm 0,2951	0,5412 \pm 0,0693

Tableau annexe 4. Gènes identifiés dans l'étude de la répression catabolique et leurs protéines associées. Les quatre avant-dernières colonnes sont exprimées en \log_2 (ratio d'expression). Les chiffres des en-têtes correspondent aux conditions expérimentales définies dans la **Figure 13**.

Transcrit	Fonction	1	2	3	4	Groupe
120117	CRE1	-0.394	0.148	-2.362	-2.666	
22459	putative carboxypeptidase A	2.561	2.671	NA	0.282	A
22912	glucose transporter HXT1	4.852	2.866	1.726	-0.971	A
23228	hypothetical protein Fuf1479	3.214	2.806	-0.215	-1.251	A
45717	GH47 α -1,2-mannosidase	2.3	2.044	-0.165	-0.655	A
49898	glucosamine-6-phosphate isomerase	3.169	2.196	0.676	-0.415	A
68924	putative cyclopropane/fatty acid synthase	2.348	2.066	NA	0.669	A
74807	hypothetical glycosyl hydrolase, not aligned with a GH family	2.509	1.821	0.722	0.133	A
77512	cytochrome P450	2.246	1.673	NA	0.204	A
79671	N-acetyl-glucosamine-6-phosphate deacetylase	2.297	0.991	NA	-0.572	A
79677	N-acetyl-glucosamine kinase	2.111	1.223	NA	-0.739	A
80833	N-acetylglucosamine permease	3.445	3.095	NA	-1.144	A
104322	unique protein	2.22	1.241	NA	-0.02	A
106248	monosaccharide transporter	2.781	1.685	NA	-0.622	A
108586	hypothetical protein	3.041	1.403	NA	-0.152	A
121415	Zn2Cys6 transcription factor	2.211	2.185	-0.043	-0.131	A
122271	Zn2Cys6 transcription factor, high similarity to <i>C. albicans</i> Fcr1	2.346	2.897	NA	0.771	A
123079	short chain dehydrogenase/reductase	3.186	1.764	NA	-0.384	A
123234	Coproporphyrinogen III oxidase	3.023	3.22	NA	0.229	A
123726	amino acid transporter	3.913	2.411	1.014	-0.622	A
1925	fatty acid hydroxylase	-2.073	-2.014	0.437	0.874	B
41895	glutathione-S-transferase	-2.627	-2.433	-0.41	-0.613	B
44967	hypothetical protein, conserved	-2.72	-2.626	-0.25	-0.176	B
44987	ERG25 methylsterol desaturase	-1.503	-2.228	0.322	-0.083	B
45250	oleate-delta12-desaturase	-1.956	-2.934	0.073	-0.963	B
49205	cytochrome C peroxidase	-2.117	-1.476	-0.409	0.669	B
49589	unique protein	-2.137	-1.848	-1.098	-0.663	B
50215	GH16 candidate β -glycosidase	-2.049	-3.941	0.943	-0.729	B
65522	unique protein	-1.033	-2.348	0.714	-0.09	B
68608	Thiazole biosynthetic enzyme (orthologue of Stress-inducible protein sti35)	-4.435	-3.889	-0.748	-0.262	B
74060	hypothetical protein with similarity to <i>Aspergillus</i> allergen	-2.696	-1.899	-1.22	-0.758	B
74854	stearic acid desaturase Sde1	-2.23	-3.066	0.439	-0.319	B
76218	protein of zinc containing alcohol dehydrogenase superfamily	-0.929	-2.102	0.663	0.098	B
76381	urate oxidase	-1.282	-2.149	0.978	0.286	B

76722	flavo-hemoglobin	-2.299	-1.748	-0.448	NA	B
77481	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	-2.806	-2.651	-0.024	0.118	B
80920	alcohol dehydrogenase ADH1	-1.511	-2.455	0.201	-1.07	B
82032	hypothetical protein	-3.934	-3.825	-0.42	-0.233	B
110620	hypothetical protein	-2.155	-1.552	-0.941	-0.191	B
111082	putative glutathione S transferase	-2.216	-3.368	-0.45	-1.763	B
121230	unique protein with similarity to phosphoproteoglycans	-2.389	-2.454	-0.027	0.139	B
121491	trehalose synthase CCG9	-2.797	-1.955	-0.912	0.135	B
121620	NMT1 thiamine biosynthesis protein	-6.66	-5.859	-1.016	-0.081	B
123475	Putative cell wall T-rich mannoprotein. Distantly related to <i>S. c.</i> Dan4p.	-1.727	-3.33	1.298	-0.268	B
3363	hypothetical protein with 9 transmembrane domains	NA	-1.541	2.531	1.769	C
4950	urea transporter	-0.092	-0.759	1.755	2.284	C
49970	H ⁺ nucleoside cotransporter	-0.074	-1.356	2.746	1.679	C
62172	amino acid permease	NA	-0.929	2.765	2.564	C
70860	allantoate permease	0.288	0.185	2.229	2.319	C
72922	GABA permease	-2.434	-1.335	0.46	1.815	C
75383	short chain dehydrogenase/reductase	-0.384	-0.101	1.187	2.133	C
76766	xanthine/uracil permease	-1.536	-1.225	1.487	2.137	C
80086	MSF peptide transporter	NA	-2.374	2.549	1.057	C
81022	allantoate permease	NA	-1.441	3.252	2.201	C
82095	high affinity ammonium transporter	-1.801	-1.623	3.913	4.612	C
120357	putative Zn-binding oxidoreductase	NA	0.718	2.999	4.246	C
121136	unique protein	0.497	-0.319	3.223	3.563	C
123718	neutral amino acid permease	0.482	-2.246	5.007	2.457	C
123795	allantoinase	-2.409	-1.265	0.099	1.746	C
123976	unique secreted protein	0.035	-2.155	2.135	1.116	C
21725	GH20 N-acetyl-glucosaminidase NAG1	2.101	0.742	0.608	-0.852	D
21960	phospholipase C	2.085	0.277	0.547	-1.337	D
37525	G-coupled receptor GprK-type	2.251	0.555	1.048	-0.832	D
38812	iron transporter	2.408	0.912	NA	-1.873	D
44175	lactate pyruvate transporter	2.428	0.594	1.152	-0.54	D
44230	GTP-binding protein EsdC	2.669	1.379	-0.065	-1.598	D
50618	MSF transporter	2.719	1.275	1.293	0.165	D
54511	hypothetical protein, contains BTB/POZ domain involved in protein binding	2.1	0.757	0.623	-0.861	D
55319	GH54, L- α -arabinofuranosidase	3.445	1.434	0.331	-1.782	D
55990	hypothetical protein	2.699	0.018	NA	-2.555	D
57749	MSF transporter	1.625	-0.355	NA	-2.914	D
58356	glycerol kinase	2.379	0.856	NA	-1.717	D
60489	CE5 cutinase	2.22	1.824	NA	-1.844	D
60635	GH92 α -1,2-mannosidase	2.124	0.684	NA	-2.383	D
65741	hypothetical transmembrane protein	2.682	0.807	1.034	-1.001	D
69245	GH2 β -mannosidase	2.261	0.923	NA	-1.825	D
72091	hypothetical protein	2.193	0.078	NA	-0.321	D
73179	GH95 α -fucosidase	2.426	0.574	NA	-3.014	D
74198	GH92 α -1,2-mannosidase	2.513	NA	NA	-3.79	D
76852	GH2 glycoside hydrolase	1.879	NA	NA	-2.613	D
77093	acid sphingomyelin phosphodiesterase (probably vacuolar)	1.73	-0.481	-0.114	-2.864	D

78683	aldehyde dehydrogenase	2.019	0.57	0.807	-0.66	D
79816	hypothetical secreted protein	2.342	0.465	0.053	-3.891	D
80019	short-chain dehydrogenase/reductase	2.39	0.766	0.401	-0.664	D
80659	alcohol oxidase AOX1	2.633	NA	NA	-3.664	D
81778	glutaminase A	1.805	-0.01	NA	-2.432	D
108143	unique protein	2.291	NA	NA	-4.145	D
112018	unique glycine-rich secreted protein	2.462	0.113	NA	-1.642	D
120784	putative cell wall mannoprotein	2.393	1.108	0.123	-1.628	D
120873	GH71 α -1,3-glucanase	3.108	0.868	-0.173	-2.514	D
120975	hypothetical protein	2.055	NA	NA	-2.058	D
121475	unique protein	2.741	1.029	0.202	-1.696	D
122147	unique protein	1.724	-0.067	-0.244	-2.468	D
123786	GH16 β -1,3, β -1,4-glucanase	3.423	0.679	0.95	-2.536	D
123978	hypothetical methanol oxidase	2.703	1.737	-0.111	-1.582	D
124043	laccase, secreted	1.973	NA	NA	-2.582	D
1885	glucamylase with starch binding domain	2.009	-0.29	2.399	-0.157	E
4999	P450 monooxygenase	2.63	1.276	2.376	1.596	E
21758	isocitrate lyase	1.633	-0.489	2.868	0.855	E
23171	peptaibol synthase	2.658	-1.359	1.231	-3.28	E
23415	amino acid permease Dip5	1.496	-0.122	2.697	1.093	E
27992	PTH11-type G-coupled receptor	2.612	-0.201	2.123	-0.715	E
32712	NACHT domain protein with ankyrin repeats	3.039	1.742	3.073	1.713	E
44747	SNF2 family helicase	4.214	0.716	3.2	-0.852	E
45445	tyrosinase	2.53	-0.032	2.385	-0.211	E
46794	MSF peptide transporter	1.571	-0.191	3.069	1.413	E
47710	monosaccharide transporter	3.649	-1.389	4.371	-0.608	E
51110	IndA1 amino acid transporter	2.153	-0.645	3.93	0.864	E
51650	hypothetical protein	0.855	-0.81	2.527	0.751	E
51893	putative peroxidase/hem containing	2.796	0.794	1.945	-0.253	E
54667	acyl-CoA synthetase	2.469	1.493	2.267	0.75	E
54768	hypothetical protein	2.299	-0.612	1.471	-1.201	E
55272	hypothetical protein	5.813	0.321	6.216	-0.097	E
58282	CE9 esterase	2.869	-0.092	3.445	0.579	E
64874	MSF toxin efflux pump	1.69	-1.269	2.527	0.06	E
65191	maltose permease	1.826	0.025	2.502	0.661	E
67275	RAS1	1.838	-0.341	2.221	-0.027	E
68812	MSF sugar transporter	2.623	1.496	2.358	1.275	E
73818	catalase C	1.889	-1.623	3.608	0.023	E
74282	hypothetical cysteine-rich protein	4.55	-2.113	3.228	-4.398	E
74580	hypothetical transmembrane protein	2.996	-1.783	4.06	-0.771	E
76897	MSF transporter	2.697	0.9	1.824	-0.317	E
80026	MSF permease	0.854	-2.631	2.23	-1.683	E
80149	hypothetical protein, conserved	2.006	0.571	2.256	0.726	E
80863	GH18 endochitinase CHI18-5	4.071	-0.69	4.615	-0.296	E
80980	porphyromonas-type peptidyl arginine deiminase	5.239	1.585	4.652	0.732	E
81149	aquaglyceroporin	2.173	-0.413	1.018	-2.029	E
81383	G-protein coupled receptor, GprK-like,	3.569	1.65	2.356	0.145	E
82235	GH31 α -glucosidase	2.343	-0.872	1.227	-2.147	E
108357	hypothetical protein	4.071	0.511	4.2	0.43	E
120017	small oligopeptide transporter	3.55	-0.945	5.417	0.67	E

120031	hypothetical protein	3.859	-0.197	3.474	-0.938	E
120156	cell morphogenesis protein PAG1	2.436	0.143	2.134	-0.452	E
120823	secreted aldose epimerase	4.271	-1.249	4.227	-1.28	E
121308	δ -1-pyrroline-5-carboxylate dehydrogenase	2.28	0.971	2.242	1.206	E
121412	amino acid polyamine transporter II	2.204	0.724	1.599	-0.021	E
121735	GH3 β -glucosidase Cel3B	3.569	0.856	2.594	-0.484	E
121746	GH55 exo- β -1,3-glucosidase ("GLUC78")	3.385	1.183	2.458	0.019	E
122127	cell wall protein PhiA	1.909	-1.245	4.523	1.447	E
122780	GH28 polygalacturonase	4.146	-0.185	3.332	-1.413	E
123009	glutamine synthetase	1.967	-1.205	2.273	-0.459	E
123207	hypothetical protein	3.545	-0.59	3.116	-1.545	E
123456	GH65 $\alpha\alpha'$ -trehalase	0.971	-0.697	2.215	0.93	E
123713	MedA	2.4	-0.248	2.755	0.081	E
123968	hypothetical protein	2.423	0.926	1.6	-0.455	E
124141	hypothetical secreted protein with similarity to phosphoproteoglycan	2.663	-0.102	2.243	-0.602	E
3327	NADH:flavin oxidoreductase/NADH oxidase	-2.289	NA	-1.9	0.102	F
23382	aldehyde reductase AKR7	-5.146	NA	-4.473	NA	F
30759	protein of zinc containing alcohol dehydrogenase superfamily	-3.835	0.307	-3.465	-0.249	F
43701	MSF multidrug transporter	-4.594	0.197	-5.722	-2.668	F
59014	ABC multidrug transporter	-2.544	-0.375	-2.036	0.264	F
60616	integral membrane protein	-2.972	-0.45	-2.362	0.363	F
66041	hypothetical protein	-2.147	-1.116	-1.389	NA	F
73618	polyketide synthase	-4.181	-0.202	-3.623	NA	F
73621	polyketide synthase	-4.415	-0.307	-3.929	NA	F
73623	FAD-monooxygenase	-3.494	-0.111	-2.821	0.056	F
73631	isoamyl alcohol oxidase	-4.418	NA	-4.09	NA	F
74194	L-xylulose reductase LXR1	-2.326	-0.438	-2.011	-0.231	F
74278	Pyridine nucleotide-disulphide oxidoreductase, class-II	-2.349	-0.923	-2.194	-0.399	F
76238	plasma membrane H ⁺ ATPase	-2.957	0.126	-1.548	1.626	F
76763	PTH11-type G-coupled receptor	-2.366	-0.225	-2.298	0.138	F
78970	mitochondrial (phosphate) carrier	-2.462	-0.715	-1.354	0.889	F
80879	H ⁺ /oligopeptide transporter	-3.313	-1.331	-1.721	1.104	F
81525	isoflavone reductase	-2.183	-0.071	-1.903	-0.024	F
81598	GH18 endochitinase CHI18-7	-2.11	-0.807	-1.606	-0.334	F
81979	glutathione-S-transferase	-2.258	0.36	-1.975	0.941	F
82204	high affinity ammonium transporter	-2.324	2.408	-2.377	2.806	F
102497	Zn2Cys6 transcription factor	-2.614	0.217	-3.466	-0.823	F
102499	Zn2Cys6-transcription factor	-2.424	-0.346	-2.099	NA	F
103015	hypothetical protein	-3.104	NA	-3.325	NA	F
105752	C4 dicarboxylate transporter	-3.547	0.411	-3.091	NA	F
106556	hypothetical protein	-5.06	-0.155	-5.095	NA	F
107639	hypothetical protein	-2.593	-0.003	-2.29	0.356	F
119735	glyceraldehyde-3-phosphate dehydrogenase	-2.956	-0.609	-1.473	0.331	F
119759	metR bZIP transcription factor	-2.189	0.333	-1.799	0.595	F
119819	G-protein coupled receptor, mPR-type	-2.039	-0.969	-1.306	0.157	F
120568	enolase	-2.101	-0.627	-1.118	0.317	F
121486	hypothetical protein	-3.612	1.325	-3.219	1.463	F
122523	hypothetical protein	-2.307	-0.649	-1.766	-0.178	F
123084	(chloro)peroxidase	-6.911	0.073	-6.581	-0.031	F

123468	IMP dehydrogenase	-2.738	-0.12	-2.446	-0.347	F
123914	unique secreted protein	-3.08	NA	-3.448	-1.207	F
21412	secreted protein, hypothetical	-0.473	-2.045	-0.274	-2.01	G
50104	phosphatidyl synthase	NA	-2.211	2.113	0.575	G
50212	distantly related to glycosyltransferases	-1.661	-4.402	1.623	-1.051	G
52315	Ctr copper transporter	-0.52	-2.001	0.473	-1.402	G
54198	hypothetical protein, conserved	-0.031	-2.072	0.869	-1.337	G
55443	hypothetical protein, secreted	0.924	-2.275	2.947	0.374	G
59151	BYS1 domain protein (=Blastomyces yeast-phase-specific protein) secreted	0.804	-1.955	2.349	0.145	G
62100	Hsp30	0.438	-1.882	2.078	-0.424	G
65406	GH16, glucanosyltransferase MWG2 (orthologue of <i>A. fumigatus</i> Crf1)	0.262	-2.338	1.432	-1.215	G
71532	GH71 α -1,3-glucanase	1.265	-2.698	3.062	-0.244	G
77547	GT1 glycosyltransferase	NA	-2.396	0.798	NA	G
81082	aquaglyceroporin	NA	-3.819	3.167	NA	G
103179	unique protein	NA	-2.926	4.849	0.678	G
105106	hypothetical protein	0.912	-2.107	1.526	-0.848	G
105224	hypothetical membrane protein	NA	-2.087	NA	NA	G
105313	unique protein	-1.394	-2.503	1.986	-0.876	G
107881	hypothetical protein	NA	-2.496	1.896	0.53	G
108914	hypothetical protein	-1.591	-2.635	-0.469	-2.006	G
109234	putative D-aminopeptidase	NA	-2.051	1.41	0.252	G
111750	ferric reductase	0.326	-2.256	1.819	-1.065	G
120120	GCN5-related N-acetyltransferase	0.538	-2.966	1.75	-0.243	G
121226	2-dehydropantoate-2-reductase	1.291	-2.386	2.537	-0.7	G
121251	putative cell wall protein (Metarhizium adhesion Mad1)	-0.802	-2.327	0.258	-1.256	G
121818	hypothetical secreted protein	-0.501	-2.471	0.724	-1.144	G
122242	hypothetical protein	NA	-2.038	1.093	0.205	G
124282	hypothetical SET and MYND domain protein	-1.3	-3.175	0.782	-0.917	G
45138	sulfite reductase, β -subunit	-0.125	2.119	-0.227	1.983	H
46209	SAM-methyltransferase	NA	1.484	0.258	2.21	H
47066	candidate sulfate adenylyltransferase	0.151	1.567	0.664	2.305	H
52055	mitochondrial NADP-dependent isocitrate dehydrogenase	-0.815	1.251	-0.236	2.203	H
54352	hypothetical protain, conserved	NA	2.457	NA	1.25	H
55886	GH16 β -glycosidase (GPI anchor)	NA	2.034	NA	2.701	H
57204	lipase, secreted	NA	1.886	NA	2.275	H
59173	MSF transporter	NA	1.654	0.184	2.159	H
65410	Phosphoadenosine phosphosulfate reductase	-0.14	3.455	-0.548	3.56	H
70520	short chain dehydrogenase/reductase	NA	0.848	-0.963	2.023	H
72379	conidiospore surface protein CMP1	NA	5.72	NA	6.19	H
76288	short chain dehydrogenase/reductase	-0.764	1.18	-0.374	2.327	H
76633	MDR transporter	-0.05	2.057	-0.964	1.694	H
79741	ER-bound Farnesyl-diphosphate farnesyltransferase	-0.658	3.335	-0.326	3.913	H
81004	aspartate protease, secreted	NA	2.439	-0.3	2.506	H
81576	hypothetical assimilatory sulfite reductase, alpha subunit	-1.024	3.903	-1.552	3.416	H
82026	trans-aconitate methyltransferase	NA	4.152	NA	5.325	H
82208	polyketide synthase	NA	3.697	1.493	5.438	H
107960	hypothetical secreted protein with WSC-CBM	NA	2.117	0.478	2.947	H
108642	hypothetical secreted protein	-0.109	1.742	-0.176	2.063	H

112258	hypothetical protein	-0.118	3.117	0.106	3.486	H
119552	unique secreted protein, S,P,R-rich	0.36	3.995	0.304	4.364	H
123979	peptide transporter MTD1	NA	3.531	NA	4.499	H
124002	MED2 nuclear component of the mediator complex	NA	1.57	NA	2.831	H
124079	hypothetical protein	NA	3.561	NA	4.867	H
124277	putative copper radical oxidase Cro1	NA	3.756	NA	4.3	H
1751	FAD monooxygenase	-3.102	-1.567	-0.847	0.244	X
21908	fatty acid elongase	-2.459	-1.654	-1.519	-1.055	X
39637	C4-dicarboxylate transporter	-0.327	1.178	1.666	3.261	X
43671	ammonium permease MEA1	-3.49	-0.017	-0.063	4.003	X
48211	hypothetical conserved protein, intracellular	0.288	1.685	1.691	3.351	X
49274	GH16 β -1,3/4-glucanase	-2.728	-1.565	-1.549	-0.101	X
58264	putative paxU orthologue involved in indole-terpene biosynthesis	-2.117	-0.743	-0.849	0.419	X
60847	ATP synthase subunit 9, proteolipid P2	-4.492	-0.778	-1.201	2.527	X
64959	phosphatidyl synthase	-2.234	-0.43	-0.581	1.661	X
76215	sulfide:quinone oxidoreductase	-0.476	1.101	1.019	3.095	X
78688	Hypothetical protein with a heat shock factor-type DNA-binding domain	-0.798	0.486	0.799	2.566	X
81690	orthologue of <i>Aspergillus nidulans</i> CreD	-2.183	-1.355	-0.709	0.542	X
107055	hypothetical protein	0.831	1.173	0.939	2.173	X
120877	protein of Zn-dependant β -lactamase family	-0.065	1.215	1.32	3.48	X
121534	pyruvate decarboxylase	-5.326	-2.263	-2.396	0.775	X
122811	glutamine synthetase	-2.431	-0.382	-0.08	2.095	X
123382	subtilisin-type serine protease	-2.708	-1.506	-0.962	0.057	X

Tableau annexe 5. Enrichissement des catégories fonctionnelles (FunCat) de l'étude de la répression catabolique. Pour chaque catégorie fonctionnelle de la **Figure 16**, le taux d'enrichissement de chaque groupe a été calculé par rapport à l'annotation fonctionnelle FunCat totale. Les p-values associées à ces enrichissements ont été calculées à l'aide d'un test hypergéométrique.

Catégories fonctionnelles (FunCat)		Biogenesis Of Cell Components	Cell Rescue, Defense And Virulence	Cellular Signalling	Transported Compounds	Protein Fate (Folding, Modification, Destination)	Transcription	Energy	Secondary Metabolites	Lipid, Fatty Acid And Isoprenoid Metabolism	C-Compound And Carbohydrate Metabolism	Aa Metabolism	General Metabolism
Groupe	FunCat ID	42	32	30	20.01	14	11	02	01.20	01.06	01.05	01.01	01
E	Enrich.	0.35		1.04	1.64	0	0.28	0.72	0.96	0.21	0.97	0.82	0.05
	p-value	4.8e-02		2.3e-01	2.2e-02		1.5e-02	1.9e-01	2.3e-01	3.1e-02	1.6e-01	2.2e-01	3.5e-02
G	Enrich.	0.48	0.41		1.03	0.62	0.39			0.56	1.33		0.15
	p-value	2.6e-01	2.0e-01		2.5e-01	2.1e-01	1.8e-01			3.0e-01	1.9e-01		3.1e-01
C	Enrich.				3.32			0.7				2.41	0.16
	p-value				2.5e-04			3.5e-01				9.2e-02	3.3e-01
H	Enrich.	0.36			1.55	0.47	0.29	0.49	0.65	0.42	0.25	0.56	0.44
	p-value	1.6e-01			1.0e-01	1.1e-01	9.5e-02	2.6e-01	3.4e-01	2.2e-01	5.6e-02	3.0e-01	1.2e-01
D	Enrich.	0.28		1.64	0.8			0.38		0.32	2.3	0.43	0.25
	p-value	8.8e-02		1.7e-01	1.9e-01			1.8e-01		1.3e-01	1.8e-03	2.3e-01	2.4e-01
F	Enrich.		0.39	0.89	1.13		0.55	1.23	1.64		0.62	0.35	0.14
	p-value		6.4e-02	2.8e-01	1.6e-01		1.1e-01	1.9e-01	1.3e-01		1.1e-01	1.6e-01	1.7e-01
B	Enrich.	0.45	1.54					1.23	2.46	2.11	0.62		
	p-value	2.4e-01	1.5e-01					2.8e-01	8.9e-02	7.9e-02	2.1e-01		
A	Enrich.				0.82	0.49	0.61		0.69	0.44	1.83	0.59	0.12
	p-value				2.3e-01	1.3e-01	2.0e-01		3.5e-01	2.3e-01	4.5e-02	3.2e-01	2.4e-01
X	Enrich.		0.39	0.89	0.65	0.29		2.45	0.82	1.06	0.62	1.41	
	p-value		1.8e-01	3.8e-01	2.2e-01	9.3e-02		5.4e-02	3.7e-01	2.9e-01	2.1e-01	2.6e-01	

Tableau annexe 6. Enrichissement fonctionnel global des gènes de l'étude de la répression catabolique. Une analyse d'enrichissement basée sur l'annotation Gene Ontology a été effectuée pour l'ensemble des gènes détectés dans l'étude comme fortement régulés. 4 977 des 9 129 gènes du génome de *T. reesei* sont annotés avec au moins un terme GO, dont 171 de nos 250 gènes. La significativité des ratios d'enrichissement a été calculée à l'aide d'un test hypergéométrique et les p-values ont été ajustées par la méthode de correction des tests multiples de Benjamini-Hochberg. Ce tableau ne présente que les fonctions obtenant des p-values inférieures à 0,05.

Description GO	Étude	Total	Enrich.	p-value
oxidoreductase activity	38	452	2.4	2.09e-05
carbohydrate metabolism	18	142	3.7	9.94e-05
membrane	32	383	2.4	9.94e-05
transport	30	350	2.5	9.97e-05
transporter activity	23	251	2.7	5.04e-04
oligopeptide transport	3	3	29.1	1.96e-03
fatty acid biosynthesis	6	24	7.3	4.66e-03
glucan endo-1,3-alpha-glucosidase activity	3	4	21.8	5.38e-03
ATP binding	6	557	0.3	5.38e-03
regulation of oxidoreductase activity	7	41	5	9.57e-03
nucleus	7	567	0.4	9.57e-03
integral to membrane	26	387	2	9.57e-03
nucleic acid binding	2	306	0.2	2.33e-02
chitinase activity	4	15	7.8	2.33e-02
endochitinase activity	4	15	7.8	2.33e-02
ammonium transporter activity	2	2	29.1	2.33e-02
metabolism	30	519	1.7	2.36e-02
hydrolase activity, hydrolyzing O-glycosyl compounds	9	83	3.2	2.42e-02
carbohydrate transport	7	53	3.8	2.42e-02
transferase activity	6	39	4.5	2.42e-02
cofactor binding	5	27	5.4	2.43e-02
electron transport	22	349	1.8	2.74e-02
DNA binding	4	370	0.3	3.02e-02
amino acid-polyamine transporter activity	6	44	4	3.36e-02
sugar porter activity	7	59	3.5	3.36e-02
amino acid transport	6	44	4	3.36e-02

Tableau annexe 7. Gènes identifiés dans l'étude du transcriptome global au cours de la production de cellulases, et leurs protéines associées. Les neuf avant-dernières colonnes sont exprimées en \log_2 (ratio d'expression). Les en-têtes indiquent les conditions expérimentales définies **Figure 19**.

Transcrit	Annotation	QM6a T0	NG14 T1h	NG14 T3h	NG14 T6h	NG14 T24h	RUTC30 T1h	RUTC30 T3h	RUTC30 T6h	RUTC30 T24h	Groupe
2392	Putative Peroxisomal hydratase-dehydrogenase-epimerase FOX2 / POX2 / MFE2	0.438	-0.45	-1.45	-1.42	-1.759	-0.531	-0.786	-1.307	-1.454	1
21824	Putative Peroxisomal hydratase-dehydrogenase-epimerase	-0.066	-0.363	-1.101	-1.46	-1.734	-0.31	NaN	-1.126	-1.325	1
45624	Putative DUF3712 family protein	-0.07	-0.377	-1.257	-1.413	-1.198	-0.955	-1.24	-1.385	-1.305	1
47710	Putative MFS sugar transporter (D-Xylose)	-0.638	-0.909	-1.5	-1.818	-1.167	-0.738	-1.481	-1.871	-1.359	1
49970	Putative pre-mRNA-splicing ATP-dependent RNA helicase prp-28	0.273	-0.733	-1.67	-1.283	-1.289	-0.731	-0.805	-0.933	-0.579	1
50249	Putative Oligopeptidase PRD1	0.124	-0.346	-1.305	-1.448	-1.662	-0.529	-0.629	-1.042	-1.265	1
56432	Putative Mitochondrial cytochrome b2 precursor	-0.385	-0.118	-1.247	-1.468	-1.429	-0.049	-0.354	-0.659	-0.541	1
57857	Putative Glycoside hydrolase family 2	-0.965	0.876	-2.169	-2.459	-2.028	-0.419	0.082	-0.359	-0.218	1
58356	Putative Glycerol kinase GUT1	0.179	-0.146	-1.58	-1.562	-2.047	-0.776	-0.838	-1.675	-1.746	1
60847	Putative Mitochondrial ATP synthase protein 9 ATP9	-1.103	-1.232	-1.565	-1.955	-2.498	-0.25	-0.192	-0.741	-0.96	1
62502	Putative MFS transporter	-0.132	-0.997	-1.386	NaN	-1.876	-0.022	NaN	-0.799	-0.831	1
62651	Putative Long fatty acid transporter and synthase protein FAT1	-0.259	-0.316	-1.575	-1.472	-1.592	-0.652	NaN	-0.254	-0.408	1
67639	Putative Acetolactate synthase	0.401	0.377	-1.299	-1.691	-1.815	-0.578	NaN	-1.302	-1.949	1
70375	Putative General amidase AMD2	-0.772	0.307	-2.233	-2.653	-2.589	-0.866	NaN	-0.955	-0.794	1
75368	Putative Peroxisomal 3-keto-acyl-CoA thiolase PTHIK / FOX3 / POX3 / POT1	0.554	-0.558	-1.438	-1.548	-2.012	-0.304	-0.402	-0.973	-1.322	1
78552	Putative Aromatic-L-amino-acid decarboxylase	-0.761	-1.354	-2.075	-1.964	-2.06	-0.507	NaN	NaN	-0.433	1
78683	Putative Aldehyde dehydrogenase	0.457	-0.564	-1.702	-1.191	-1.94	-0.608	-1.114	-1.418	-2.177	1
80645	Putative Amidase	0.314	0.334	-1.727	-1.962	-1.953	-0.847	-1.053	-1.326	-1.787	1
80762	Putative Ubiquitin C-terminal hydrolase family protein	-0.76	-0.734	-1.405	-1.438	-1.623	NaN	NaN	-0.574	-1.035	1
80863	Putative MFS transporter	0.4	-0.711	-1.049	-1.11	-0.802	-0.459	-0.703	-1.422	-1.231	1
81303	Putative Fumarate reductase	0.424	-0.33	-1.42	-1.532	-1.745	-0.857	-1.151	-1.686	-1.938	1
82037	Putative MFS transporter	0.283	-0.835	-1.669	-1.598	-1.287	-0.573	-0.75	-0.336	0.042	1
82235	Putative Glycoside hydrolase family 31	-0.039	0.285	-0.527	-1.528	-1.886	-0.384	-0.758	-1.425	-1.886	1
82403	Putative Mitochondrial very long-chain specific acyl-CoA dehydrogenase	-0.083	-0.363	-1.28	-1.306	-1.796	-0.422	-0.491	-0.951	-1.03	1
103147	Putative Cytochrome P450	0.339	0.374	-0.923	-1.175	-1.443	-0.816	-1.178	-1.747	-1.839	1
106575	Putative Glycoside hydrolase family 79	-0.594	-1.105	-2.027	-2.763	-1.431	-0.924	NaN	-0.8	-0.922	1
110414	Putative 4-aminobutyrate aminotransferase	-0.726	0.0090	-1.76	-2.404	-2.299	-0.597	-0.915	-1.311	-1.262	1
120110	Putative 2-nitropropane dioxygenase family oxidoreductase	0.334	-0.422	-1.336	-1.344	-1.762	-0.332	NaN	-1.166	-1.341	1
121298	Putative Formate dehydrogenase FDH1	-0.014	-0.882	-1.964	NaN	NaN	-0.874	-1.147	-1.009	-1.546	1
121405	Putative 4-aminobutyrate aminotransferase UGA1	-0.166	-0.487	-1.731	-1.681	-1.342	-0.461	-0.369	-0.595	-0.736	1

123179	Putative GTP cyclohydrolase II	0.0090	-0.238	-1.13	-1.005	-1.276	-0.733	-1.047	-1.063	-1.557	1
123720	Putative Peroxisomal 3-ketoacyl-CoA thiolase	0.378	-0.503	-1.446	-1.238	-1.768	-0.535	NaN	-1.329	-1.718	1
123738	Putative Arginase ARG1	0.292	-0.363	-2.248	-1.726	-1.784	-1.021	NaN	NaN	-1.71	1
123827	Putative Catalase peroxydase CAT2	0.242	-0.296	-1.37	-0.607	-1.086	-1.001	-1.008	-1.003	-1.47	1
124051	Putative Secreted serine peptidase	-0.729	0.034	-2.898	-1.81	-2.259	-0.53	-0.784	-1.119	-1.663	1
4146	Putative Salicylate monooxygenase	2.199	0.135	0.18	-0.176	-0.425	NaN	NaN	-0.053	0.0040	2
21725	Glycoside hydrolase family 20 NAG1	1.046	0.036	-0.946	-0.883	-0.859	-0.042	NaN	-0.178	-0.065	2
22093	Putative Farnesyltransferase/geranylgeranyltransferase RAM2	1.913	0.877	0.102	-0.8	-1.457	0.225	NaN	-0.548	-0.827	2
27689	Putative MFS transporter	1.496	-0.165	0.69	0.316	0.09	0.384	NaN	NaN	0.331	2
38640	Putative protein of unknown function	1.271	0.741	-0.244	-0.645	-0.63	0.029	NaN	NaN	-0.282	2
44175	Putative MFS transporter (Lactate/pyruvate) JEN1	1.206	-0.473	-0.669	-0.63	-1.131	-0.0050	-0.139	-0.224	-0.662	2
44628	Putative Kynureninase BNA5	1.363	0.145	-0.518	-0.571	-0.908	0.107	NaN	-0.067	-0.33	2
47286	Putative Zinc metalloproteinase	1.326	0.541	0.382	-0.033	-0.563	0.436	-0.128	-0.229	-0.742	2
47603	Putative Succinyl-CoA:3-ketoacid-coenzyme A transferase	1.369	0.017	-1.46	-1.212	-1.375	0.032	-0.294	-0.14	-0.124	2
49589	Putative protein of unknown function	2.098	0.551	-0.154	0.08	0.012	0.356	0.0070	0.691	1.206	2
61212	Putative Cytochrome P450 : isotrichodermin C-15 hydroxylase	3.83	1.008	-1.692	-0.433	-1.013	-0.376	NaN	-0.252	-0.485	2
62611	Putative ABC multidrug transporter	1.027	0.046	0.066	0.287	0.255	-0.105	-0.232	0.139	0.246	2
62693	Putative ABC transporter STE6	1.85	0.845	0.56	0.074	-0.978	0.435	NaN	0.03	-0.172	2
64925	Putative Glycosyltransferase family 32	0.231	-0.282	-1.062	-1.329	-0.969	0.23	NaN	0.598	0.808	2
70373	Putative Secreted DUF1237 domain protein	1.452	0.449	-0.971	-0.774	-0.811	-0.424	NaN	NaN	-0.686	2
70630	Putative Homoserine acetyltransferase family protein	1.719	0.048	-0.044	0.203	-0.095	0.681	NaN	0.748	0.911	2
74020	Putative Orotidine-5'-phosphate (OMP) decarboxylase URA3	1.253	0.46	-0.013	-0.817	-0.639	0.104	NaN	-0.234	1.029	2
76784	Putative Zinc binding dehydrogenase family protein	1.339	0.47	-0.556	-0.762	-1.137	0.347	NaN	-0.268	-0.451	2
77739	Putative Homogentisate 1,2-dioxygenase	1.23	0.122	-0.892	-0.91	-0.916	-0.279	NaN	-0.365	-0.433	2
77749	Putative Fumarylacetoacetase	1.423	0.742	-0.839	-0.931	-1.021	-0.345	NaN	-0.247	-0.418	2
79396	Putative Glycosyltransferase family 2	1.156	0.545	-0.046	-0.331	-0.333	0.178	-0.286	-0.225	0.281	2
80920	Putative Alcohol dehydrogenase ADH5	0.182	-0.036	-0.418	-0.399	-0.582	1.185	0.606	0.629	0.354	2
81089	Putative Cystathionine beta-synthase CYS4	1.081	0.349	-0.547	-0.355	-0.454	0.103	NaN	-0.114	-0.082	2
81149	Putative F-Box domain protein	1.507	0.49	0.105	0.214	0.333	0.05	NaN	0.173	0.225	2
81430	Putative Sodium-transporting Na+-ATPase ENA5	1.023	0.732	-0.118	-0.085	-0.412	-0.226	-0.384	-0.533	-0.989	2
81522	Putative Selenocysteine lyase-like protein	1.076	0.356	-0.655	-0.482	-0.644	-0.105	NaN	-0.384	-0.399	2
82017	Putative MFS siderochrome iron transporter SIT1	0.296	1.411	0.433	-0.41	-0.019	-0.703	-0.346	-0.967	-0.123	2
102735	Putative WSC carbohydrate binding domain protein	1.884	-0.318	-0.391	-0.64	-0.965	0.213	NaN	0.234	0.069	2
102820	Putative Secreted multicopper oxidase (laccase) FET3	1.043	0.879	1.436	-0.148	-0.607	0.2	-0.029	-0.042	0.1	2
103275	Putative Zn-finger domain protein AZF1	2.038	0.595	-0.1	-0.485	-0.777	-0.0030	NaN	0.125	-0.0010	2

104140	Putative Mitochondrial 3-hydroxyisobutyrate dehydrogenase	0.902	0.14	-1.194	-0.669	-1.204	NaN	NaN	0.298	-0.064	2
109078	Putative Farnesyltransferase beta subunit RAM1	1.942	0.859	0.427	-0.174	-0.621	0.243	NaN	-0.104	-0.256	2
119551	Putative Mitochondrial carrier protein LEU5	0.761	0.196	-1.044	-0.356	-0.426	-0.325	NaN	-0.137	-0.116	2
119694	Putative protein of unknown function	1.254	0.719	-0.511	-0.2	0.076	-0.145	NaN	0.072	0.13	2
120060	Putative Peptidase family M20/M25/M40 protein	1.111	0.564	0.374	-0.01	0.234	0.088	NaN	-0.128	0.049	2
120698	Putative pH-response transcription factor pacC/RIM101	1.145	0.726	0.436	-0.77	-0.761	0.311	-0.395	-0.716	-0.576	2
123260	Putative protein of unknown function	0.918	0.41	-0.274	0.255	0.159	-0.463	NaN	-0.631	-1.008	2
123327	Putative SWI-SNF chromatin remodeling complex protein	1.439	0.353	-0.469	-0.81	-0.938	0.111	-0.019	-0.426	-0.516	2
123502	Putative Carbohydrate-Binding Module Family 21 GAC1	1.064	-0.026	-0.415	-0.54	-0.553	-0.238	NaN	-0.441	-0.228	2
123702	Putative MFS transporter	1.195	0.31	-0.789	-0.941	-0.9	0.207	0.048	0.236	0.593	2
124115	Putative Phosphoenolpyruvate carboxykinase PCK1	2.189	0.83	-0.592	-0.869	-1.363	-0.185	NaN	-0.946	-1.127	2
3434	Putative Patatin-like phospholipase	-1.009	0.071	-0.15	-0.4	-0.124	-0.392	-0.653	-0.839	-0.875	3
3464	Putative Acyl-CoA dehydrogenase family protein	-0.379	0.265	-0.703	-0.786	-0.933	-0.467	NaN	-0.986	-1.103	3
4480	Putative Dihydroxyacetone kinase 1 DAK2	-0.032	0.185	-0.14	-0.22	-0.461	-0.471	-0.57	-1.012	-1.327	3
5016	Putative protein of unknown function	0.302	0.031	-0.781	-0.029	-0.372	-0.747	NaN	-1.298	-1.156	3
5270	Putative Non-specific lipid-transfer protein	0.044	-0.229	-0.8	-1.093	-1.258	-0.091	NaN	-0.255	-0.584	3
21373	Putative Serine/threonine protein kinase SPS1	0.013	0.245	-0.127	-0.488	-0.56	-0.372	-0.535	-0.913	-1.087	3
21836	Putative Phosphoglucomutase PGM2 /PGM1	-0.546	-0.361	-1.005	-0.731	-0.935	-0.679	-0.869	-0.844	-1.01	3
21873	Putative Dihydroxy-acid dehydratase ILV3	-0.173	-0.18	-0.773	-0.987	-1.213	-0.3	NaN	-0.729	-0.698	3
22030	Putative Phenylacetyl-CoA ligase	-0.398	-0.263	-0.9	-0.905	-1.111	-0.261	NaN	-0.499	-0.616	3
22341	Putative Microtubule-associated protein ASE1	-0.337	0.575	0.259	-0.243	-0.553	-0.452	-0.316	-0.844	-1.163	3
22415	Aldose 1 epimerase AEP1	-0.475	-0.493	-1.05	-1.05	-1.333	-0.273	NaN	-0.287	-0.753	3
22654	Putative Ferredoxin reductase-like	0.187	-0.468	-1.148	-0.635	-0.925	-0.407	NaN	-0.892	-1.196	3
22945	Putative Uracil phosphoribosyltransferase FUR1	-0.15	-0.254	-0.727	-0.609	-0.718	-0.611	NaN	-0.882	-1.069	3
23367	Putative Choline oxidase	0.435	-0.446	-0.537	-1.086	-1.389	-0.457	-0.246	-0.8	-0.889	3
23408	Putative protein of unknown function	0.305	0.102	-0.961	-0.881	-1.144	-0.628	NaN	-0.626	-1.246	3
31415	Putative protein of unknown function	0.013	-0.051	-0.339	-1.034	-0.859	NaN	-0.048	NaN	-0.41	3
41035	Putative GPCR receptor	-1.099	-0.045	0.051	-1.718	-1.105	-1.286	-1.182	-2.376	-2.328	3
42866	Putative Thioesterase	-0.224	-0.336	-0.468	-0.585	-1.09	0.258	NaN	NaN	-0.216	3
43129	Putative protein of unknown function	-0.774	-0.056	-1.02	-0.83	-0.937	-0.929	-0.902	-0.745	-1.272	3
44214	Carbohydrate Esterase family 5 AXE2	-0.584	-0.702	-1.463	-0.959	-0.852	-0.656	NaN	-0.481	-0.921	3
46002	Putative protein of unknown function	-0.231	0.763	-0.075	-0.162	0.0060	-0.367	-0.489	-0.87	-1.439	3
46936	Putative 2-deoxy-D-gluconate 3-dehydrogenase	0.177	-0.038	-0.408	-0.632	-1.159	NaN	NaN	-0.258	-0.59	3
47153	D-galacturonic acid reductase GAR2	-0.461	-0.401	-0.713	-0.886	-1.024	-0.193	-0.0070	-0.537	-0.792	3
47221	Putative Nucleoside diphosphate kinase YNK1	-1.393	-0.838	-1.054	-1.05	-0.739	-0.574	-0.549	-0.771	-1.244	3

48211	Putative protein of unknown function	0.16	0.054	-0.074	NaN	0.475	-0.794	NaN	-1.14	-1.213	3
48482	Putative Protein similar to MEMO1	-0.759	-0.041	-0.535	-0.403	-0.563	-0.644	-0.665	-0.924	-1.452	3
50491	Putative protein of unknown function	-0.488	-0.361	-0.636	-0.953	-1.001	-0.358	-0.297	-0.594	-0.806	3
51365	Subtilisin-like P8048	0.107	1.429	-0.113	-0.932	-0.959	-0.172	-0.861	-2.223	-2.062	3
51650	Putative Phosphate metabolism protein PHM7	-0.254	0.378	0.484	-0.25	0.356	-0.554	-0.737	-1.346	-1.46	3
51893	Putative Linoleate diol synthase	0.205	0.089	-0.134	-1.19	-0.939	-0.342	-0.59	-1.365	-1.084	3
52375	Putative Betaine aldehyde dehydrogenase	-0.092	-0.358	-0.699	-0.948	-1.079	-0.213	NaN	NaN	-0.655	3
52463	Putative Zinc metalloprotease	-0.164	-0.07	-0.0090	0.465	-0.0040	-0.403	NaN	-0.593	-1.33	3
54991	Putative Dehydrogenase	0.147	0.033	-0.74	-1.194	-1.187	NaN	NaN	-0.492	-0.488	3
55644	Putative 20S proteasome alpha 5 subunit involved in ubiquitin-dependent catabolism PUP2	-0.876	-0.537	-0.918	-0.901	-1.061	-0.588	NaN	-0.653	-1.167	3
58282	Putative Carbohydrate Esterase family 9	-0.486	0.205	-0.851	-1.011	-0.927	-0.303	-0.551	-0.818	-0.897	3
61912	Putative Zn-dependent aspartyl aminopeptidase	-0.515	-0.586	-1.308	-1.012	-0.939	-0.527	-0.399	-0.623	-0.885	3
62271	Putative Acetyltransferase	-0.277	0.21	-0.245	-0.354	-0.531	NaN	NaN	-0.428	-1.001	3
63813	Putative Cytochrome b5	0.412	0.013	-1.019	-0.533	-0.592	-0.6	NaN	-0.807	-0.763	3
65406	Putative Glycoside hydrolase family 16	-0.402	-0.13	-0.106	-1.013	-0.577	-0.094	NaN	-0.434	-0.145	3
65483	Putative Secreted superoxide dismutase	-0.597	-0.396	-0.459	-1.0	-0.927	-0.182	NaN	-0.864	-0.877	3
66016	Putative Alpha/beta-hydrolase	-0.358	-0.334	-0.43	0.401	-0.029	-0.641	NaN	-0.442	-1.099	3
66175	Putative Rhamnolipids biosynthesis 3-oxoacyl-[acyl-carrier-protein] reductase	-0.225	-0.088	-1.004	-0.904	-1.286	-0.17	NaN	-0.727	-0.794	3
66647	Putative HAD superfamily hydrolase	-0.051	0.138	-0.648	-1.038	-1.249	-0.187	NaN	-0.547	-0.924	3
67752	Putative MFS transporter	-1.125	1.923	-1.763	NaN	-1.514	-0.564	-0.45	-0.832	-0.092	3
68036	Putative Cystathionine gamma-synthase	-1.752	0.204	0.213	-0.285	-0.27	-0.339	-0.553	-1.112	-1.475	3
68338	Putative NipSnap family protein	-0.1	-0.061	-0.864	-1.041	-1.177	-0.517	NaN	-0.541	-0.943	3
72012	Putative Glycosyltransferase Family 8 GLG2	-1.014	-0.214	-0.398	-0.526	-0.591	0.033	-0.167	0.0070	-0.341	3
72859	Putative RNA-binding protein RTC3	-0.534	-0.238	-0.305	NaN	0.16	-1.176	NaN	-1.01	-1.707	3
72922	Putative GABA permease	-0.785	-1.106	-1.461	-1.08	-0.822	-0.306	-0.204	-0.226	-0.293	3
73818	Putative Peroxisomal catalase CATC	-0.672	0.193	-0.919	-0.887	-1.153	-0.574	NaN	-1.159	-1.324	3
75159	Putative Aminopeptidase APE2	-0.185	-0.151	-1.003	-0.705	-0.89	-0.309	-0.255	-0.508	-0.782	3
76010	20S proteasome alpha subunit 7 PSA7 / PRE6	-0.557	-0.48	-0.822	-0.782	-1.065	-0.504	-0.0050	-0.465	-0.703	3
76218	Putative Zinc-containing alcohol dehydrogenase	-0.757	-0.158	-0.614	-0.915	-1.045	-0.589	NaN	-0.983	-1.357	3
76288	Putative Aldose reductase	-1.157	-0.567	-0.567	-0.31	-0.702	-0.524	-0.127	-0.62	-1.003	3
76620	Putative Mitochondrial glycerol-3-phosphate dehydrogenase GUT2	-0.36	-0.724	-1.083	-0.515	-0.799	-0.263	-0.345	-0.489	-0.59	3
76897	Putative MFS nicotinic acid transporter TNA1	-0.972	0.527	-1.119	-1.218	-1.484	-0.943	-0.989	-1.453	-1.518	3
77284	Putative Glycoside hydrolase family 12	0.256	-0.387	-0.76	-0.64	-0.616	-0.595	NaN	-0.809	-1.02	3
77579	Putative Secreted vacuolar pepsin PEP4	-0.015	-0.303	-0.536	-0.399	-0.557	-0.567	-0.729	-0.794	-1.136	3
77747	Putative Actin binding protein	-0.316	0.0090	-0.632	-0.376	-0.477	-1.114	-0.712	-0.449	-1.254	3

78611	Putative Phospholipase D1	0.2	-0.248	-0.554	-0.472	-0.807	-0.423	-0.453	-0.807	-1.111	3
78626	Putative protein of unknown function	-1.23	-0.118	-1.033	-1.272	-1.276	-0.365	-0.714	-1.013	-1.186	3
78650	Putative Homedomain like protein	-0.173	-0.342	-0.24	-0.717	-1.058	0.204	NaN	-0.2	NaN	3
81598	Putative Glycoside hydrolase family 18 CHI18-7	-0.229	0.167	-0.413	-0.343	-0.365	-0.566	NaN	-0.74	-1.267	3
81756	Putative Secreted protein of unknown function	-0.019	-0.389	-0.86	-1.358	-0.98	-1.092	-1.038	-0.989	-1.128	3
82374	Putative protein of unknown function	0.209	NaN	-0.312	0.699	0.433	-0.978	-0.793	-0.876	-1.127	3
82623	Putative Secreted subtilisin	-2.015	-0.067	-3.492	-0.346	-0.157	-1.047	-0.221	0.054	-0.606	3
102437	Putative Secreted protein of unknown function	-0.112	0.13	-1.122	-1.372	-1.16	-0.282	-0.168	-0.48	-0.226	3
103625	Putative 20S proteasome maturation protein UMP1	-0.38	-0.419	-0.844	-0.867	-1.254	NaN	NaN	-0.586	-0.454	3
103886	Putative DUF1774 family protein	0.293	0.455	-0.298	0.563	0.363	-0.981	-1.262	-1.166	-1.796	3
104277	Putative Cell wall protein	-0.526	0.778	-1.434	-1.13	-1.114	-0.125	-0.13	-0.423	-0.516	3
105070	Putative Cystathionine gamma-synthase	-0.227	-0.143	-0.351	NaN	-1.02	0.049	NaN	-0.272	-0.296	3
105251	Putative protein of unknown function	-0.846	-0.033	-0.667	-0.879	-0.704	-0.575	-0.549	-0.693	-1.004	3
105313	Putative Apple like carbohydrate binding domain protein	-1.425	-0.734	-0.825	-1.177	-0.983	-0.494	-0.629	-0.952	-1.109	3
105968	Putative PcbC family protein	0.115	-0.26	-1.107	-0.651	-0.551	-0.307	NaN	-0.561	-0.34	3
106356	Putative Sphingoid long-chain base transporter	-1.188	-0.365	-0.82	-0.475	-0.548	-0.648	NaN	-0.624	-0.949	3
107294	Putative RTA1 domain protein	-1.048	-0.592	-0.725	-1.217	-1.251	-0.409	NaN	-0.802	-1.425	3
107475	Putative protein of unknown function	-0.575	-0.444	-0.268	0.101	-0.559	0.091	NaN	NaN	-1.0	3
107670	Putative Alpha/beta-hydrolase	-0.199	-0.16	-0.864	-0.947	-1.065	-0.157	NaN	-0.292	-0.505	3
107881	Putative protein of unknown function	-2.322	-0.378	-0.89	-0.924	0.192	-1.19	-1.042	-1.077	-1.409	3
108201	Putative Levodione reductase	0.042	-0.253	-0.92	-0.858	-1.121	-0.459	NaN	-0.605	-0.556	3
108583	Putative protein of unknown function	-1.073	-0.053	-0.54	-1.023	-0.825	-0.71	-0.685	-1.316	-1.649	3
108909	Putative protein of unknown function	-0.494	0.126	-0.301	-0.66	-0.904	-0.413	-0.672	-0.925	-1.084	3
109122	Putative Amino acid transporter	0.0040	-1.266	-0.888	-0.948	-0.985	NaN	-0.291	-0.129	-0.128	3
109998	Putative Glutathione S-transferase	-0.965	-0.576	-0.32	0.249	-0.573	-0.243	NaN	-0.037	-1.014	3
110170	Putative Glyoxylate reductase GOR1	-0.41	-0.218	-0.921	-1.075	-1.17	-0.067	-0.233	-0.432	-0.3	3
110452	Putative protein of unknown function	-0.749	-0.168	-0.742	-0.714	-0.657	-0.659	-0.366	-0.588	-1.256	3
111085	Putative Enoyl-CoA hydratase/isomerase	0.038	-0.092	-0.578	-0.452	-0.395	-0.045	NaN	-0.744	-1.015	3
112018	Putative Secreted protein of unknown function	0.1	0.043	-0.483	-0.966	-0.773	-0.346	-0.553	-0.999	-1.365	3
119813	Putative protein of unknown function	-0.316	-0.239	-0.206	-0.537	-0.812	-0.281	NaN	-0.7	-1.046	3
119876	Secreted pepsin P7959	-1.227	-0.31	-0.334	-1.59	-1.466	-0.693	-0.698	-1.589	-1.752	3
120198	Putative Glycosyltransferase Family 35 GPH1	-0.381	0.081	-0.599	-0.324	-0.506	-0.6	-0.804	-0.781	-1.258	3
120424	Putative Lactoylglutathione lyase GLO1	-0.55	-0.109	-0.573	-0.108	-0.151	-0.44	NaN	-0.687	-1.12	3
120568	Putative Enolase ENO1	-1.591	-0.38	-0.276	-0.286	-0.377	0.066	-0.205	-0.478	-0.682	3
120675	Putative Glutamate carboxypeptidase-like protein DUG1	0.444	0.13	-1.274	-0.84	-0.957	-0.385	-0.606	-0.742	-1.065	3

121343	Putative Proteasome alpha subunit PRE5	-0.744	-0.551	-0.817	-0.626	-1.011	-0.398	NaN	-0.552	-0.945	3
121427	Putative protein of unknown function	-0.699	0.012	-0.188	-0.779	-0.715	-0.459	NaN	-0.65	-1.153	3
121522	Putative Histone H2A HTA1	-0.581	-0.252	0.158	-0.195	-0.41	-0.374	-0.853	-1.256	-1.793	3
121620	Putative protein involved in synthesis of the thiamine precursor hydroxymethylpyrimidine (HMP) THI11	0.469	-0.192	-0.381	-0.346	-0.705	0.178	-0.102	-0.102	-2.006	3
121717	Putative protein of unknown function	NaN	0.285	-0.626	-0.084	-0.359	-0.759	-1.023	NaN	-1.9	3
121746	Putative Glycoside hydrolase family 55	-1.325	-0.354	-0.655	-0.77	-0.751	-0.15	NaN	-0.651	-0.697	3
121789	Putative Triosephosphate isomerase TPI1	-0.574	-0.536	-0.63	-0.361	-0.753	-0.329	-0.518	-0.703	-1.003	3
121800	Putative SPFH domain protein	-0.203	-0.957	-0.826	0.047	-0.119	-1.223	-1.025	-0.538	-1.251	3
121818	Putative Secreted protein of unknown function	-0.054	-0.281	-0.043	-0.779	-1.045	-0.211	0.017	-0.248	-0.23	3
121839	Putative Chaperone protein related to HSP70	-0.352	-0.022	-0.491	-1.014	-0.65	-0.455	-0.534	-0.685	-0.775	3
121977	Putative Mitochondrial peptidyl-prolyl cis-trans isomerase (Cyclophilin) CPR3 / CYPD	-0.648	-0.254	-0.551	-0.819	-0.834	-0.385	-0.909	-1.047	-1.482	3
122087	Putative protein of unknown function	-1.347	0.095	-0.04	0.102	-0.203	-0.403	-0.559	-0.935	-1.073	3
122242	Putative Secreted protein of unknown function	-1.238	0.166	-0.045	-1.167	-0.922	-0.729	NaN	-1.124	-1.344	3
122363	Chaperone protein HSP30-like	-1.662	-0.069	-0.483	-0.354	-0.879	NaN	NaN	-0.662	-0.959	3
123056	Woronin body protein HEX1	0.082	0.047	-0.044	-0.665	-0.617	-0.015	-0.591	-0.825	-1.212	3
123261	Putative protein of unknown function	0.21	0.097	-0.703	-1.183	-1.315	-0.385	NaN	-0.144	-1.217	3
123456	Putative Glycoside hydrolase family 65 ATH1	-1.449	-0.916	-1.11	-0.763	-0.759	0.069	-0.178	-0.136	-0.108	3
123475	Putative Secreted protein of unknown function	-2.287	-0.366	-0.485	-0.528	0.306	-0.812	-1.437	-1.96	-2.706	3
123968	Putative Alpha/beta-hydrolase	0.022	-0.461	-0.598	-1.021	-0.978	0.0080	NaN	-0.3	-0.46	3
124031	Putative 20S proteasome alpha 3 subunit PRE9	-0.941	-0.553	-0.841	-0.804	-1.134	-0.585	NaN	-0.408	-0.8	3
2322	Putative MOSC domain protein	-0.215	0.553	1.052	0.794	0.533	0.674	NaN	0.0060	0.324	4
2687	Putative Heat shock protein family Hsp78	-1.017	0.049	-0.122	0.395	0.169	-0.36	-0.045	-0.237	-0.54	4
4626	Putative Ca ²⁺ -dependent lipid-binding protein CLB1/vesicle protein vp115/Granuphilin A	0.21	0.051	0.21	1.069	0.787	-0.127	NaN	-0.157	-0.28	4
5182	Putative Dimeric alpha+beta barrel protein	-0.295	-0.092	-0.314	-0.275	-0.115	-0.198	NaN	0.294	1.099	4
21170	Putative 40S ribosomal protein S12 RSP12	-1.268	-0.157	0.685	0.288	0.545	0.358	-0.026	-0.059	0.228	4
21658	Putative 60S ribosomal protein RPL6	-1.305	-0.092	0.696	0.124	0.397	0.519	-0.036	-0.035	0.499	4
21890	Putative 60S ribosomal protein L32 RPL32	-1.395	-0.211	0.772	0.366	0.616	0.395	-0.0040	0.038	0.392	4
22830	Putative Glutathione-S-transferase	0.013	0.104	0.698	1.193	0.712	NaN	NaN	0.354	-0.139	4
22831	Putative Flavoprotein domain protein	0.153	-0.0070	0.282	0.875	1.046	-0.106	NaN	NaN	0.111	4
23059	Putative 60S ribosomal protein L26 RPL26A	-1.024	-0.074	0.677	0.246	0.295	0.223	NaN	0.013	0.43	4
25225	Putative Replication factor C subunit 1 RFC1	-0.157	-0.058	0.164	0.209	0.192	0.065	NaN	0.624	1.066	4
26642	Putative MFS transporter	-0.05	-0.248	0.137	0.653	1.027	-0.013	NaN	0.183	0.727	4
39942	Glycoside hydrolase family 17 EGLC	-0.951	0.291	0.266	1.094	0.926	0.073	0.239	0.208	0.173	4

43906	Putative 60S ribosomal protein L10a RPL1	-1.01	0.103	0.655	0.326	0.65	0.11	0.066	0.0010	0.288	4
44434	Putative 60S ribosomal protein L4 RPL4	-1.06	-0.321	0.543	0.0090	0.165	0.313	0.102	-0.055	0.233	4
46285	Putative Chaperone HSP30	-1.681	-0.07	0.019	0.211	-0.306	NaN	NaN	-0.662	-0.876	4
46490	Putative 40S ribosomal protein S16 RPS16B	-1.456	-0.184	0.658	0.102	0.365	0.355	0.162	-0.039	0.227	4
46958	Putative Translation elongation factor 1a TEF1a	-1.092	-0.224	0.179	-0.239	-0.063	0.199	-0.156	-0.324	-0.251	4
48170	Putative 60S ribosomal protein RPL27B	-1.573	-0.161	0.708	0.422	0.59	0.367	-0.011	0.0070	0.147	4
51212	Putative 60S ribosomal protein L5 RPL5	-1.262	-0.0010	0.83	0.349	0.501	0.175	0.051	-0.076	0.284	4
51499	Putative Mitochondrial acetoacetyl-CoA synthetase ILV5	-0.569	-0.241	1.015	0.01	0.027	0.52	0.399	0.281	0.704	4
52267	Putative Mannitol-1-phosphate dehydrogenase DSF1	-0.538	-0.205	0.731	0.406	0.433	1.328	0.64	0.356	0.021	4
53025	Putative S-adenosylmethionine-dependent methyltransferase	0.021	0.153	0.605	1.045	1.073	-0.044	-0.384	0.114	0.233	4
54366	Putative Neutral/alkaline non-lysosomal ceramidase	-1.069	-0.379	-0.521	-0.388	-0.325	0.223	NaN	0.477	0.648	4
55747	Putative ABC transporter	0.052	0.042	0.206	0.758	1.111	0.0080	0.075	0.451	0.379	4
56176	Putative ABC transporter	-0.925	-0.171	0.015	0.84	1.042	-0.17	0.385	0.808	0.586	4
56863	Putative protein of unknown function	-1.056	-0.05	0.031	-0.581	-0.376	0.141	0.463	0.329	0.336	4
61595	Putative Cytosolic 60S ribosomal protein RPL23a	-1.095	0.118	0.861	0.472	0.744	0.319	NaN	0.147	0.423	4
62100	Chaperone HSP23	-1.074	-0.22	-0.056	1.352	1.399	0.085	-0.028	0.145	-0.069	4
62820	Putative 60S ribosomal protein L15 RPL15	-1.365	0.0010	0.699	0.135	0.479	0.436	0.175	-0.089	0.344	4
62985	Putative Lipase/esterase	-0.355	0.164	0.512	1.053	0.949	-0.253	NaN	-0.231	-0.398	4
64874	Putative MFS transporter	-0.448	0.309	1.251	0.132	0.381	-0.332	NaN	-0.288	-0.235	4
66041	Putative Glycoside hydrolase family 18	-1.331	-0.263	-0.39	-0.33	-0.025	-0.229	0.286	0.636	1.096	4
66303	Putative Glycoside hydrolase	0.023	-0.163	0.41	NaN	1.017	-0.071	NaN	-0.093	0.071	4
68755	Putative Short-chain dehydrogenase / reductase	-1.159	0.088	0.995	1.359	0.4	0.088	NaN	-0.261	-1.033	4
68909	Putative Ribosomal protein RPS26	-1.606	-0.053	0.679	0.397	0.514	0.317	0.345	-0.086	0.164	4
72643	Putative Arginosuccinate synthase ARG1	-1.039	-0.275	0.54	0.587	0.605	0.396	-0.011	0.213	0.61	4
73623	Putative Salicylate hydroxylase	-1.254	1.263	1.692	-0.018	-0.249	-0.017	NaN	-0.588	-0.842	4
73765	Putative 60S ribosomal protein L3 RPL3	-1.277	-0.151	0.67	0.112	0.379	0.226	0.031	-0.077	0.271	4
73840	Putative 60S ribosomal protein L8 RPL2B	-1.196	-0.285	0.454	-0.163	0.137	0.321	0.084	-0.036	0.268	4
73985	Putative 40S ribosomal subunit component RPS10A	-1.669	-0.066	0.698	0.243	0.551	0.309	0.208	-0.072	0.114	4
74568	Putative 40S ribosomal protein S4 RPS4A	-1.265	-0.098	0.61	0.116	0.435	0.289	0.151	-0.048	0.409	4
75152	Putative FAD/NAD(P)-binding domain protein	0.135	0.072	0.353	1.044	0.561	-0.135	NaN	NaN	0.051	4
75564	Putative 60S ribosomal protein L11 RPL11A	-1.215	0.147	0.65	0.388	0.641	0.321	0.077	-0.029	0.174	4
75604	Putative protein of unknown function	-0.112	-0.203	0.352	1.109	0.839	0.252	NaN	-0.223	-0.128	4
75609	Putative Cytosolic 60S ribosomal protein RPL18	-1.202	-0.175	0.741	0.264	0.542	0.36	0.084	-0.046	0.489	4
76713	Putative MFS transporter	-0.176	0.276	0.4	0.581	1.008	-0.245	NaN	-0.084	-0.122	4
76766	Putative Purine transporter	-0.88	0.092	0.718	1.144	1.36	-0.144	0.054	0.235	0.374	4

76939	Putative 40S ribosomal protein S9 RPS9B	-1.325	-0.246	0.535	0.084	0.361	0.351	0.0080	-0.182	0.24	4
77557	Putative Glycosyltransferase	-1.069	-0.142	-0.646	-0.439	-0.024	-0.227	0.162	-0.045	0.033	4
77699	Putative DUF833 domain ser/thr-rich protein	0.085	0.372	0.905	1.075	0.431	0.458	NaN	-0.603	-0.474	4
77906	Putative 40S ribosomal protein S0 RPS0B	-1.242	-0.155	0.659	0.101	0.386	0.384	-0.0010	-0.148	0.244	4
78233	Putative 40s ribosomal protein RPS21B	-1.109	-0.351	0.415	-0.067	0.3	NaN	NaN	0.106	0.3	4
78377	Putative GTP-binding protein NOG2	-0.124	1.055	1.238	NaN	0.815	-0.043	-0.057	-0.012	0.165	4
78688	Putative Heat shock factor-type DNA-binding domain protein	-0.778	0.036	0.47	1.228	1.348	-0.243	-0.443	-0.23	-0.211	4
78695	Putative Zn-finger domain protein ZPR1	-1.103	-0.275	0.0050	0.425	0.084	-0.51	NaN	-0.247	-0.318	4
78757	Putative Plasma membrane H ⁺ -ATPase PMA2	-0.215	-0.637	0.76	0.16	0.208	0.545	0.332	0.474	1.171	4
78870	Putative 60S ribosomal protein L9 RPL9A	-1.372	0.141	0.863	0.409	0.912	0.304	0.26	0.056	0.403	4
79106	Putative 40s ribosomal protein RPL44 / RPL42A	-1.583	-0.238	0.476	0.216	0.21	0.069	-0.023	0.062	0.088	4
79169	Putative Voltage-gated potassium channel subunit beta	-1.557	0.392	0.783	1.08	0.768	0.326	-0.025	0.066	-0.056	4
80211	Putative 60S ribosomal protein L14 RPL14B	-1.236	-0.227	0.748	0.497	0.406	0.209	-0.038	-0.037	0.312	4
80291	Putative C6 transcription factor	-0.175	0.024	0.231	1.045	0.89	0.095	-0.115	0.037	0.241	4
80484	Putative 40S ribosomal protein S22 RPS1A	-1.258	-0.09	0.562	0.22	0.486	0.34	0.169	-0.041	0.423	4
81609	Glycoside Hydrolase Family 15 Cellulose signaling associated protein ENV1	-0.226	0.439	0.169	1.013	1.014	-0.475	-0.061	0.237	0.227	4
81713	Putative Cytosolic 40S ribosomal protein RPS24A	-1.091	-0.184	0.584	0.088	0.407	0.354	NaN	-0.062	0.458	4
81803	Putative Mitochondrial protein sorting UPS2 / MSF1	-0.676	1.387	0.621	1.03	0.88	-0.536	-0.31	-0.195	0.355	4
81906	Putative S-adenosylmethionine (AdoMet)-dependent methyltransferase, Actin binding protein ABP140	-0.51	0.771	0.887	0.805	1.115	0.103	NaN	-0.0080	0.275	4
82633	Putative Glycoside hydrolase family 72	-1.438	-0.634	-0.278	1.343	1.66	-0.411	0.207	1.158	0.828	4
102966	Putative DUF1349 family protein	-2.256	1.991	1.516	0.723	-0.456	2.287	1.047	0.0040	-1.564	4
103651	Putative protein of unknown function	-1.137	-0.021	-0.163	-0.094	0.0050	-0.012	NaN	0.383	0.394	4
105260	Putative MFS transporter	-0.577	-0.216	0.675	1.002	1.274	-0.157	NaN	-0.527	-0.621	4
105565	Putative MFS transporter	0.14	0.0090	0.537	0.938	1.102	-0.35	NaN	0.186	0.301	4
107055	Putative protein of unknown function	-1.007	-0.22	-0.046	0.94	0.726	-0.695	NaN	-0.36	-0.882	4
107784	Putative 60S ribosomal protein L20 RPL20A	-1.532	-0.046	0.799	0.37	0.735	0.35	0.337	-0.087	0.209	4
109234	Putative D-aminopeptidase	-1.52	-0.039	0.561	-0.156	-0.121	0.02	0.212	-0.448	-0.712	4
110653	Putative protein of unknown function	-1.368	-0.651	-0.086	1.339	1.073	-0.378	NaN	0.58	0.67	4
119710	Putative MFS transporter	-0.175	-0.204	0.899	1.456	1.032	-0.086	0.0010	0.016	-0.128	4
119735	Putative Glyceraldehyde-3-phosphate dehydrogenase GPD1	-0.948	-0.148	0.446	1.128	0.95	0.321	0.133	0.245	-0.0090	4
119855	Putative 40S ribosomal protein S3 RPS3	-1.025	-0.301	0.488	0.039	-0.037	0.403	NaN	0.015	0.355	4
120189	Putative GPI anchored protein	-1.257	0.379	0.243	0.359	0.332	0.239	0.168	0.259	-0.117	4
120621	Putative 60S ribosomal protein L33a RPL33A	-1.073	0.067	1.183	0.922	1.143	0.183	0.288	-0.101	0.475	4
120635	Putative Transketolase TKL1	-1.085	-0.126	0.189	0.578	0.384	0.235	0.122	0.148	-0.082	4
120868	Putative 40S ribosomal protein S13 RPS13	-1.101	-0.079	0.711	0.291	0.654	0.458	0.186	0.04	0.674	4

120978	Putative 60S ribosomal protein L34 RPL24A	-1.452	0.069	0.852	0.497	0.713	0.252	0.183	0.012	0.144	4
121003	Putative 40S ribosomal protein S18 RPS18B	-1.451	-0.251	0.543	0.248	0.577	0.193	0.096	-0.187	0.053	4
121004	Putative protein of unknown function	0.134	1.025	0.155	0.781	0.446	NaN	NaN	-0.65	-0.646	4
121278	Putative Cytosolic 60S ribosomal protein RPL21B	-1.205	-0.056	0.695	0.359	0.693	0.203	0.047	0.041	0.344	4
121686	Putative Serine hydroxymethyltransferase SHM2	-0.549	0.318	1.18	0.453	0.596	0.367	0.238	0.258	0.581	4
121706	Putative Phospho-2-dehydro-3-deoxyheptonate aldolase ARO4	-0.255	0.215	1.115	0.401	0.64	0.503	NaN	0.252	0.52	4
121741	Putative 40S ribosomal protein S7 RPS7B	-1.018	-0.114	0.587	0.16	0.449	0.282	0.059	0.025	0.325	4
121837	Putative 40S ribosomal protein S5 RPS5	-1.404	-0.144	0.708	0.186	0.456	0.363	0.16	-0.06	0.364	4
121904	Putative 60S ribosomal protein L7 RPL7B	-1.323	-0.137	0.778	0.308	0.642	0.254	-0.122	-0.016	0.397	4
121989	Putative Secreted oxalate decarboxylase	-1.001	-0.098	-0.193	-0.223	-0.086	-0.405	NaN	-0.508	-0.52	4
122001	Putative 60S ribosomal protein L23 RPL23B	-1.401	-0.041	0.681	0.351	0.495	0.289	0.036	-0.047	0.459	4
122036	Putative 40S ribosomal protein S2 RPS2	-1.422	-0.199	0.644	0.203	0.351	0.461	0.043	0.113	0.336	4
122040	Putative 60S ribosomal protein L31 RPL31B	-1.639	0.081	0.901	0.388	0.791	0.178	0.072	-0.063	0.115	4
122079	Putative Short chain dehydrogenase	-0.506	1.923	0.943	0.774	0.815	-0.376	-0.578	-0.796	-1.166	4
122108	Putative protein of unknown function	-0.459	1.267	0.949	0.828	0.694	-0.118	-0.15	-0.386	-0.739	4
122141	Putative 60S ribosomal protein L28 RPL28	-1.539	-0.09	0.73	0.27	0.585	0.267	0.121	-0.074	0.093	4
122301	Putative O-acetylhomoserine MET17	-0.311	-0.106	0.3	-0.281	-0.321	0.521	0.184	0.465	1.076	4
122506	Putative Secreted protein of unknown function	-1.056	0.021	NaN	-0.036	-0.089	-0.098	0.147	-0.688	-0.528	4
122523	Putative C6 transcription factor	0.264	-0.815	0.414	1.093	0.772	-0.0090	NaN	0.388	0.055	4
122587	Putative 40S ribosomal protein S11 RPS11B	-1.475	-0.064	0.613	0.326	0.692	0.387	0.134	-0.152	0.121	4
122590	Putative protein of unknown function	-0.606	0.744	0.434	0.813	0.975	-1.192	-0.91	-0.569	-0.406	4
122679	Putative AIM38 protein	-0.828	0.452	0.908	1.294	1.3	0.392	0.283	0.059	-0.07	4
122870	Putative Secreted cell wall	-1.183	0.176	0.033	NaN	0.194	0.08	NaN	0.355	0.144	4
122952	Putative protein of unknown function	-0.364	0.909	1.052	0.763	1.48	-0.489	-0.104	-0.158	0.213	4
123145	Putative 60S ribosomal protein L17 RPL17A	-1.215	-0.208	0.723	0.203	0.369	0.264	NaN	-0.072	0.356	4
123176	Putative 60S ribosomal protein L12 RPL12A	-1.377	-0.026	0.951	0.516	0.968	0.348	0.405	0.302	0.57	4
123251	Putative Flavohemoglobin	-0.497	0.442	0.876	0.645	0.059	1.254	0.877	0.115	-0.107	4
123431	Putative Clavamate synthase	-1.413	-0.051	-0.334	-0.333	NaN	-0.29	NaN	0.528	0.951	4
123471	Putative Mitochondrial homocitrate synthase LYS21	-0.63	-0.277	0.941	0.42	0.845	0.29	0.118	0.323	1.121	4
123499	Putative HET domain protein	0.219	0.198	0.301	0.552	1.107	-0.411	NaN	0.332	0.309	4
123718	Putative Amino acid transporter	-1.704	-1.195	-2.901	0.873	1.18	-0.71	0.515	1.134	1.058	4
124010	Putative 60S ribosomal protein L13a RPL16A	-1.266	-0.305	0.566	0.198	0.454	0.361	0.181	-0.041	0.209	4
124016	Glycoside hydrolase family 36 AGL2	-0.034	-0.434	-0.41	1.678	1.01	-0.498	NaN	-0.674	-0.634	4
124110	Putative 60S ribosomal protein L17a RPL8A	-1.319	-0.154	0.757	0.307	0.543	0.248	-0.0060	-0.04	0.374	4
124119	Putative protein of unknown function	-0.161	0.106	0.362	1.023	0.85	-0.072	0.075	0.278	0.27	4

1885	Putative Glycoside hydrolase family 15 SGA1	0.91	-1.362	-2.887	-2.944	-3.042	-0.075	-0.209	-0.707	-0.788	5
3716	Putative YjgF-like domain protein	-1.729	-1.099	-4.47	-4.167	-3.76	-0.843	-0.607	-1.368	-1.31	5
3717	Putative 2-keto-3-deoxy-L-galactonate aldolase	-0.15	-0.717	-2.408	-2.403	-2.63	0.068	NaN	-0.011	0.06	5
4040	Putative DNA-directed RNA polymerase III subunit RPC34	0.023	-0.471	-0.71	-1.018	-1.163	0.104	-0.197	-0.034	0.079	5
4308	Putative Aminopeptidase N	0.722	0.237	-1.178	-0.629	-0.994	0.016	-0.198	-0.221	-0.592	5
4561	Putative Glycosyltransferase family 32	0.149	-0.628	-1.832	-2.093	-2.351	-0.0040	NaN	-0.44	-0.487	5
21659	Putative Dipeptidyl-peptidase	0.186	-0.282	-1.391	-0.98	-1.29	-0.475	-0.481	-0.576	-0.981	5
21758	Putative Isocitrate lyase ICL1 / ACU3	1.142	0.114	-1.327	-1.38	-1.874	-0.379	-0.781	-1.286	-1.68	5
22912	Putative Glucose transporter HXT1	2.743	-1.028	-2.481	-3.67	-3.473	-0.205	NaN	-1.572	-1.601	5
23171	Putative 70 kb peptaibol synthase	0.49	-0.178	-0.325	-0.99	-1.369	0.184	NaN	-0.012	NaN	5
43398	Putative protein of unknown function	0.464	-1.027	-1.438	-1.395	-1.671	0.718	NaN	0.079	NaN	5
46794	Putative MFS peptide transporter	1.146	-1.23	-1.784	-2.783	-3.33	-1.577	-2.204	-3.568	-3.875	5
51110	Putative Amino acid transporter GAP1	0.176	-1.26	-1.022	-1.122	-0.962	-0.169	-0.221	-0.172	-0.0080	5
54198	Putative Glycoside hydrolase	0.968	-0.438	-0.367	-0.626	-0.724	0.037	NaN	-0.637	-1.008	5
54525	Putative Aldehyde dehydrogenase	0.945	0.0030	-2.341	-2.473	-2.678	-0.392	NaN	-0.792	-0.961	5
56934	Putative Gluconokinase	0.514	-1.277	-1.185	-1.026	-0.76	NaN	NaN	0.033	0.588	5
60490	Putative 2-nitropropane dioxygenase	0.054	0.067	-0.917	-1.216	-1.755	-0.242	NaN	-0.598	-0.562	5
60616	Putative Integral membrane protein	0.6	-1.164	-1.167	0.276	-0.306	-0.256	NaN	NaN	-0.016	5
61374	Putative MFS transporter	-1.041	-1.188	-4.225	-4.423	-3.963	-0.493	NaN	-0.184	-0.35	5
62439	Putative Dehydrogenase	0.616	-0.321	-1.013	-1.239	-1.39	-0.188	NaN	-0.807	-0.648	5
63882	Putative Mitochondrial 3-hydroxyisobutyryl-CoA hydrolase EHD3	0.579	0.0050	-1.015	-0.971	-1.06	-0.113	NaN	-0.468	-0.428	5
64066	Putative Acetyltransferase	0.144	-0.273	-1.148	-1.2	-1.481	-0.191	-0.35	-0.239	-0.467	5
64345	Putative Mitochondrial methylcrotonoyl-CoA carboxylase beta chain	1.403	-0.0050	-2.022	-1.837	-2.012	-0.013	-0.085	-0.326	-0.365	5
65153	Putative MFS transporter	0.99	0.386	-1.268	-1.81	NaN	0.0070	NaN	-0.62	-0.71	5
65292	Putative Acetyl-CoA synthetase-like	0.184	0.11	-1.213	-1.378	-1.513	-0.061	NaN	-0.432	-0.375	5
65508	Putative Fructose-1,6-bisphosphatase FBP1	0.637	0.035	-1.161	-0.999	-1.391	-0.36	NaN	-1.333	-1.467	5
65921	Putative Mitochondrial methylcrotonoyl-CoA carboxylase alpha chain	1.316	0.166	-1.812	-1.574	-1.743	-0.299	-0.407	-0.388	-0.563	5
65992	Putative Dimeric alpha+beta barrel protein	0.343	0.438	-0.95	-0.953	-1.321	-0.287	-0.858	-0.862	-1.36	5
66541	Putative Mitochondrial Isocitrate lyase ICL2	0.789	-0.608	-1.445	-1.374	-1.599	0.201	NaN	-0.119	-0.345	5
69026	Putative MFS transporter	0.091	-0.276	-1.012	-1.081	-1.545	-0.058	NaN	-0.226	-0.72	5
71094	Putative Extracellular dioxygenase	0.868	-0.402	-1.279	-2.271	-2.781	NaN	NaN	-0.36	-0.606	5
72611	Putative C6 transcription factor	0.735	0.152	-1.293	-1.483	-1.652	-0.189	NaN	NaN	-0.308	5
73559	Putative C6 transcription factor	0.371	0.116	-0.981	-1.249	-1.129	-0.038	-0.184	0.011	0.04	5
74194	L-xylulose reductase LXR1	1.566	-0.815	-1.364	-0.459	-0.886	-0.387	NaN	-0.55	-0.904	5
74453	Putative 3-hydroxybutyryl-CoA dehydrogenase	1.434	0.508	-1.161	-1.391	-1.408	-0.053	-0.225	-0.631	-0.711	5

75175	Putative Mitochondrial short/branched chain specific acyl-CoA dehydrogenase	0.874	-0.254	-1.605	-1.344	-1.586	-0.184	-0.549	-0.569	-0.641	5
75290	Putative protein of unknown function	0.402	-0.148	-0.961	-1.029	-1.022	-0.071	0.159	0.042	-0.223	5
76155	Putative Acid phosphatase	2.735	-0.148	-1.666	-1.848	-2.308	NaN	NaN	-0.323	-0.144	5
77093	Putative Acid sphingomyelinase	1.618	0.711	-0.967	-2.124	-2.328	-0.244	NaN	-0.708	-0.838	5
77234	Putative C2H2 Zn-finger domain protein	0.01	-0.721	-1.416	-1.384	-1.276	0.246	-0.069	-0.188	0.042	5
78773	Putative Long-chain acyl-CoA synthetase FAA2 / FAM1	0.287	-0.397	-1.068	-0.929	-1.249	-0.38	-0.456	-0.929	-1.157	5
79271	Putative Peroxisomal Malate synthase MLS1	1.423	0.342	-0.85	-0.747	-1.212	-0.46	NaN	-1.26	-1.261	5
79426	Putative Clavaminate synthase	0.104	-0.071	-0.651	-0.696	-1.081	NaN	-0.124	-0.074	-0.332	5
80026	Putative protein of unknown function	-0.893	0.146	-0.085	-2.388	-2.664	-0.808	NaN	-2.478	-2.72	5
80621	Putative 2-methylcitrate synthase	1.415	0.211	-2.018	-1.893	-1.921	-0.147	NaN	-0.517	-0.622	5
80716	Putative L-ornithine transaminase CAR2	1.414	-0.131	-1.646	-1.417	-1.692	-0.355	NaN	-0.589	-0.796	5
80879	Putative MFS H+/oligopeptide transporter	0.354	-1.442	-2.221	-1.022	-0.891	NaN	NaN	-0.01	0.025	5
81096	Putative Ferric reductase transmembrane component	0.057	-0.283	-0.113	-0.5	-1.106	NaN	NaN	-0.074	0.024	5
81115	Putative X-Pro aminopeptidase	0.16	0.047	-1.333	-1.118	-1.314	-0.305	-0.207	-0.449	-0.726	5
81410	Putative SAM methyltransferase	0.856	0.078	-0.977	-1.362	-1.855	-0.557	-0.776	-1.153	-1.575	5
81778	Putative Glutaminase GTAA	1.29	0.355	-1.156	-1.282	-1.669	-0.46	NaN	-0.883	-1.017	5
82601	Putative Mitochondrial delta-1-pyrroline-5-carboxylate dehydrogenase precursor PUT2	1.886	0.05	-2.339	-1.912	-2.227	0.057	NaN	NaN	-0.694	5
103853	Putative Mitochondrial inner membrane carnitine transporter CRC1	0.171	-0.448	-1.083	-1.27	-1.585	NaN	NaN	-0.233	-0.384	5
106314	Putative Ankryn repeat protein	1.181	0.354	-0.709	-0.959	-1.248	-0.672	NaN	-0.982	-0.94	5
108555	Putative Cytochrome b5	0.895	-0.091	-0.539	-1.142	-1.144	-0.196	NaN	-0.599	-1.079	5
108732	Putative Mitochondrial isovaleryl-CoA dehydrogenase precursor	1.451	0.112	-1.942	-1.758	-1.865	-0.049	NaN	-0.166	-0.398	5
111251	Putative protein of unknown function	1.211	-0.014	-0.802	-1.601	-1.891	-0.435	NaN	-0.61	-0.89	5
112460	Putative Carnitine O-acetyltransferase CAT2 / YCAT	0.135	-0.581	-1.109	-1.158	-1.321	-0.513	-0.413	-0.925	-1.131	5
120017	Putative Oligopeptide transporter OPT-like	1.988	0.797	-1.768	-0.702	-1.02	-0.936	-1.305	-1.745	-1.594	5
120079	Putative Mitochondrial Acetyl-CoA acetyltransferase	1.762	0.188	-2.738	-2.522	-2.761	NaN	NaN	-0.488	-0.758	5
120300	Putative Electron transfer flavoprotein alpha subunit AIM45	0.773	0.172	-1.284	-1.093	-1.306	-0.171	NaN	-0.61	-0.854	5
120473	Putative Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex	1.342	0.124	-1.648	-1.331	-1.427	0.093	-0.18	-0.1	-0.322	5
120765	Putative Mitochondrial enoyl-CoA hydratase	0.722	0.307	-1.128	-0.907	-1.343	-0.24	NaN	-0.544	-0.784	5
120784	Aldose 1-epimerase AEP3	1.086	0.203	-1.388	-0.707	-1.466	-0.928	-1.208	-1.904	-2.159	5
120786	Putative Mitochondrial hydroxymethylglutaryl-CoA lyase	0.508	-0.414	-1.165	-0.749	-0.958	-0.249	NaN	-0.367	-0.496	5
120837	Putative Secreted protein of unknown function	1.2	0.932	0.305	-1.34	-1.547	-0.239	-0.501	-1.474	-1.604	5
120943	Putative NAD-specific glutamate dehydrogenase GDH2	0.931	-0.084	-1.152	-0.599	-0.983	-0.761	-0.877	-1.587	-1.689	5
121255	Putative 2-methylcitrate dehydratase PDH1	1.158	0.196	-1.859	-1.625	-1.791	-0.363	-0.674	-0.845	-1.026	5
121629	Putative Isovaleryl-CoA dehydrogenase	0.332	-0.491	-1.011	-0.932	-1.423	-0.452	-0.369	-0.83	-0.8	5

121735	Glycoside hydrolase family 3 CEL3B	0.194	-1.309	-3.791	-4.795	-4.963	-0.114	NaN	-0.847	-1.244	5
122745	Putative Mitochondrial 2-oxoisovalerate dehydrogenase beta subunit	1.745	0.419	-1.813	-1.25	-1.329	NaN	NaN	-0.127	-0.506	5
122774	Putative 2-nitropropane dioxygenase	0.251	-0.47	-1.116	-0.956	-1.29	-0.207	NaN	-0.385	-0.789	5
122963	Putative Copper amine oxidase	-0.857	-2.26	-3.922	-3.796	-3.854	-0.594	-0.462	-0.822	-0.954	5
123274	Putative Aldehyde dehydrogenase ALD2	-0.706	-0.961	-3.171	-3.827	-3.805	-1.204	-1.383	-2.341	-2.397	5
123611	Putative Proline permease PUT4	0.844	-1.86	-1.792	-1.623	-2.069	NaN	NaN	-0.36	-0.42	5
123771	Putative Mitochondrial glutaryl-CoA dehydrogenase precursor	0.74	0.139	-0.897	-1.26	-1.489	-0.418	-0.514	-1.089	-1.257	5
123820	Putative Branched chain amino acid aminotransferase	2.193	0.055	-1.461	-1.485	-1.879	-0.103	NaN	-0.222	-0.315	5
123978	Putative Alcohol oxidase AOD1	1.076	0.289	-1.425	-2.735	-3.776	-0.038	-0.477	-1.171	-1.353	5
49081	Glycoside hydrolase family 74 CEL74A	-0.704	1.881	4.922	6.601	6.652	1.8	1.692	3.774	5.144	6
49976	Putative Glycoside hydrolase family 45	-0.572	NaN	4.337	5.307	5.692	2.574	2.581	4.521	5.926	6
69276	Putative Glycoside hydrolase family 30	-0.07	NaN	NaN	6.263	6.783	1.103	1.549	2.71	3.336	6
72567	Glycoside hydrolase family 6 CBH2/CEL6A	-3.034	4.873	7.898	8.967	9.322	4.087	4.823	NaN	NaN	6
73643	Glycoside hydrolase family 61 EGL4/CEL61A	-2.544	2.025	3.662	4.763	5.173	1.761	1.488	2.203	2.898	6
120312	Glycoside hydrolase family 5 EGL2/CEL5A	-2.038	NaN	7.248	9.067	9.459	3.554	4.075	5.218	6.302	6
120504	Putative protein of unknown function	-0.119	2.006	3.715	4.328	4.721	1.256	NaN	2.838	3.679	6
121127	Glycoside hydrolase family 3 BXL1	-1.397	1.874	3.568	5.059	5.363	2.174	1.667	2.955	3.067	6
122081	Glycoside hydrolase family 7 EGL1/CEL7B	-1.374	2.768	6.051	7.286	7.69	3.034	3.833	NaN	NaN	6
123992	Endoglucanase Swollenin SWO1	-2.025	3.192	5.398	6.467	7.061	2.921	3.027	4.29	4.92	6
3405	Putative Lactose permease	-4.018	1.899	2.794	2.679	2.92	2.065	2.025	NaN	NaN	7
4213	Putative Secreted Ribonuclease T2	NaN	2.082	2.02	1.613	1.573	1.727	1.078	1.373	0.945	7
5119	Putative Fet3 ferroxidase-like protein	-1.958	0.256	0.277	1.94	2.553	-0.16	0.215	0.725	0.727	7
22197	Glycoside hydrolase family 1 CEL1B	-1.912	1.292	1.511	1.89	1.64	2.088	1.382	2.514	2.582	7
22284	Putative Gluconolactonase	-1.333	1.559	2.25	1.986	1.868	2.156	1.461	1.99	1.85	7
23062	Putative protein of unknown function	1.022	1.2	0.99	1.176	1.472	NaN	NaN	0.971	1.604	7
23271	Putative COPII-coated vesicle protein ERV41	-0.1	-0.0060	0.506	0.835	0.927	0.795	0.981	1.053	0.935	7
28563	Putative Folliculin domain protein	-0.129	0.387	1.299	1.325	1.438	0.513	NaN	0.678	1.065	7
38812	Putative Iron regulated transporter	-0.691	1.249	2.586	2.671	2.871	NaN	-0.06	0.021	0.659	7
44967	Putative protein of unknown function	-0.221	0.414	1.052	1.871	1.539	0.844	0.898	0.661	0.978	7
45250	Putative delta-12 fatty acid desaturase	0.429	0.892	1.637	0.985	1.099	1.248	1.275	1.288	1.986	7
50618	Putative MFS sugar transporter	1.177	-0.933	1.432	2.226	2.991	-0.092	NaN	1.21	2.232	7
50894	Putative Sugar transporter	-1.073	1.103	1.275	2.856	3.577	1.827	1.107	2.399	3.283	7
53238	Putative GPCR family protein	-0.168	-0.066	0.469	0.123	0.716	0.398	0.265	0.699	1.337	7
53673	Putative NADP:D-xylose dehydrogenase	-0.648	0.288	1.314	2.261	2.149	0.676	NaN	1.036	0.956	7
54632	Putative Sugar transporter (Fucose)	-0.699	1.348	2.558	NaN	2.872	1.846	0.95	2.013	2.126	7

55272	Putative protein of unknown function	1.592	0.802	1.589	1.139	2.012	0.189	NaN	0.192	0.548	7
55999	Putative Glycoside hydrolase family 31	-0.455	0.192	0.562	0.668	0.829	0.522	NaN	1.016	1.3	7
56684	Putative MFS sugar transporter (Lactose)	NaN	1.486	2.236	2.474	2.919	0.782	NaN	1.763	2.173	7
58114	Putative protein of unknown function	0.044	0.171	0.69	0.638	0.857	NaN	NaN	0.951	1.531	7
59272	Putative MFS transporter	-0.226	-0.092	1.38	1.196	1.755	0.448	0.407	1.276	0.769	7
65380	Putative Glycoside hydrolase family 47	0.061	-0.074	0.741	0.71	0.772	NaN	NaN	0.637	1.001	7
65410	Putative 3'-phosphoadenylylsulfate reductase MET16	-0.351	0.486	0.599	0.276	0.591	0.248	NaN	0.792	1.219	7
69944	Putative Glycoside hydrolase family 31	-0.106	-0.106	0.151	2.314	2.404	0.535	NaN	1.569	1.842	7
73678	Putative Calnexin CNE1	-0.1	-0.331	0.624	0.797	0.739	1.262	1.847	2.069	2.313	7
74854	Putative Delta-9 fatty acid desaturase OLE1	0.244	0.701	1.53	1.35	1.188	1.234	1.242	1.252	1.319	7
75450	Putative SNF2 family helicase	-0.107	0.361	1.358	1.66	1.882	0.295	0.371	0.58	0.42	7
76227	Glycoside hydrolase family 3 CEL3E	0.186	0.679	1.528	1.603	1.31	0.753	NaN	0.779	0.455	7
76238	Putative Plasma membrane H ⁺ -ATPase PMA1	-0.847	-1.674	-0.241	1.308	1.101	-0.021	0.44	1.12	1.229	7
76696	Putative Flavin-containing monooxygenase	-0.085	1.206	1.793	2.082	1.962	0.387	0.163	0.462	0.451	7
76722	Putative Flavohaemoglobin YHB1	-0.32	1.071	2.028	2.291	2.391	0.699	NaN	1.459	1.89	7
77481	Putative D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	-0.267	0.655	2.056	2.053	2.425	0.838	0.66	0.996	0.8	7
77513	Putative Fungal C6 transcription factor	-0.388	0.818	1.058	1.055	1.173	0.649	0.647	1.017	1.277	7
79598	Putative protein of unknown function	0.028	0.247	0.824	1.053	1.022	1.303	0.951	1.514	1.524	7
79816	Putative Transcriptional regulator GAL80	3.744	-0.205	1.217	1.591	1.249	0.161	0.228	1.966	1.245	7
80240	Glycoside hydrolase family 35 BGA1	-0.043	1.347	2.201	3.275	3.114	0.259	0.183	0.964	1.606	7
81576	Putative Assimilatory sulfite reductase alpha subunit MET10	0.0050	0.229	0.722	0.494	0.715	0.943	NaN	1.292	1.571	7
81690	Putative Carbon catabolite repression protein ART5 / CreD	-0.041	0.934	1.305	1.521	1.686	0.143	0.357	0.537	0.687	7
82227	Glycoside hydrolase family 3 CEL3C	-0.711	1.767	1.761	1.55	1.692	2.576	1.567	2.635	2.329	7
104072	Xylose permease XLT1	-2.026	NaN	NaN	3.515	2.996	0.014	-0.386	-0.495	0.026	7
105444	Putative Secreted protein of unknown function	-0.646	-0.531	-0.634	1.034	1.619	0.448	NaN	1.864	2.575	7
106171	Putative HET domain protein	0.642	0.933	1.244	1.614	1.937	0.423	NaN	1.598	3.081	7
107715	Putative protein of unknown function (Orphan)	-0.251	0.284	0.15	-0.061	0.413	0.08	0.555	1.226	1.607	7
107776	D-xylose reductase XYL1	-0.702	0.116	0.736	0.948	0.879	1.363	1.045	1.243	1.494	7
108357	Putative Zn-finger domain protein	0.113	1.005	1.7	NaN	NaN	0.817	0.333	0.389	0.706	7
108914	Putative Betaine lipid synthase	-0.323	0.261	0.643	0.203	0.123	0.787	0.798	0.759	1.291	7
109806	Putative Metallo-phosphoesterase	-0.149	0.332	0.528	1.068	1.23	0.465	NaN	1.2	1.665	7
112134	Putative protein of unknown function	-0.556	NaN	0.641	1.847	1.744	0.795	0.565	1.287	1.281	7
112390	Putative WD40 repeat domain protein	0.061	0.312	0.671	0.514	0.885	0.763	NaN	1.387	2.229	7
119578	Putative ER-derived vesicles protein ERV29	0.0	-0.327	0.568	0.669	0.759	0.68	NaN	1.214	1.495	7
119806	Putative protein of unknown function	-0.444	0.146	0.368	0.245	0.471	0.219	0.363	0.802	1.353	7

119903	Putative ER localized derlin-like family member involved in ER stress and homeostasis DFM1	-0.149	0.041	0.895	1.401	1.409	0.462	NaN	1.114	1.304	7
120688	Putative Secreted protein of unknown function	0.509	1.095	0.679	1.18	1.164	0.545	NaN	1.22	1.631	7
120749	Glycoside hydrolase family 1 BGL2	-3.044	2.123	2.375	2.422	2.39	2.616	2.142	2.569	2.728	7
120821	Putative Secreted histidine acid phosphatase	-0.785	0.208	0.663	-0.088	0.275	0.527	1.111	0.821	1.255	7
120953	Putative Glycoside hydrolase family 18	-0.172	-0.08	-0.024	1.703	2.428	0.458	1.248	2.469	3.459	7
121087	Putative 4Fe-4S ferredoxin domain protein	0.397	0.433	1.09	1.319	1.431	0.624	0.341	1.017	1.223	7
121126	Putative Integral membrane protein	-0.315	0.214	1.11	2.398	1.877	0.652	NaN	0.819	0.382	7
121232	Putative Secreted DUF1772 family protein	-0.576	0.713	1.877	2.56	2.72	NaN	NaN	0.776	0.548	7
121294	Putative Glycoside hydrolase family 16	-0.71	0.148	0.589	1.477	1.984	NaN	NaN	1.34	1.951	7
121397	Putative Translocation protein SEC61	-0.27	-0.385	0.331	0.27	0.309	0.852	1.244	1.279	1.404	7
121491	Putative Glycosyltransferase family 4	-0.211	0.765	1.465	2.273	1.938	0.584	0.228	0.208	-0.025	7
121534	Putative Pyruvate decarboxylase	-0.766	0.9	1.872	2.654	2.366	0.726	0.651	0.41	0.291	7
121754	Putative Translocation protein SEC63	-0.069	0.163	0.483	0.602	0.636	0.78	0.93	1.121	1.361	7
122208	Xylanase regulator 1 Xyr1	-0.834	2.008	2.092	2.17	2.255	2.075	1.657	2.572	3.131	7
122287	Putative Glutamate synthase GLT1	-0.149	0.03	1.592	0.914	1.1	0.69	0.763	1.37	1.977	7
122415	Disulfide-isomerase PDI1	0.122	-0.291	0.898	0.824	0.895	1.77	1.903	1.886	2.519	7
122811	Putative Glutamine synthetase GLN1	-0.13	0.092	0.95	0.496	0.65	0.849	0.594	0.853	1.466	7
122914	Putative DUF3295 family protein	0.044	1.405	1.047	1.184	1.362	0.949	0.906	1.577	1.982	7
122920	ER chaperone BIP1 (=KAR2)	-0.508	-0.328	0.85	1.032	1.148	1.503	1.964	2.095	2.361	7
123473	Putative MFS sugar transporter	0.223	1.147	1.778	2.043	2.524	0.388	0.492	0.989	1.602	7
123914	Putative Secreted protein of unknown function	2.389	0.525	0.236	0.717	0.704	0.311	0.811	1.551	2.162	7
123967	Putative Hydrophobin	-0.394	NaN	2.797	NaN	2.507	1.194	0.804	0.418	0.553	7
1818	Putative protein of unknown function	-0.08	1.155	0.812	0.932	1.137	0.309	0.0060	-0.013	0.583	8
3591	Putative Formyltetrahydrofolate deformylase	-0.775	-0.461	0.666	0.738	1.502	0.076	NaN	0.588	0.769	8
3671	Putative H/ACA ribonucleoprotein complex subunit GAR1	-0.277	0.091	1.158	0.469	0.697	0.495	NaN	0.318	0.741	8
3976	Putative USP (universal stress protein) domain protein	0.636	1.032	1.03	0.253	0.485	0.074	NaN	-0.442	-0.301	8
5011	Putative protein of unknown function	0.132	0.327	0.483	1.458	1.433	-0.39	0.095	0.241	0.834	8
6108	Putative S-adenosyl-L-methionine-dependent methyltransferase binding domain protein	-0.292	0.395	0.767	1.13	0.868	0.53	0.375	1.039	0.927	8
21433	Putative Transcription factor RFEC	-0.119	0.844	0.604	0.755	0.559	0.479	0.321	0.58	1.029	8
21452	Putative protein of unknown function	-0.199	0.013	0.745	1.268	1.233	-0.094	NaN	0.515	0.692	8
21716	Putative Acetate kinase	-0.172	0.477	0.762	1.517	1.139	0.268	-0.232	0.393	0.311	8
21908	Putative Fatty acid elongation protein GNS1/SUR4	0.058	0.599	1.35	1.761	1.448	0.45	0.215	0.379	0.043	8
21960	Putative Phospholipase E	0.426	0.496	0.16	1.005	1.276	0.108	NaN	0.408	1.256	8
23368	Putative L-ornithine 5-monooxygenase	0.108	0.494	1.507	0.961	1.012	NaN	NaN	0.253	0.167	8

44449	Putative Centromere/microtubule-binding protein CBF5	-0.116	0.372	1.256	0.701	0.959	NaN	NaN	0.479	0.677	8
47317	Putative N-terminal binuclear Zn cluster protein	0.209	-0.016	1.008	1.317	1.447	NaN	NaN	0.385	0.577	8
47733	Putative Secreted protein of unknown function	-0.17	0.592	0.883	0.595	1.545	0.678	NaN	0.154	0.822	8
47897	Putative ABC transporter SNQ2	0.248	-0.0080	0.82	0.864	1.122	0.081	0.137	0.377	0.915	8
48788	Putative Fatty acid synthase subunit alpha FAS2	0.911	0.445	1.478	1.019	0.902	0.901	0.305	0.553	0.801	8
51603	Putative Protein similar to MU182 (meiotically up regulated) of Sc. pombe	0.332	0.087	0.242	1.31	1.074	NaN	NaN	0.017	0.295	8
51686	Putative Dienelactone hydrolase AIM2	-0.176	0.468	0.862	1.599	1.421	NaN	NaN	0.494	0.516	8
53893	Putative protein of unknown function	-0.032	1.056	0.681	0.966	1.174	0.041	0.166	0.311	0.913	8
54086	Putative Sorbitol utilization protein SOU2	-0.651	0.543	1.268	1.532	1.633	NaN	0.351	0.552	0.452	8
54694	Putative Norsolorinic acid reductase	-0.282	0.224	0.592	2.267	1.161	0.055	NaN	0.077	-0.283	8
55766	Putative Nucleolar GTP-binding protein 1 NOG1	0.241	1.095	1.69	1.226	1.565	0.246	NaN	0.279	0.621	8
55802	Putative Glycoside hydrolase family 76	-0.0020	0.759	0.95	0.878	0.727	NaN	NaN	0.681	1.078	8
56077	Putative Transcription factor	-0.032	1.108	1.133	0.639	1.026	1.08	NaN	0.822	0.836	8
56384	Putative Nucleolar protein required for pre-rRNA processing and ribosome biogenesis NSR1	0.19	0.632	1.246	NaN	0.699	-0.116	NaN	0.215	0.741	8
56434	Putative FKBP-type peptidyl-prolyl isomerase FPR3	0.23	0.906	1.449	0.834	1.291	0.09	NaN	0.0090	0.409	8
58427	Putative ATP-dependent RNA helicase DBP2	0.061	1.344	1.41	0.85	1.31	0.138	NaN	0.389	1.108	8
61750	Putative Sphinganine C4 hydrolase SUR2	0.121	0.534	0.672	1.04	0.962	0.119	0.231	0.085	0.318	8
61798	Putative protein of unknown function	-0.0080	0.669	0.65	0.794	1.105	0.425	NaN	NaN	0.806	8
63001	Putative Pyridoxamine phosphate oxidase	-0.31	0.339	0.905	1.059	1.188	NaN	NaN	0.8	1.212	8
63413	Putative protein of unknown function	-0.104	1.265	0.943	0.929	1.119	0.129	NaN	0.259	0.731	8
66436	Putative SSU processome subunit involved in processing of pre-18S rRNA NOP1	-0.1	0.597	1.24	0.659	1.029	0.419	NaN	0.342	0.615	8
67275	Putative Guanine nucleotide exchange factor	0.482	0.532	0.924	0.839	1.01	-0.078	NaN	0.283	0.245	8
67334	Putative MFS L-fucose transporter	0.18	0.835	0.765	0.734	1.129	0.208	NaN	0.68	0.886	8
69465	Putative FAD monooxygenase	-0.275	0.047	1.081	1.617	1.623	-0.295	0.05	NaN	-0.085	8
69611	Putative MFS transporter	0.473	0.574	1.388	0.706	0.989	NaN	NaN	0.387	0.815	8
70847	Putative Lipase	-0.38	0.027	0.554	1.16	1.199	0.173	NaN	0.411	0.494	8
71092	Putative Secreted histidine acid phosphatase	-0.298	0.337	1.011	1.124	1.572	0.481	NaN	0.493	0.763	8
72259	Putative Regulator of G-protein signaling	-0.129	0.554	0.681	0.972	1.253	-0.16	0.091	0.332	1.101	8
75258	Putative Protease ULP1	0.262	0.684	0.856	1.074	1.002	0.11	NaN	0.258	0.68	8
75383	Putative Oxidoreductase	0.131	0.318	0.442	0.404	0.829	-0.071	0.065	0.164	1.121	8
75713	Putative Cytochrome P450	-0.111	0.077	0.659	1.829	1.574	0.317	NaN	0.698	0.557	8
75949	Putative G protein gamma subunit GBG1 / STE18	0.37	0.2	0.576	0.717	1.022	0.281	NaN	0.316	0.359	8
76172	Putative ER-derived vesicles protein ERV46	-0.061	0.096	0.815	0.98	0.918	NaN	NaN	0.922	1.276	8
76359	Putative protein of unknown function	0.76	0.427	0.225	0.777	1.294	-0.105	NaN	0.704	0.922	8

76641	Putative MFS transporter	-0.128	0.34	0.7	0.908	1.127	0.178	0.453	0.933	1.009	8
76949	Putative DUF250 domain protein	0.424	0.317	0.502	1.144	1.508	0.259	0.36	1.029	1.184	8
77309	Putative S-adenosylmethionine decarboxylase SPE2	0.064	0.375	1.048	0.687	0.781	0.248	NaN	0.334	0.468	8
78465	Putative MFS siderophore iron transporter ENB1	-0.168	0.483	0.875	1.201	1.103	0.022	0.227	0.366	0.122	8
78591	Putative Fatty acid synthase beta subunit	0.751	0.311	1.238	0.821	0.798	0.786	NaN	0.545	0.849	8
79202	Putative MFS transporter	-2.444	2.572	3.344	3.147	3.142	2.797	1.903	2.627	2.304	8
79258	Putative Magnesium-transporting Mg ²⁺ -ATPase	0.124	-0.164	0.593	1.087	1.076	0.068	0.383	0.628	0.636	8
79448	Putative Secreted protein of unknown function	0.173	0.199	0.503	1.23	1.519	0.338	0.323	0.537	0.929	8
79545	Putative Inosine-5'-monophosphate dehydrogenase IMD2	-0.497	0.5	1.063	0.511	0.923	0.181	0.24	0.292	0.697	8
79686	Putative Poly(A) binding protein PAB1	0.041	0.653	1.059	0.614	0.886	0.152	0.09	0.092	0.715	8
80003	Putative Phosphoglycerate dehydrogenase	0.248	1.09	1.478	0.503	0.585	0.259	NaN	0.813	0.956	8
80007	Putative protein of unknown function	0.715	1.615	1.199	0.397	-0.267	NaN	NaN	-0.151	0.067	8
80340	Putative Glycosyltransferase family 32	-0.148	0.448	1.127	0.875	0.667	0.867	NaN	0.461	0.678	8
80639	Putative Plasma membrane iron permease FTR1	-1.869	0.154	0.352	1.7	2.305	-0.195	NaN	0.709	0.699	8
80867	Putative Mitochondrial ribosomal protein of the small subunit YMR31	0.065	0.491	0.817	1.0	0.796	0.268	NaN	0.274	0.475	8
81148	Putative Aflatoxin B1 aldehyde reductase member 2	-0.669	0.24	0.888	1.472	1.369	0.305	NaN	0.65	0.611	8
81599	Putative Nucleolar Box C/D snoRNP complexes component NOP56 / SIK1	-0.067	0.589	1.301	0.719	0.984	0.251	NaN	0.247	0.461	8
81670	Putative MFS transporter	0.6	1.313	0.943	0.834	1.163	0.407	NaN	0.727	0.865	8
82032	Putative protein of unknown function	-0.616	1.416	1.382	1.847	1.146	0.257	0.176	-0.065	-0.274	8
82296	Putative MFS transporter	0.063	-0.337	0.875	0.916	0.993	NaN	NaN	0.839	1.079	8
82512	Putative 26S proteasome Transcription factor RPN4	0.129	1.03	0.727	1.191	1.263	0.458	0.265	0.131	1.026	8
102737	Putative Velvet domain protein	0.709	1.033	1.288	1.058	1.368	0.151	NaN	0.24	0.656	8
103048	Putative protein of unknown function	-0.023	0.623	0.787	1.373	1.199	-0.411	NaN	NaN	0.43	8
103799	Putative Amidohydrolase	0.207	0.62	0.884	1.016	0.793	0.242	NaN	0.504	0.31	8
104115	Putative Cytochrome P450	-0.096	-0.379	0.718	0.896	1.309	NaN	NaN	0.386	0.305	8
105106	Putative protein of unknown function	1.439	1.012	0.81	0.891	0.615	0.437	NaN	0.281	0.232	8
109088	Putative Heterochromatin factor HP1	-0.091	0.626	0.857	1.095	0.86	0.51	NaN	0.517	0.693	8
109275	Putative protein of unknown function	-0.371	1.461	0.421	0.725	0.68	0.045	NaN	0.484	1.012	8
119619	Putative protein of unknown function	-0.345	0.569	0.822	1.035	1.086	0.411	NaN	0.211	0.177	8
120064	Putative ATP-dependent RNA helicase MSS116	-0.072	1.314	1.185	NaN	1.35	-0.08	NaN	0.067	0.595	8
120370	Putative RmlC-like cupin domain protein	-0.352	0.27	1.382	1.481	1.315	0.222	NaN	NaN	0.191	8
120823	Putative Ser/thr-rich cell wall mannoprotein	0.049	0.18	1.071	0.255	1.007	0.793	0.133	0.256	0.43	8
121396	Putative protein of unknown function	0.028	1.123	0.546	0.926	1.122	0.0090	-0.157	0.185	0.134	8
121801	Putative Nucleolar protein NOP58	-0.18	0.867	1.324	0.774	1.178	0.213	NaN	0.277	0.57	8
122381	Putative Cyclin like protein CLG1	0.0080	0.619	0.59	0.589	0.931	0.234	0.083	0.532	1.234	8

122499	Putative Thiamine triphosphatase	-1.097	0.831	1.399	1.52	1.446	0.542	NaN	0.444	0.165	8
122572	Putative Heat shock protein related to HSP70	-0.476	0.541	1.132	0.26	0.718	0.29	0.193	0.205	1.333	8
122644	Putative Choline/ethanolaminephosphotransferase	-0.011	0.268	0.86	1.262	1.358	0.287	NaN	0.759	0.779	8
122653	Putative Squalene synthetase ERG9	-0.137	0.589	0.653	0.568	1.04	0.143	NaN	0.453	0.929	8
122879	Putative protein of unknown function	0.266	1.304	1.102	0.926	1.022	0.049	0.13	0.427	0.826	8
123009	Putative protein of unknown function	0.364	0.623	0.989	0.614	1.019	0.466	NaN	0.716	0.828	8
123220	Putative RNA polymerase I subunit RPA135	0.033	0.506	1.199	NaN	1.098	NaN	0.072	0.229	0.262	8
123293	Putative ABC transporter	0.333	-0.611	0.699	1.205	0.966	-0.03	0.354	0.711	0.714	8
123795	Putative Allantoinase DAL1	0.182	0.516	0.702	1.063	1.076	0.093	NaN	0.082	0.148	8
23115	Expansin like protein EEL2	0.777	0.118	0.494	-0.093	-0.092	0.644	0.558	0.413	1.161	9
47971	Putative F-Box and WD40 repeat protein	0.603	1.023	0.771	0.168	0.759	0.112	NaN	0.355	1.688	9
54511	Putative protein of unknown function	1.363	0.594	0.548	0.712	0.852	0.448	0.573	1.065	1.42	9
54675	Putative Transcription factor VIB1	0.362	0.176	-0.181	0.02	0.133	NaN	NaN	0.619	1.072	9
61408	Putative GTPase-activating protein SAR1	0.645	0.372	0.251	-0.024	0.296	0.38	NaN	0.457	1.015	9
62377	Putative Glycosyltransferase family 47	0.406	0.126	0.287	0.191	0.31	0.474	NaN	0.693	1.08	9
64710	Putative AAA+-type ATPase	1.206	0.528	0.04	0.081	0.507	0.06	0.252	0.826	0.481	9
64790	Putative Acetate permease ADY2 ATO2	0.559	-1.016	-0.259	-0.322	0.066	0.429	NaN	1.344	2.073	9
69115	Putative Dienelactone hydrolase	0.39	-0.354	0.178	0.366	0.061	NaN	0.85	1.168	0.253	9
76505	Putative Stress response element binding protein SEB1	1.025	0.096	0.205	0.044	0.284	0.464	0.253	0.695	1.359	9
77154	Putative Homeobox protein PAH1	0.365	0.142	0.54	0.31	0.445	0.657	0.474	0.804	1.302	9
77577	Putative Wall stress component sensor SLG1 / WSC1 (involved in arrest secretion response in yeast)	0.385	0.192	0.295	0.183	0.336	0.038	0.279	0.613	1.209	9
78496	Putative Flavin-containing monooxygenase	0.559	-0.046	0.262	-0.076	0.057	0.744	0.4	0.871	1.118	9
82619	Putative Arginosuccinate synthetase-like protein	1.369	0.623	0.305	0.578	0.244	0.396	0.187	0.619	0.079	9
107571	Putative protein of unknown function	0.09	-0.049	0.299	0.014	0.23	NaN	NaN	0.705	1.373	9
45445	Secreted tyrosinase TYR2	-0.208	0.018	0.06	0.306	0.567	0.102	-0.596	-1.038	-0.914	10
61078	Putative protein of unknown function	-0.319	-0.364	-0.414	1.08	0.811	-1.062	NaN	-1.306	-1.613	10
65333	Putative Glycoside hydrolase family 15	-0.211	-0.521	-0.461	1.234	0.667	-0.826	-0.477	-0.704	-1.095	10
73618	Putative Polyketide synthase	-1.666	1.832	2.302	0.0040	-0.604	-0.176	-0.141	-0.88	-1.305	10
73621	Putative Polyketide synthase	-1.935	1.856	2.466	0.119	-0.633	-0.144	-0.175	-0.903	-1.552	10
73631	Putative FAD binding domain protein	-1.616	1.177	1.461	-0.026	-0.618	0.282	-0.083	-0.782	-1.361	10
80142	Putative Chaperone HSP98/HSP104	-0.879	-0.181	0.063	1.09	0.608	-0.26	-0.319	-0.467	-1.292	10
105860	Putative Secreted protein of unknown function	-0.887	-0.284	-1.659	-0.17	0.101	-0.137	NaN	0.313	0.387	10
107639	Putative NAD(P)-binding Rossmann-fold domain protein	-0.79	-0.503	-0.0010	0.69	0.357	-0.208	0.114	-0.246	-1.016	10
109945	Putative protein of unknown function	-1.087	0.805	0.168	-0.281	0.404	-0.449	-0.669	-0.537	-1.292	10
111875	Putative protein of unknown function	-1.118	-0.301	-0.287	-0.322	-0.563	0.104	0.357	0.53	0.068	10

112568	Putative protein of unknown function	-1.026	-0.541	-0.575	-0.645	-0.458	-0.102	NaN	0.359	0.505	10
121693	Putative Glutathione-dependent formaldehyde-activating GFA domain protein	-2.02	0.34	0.641	NaN	NaN	-0.416	-0.239	-1.056	-1.716	10
122095	Putative Secreted protein of unknown function	-0.525	-0.0060	0.414	NaN	0.619	-0.544	NaN	-1.096	-1.078	10
122240	Putative Mitochondrial carnitine O-acetyltransferase YAT1	-0.485	-0.629	-1.028	-0.111	-0.077	-0.069	0.036	0.37	0.0010	10
123084	Putative NAD-binding Rossmann fold oxidoreductase	-0.13	0.329	-0.024	1.245	0.159	0.346	1.565	0.2	-2.797	10
123538	Putative Glycoside hydrolase family 72 GAS5	-0.094	0.072	0.395	0.078	0.172	0.05	-0.439	-1.075	-0.602	10
124282	Putative Copper radical oxidase	0.235	0.812	1.123	-0.211	0.465	-0.174	-0.504	-0.962	-0.995	10

Tableau annexe 8. Enrichissement fonctionnel par groupe de gènes de l'étude du transcriptome au cours de la production de cellulases. Pour chaque groupe de gènes significativement régulés, une analyse d'enrichissement basée sur l'annotation Gene Ontology a été effectuée. 4 977 des 9 129 gènes du génome de *T. reesei* sont annotés avec au moins un terme GO, dont 426 de nos 592 gènes. La significativité des ratios d'enrichissement a été calculée à l'aide d'un test exact de Fisher basé sur la distribution hypergéométrique. Ce tableau ne présente que les fonctions comprenant au moins 10 représentants sur l'ensemble du génome et obtenant une p-value inférieure à 0,05.

Groupe	Description GO	Étude	Total	p-value
1	pyridoxal phosphate binding	2	10	1,75E-03
1	amidase activity	2	14	3,47E-03
1	transaminase activity	2	18	5,75E-03
1	peroxidase activity	2	23	9,31E-03
1	response to oxidative stress	2	26	1,18E-02
1	electron transport	6	349	2,18E-02
1	oxidoreductase activity	7	452	2,21E-02
1	L-arabinose isomerase activity	2	49	3,90E-02
1	carbohydrate transport	2	53	4,50E-02
2	transport	6	350	9,68E-03
2	integral to membrane	6	387	1,55E-02
3	proteasome core complex (sensu Eukaryota)	4	14	4,71E-05
3	proteolysis and peptidolysis	12	207	6,57E-05
3	proteasome endopeptidase activity	4	22	3,14E-04
3	threonine endopeptidase activity	4	23	3,75E-04
3	ubiquitin-dependent protein catabolism	4	34	1,73E-03
3	metallopeptidase activity	4	42	3,80E-03
3	metabolism	15	519	1,19E-02
3	3-oxoacyl-[acyl-carrier protein] reductase activity	2	14	1,91E-02
3	pepsin A activity	2	17	2,77E-02
3	aspartic-type endopeptidase activity	2	18	3,08E-02
3	oxidoreductase activity, acting on CH-OH group of donors	2	19	3,41E-02
3	oxidoreductase activity	12	452	4,33E-02
4	ribosome	44	136	8,93E-48
4	structural constituent of ribosome	44	137	1,30E-47
4	protein biosynthesis	45	161	1,11E-45
4	intracellular	45	223	1,15E-38
4	small ribosomal subunit	5	10	3,62E-07
4	tetracycline transport	2	16	3,16E-02
4	tetracycline:hydrogen antiporter activity	2	16	3,16E-02
4	O-methyltransferase activity	2	16	3,16E-02
4	RNA binding	4	69	3,30E-02
5	oxidoreductase activity	15	452	3,80E-04
5	metabolism	16	519	5,36E-04
5	ligase activity	3	18	1,38E-03
5	enoyl-CoA hydratase activity	2	10	6,65E-03
5	tricarboxylic acid cycle	2	12	9,59E-03
5	transferase activity, transferring groups other than amino-acyl groups	2	13	1,12E-02
5	electron transport	10	349	1,14E-02
5	catalytic activity	11	404	1,15E-02
5	copper ion binding	2	21	2,84E-02
5	acyltransferase activity	2	21	2,84E-02
5	carbohydrate transport	3	53	2,89E-02
5	sugar porter activity	3	59	3,80E-02
6	cellulose binding	7	14	2,75E-17

6	extracellular region	7	36	6,60E-14
6	hydrolase activity, hydrolyzing O-glycosyl compounds	8	83	1,86E-13
6	carbohydrate metabolism	8	142	1,55E-11
6	beta-glucosidase activity	1	13	2,58E-02
7	hydrolase activity, hydrolyzing O-glycosyl compounds	6	83	1,59E-04
7	carbohydrate transport	5	53	1,67E-04
7	beta-glucosidase activity	3	13	2,54E-04
7	sugar porter activity	5	59	2,78E-04
7	carbohydrate metabolism	7	142	4,75E-04
7	membrane	11	383	1,17E-03
7	oxidoreductase activity	12	452	1,34E-03
7	L-arabinose isomerase activity	4	49	1,37E-03
7	transporter activity	8	251	3,18E-03
7	endoplasmic reticulum	2	11	5,13E-03
7	electron transport	9	349	7,10E-03
7	transport	9	350	7,23E-03
7	integral to membrane	9	387	1,36E-02
7	nitrogen compound metabolism	2	20	1,67E-02
7	calcium ion binding	3	66	2,80E-02
7	iron ion binding	2	28	3,17E-02
7	disulfide oxidoreductase activity	2	31	3,82E-02
8	carbonyl reductase (NADPH) activity	2	11	7,33E-03
8	integral to membrane	10	387	1,61E-02
8	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	2	18	1,93E-02
8	signal transducer activity	2	21	2,59E-02
8	fatty acid biosynthesis	2	24	3,33E-02
8	unspecific monooxygenase activity	2	24	3,33E-02
8	RNA binding	3	69	4,98E-02

Tableau annexe 9. Enrichissement fonctionnel global des gènes de l'étude du transcriptome au cours de la production de cellulases. Une analyse d'enrichissement basée sur l'annotation Gene Ontology a été effectuée pour l'ensemble des gènes détectés comme significativement régulés. La significativité des ratios d'enrichissement a été calculée à l'aide d'un test exact de Fisher basé sur la distribution hypergéométrique. Ce tableau ne présente que les fonctions obtenant une p-value inférieure à 0,05.

GO Description	Étude	Total	Enrich.	p-value
structural constituent of ribosome	45	137	3,84	4,00E-09
ribosome	45	136	3,87	4,00E-09
protein biosynthesis	46	161	3,34	7,21E-08
intracellular	48	223	2,51	5,40E-05
carbohydrate metabolism	31	142	2,55	3,00E-03
nucleus	23	567	0,47	1,84E-02
acyl-CoA dehydrogenase activity	6	8	8,76	2,43E-02
hydrolase activity, hydrolyzing O-glycosyl compounds	20	83	2,82	1,63E-02
oxidoreductase activity	64	452	1,65	3,64E-02
protein amino acid phosphorylation	1	145	0,08	1,63E-02
protein kinase activity	1	139	0,08	1,84E-02
regulation of transcription, DNA-dependent	11	333	0,39	3,64E-02
cellulose binding	7	14	5,84	3,71E-02

Tableau annexe 10. Gènes régulés au cours de la production de cellulases et touchés par des mutations.

Groupe	Transcrit	Fonction	Souches mutées	Type mut.
1	82037	Putative MFS transporter	RUT C30 &NG14	snp C->G
1	82037	Putative MFS transporter	RUT C30 &NG14	snp T->G
3	107475	Putative protein of unknown function	RUT C30 &NG14	snp T->A
3	107475	Putative protein of unknown function	RUT C30 &NG14	snp T->G
5	46794	Putative MFS peptide transporter	RUT C30 &NG14	snp C->T
5	103853	Putative Mitochondrial inner membrane carnitine transporter CRC1	RUT C30 &NG14	délétion
5	103853	Putative Mitochondrial inner membrane carnitine transporter CRC1	RUT C30 &NG14	délétion
7	77513	Putative Fungal C6 transcription factor	RUT C30 &NG14	snp C->T
7	121087	Putative 4Fe-4S ferredoxin domain protein	RUT C30 &NG14	snp G->A
non sélect.	3718	Putative D-serine dehydratase DSD1	RUT C30 &NG14	snp G->T
non sélect.	36543	Putative Transcription factor STE12	RUT C30 &NG14	translocation
non sélect.	104599	Putative L-galactonate dehydratase	RUT C30 &NG14	snp A->G
8	56077	Putative Transcription factor	RUT C30	snp G->A
8	78465	Putative MFS Siderophore iron transporter ENB1	RUT C30	snp G->A
9	54511	Putative protein of unknown function	RUT C30	snp C->A
non sélect.	28731	Putative cAMP receptor GPR2	RUT C30	snp G->A
non sélect.	56726	Putative 2-oxoisovalerate dehydrogenase alpha subunit	RUT C30	snp G->C
non sélect.	120117	Cre1	RUT C30	délétion
non sélect.	55887	Putative protein of unknown function	RUT C30	snp C->T
non sélect.	122689	Putative Cytoplasmic protein with a role in regulation of Ty1 transposition RTT10	RUT C30	snp C->T

7.2 Figures annexes

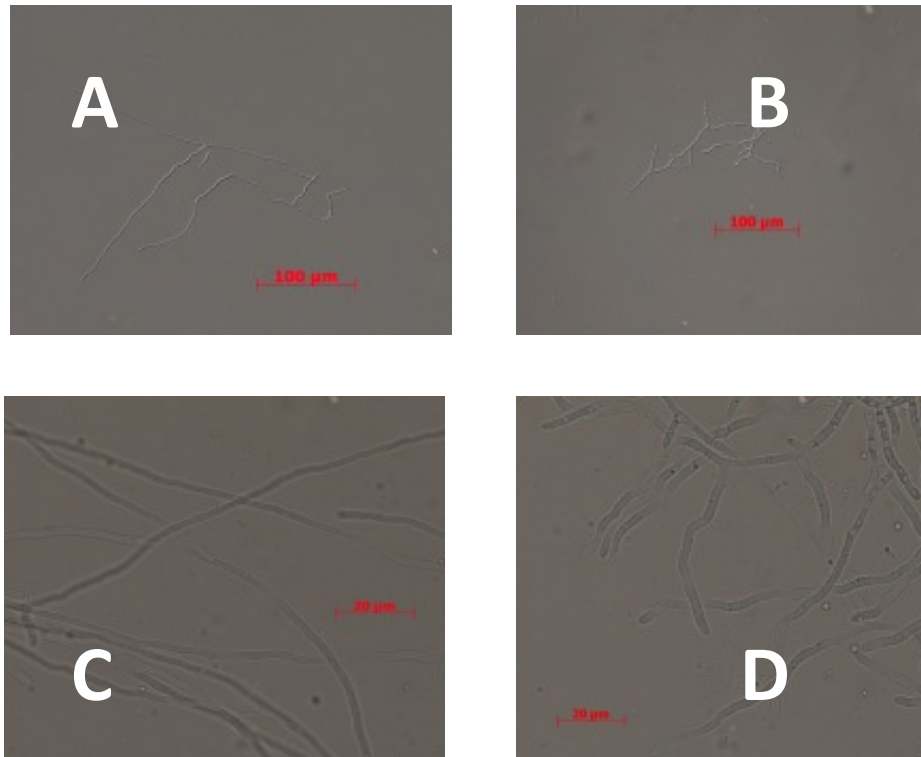


Figure annexe 1. Modifications morphologiques de la souche QM 9414 $\Delta cre1$. Morphologie des hyphes de la souche QM 9414 (A et C) et de son mutant $\Delta cre1$ (B et D). Les clichés ont été réalisés à Vienne (Autriche) au moyen d'un Zeiss AxioImager équipé d'une caméra AxioCam MRc5 et du système d'analyse d'images AxioVision AC, à un grossissement 20X ou 100X (immersion). Par rapport à son parent sauvage, la souche $\Delta cre1$ présente des hyphes raccourcis mais plus robustes, pourvus d'une paroi très épaisse et moins de septums.

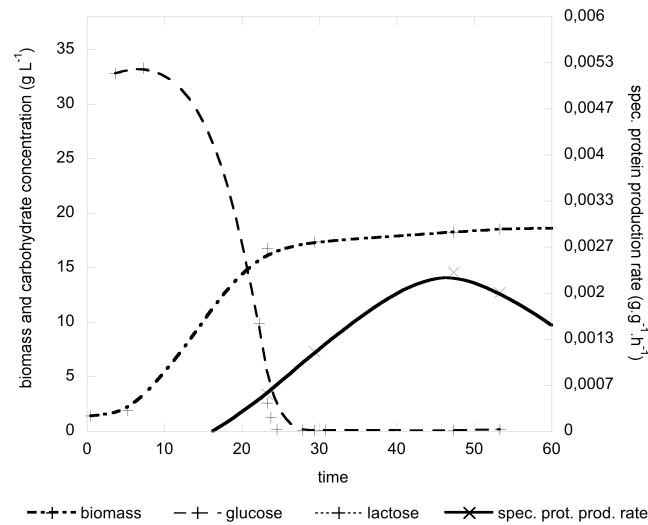


Figure annexe 2. Profil de culture de la seconde culture de RUT C30. Après une phase de croissance en glucose, la souche RUT C30 a été cultivée en fed-batch lactose induisant une production de cellulases. La concentration en biomasse sèche, la consommation de sucres et les taux spécifiques de production de protéines ont été enregistrés. Contrairement à la première culture de la même souche, le taux spécifique de production de protéines par les cellules diminue ici après une cinquantaine d'heures. Les niveaux de protéines extracellulaires enregistrés au cours des 30 premières heures du fed-batch sont néanmoins similaires pour les deux cultures ; ils atteignent 1g.L⁻¹ à T24h (dernier point de la cinétique étudiée), soit une concentration deux fois supérieure à celle enregistrée pour la culture NG14 au même instant.

Figure annexe 3. (Ci-après)

Séquence nucléotidique de *cre1*, scaffold 2, de +784322 à +789378. L'alignement entre les souches QM6a, RUT C30 et CL847 a été réalisé au moyen de GeneDoc, version 2.7.0 (2006).

Séquence nucléotidique de *xyr1*, scaffold 11, de -205723 à -200774. L'alignement entre les souches QM6a et CL847 a été réalisé au moyen de GeneDoc, version 2.7.0 (2006).

Séquence nucléotidique de *ace1*, scaffold 3, de -1211313 à -1208983. L'alignement entre les souches QM6a et CL847 a été réalisé au moyen de GeneDoc, version 2.7.0 (2006).

Séquence nucléotidique de *ace2*, scaffold 10, de -889203 à -886178. L'alignement entre les souches QM6a et CL847 a été réalisé au moyen de GeneDoc, version 2.7.0 (2006).

cre1

QM6a cre1 : GATGGCCATTCTTTTCCTCTTTGCTCTGTTTTTTTTTTTTTCCCTTCTCTTCCTTCTGCCTCCTTTCTTTCTTTTCTGTTTCTCGGGGACTTGATCTTCTGTTGCCATCT : 117
RUTC30 cre : GATGGCCATTCTTTTCCTCTTTGCTCTGTTTTTTTTTTTTTCCCTTCTCTTCCTTCTGCCTCCTTTCTTTCTTTTCTGTTTCTCGGGGACTTGATCTTCTGTTGCCATCT : 117
CL847 cre1 : GATGGCCATTCTTTTCCTCTTTGCTCTGTTTTTTTTTTTTTCCCTTCTCTTCCTTCTGCCTCCTTTCTTTCTTTTCTGTTTCTCGGGGACTTGATCTTCTGTTGCCATCT : 117
GATGGCCATTCTTTTCCTCTTTGCTCTGTTTTTTTTTTTTTCCCTTCTCTTCCTTCTGCCTCCTTTCTTTCTTTTCTGTTTCTCGGGGACTTGATCTTCTGTTGCCATCT

120 * 140 * 160 * 180 * 200 * 220 *
QM6a cre1 : TCTTCTCTTTCTGATCTTGTGAGAATGCTCGCCTGAGTGTAGAATTCACCCTGGCAATCGCATCATCCACCCTAGCCGGAGGAGTGCACGACACGGAGCATTCTAGTGCATATAA : 234
RUTC30 cre : TCTTCTCTTTCTGATCTTGTGAGAATGCTCGCCTGAGTGTAGAATTCACCCTGGCAATCGCATCATCCACCCTAGCCGGAGGAGTGCACGACACGGAGCATTCTAGTGCATATAA : 234
CL847 cre1 : TCTTCTCTTTCTGATCTTGTGAGAATGCTCGCCTGAGTGTAGAATTCACCCTGGCAATCGCATCATCCACCCTAGCCGGAGGAGTGCACGACACGGAGCATTCTAGTGCATATAA : 234
TCTTCTCTTTCTGATCTTGTGAGAATGCTCGCCTGAGTGTAGAATTCACCCTGGCAATCGCATCATCCACCCTAGCCGGAGGAGTGCACGACACGGAGCATTCTAGTGCATATAA

240 * 260 * 280 * 300 * 320 * 340 *
QM6a cre1 : GCATATCTAAACAATGGTTAGCATGCTGTAAGTGTGTGCGTGAAGCAAGCAAGTTTGGCACCAGGGCACCTTGGTCTCGTAGCGGTTAGCGTTCGGCCCCACGACACGGGAAATTGA : 351
RUTC30 cre : GCATATCTAAACAATGGTTAGCATGCTGTAAGTGTGTGCGTGAAGCAAGCAAGTTTGGCACCAGGGCACCTTGGTCTCGTAGCGGTTAGCGTTCGGCCCCACGACACGGGAAATTGA : 351
CL847 cre1 : GCATATCTAAACAATGGTTAGCATGCTGTAAGTGTGTGCGTGAAGCAAGCAAGTTTGGCACCAGGGCACCTTGGTCTCGTAGCGGTTAGCGTTCGGCCCCACGACACGGGAAATTGA : 351
GCATATCTAAACAATGGTTAGCATGCTGTAAGTGTGTGCGTGAAGCAAGCAAGTTTGGCACCAGGGCACCTTGGTCTCGTAGCGGTTAGCGTTCGGCCCCACGACACGGGAAATTGA

360 * 380 * 400 * 420 * 440 * 460
QM6a cre1 : GTCGTGCCCAGAAGCAAAGCCAGCGTCCAAGTTGAATCGATAGAGTAATGGCTTCATCGCTGGCAAAGAGTCGCCGACGATCATCGGTGCGTCAAACGACTTTCGCTTCTAGCAC : 468
RUTC30 cre : GTCGTGCCCAGAAGCAAAGCCAGCGTCCAAGTTGAATCGATAGAGTAATGGCTTCATCGCTGGCAAAGAGTCGCCGACGATCATCGGTGCGTCAAACGACTTTCGCTTCTAGCAC : 468
CL847 cre1 : GTCGTGCCCAGAAGCAAAGCCAGCGTCCAAGTTGAATCGATAGAGTAATGGCTTCATCGCTGGCAAAGAGTCGCCGACGATCATCGGTGCGTCAAACGACTTTCGCTTCTAGCAC : 468
GTCGTGCCCAGAAGCAAAGCCAGCGTCCAAGTTGAATCGATAGAGTAATGGCTTCATCGCTGGCAAAGAGTCGCCGACGATCATCGGTGCGTCAAACGACTTTCGCTTCTAGCAC

* 480 * 500 * 520 * 540 * 560 * 580
QM6a cre1 : CTAGCCGGCCTCTTGCCTGCGCCGACTCTCGACGAAGCCGGCCACACCAAAATGTATTGCGTGCCTCCCATGTGTGGGCAGCTCGCGAAGAATAAGCATAGAGTTGAGACCTTCCC : 585
RUTC30 cre : CTAGCCGGCCTCTTGCCTGCGCCGACTCTCGACGAAGCCGGCCACACCAAAATGTATTGCGTGCCTCCCATGTGTGGGCAGCTCGCGAAGAATAAGCATAGAGTTGAGACCTTCCC : 585
CL847 cre1 : CTAGCCGGCCTCTTGCCTGCGCCGACTCTCGACGAAGCCGGCCACACCAAAATGTATTGCGTGCCTCCCATGTGTGGGCAGCTCGCGAAGAATAAGCATAGAGTTGAGACCTTCCC : 585
CTAGCCGGCCTCTTGCCTGCGCCGACTCTCGACGAAGCCGGCCACACCAAAATGTATTGCGTGCCTCCCATGTGTGGGCAGCTCGCGAAGAATAAGCATAGAGTTGAGACCTTCCC

* 600 * 620 * 640 * 660 * 680 * 700
QM6a cre1 : CAGGCGCTGCGGCCAAAGGGGAAAAAGTATCCATTTTCATAAGAGACATCTTCTCGGGATTTTCGACGGCCGCACGCAGCGCCATGCTGCTCAAACCGTCTAGGCCGATGTTTCTCC : 702
RUTC30 cre : CAGGCGCTGCGGCCAAAGGGGAAAAAGTATCCATTTTCATAAGAGACATCTTCTCGGGATTTTCGACGGCCGCACGCAGCGCCATGCTGCTCAAACCGTCTAGGCCGATGTTTCTCC : 702
CL847 cre1 : CAGGCGCTGCGGCCAAAGGGGAAAAAGTATCCATTTTCATAAGAGACATCTTCTCGGGATTTTCGACGGCCGCACGCAGCGCCATGCTGCTCAAACCGTCTAGGCCGATGTTTCTCC : 702
CAGGCGCTGCGGCCAAAGGGGAAAAAGTATCCATTTTCATAAGAGACATCTTCTCGGGATTTTCGACGGCCGCACGCAGCGCCATGCTGCTCAAACCGTCTAGGCCGATGTTTCTCC

* 720 * 740 * 760 * 780 * 800 * 82
QM6a cre1 : TGTCTGATTTTGGTTCCGCTCTCCAAGTTAGGTAAGTCTCAAGTTCGCGCGTGGTACTTGTCTCGCTTGTGTTGGTACCTTGGCGCTCGAATGGTCTCCGTCGCCCTCAAAGCCACGCAAA : 819
RUTC30 cre : TGTCTGATTTTGGTTCCGCTCTCCAAGTTAGGTAAGTCTCAAGTTCGCGCGTGGTACTTGTCTCGCTTGTGTTGGTACCTTGGCGCTCGAATGGTCTCCGTCGCCCTCAAAGCCACGCAAA : 819
CL847 cre1 : TGTCTGATTTTGGTTCCGCTCTCCAAGTTAGGTAAGTCTCAAGTTCGCGCGTGGTACTTGTCTCGCTTGTGTTGGTACCTTGGCGCTCGAATGGTCTCCGTCGCCCTCAAAGCCACGCAAA : 819
TGTCTGATTTTGGTTCCGCTCTCCAAGTTAGGTAAGTCTCAAGTTCGCGCGTGGTACTTGTCTCGCTTGTGTTGGTACCTTGGCGCTCGAATGGTCTCCGTCGCCCTCAAAGCCACGCAAA

0 * 840 * 860 * 880 * 900 * 920 *
QM6a cre1 : TAGCTTTTCCGGGGAACGGCGGCCAATCAGTCGAGCAGCAGCAGCGCTGCGCCTCATCTGCGCCGATACCGTTCACAGTGGCAGCCGGGTTGACGTGGGCACGGCTGGTTTAGCTGA : 936
RUTC30 cre : TAGCTTTTCCGGGGAACGGCGGCCAATCAGTCGAGCAGCAGCAGCGCTGCGCCTCATCTGCGCCGATACCGTTCACAGTGGCAGCCGGGTTGACGTGGGCACGGCTGGTTTAGCTGA : 936
CL847 cre1 : TAGCTTTTCCGGGGAACGGCGGCCAATCAGTCGAGCAGCAGCAGCGCTGCGCCTCATCTGCGCCGATACCGTTCACAGTGGCAGCCGGGTTGACGTGGGCACGGCTGGTTTAGCTGA : 936
TAGCTTTTCCGGGGAACGGCGGCCAATCAGTCGAGCAGCAGCAGCGCTGCGCCTCATCTGCGCCGATACCGTTCACAGTGGCAGCCGGGTTGACGTGGGCACGGCTGGTTTAGCTGA

QM6a_cre1 : TGGCGGGCATTGAGTGCAGACGTGTGTGTAATCTTGAAGGTGAAGGATGCTTGAGCGGACAGTGGATCCAGACTCGGTGAGAAGCGTGGAGGTGACGAGAAGAAAAATTCAGGTAA : 1053
 RUTC30_cre : TGGCGGGCATTGAGTGCAGACGTGTGTGTAATCTTGAAGGTGAAGGATGCTTGAGCGGACAGTGGATCCAGACTCGGTGAGAAGCGTGGAGGTGACGAGAAGAAAAATTCAGGTAA : 1053
 CL847_cre1 : TGGCGGGCATTGAGTGCAGACGTGTGTGTAATCTTGAAGGTGAAGGATGCTTGAGCGGACAGTGGATCCAGACTCGGTGAGAAGCGTGGAGGTGACGAGAAGAAAAATTCAGGTAA : 1053
 TGGCGGGCATTGAGTGCAGACGTGTGTGTAATCTTGAAGGTGAAGGATGCTTGAGCGGACAGTGGATCCAGACTCGGTGAGAAGCGTGGAGGTGACGAGAAGAAAAATTCAGGTAA

QM6a_cre1 : GGTAGTAGTGAGTAGAAAAGAGGACACGGCGTACTCGTACGTGTAGGAGAGGCGGGTACGATGTAGGTACCTACTGCAAAAAAAAAAAAAAAAAAACCTTCTCCTCCTCCCTCACCG : 1170
 RUTC30_cre : GGTAGTAGTGAGTAGAAAAGAGGACACGGCGTACTCGTACGTGTAGGAGAGGCGGGTACGATGTAGGTACCTACTGCAAAAAAAAAAAAAAAAAAACCTTCTCCTCCTCCCTCACCG : 1170
 CL847_cre1 : GGTAGTAGTGAGTAGAAAAGAGGACACGGCGTACTCGTACGTGTAGGAGAGGCGGGTACGATGTAGGTACCTACTGCAAAAAAAAAAAAAAAAAAACCTTCTCCTCCTCCCTCACCG : 1170
 GGTAGTAGTGAGTAGAAAAGAGGACACGGCGTACTCGTACGTGTAGGAGAGGCGGGTACGATGTAGGTACCTACTGCAAAAAAAAAAAAAAAAAAACCTTCTCCTCCTCCCTCACCG

QM6a_cre1 : TCCGACCGTGGTTTCCACCGTGGTCCGCGCCGAGGGGGCGTCCCAATAGCGGCAAAACGCTCTGTTTGGTGCCCTTGTAGCAGCGCTTGCATCTGAGAAAGAGCAGGGGAGAAATGG : 1287
 RUTC30_cre : TCCGACCGTGGTTTCCACCGTGGTCCGCGCCGAGGGGGCGTCCCAATAGCGGCAAAACGCTCTGTTTGGTGCCCTTGTAGCAGCGCTTGCATCTGAGAAAGAGCAGGGGAGAAATGG : 1287
 CL847_cre1 : TCCGACCGTGGTTTCCACCGTGGTCCGCGCCGAGGGGGCGTCCCAATAGCGGCAAAACGCTCTGTTTGGTGCCCTTGTAGCAGCGCTTGCATCTGAGAAAGAGCAGGGGAGAAATGG : 1287
 TCCGACCGTGGTTTCCACCGTGGTCCGCGCCGAGGGGGCGTCCCAATAGCGGCAAAACGCTCTGTTTGGTGCCCTTGTAGCAGCGCTTGCATCTGAGAAAGAGCAGGGGAGAAATGG

QM6a_cre1 : AAGCCATTGGAGCAGCTAGTCTGTCCAAGTACCCTCCGTCGCGAGTACTTTGGCCCTCGCTGAGCATCTCCCGTGGGCCGAGCTGGCATCGGCCATCTCACTTGGTGAGCCCTG : 1404
 RUTC30_cre : AAGCCATTGGAGCAGCTAGTCTGTCCAAGTACCCTCCGTCGCGAGTACTTTGGCCCTCGCTGAGCATCTCCCGTGGGCCGAGCTGGCATCGGCCATCTCACTTGGTGAGCCCTG : 1404
 CL847_cre1 : AAGCCATTGGAGCAGCTAGTCTGTCCAAGTACCCTCCGTCGCGAGTACTTTGGCCCTCGCTGAGCATCTCCCGTGGGCCGAGCTGGCATCGGCCATCTCACTTGGTGAGCCCTG : 1404
 AAGCCATTGGAGCAGCTAGTCTGTCCAAGTACCCTCCGTCGCGAGTACTTTGGCCCTCGCTGAGCATCTCCCGTGGGCCGAGCTGGCATCGGCCATCTCACTTGGTGAGCCCTG

QM6a_cre1 : TCTATCTCACCCACCATCTGCTGGGGTGAAGGCTGTAGGCAATACTTGGCACTATCTGCTTGGGGCAGCCCCGCCATCTGCACACCCACCCCTCGCGCCACTCCCCTGGC : 1521
 RUTC30_cre : TCTATCTCACCCACCATCTGCTGGGGTGAAGGCTGTAGGCAATACTTGGCACTATCTGCTTGGGGCAGCCCCGCCATCTGCACACCCACCCCTCGCGCCACTCCCCTGGC : 1521
 CL847_cre1 : TCTATCTCACCCACCATCTGCTGGGGTGAAGGCTGTAGGCAATACTTGGCACTATCTGCTTGGGGCAGCCCCGCCATCTGCACACCCACCCCTCGCGCCACTCCCCTGGC : 1521
 TCTATCTCACCCACCATCTGCTGGGGTGAAGGCTGTAGGCAATACTTGGCACTATCTGCTTGGGGCAGCCCCGCCATCTGCACACCCACCCCTCGCGCCACTCCCCTGGC

QM6a_cre1 : TCCTTCCCTGTGGCCCTGCCCTTTCGCCCCAGCTGGGGCTCCCACTGCAACTGAGGCCAAGGGCTGGGGGGCTCTGAAATGCTGGGGCGGAGGTACAACCGTCTGAGAGCCGACA : 1638
 RUTC30_cre : TCCTTCCCTGTGGCCCTGCCCTTTCGCCCCAGCTGGGGCTCCCACTGCAACTGAGGCCAAGGGCTGGGGGGCTCTGAAATGCTGGGGCGGAGGTACAACCGTCTGAGAGCCGACA : 1638
 CL847_cre1 : TCCTTCCCTGTGGCCCTGCCCTTTCGCCCCAGCTGGGGCTCCCACTGCAACTGAGGCCAAGGGCTGGGGGGCTCTGAAATGCTGGGGCGGAGGTACAACCGTCTGAGAGCCGACA : 1638
 TCCTTCCCTGTGGCCCTGCCCTTTCGCCCCAGCTGGGGCTCCCACTGCAACTGAGGCCAAGGGCTGGGGGGCTCTGAAATGCTGGGGCGGAGGTACAACCGTCTGAGAGCCGACA

QM6a_cre1 : AGTACCCCGTGCCCCCTTCTGGCCGTCTCTGAGCCCTGGCGACGTGTCCCGTCCCTCTGCCCCCTCGTCCCTTCCAGCGCCCCATCGGCCATCGAGCCACTGGACGGTGGCATGC : 1755
 RUTC30_cre : AGTACCCCGTGCCCCCTTCTGGCCGTCTCTGAGCCCTGGCGACGTGTCCCGTCCCTCTGCCCCCTCGTCCCTTCCAGCGCCCCATCGGCCATCGAGCCACTGGACGGTGGCATGC : 1755
 CL847_cre1 : AGTACCCCGTGCCCCCTTCTGGCCGTCTCTGAGCCCTGGCGACGTGTCCCGTCCCTCTGCCCCCTCGTCCCTTCCAGCGCCCCATCGGCCATCGAGCCACTGGACGGTGGCATGC : 1755
 AGTACCCCGTGCCCCCTTCTGGCCGTCTCTGAGCCCTGGCGACGTGTCCCGTCCCTCTGCCCCCTCGTCCCTTCCAGCGCCCCATCGGCCATCGAGCCACTGGACGGTGGCATGC

QM6a_cre1 : AGCGCTGTTGCACCCCCAAATGTCGCTGCCTAAGCCCTGCGTGACGCCCTGATTTACGCCCCCGTGTAGCGACCCGAGCATGTGCTTCGCGGTGGTGCGGTGCGGGTACTTGCCCA : 1872
 RUTC30_cre : AGCGCTGTTGCACCCCCAAATGTCGCTGCCTAAGCCCTGCGTGACGCCCTGATTTACGCCCCCGTGTAGCGACCCGAGCATGTGCTTCGCGGTGGTGCGGTGCGGGTACTTGCCCA : 1872
 CL847_cre1 : AGCGCTGTTGCACCCCCAAATGTCGCTGCCTAAGCCCTGCGTGACGCCCTGATTTACGCCCCCGTGTAGCGACCCGAGCATGTGCTTCGCGGTGGTGCGGTGCGGGTACTTGCCCA : 1872
 AGCGCTGTTGCACCCCCAAATGTCGCTGCCTAAGCCCTGCGTGACGCCCTGATTTACGCCCCCGTGTAGCGACCCGAGCATGTGCTTCGCGGTGGTGCGGTGCGGGTACTTGCCCA

```

1880          *          1900          *          1920          *          1940          *          1960          *          1980
QM6a_cre1 : CAAGTACCGCCCTCCCAGGTACAGCACGGTACGCACACCACGTA  

RUTC30_cre : CAAGTACCGCCCTCCCAGGTACAGCACGGTACGCACACCACGTA  

CL847_cre1 : CAAGTACCGCCCTCCCAGGTACAGCACGGTACGCACACCACGTA  

CAAGTACCGCCCTCCCAGGTACAGCACGGTACGCACACCACGTA  

*          2000          *          2020          *          2040          *          2060          *          2080          *          2100
QM6a_cre1 : AGCGATTGCCTCCCCCGTCCCTCCTTCTCCCGGAGGAAAA  

RUTC30_cre : AGCGATTGCCTCCCCCGTCCCTCCTTCTCCCGGAGGAAAA  

CL847_cre1 : AGCGATTGCCTCCCCCGTCCCTCCTTCTCCCGGAGGAAAA  

AGCGATTGCCTCCCCCGTCCCTCCTTCTCCCGGAGGAAAA  

*          2120          *          2140          *          2160          *          2180          *          2200          *          2220
QM6a_cre1 : CACATCGCATCTCGATCCTGCAACTACCGACGACAAGAGCCTCGCTC  

RUTC30_cre : CACATCGCATCTCGATCCTGCAACTACCGACGACAAGAGCCTCGCTC  

CL847_cre1 : CACATCGCATCTCGATCCTGCAACTACCGACGACAAGAGCCTCGCTC  

CACATCGCATCTCGATCCTGCAACTACCGACGACAAGAGCCTCGCTC  

*          2240          *          2260          *          2280          *          2300          *          2320          *          2340
QM6a_cre1 : TCTCTCCTCCCCGCAAACGAGGACCACCGCATACTAGCAGAGCCTG  

RUTC30_cre : TCTCTCCTCCCCGCAAACGAGGACCACCGCATACTAGCAGAGCCTG  

CL847_cre1 : TCTCTCCTCCCCGCAAACGAGGACCACCGCATACTAGCAGAGCCTG  

TCTCTCCTCCCCGCAAACGAGGACCACCGCATACTAGCAGAGCCTG  

*          2360          *          2380          *          2400          *          2420          *          2440          *          2460
QM6a_cre1 : CGCGTACTTGCTTTTTGACTGCCGCTCGATC  

RUTC30_cre : CGCGTACTTGCTTTTTGACTGCCGCTCGATC  

CL847_cre1 : CGCGTACTTGCTTTTTGACTGCCGCTCGATC  

CGCGTACTTGCTTTTTGACTGCCGCTCGATC  

460          *          2480          *          2500          *          2520          *          2540          *          2560          *
QM6a_cre1 : ACCGCCGCGGTACCGTCATCAAGCCCAATGGGCCCATTCAGGAACACAGT  

RUTC30_cre : ACCGCCGCGGTACCGTCATCAAGCCCAATGGGCCCATTCAGGAACACAGT  

CL847_cre1 : ACCGCCGCGGTACCGTCATCAAGCCCAATGGGCCCATTCAGGAACACAGT  

ACCGCCGCGGTACCGTCATCAAGCCCAATGGGCCCATTCAGGAACACAGT  

2580          *          2600          *          2620          *          2640          *          2660          *          2680          *
QM6a_cre1 : CGCCTGGAGCACCAGACCAGGCACATTTCGCACCCACACGGGGC  

RUTC30_cre : CGCCTGGAGCACCAGACCAGGCACATTTCGCACCCACACGGGGC  

CL847_cre1 : CGCCTGGAGCACCAGACCAGGCACATTTCGCACCCACACGGGGC  

CGCCTGGAGCACCAGACCAGGCACATTTCGCACCCACACGGGGC  

2700          *          2720          *          2740          *          2760          *          2780          *          2800
QM6a_cre1 : AGGATACACAGCAACCCCAACTCAAGGCGCGGCAACAAGGGCCAGCAGCAGCACCAGCTT  

RUTC30_cre : ----- : 2808  

CL847_cre1 : ----- : -

```

```
      *      2820      *      2840      *      2860      *      2880      *      2900      *      2920
QM6a_crel : GCGCCAAAGGCCATCCGCTCTGCGCCCCCTTCGACCCTCGTCTCTCCCAACGTCTCGCCTCCGCACTCCTACTCGTCCTTTGTCATGCCTCAGGTCCCATTCTCACTATGGCCGT : 2925
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

```
      *      2940      *      2960      *      2980      *      3000      *      3020      *      3040
QM6a_crel : GGCAACGACATTACGATGCTTGCCAAGGCGGCAAATCAGATTGAGCGCGAGACGCTTCTGGCGGGCCGTCCAACCACAACCTCGAGGCACCACCTTACTTTGGCCAGGGTGTTCGG : 3042
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

```
      *      3060      *      3080      *      3100      *      3120      *      3140      *      316
QM6a_crel : GGTTCGAGGCCACCCCTCGCTTCTTCGTACCACATGGCGAGAGCTCACTCCAACGACGAGGATGACCACTACCATGGCAGCTTGAGGCACGCCAAGAGGTCAAGCCCCAACTCG : 3159
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

```
      0      *      3180      *      3200      *      3220      *      3240      *      3260      *
QM6a_crel : CCCAACTCTACGGCTCCTTCTTCTCCACCTTTTCGACGACTCTGTGCCCCACCCCGGATCACACTCCCATCGCAACTCCCGCTCACTCCCCCGTCTCCGTCCCTTTTCGGGC : 3276
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

```
      3280      *      3300      *      3320      *      3340      *      3360      *      3380      *
QM6a_crel : TACGAGCTGCCGAGTCTGAGAAACCTGTCCCTGCAGCACAATACGACTCCGGCGCTGGCCCCATGGAGCCTCACCTGGACGCTCCGAGTTCACCCTCAGCTGCAGGCAAACACC : 3393
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

```
      3400      *      3420      *      3440      *      3460      *      3480      *      3500      *
QM6a_crel : ACCCGCAGCCCCGGCATGTCACTTACCACATCATTAGCCGCCCCGACGGCAGTCAGAGGAAACTGCCTGTCCCTCAGGTCCCCAAGGTGGCGGTTTCAGGACCTTCTCTCCGACGGC : 3510
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

```
      3520      *      3540      *      3560      *      3580      *      3600      *      3620
QM6a_crel : GTCTTCCCCAACAGCGGCAGAAGTCCACCACGGGCAGTCTTGCCGGTGGCGACCTCATGGATCGGATGTAGAATGTCCGGTACTCATGGCGGCATCGAATGACTTACGACTTTGG : 3627
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

```
      *      3640      *      3660      *      3680      *      3700      *      3720      *      3740
QM6a_crel : GCTATAGATGGATAGAAAGAGTTGGCGTTAATGGTTTGGAAATGGCAAGGGAATGGATACGGAAATGGAAATGGACAGAGTCGTCCCCCGCAGGACTTGACACGGGCTATTTTGGT : 3744
RUTC30_cre : ----- : -
CL847_crel : ----- : -
```

Q M6a_cre1 : * 3760 * 3780 * 3800 * 3820 * 3840 * 3860 : 3861
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

Q M6a_cre1 : * 3880 * 3900 * 3920 * 3940 * 3960 * 39 : 3978
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

Q M6a_cre1 : 80 * 4000 * 4020 * 4040 * 4060 * 4080 * : 4095
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

Q M6a_cre1 : 4100 * 4120 * 4140 * 4160 * 4180 * 4200 * : 4212
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

Q M6a_cre1 : 4220 * 4240 * 4260 * 4280 * 4300 * 4320 : 4329
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

Q M6a_cre1 : * 4340 * 4360 * 4380 * 4400 * 4420 * 4440 : 4446
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

Q M6a_cre1 : * 4460 * 4480 * 4500 * 4520 * 4540 * 4560 : 4563
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

Q M6a_cre1 : * 4580 * 4600 * 4620 * 4640 * 4660 * 4680 : 4680
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

QM6a_cre1 : CCGACGTTTCCCGCCGTTATCTCCTGATGATAATGGTTCGACGTCTGGGGGAGGGGGGCTGATTGCTGTGGTACGGAGGAAGAAGGGGGAGGGGAACGTCACAAAGAAGGGGTACT : 4797
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

QM6a_cre1 : TGGGTACACGCAATCGCGCACTTTTGGGCCCAAAAAGCGAAATGAGCGAGATTGCGGGGGATTTTTGCCCTAGGTAAGATGTCTTTATCCGGGGGAGGGGAGAGAGAGAGACTTCT : 4914
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

QM6a_cre1 : AAATGTTAGTAAGTTTCTGTATCTGTGCTCACATATCGTAAGACGCGGTGTTTGATGTGCTAGGGGAGATAATGATGGTGAGATAAAGACAAGATGGAGAGCAGATGGGGTCATAGA : 5031
RUTC30_cre : ----- : -
CL847_cre1 : ----- : -

QM6a_cre1 : TAGGTACCTACTTTAGATGCTTTAGTGGTAGAAGCTTGGCTTGGGGGTGAGCGAAGCCCAACCATCTTCTTCATGTACGCACCTCCATTACCTGCTTTTTTTTCTAATTTACCTGCC : 5148
RUTC30_cre : -----ACCTCCATTACCTGCTTTTTTTTCTAATTTACCTGCC : 2671
CL847_cre1 : -----ACCTCCATTACCTGCTTTTTTTTCTAATTTACCTGCC : 2671
ACCTCCATTACCTGCTTTTTTTTCTAATTTACCTGCC

QM6a_cre1 : AATGTAGGTAAGTAGTAAGGGTAACTGATCTAGCAGTTGCAGATTGCATGTGCG : 5203
RUTC30_cre : AATGTAGGTAAGTAGTAAGGGTAACTGATCTAGCAGTTGCAGATTGCATGTGCG : 2726
CL847_cre1 : AATGTAGGTAAGTAGTAAGGGTAACTGATCTAGCAGTTGCAGATTGCATGTGCG : 2726
AATGTAGGTAAGTAGTAAGGGTAACTGATCTAGCAGTTGCAGATTGCATGTGCG

xyr1

QM6a_XYR1 : ACGCACCAATTGTGAGCGCATCACTGCACCCAGCATCTCCACTCGCCAGTCCACCTCCACAAGGTGCTCCATCCCCATCCCGTTCTCCATCCATCCATGACCATCACTTGGCCACC : 117
CL847_XYR1 : ACGCACCAATTGTGAGCGCATCACTGCACCCAGCATCTCCACTCGCCAGTCCACCTCCACAAGGTGCTCCATCCCCATCCCGTTCTCCATCCATCCATGACCATCACTTGGCCACC : 117
ACGCACCAATTGTGAGCGCATCACTGCACCCAGCATCTCCACTCGCCAGTCCACCTCCACAAGGTGCTCCATCCCCATCCCGTTCTCCATCCATCCATGACCATCACTTGGCCACC

QM6a_XYR1 : ATCAACTCCGCCTCCACACCCATCGACGGCTTCCCTGTCTGGCTCGCCACCCCTAGTGAGGTGCTCCTACGGGGCCCTCGCAAGCCTTGTCTGATGGAGGCCCTCGGTGCGGTGCGA : 234
CL847_XYR1 : ATCAACTCCGCCTCCACACCCATCGACGGCTTCCCTGTCTGGCTCGCCACCCCTAGTGAGGTGCTCCTACGGGGCCCTCGCAAGCCTTGTCTGATGGAGGCCCTCGGTGCGGTGCGA : 234
ATCAACTCCGCCTCCACACCCATCGACGGCTTCCCTGTCTGGCTCGCCACCCCTAGTGAGGTGCTCCTACGGGGCCCTCGCAAGCCTTGTCTGATGGAGGCCCTCGGTGCGGTGCGA

QM6a_XYR1 : GAAGCGAACAGCAATAAAAATTGCGAGAGTCCCCATATATTGCTTCCATGATGGAGCTCTCTGGTCTTTTAACTAACAAACCCCAAGTCTCTGCCCAACCTGCCATCTCTTC : 351
CL847_XYR1 : GAAGCGAACAGCAATAAAAATTGCGAGAGTCCCCATATATTGCTTCCATGATGGAGCTCTCTGGTCTTTTAACTAACAAACCCCAAGTCTCTGCCCAACCTGCCATCTCTTC : 351
GAAGCGAACAGCAATAAAAATTGCGAGAGTCCCCATATATTGCTTCCATGATGGAGCTCTCTGGTCTTTTAACTAACAAACCCCAAGTCTCTGCCCAACCTGCCATCTCTTC

QM6a_XYR1 : TGCTGCATCCTCTCTGTTTGTGTCATTGTCTGCTGGTTTGTGCGCGTGGCCGGCAGCCGTGTAGCTTGTCAAGGCGCGCTCGCGATAGGGAGTTTCGCCTGCTTTCTCTTTTCCCCT : 468
CL847_XYR1 : TGCTGCATCCTCTCTGTTTGTGTCATTGTCTGCTGGTTTGTGCGCGTGGCCGGCAGCCGTGTAGCTTGTCAAGGCGCGCTCGCGATAGGGAGTTTCGCCTGCTTTCTCTTTTCCCCT : 468
TGCTGCATCCTCTCTGTTTGTGTCATTGTCTGCTGGTTTGTGCGCGTGGCCGGCAGCCGTGTAGCTTGTCAAGGCGCGCTCGCGATAGGGAGTTTCGCCTGCTTTCTCTTTTCCCCT

QM6a_XYR1 : CTTTATACCCTCGGCCCCAGCCCGTGAGAGCTGTGCGCGGCGAGTCAATCTTCGTCAAGAGGCGAGCGGTTTGATTCAGCCGAGCAGTTTCACTTTAGAAGGATGCCGACTTAAACGA : 585
CL847_XYR1 : CTTTATACCCTCGGCCCCAGCCCGTGAGAGCTGTGCGCGGCGAGTCAATCTTCGTCAAGAGGCGAGCGGTTTGATTCAGCCGAGCAGTTTCACTTTAGAAGGATGCCGACTTAAACGA : 585
CTTTATACCCTCGGCCCCAGCCCGTGAGAGCTGTGCGCGGCGAGTCAATCTTCGTCAAGAGGCGAGCGGTTTGATTCAGCCGAGCAGTTTCACTTTAGAAGGATGCCGACTTAAACGA

QM6a_XYR1 : ACAGTCAATCAGTCTCTCCTCCACTGGATACCTACCTAGGTAGTCTGCAGCAGAAGCAGCTCCTATCCTCAACCTTTTTCTGCTCAGAGCTTGGCGGATTTCTGCGCCCTTACAG : 702
CL847_XYR1 : ACAGTCAATCAGTCTCTCCTCCACTGGATACCTACCTAGGTAGTCTGCAGCAGAAGCAGCTCCTATCCTCAACCTTTTTCTGCTCAGAGCTTGGCGGATTTCTGCGCCCTTACAG : 702
ACAGTCAATCAGTCTCTCCTCCACTGGATACCTACCTAGGTAGTCTGCAGCAGAAGCAGCTCCTATCCTCAACCTTTTTCTGCTCAGAGCTTGGCGGATTTCTGCGCCCTTACAG

QM6a_XYR1 : TTTACTGTTGTGGTTTGTCTCCTCTGTGAGGGTTATAGCCCCCTTCTCTCAACACAGCTTTCATCTAAAACACAAGCTTGCACCGACAGCAGCAGTGTAGTCAAGTTTTTATACAACC : 819
CL847_XYR1 : TTTACTGTTGTGGTTTGTCTCCTCTGTGAGGGTTATAGCCCCCTTCTCTCAACACAGCTTTCATCTAAAACACAAGCTTGCACCGACAGCAGCAGTGTAGTCAAGTTTTTATACAACC : 819
TTTACTGTTGTGGTTTGTCTCCTCTGTGAGGGTTATAGCCCCCTTCTCTCAACACAGCTTTCATCTAAAACACAAGCTTGCACCGACAGCAGCAGTGTAGTCAAGTTTTTATACAACC

QM6a_XYR1 : GACAAGAATCAGGTCGCTCTGAGTTTCTCGGTTTGTGATCGAAGGCGGTCTGCTATACCGCTCAGACGTTTCGTTCTTATACTCAACGGCCGGATCAAGCGCTCGCCGCTGATCG : 936
CL847_XYR1 : GACAAGAATCAGGTCGCTCTGAGTTTCTCGGTTTGTGATCGAAGGCGGTCTGCTATACCGCTCAGACGTTTCGTTCTTATACTCAACGGCCGGATCAAGCGCTCGCCGCTGATCG : 936
GACAAGAATCAGGTCGCTCTGAGTTTCTCGGTTTGTGATCGAAGGCGGTCTGCTATACCGCTCAGACGTTTCGTTCTTATACTCAACGGCCGGATCAAGCGCTCGCCGCTGATCG

QM6a_XYR1 : CCCCTTTTGCAACAACCTGGTCAACCATACGTTCAACAGCTTTCACATCGCACACAGCGCGCCACAATGTTGTCCAATCCTCTCCGTCGCTATTCTGCCTACCCCGACATCTCCTCGGC : 1053
CL847_XYR1 : CCCCTTTTGCAACAACCTGGTCAACCATACGTTCAACAGCTTTCACATCGCACACAGCGCGCCACAATGTTGTCCAATCCTCTCCGTCGCTATTCTGCCTACCCCGACATCTCCTCGGC : 1053
CCCCTTTTGCAACAACCTGGTCAACCATACGTTCAACAGCTTTCACATCGCACACAGCGCGCCACAATGTTGTCCAATCCTCTCCGTCGCTATTCTGCCTACCCCGACATCTCCTCGGC

QM6a_XYR1 : GTCATTTGACCCGAACCTACCATGGCTCACAGTGCATCTCCACTCGATCAACGTCAACACATTCGGCAACAGCCACCCCTATCCCATGCAGCACCTCGCACAGCATGCGGAGCTTTC : 1170
CL847_XYR1 : GTCATTTGACCCGAACCTACCATGGCTCACAGTGCATCTCCACTCGATCAACGTCAACACATTCGGCAACAGCCACCCCTATCCCATGCAGCACCTCGCACAGCATGCGGAGCTTTC : 1170
GTCATTTGACCCGAACCTACCATGGCTCACAGTGCATCTCCACTCGATCAACGTCAACACATTCGGCAACAGCCACCCCTATCCCATGCAGCACCTCGCACAGCATGCGGAGCTTTC

QM6a_XYR1 : GAGTTCACGCATGATAAGGGCCAGTCCGGTGCAGCCAAAGCAGCGCCAGGGCTCTCTTATTGCTGCCAGGAAGAATTCAACGGGTACTGCTGGGCCCATTCGGCGGAGGATCAGTCG : 1287
 CL847_XYR1 : GAGTTCACGCATGATAAGGGCCAGTCCGGTGCAGCCAAAGCAGCGCCAGGGCTCTCTTATTGCTGCCAGGAAGAATTCAACGGGTACTGCTGGGCCCATTCGGCGGAGGATCAGTCG : 1287
 GAGTTCACGCATGATAAGGGCCAGTCCGGTGCAGCCAAAGCAGCGCCAGGGCTCTCTTATTGCTGCCAGGAAGAATTCAACGGGTACTGCTGGGCCCATTCGGCGGAGGATCAGTCG

QM6a_XYR1 : CGCTTGTGACCAGTGC AACCAGCTTCGTACCAAGTGGCATGGCTTACACCCATGTGCCCATTTGTATAGGTATGTCCCTTTTCTCTACACAGTGTATGCTGCGCTCAAGCACATGTAC : 1404
 CL847_XYR1 : CGCTTGTGACCAGTGC AACCAGCTTCGTACCAAGTGGCATGGCTTACACCCATGTGCCCATTTGTATAGGTATGTCCCTTTTCTCTACACAGTGTATGCTGCGCTCAAGCACATGTAC : 1404
 CGCTTGTGACCAGTGC AACCAGCTTCGTACCAAGTGGCATGGCTTACACCCATGTGCCCATTTGTATAGGTATGTCCCTTTTCTCTACACAGTGTATGCTGCGCTCAAGCACATGTAC

QM6a_XYR1 : TGATCGATCTTGTTTAGAATTCGGCCTTGGATGCGAATATGTCCGAGAGAGAAAGAAGCGTGGCAAAGCTTCGCGCAAGGATATTGCTGCCAGCAAGCCGCGGGCTGCAGCACA : 1521
 CL847_XYR1 : TGATCGATCTTGTTTAGAATTCGGCCTTGGATGCGAATATGTCCGAGAGAGAAAGAAGCGTGGCAAAGCTTCGCGCAAGGATATTGCTGCCAGCAAGCCGCGGGCTGCAGCACA : 1521
 TGATCGATCTTGTTTAGAATTCGGCCTTGGATGCGAATATGTCCGAGAGAGAAAGAAGCGTGGCAAAGCTTCGCGCAAGGATATTGCTGCCAGCAAGCCGCGGGCTGCAGCACA

QM6a_XYR1 : AACTCCGGCCAGGTCCAGGATGGTCCAGAGGATCAACATCGCAAATCTCACGCCAGCAAAGCGAATCTTCGCGTGGCAGCGCTGAGCTTGCCAGCCTGCCACGACCCGCTCA : 1638
 CL847_XYR1 : AACTCCGGCCAGGTCCAGGATGGTCCAGAGGATCAACATCGCAAATCTCACGCCAGCAAAGCGAATCTTCGCGTGGCAGCGCTGAGCTTGCCAGCCTGCCACGACCCGCTCA : 1638
 AACTCCGGCCAGGTCCAGGATGGTCCAGAGGATCAACATCGCAAATCTCACGCCAGCAAAGCGAATCTTCGCGTGGCAGCGCTGAGCTTGCCAGCCTGCCACGACCCGCTCA

QM6a_XYR1 : TGGCCACATTGAGGGCTCTGTGAGCTCCTTCAGCGACAATGGCCTTTCCAGCATGCTGCCATGGGGCGCATGGATGGCCTGGAAGATCACCATGGCCACGTCGGAGTTGATCCTGC : 1755
 CL847_XYR1 : TGGCCACATTGAGGGCTCTGTGAGCTCCTTCAGCGACAATGGCCTTTCCAGCATGCTGCCATGGGGCGCATGGATGGCCTGGAAGATCACCATGGCCACGTCGGAGTTGATCCTGC : 1755
 TGGCCACATTGAGGGCTCTGTGAGCTCCTTCAGCGACAATGGCCTTTCCAGCATGCTGCCATGGGGCGCATGGATGGCCTGGAAGATCACCATGGCCACGTCGGAGTTGATCCTGC

QM6a_XYR1 : CCTGGGCGCAACTCAGCTGGAAGCGTCAATCAGCAATGGCCCTGGGGCGCATACGGTGAAGTCCACCCCGGCTATGAGAGCCCCGGCATGAATGGCCATGTGATGGTGCACCCCGTGT : 1872
 CL847_XYR1 : CCTGGGCGCAACTCAGCTGGAAGCGTCAATCAGCAATGGCCCTGGGGCGCATACGGTGAAGTCCACCCCGGCTATGAGAGCCCCGGCATGAATGGCCATGTGATGGTGCACCCCGTGT : 1872
 CCTGGGCGCAACTCAGCTGGAAGCGTCAATCAGCAATGGCCCTGGGGCGCATACGGTGAAGTCCACCCCGGCTATGAGAGCCCCGGCATGAATGGCCATGTGATGGTGCACCCCGTGT

QM6a_XYR1 : TGGCGCGCAGACCACCATGGCCGGGTATTCGGTATCTCGTATGCTGCGCAAGCCCCGAGTCCGGCTACGTATAGCAGCGACGGTAACCTTCGACTCACCCGGTCACATCCATGATTA : 1989
 CL847_XYR1 : TGGCGCGCAGACCACCATGGCCGGGTATTCGGTATCTCGTATGCTGCGCAAGCCCCGAGTCCGGCTACGTATAGCAGCGACGGTAACCTTCGACTCACCCGGTCACATCCATGATTA : 1989
 TGGCGCGCAGACCACCATGGCCGGGTATTCGGTATCTCGTATGCTGCGCAAGCCCCGAGTCCGGCTACGTATAGCAGCGACGGTAACCTTCGACTCACCCGGTCACATCCATGATTA

QM6a_XYR1 : CCGCTGGCAAATGGGAGCTCGCCCTCATGGGGAGTCTCGCTGGCCTCGCCTTCGAACCAAGTCCAGCTTCAGCTCTCGCAGCCATCTTCAAGCAAAGCGATTTGCGATATCCTGT : 2106
 CL847_XYR1 : CCGCTGGCAAATGGGAGCTCGCCCTCATGGGGAGTCTCGCTGGCCTCGCCTTCGAACCAAGTCCAGCTTCAGCTCTCGCAGCCATCTTCAAGCAAAGCGATTTGCGATATCCTGT : 2106
 CCGCTGGCAAATGGGAGCTCGCCCTCATGGGGAGTCTCGCTGGCCTCGCCTTCGAACCAAGTCCAGCTTCAGCTCTCGCAGCCATCTTCAAGCAAAGCGATTTGCGATATCCTGT

QM6a_XYR1 : GCTTGTGAGCCTCTGCTGCCTCACCTGGGAAACATCCTCCCCGTGCTTTGGCGTGGCATCTGATTGACCTGTACTTCTCCTCGTCTTTCATCAGCACAGATGCACCCAAATGTCCCCATA : 2223
 CL847_XYR1 : GCTTGTGAGCCTCTGCTGCCTCACCTGGGAAACATCCTCCCCGTGCTTTGGCGTGGCATCTGATTGACCTGTACTTCTCCTCGTCTTTCATCAGCACAGATGCACCCAAATGTCCCCATA : 2223
 GCTTGTGAGCCTCTGCTGCCTCACCTGGGAAACATCCTCCCCGTGCTTTGGCGTGGCATCTGATTGACCTGTACTTCTCCTCGTCTTTCATCAGCACAGATGCACCCAAATGTCCCCATA

QM6a_XYR1 : CGTTCTGGGCTTCGCTCTCCGGAAGCGCTCCTTCTTGCACCCACGAACCCACGAAGGTGCCAGCCCGCGCTGCTTGGGAGCATGCTGTGGGTGGCGGCACAGACTAGCGAAGCGTC : 2340
 CL847_XYR1 : CGTTCTGGGCTTCGCTCTCCGGAAGCGCTCCTTCTTGCACCCACGAACCCACGAAGGTGCCAGCCCGCGCTGCTTGGGAGCATGCTGTGGGTGGCGGCACAGACTAGCGAAGCGTC : 2340
 CGTTCTGGGCTTCGCTCTCCGGAAGCGCTCCTTCTTGCACCCACGAACCCACGAAGGTGCCAGCCCGCGCTGCTTGGGAGCATGCTGTGGGTGGCGGCACAGACTAGCGAAGCGTC

QM6a_XYR1 : CTCTTTGACGAGCCTGCCGTCGGCGAGGAGCAAGGTCTGCCAGAAGCTGCTCGAGCTGACCGTTGGGCTTCTTCAGCCCTGATCCACACCGGCACCAACAGCCGCTCTCCAAGAC : 2457
 CL847_XYR1 : CTCTTTGACGAGCCTGCCGTCGGCGAGGAGCAAGGTCTGCCAGAAGCTGCTCGAGCTGACCGTTGGGCTTCTTCAGCCCTGATCCACACCGGCACCAACAGCCGCTCTCCAAGAC : 2457
 CTTCTTGACGAGCCTGCCGTCGGCGAGGAGCAAGGTCTGCCAGAAGCTGCTCGAGCTGACCGTTGGGCTTCTTCAGCCCTGATCCACACCGGCACCAACAGCCGCTCTCCAAGAC

QM6a_XYR1 : TAGCCCCGTCGTCGGTGTCTGCCCTGGGAGTTCTTGGGGTGGCCATGCCGGGCTCGTGAACATGGATTCACTGGCCGGCGAAACGGGTGCTTTTGGGGCCATAGGGAGCCTTGA : 2574
 CL847_XYR1 : TAGCCCCGTCGTCGGTGTCTGCCCTGGGAGTTCTTGGGGTGGCCATGCCGGGCTCGTGAACATGGATTCACTGGCCGGCGAAACGGGTGCTTTTGGGGCCATAGGGAGCCTTGA : 2574
 TAGCCCCGTCGTCGGTGTCTGCCCTGGGAGTTCTTGGGGTGGCCATGCCGGGCTCGTGAACATGGATTCACTGGCCGGCGAAACGGGTGCTTTTGGGGCCATAGGGAGCCTTGA

QM6a_XYR1 : CGACGTCATCACCTATGTGCACCTCGCCACGGTCTCGGTCGAGTACAAGGGCGCCAGCCTGCGGTGGTGGGGTGGCCATGGTCTCTCGCCAGAGAGCTCAAGCTTGGCCG : 2691
 CL847_XYR1 : CGACGTCATCACCTATGTGCACCTCGCCACGGTCTCGGTCGAGTACAAGGGCGCCAGCCTGCGGTGGTGGGGTGGCCATGGTCTCTCGCCAGAGAGCTCAAGCTTGGCCG : 2691
 CGACGTCATCACCTATGTGCACCTCGCCACGGTCTCGGTCGAGTACAAGGGCGCCAGCCTGCGGTGGTGGGGTGGCCATGGTCTCTCGCCAGAGAGCTCAAGCTTGGCCG

QM6a_XYR1 : TGAGCTGCCGCCTGGCAATCCACCTGCCAACCAGGAGGACGGCGAGGGCCTTAGCGAAGACGTGGATGAGCAGCACTTGAACAGAAACAACACTCGCTTCGTGACGGAAGAGGAGCG : 2808
 CL847_XYR1 : TGAGCTGCCGCCTGGCAATCCACCTGCCAACCAGGAGGACGGCGAGGGCCTTAGCGAAGACGTGGATGAGCAGCACTTGAACAGAAACAACACTCGCTTCGTGACGGAAGAGGAGCG : 2808
 TGAGCTGCCGCCTGGCAATCCACCTGCCAACCAGGAGGACGGCGAGGGCCTTAGCGAAGACGTGGATGAGCAGCACTTGAACAGAAACAACACTCGCTTCGTGACGGAAGAGGAGCG

QM6a_XYR1 : CGAAGAGCGACGGCGAGCATGGTGGCTCGTTTACATCGTCGACAGGCACCTGGCGCTCTGCTACAACCGCCCTTGTTCCTTCTGGACAGCGAGTGCAGCGACTTGTACCACCCGAT : 2925
 CL847_XYR1 : CGAAGAGCGACGGCGAGCATGGTGGCTCGTTTACATCGTCGACAGGCACCTGGCGCTCTGCTACAACCGCCCTTGTTCCTTCTGGACAGCGAGTGCAGCGACTTGTACCACCCGAT : 2925
 CGAAGAGCGACGGCGAGCATGGTGGCTCGTTTACATCGTCGACAGGCACCTGGCGCTCTGCTACAACCGCCCTTGTTCCTTCTGGACAGCGAGTGCAGCGACTTGTACCACCCGAT

QM6a_XYR1 : GGACGACATCAAGTGGCAGGCAGGCAAATTTCCGAGCCACGATGCAGGGAACCTCCAGCATCAACATCGATAGCTCCATGACGGACAGGTTTGGCGATAGTCCCAGGGCGCTCGCGG : 3042
 CL847_XYR1 : GGACGACATCAAGTGGCAGGCAGGCAAATTTCCGAGCCACGATGCAGGGAACCTCCAGCATCAACATCGATAGCTCCATGACGGACAGGTTTGGCGATAGTCCCAGGGCGCTCGCGG : 3042
 GGACGACATCAAGTGGCAGGCAGGCAAATTTCCGAGCCACGATGCAGGGAACCTCCAGCATCAACATCGATAGCTCCATGACGGACAGGTTTGGCGATAGTCCCAGGGCGCTCGCGG

QM6a_XYR1 : CGCACACTACGAGTGCCGCGGTCTGATGATGCAATCTCGGGCGAGATTGTGATGTCACCATGCTAAAAGCCACCCCGGTTCCGGCTTGG : 3159
 CL847_XYR1 : CGCACACTACGAGTGCCGCGGTCTGATGATGCAATCTCGGGCGAGATTGTGATGTCACCATGCTAAAAGCCACCCCGGTTCCGGCTTGG : 3159
 CGCACACTACGAGTGCCGCGGTCTGATGATGCAATCTCGGGCGAGATTGTGATGTCACCATGCTAAAAGCCACCCCGGTTCCGGCTTGG

QM6a_XYR1 : ATTCCGCTCCGCGCGGGATTGGGACGAGCAGGTTGCTGAAATCACCCGACACCTGGACATGATGAGGAGAGCCTCAAGAGGTTTCGTGGCCAAAGCATCTGCCATTGTCTCAAAGGA : 3276
 CL847_XYR1 : ATTCCGCTCCGCGCGGGATTGGGACGAGCAGGTTGCTGAAATCACCCGACACCTGGACATGATGAGGAGAGCCTCAAGAGGTTTCGTGGCCAAAGCATCTGCCATTGTCTCAAAGGA : 3276
 ATTCCGCTCCGCGCGGGATTGGGACGAGCAGGTTGCTGAAATCACCCGACACCTGGACATGATGAGGAGAGCCTCAAGAGGTTTCGTGGCCAAAGCATCTGCCATTGTCTCAAAGGA

QM6a_XYR1 : CAAGGAGCAGCATGAGATGCACGACAGTGGAGCGGTAACAGACATGCAATCTCCACTCTCGGTGCGGACCAACCGCTCCAGCCGCATGACGGAGAGCGAGATCCAGGCCAGCATCGT : 3393
 CL847_XYR1 : CAAGGAGCAGCATGAGATGCACGACAGTGGAGCGGTAACAGACATGCAATCTCCACTCTCGGTGCGGACCAACCGCTCCAGCCGCATGACGGAGAGCGAGATCCAGGCCAGCATCGT : 3393
 CAAGGAGCAGCATGAGATGCACGACAGTGGAGCGGTAACAGACATGCAATCTCCACTCTCGGTGCGGACCAACCGCTCCAGCCGCATGACGGAGAGCGAGATCCAGGCCAGCATCGT

QM6a_XYR1 : GGTGGCTTACAGCACCCATGTGATGCATGTCTCCACATCTCTTGGGATAAGTGGGATCCCATCAACCTTCTAGACGACGACGACTTGTGGATCTCGTCGGAAGGATTCGTGAC : 3510
 CL847_XYR1 : GGTGGCTTACAGCACCCATGTGATGCATGTCTCCACATCTCTTGGGATAAGTGGGATCCCATCAACCTTCTAGACGACGACGACTTGTGGATCTCGTCGGAAGGATTCGTGAC : 3510
 GGTGGCTTACAGCACCCATGTGATGCATGTCTCCACATCTCTTGGGATAAGTGGGATCCCATCAACCTTCTAGACGACGACGACTTGTGGATCTCGTCGGAAGGATTCGTGAC

QM6a_XYR1 : 3520 * 3540 * 3560 * 3580 * 3600 * 3620 : 3627
 CL847_XYR1 : GCGCAGCAGCCACGCGGTATCGGCTGCCGAAGCTATTAGCCAGATTCTCGAGTTTGACCCTGGCCTGGAGTTTATGCCATTCTTCTACGGCGTCTATCTCCTGCAGGGTTCCCTTCCT : 3627
 GCGCAGCAGCCACGCGGTATCGGCTGCCGAAGCTATTAGCCAGATTCTCGAGTTTGACCCTGGCCTGGAGTTTATGCCATTCTTCTACGGCGTCTATCTCCTGCAGGGTTCCCTTCCT

QM6a_XYR1 : * 3640 * 3660 * 3680 * 3700 * 3720 * 3740 : 3744
 CL847_XYR1 : CCTCCTGCTCATCGCCGACAAGCTGCAGGCCGAAGCGTCTCCAAGCGTCATCAAGGCTTGGCAGACCATTGTTAGGGCACACGAAGCTTGCCTTGTGACGCTGAGCACAGAGTATCA : 3744
 CCTCCTGCTCATCGCCGACAAGCTGCAGGCCGAAGCGTCTCCAAGCGTCATCAAGGCTTGGCAGACCATTGTTAGGGCACACGAAGCTTGCCTTGTGACGCTGAGCACAGAGTATCA

QM6a_XYR1 : * 3760 * 3780 * 3800 * 3820 * 3840 * 3860 : 3861
 CL847_XYR1 : GGTAAGCCCTATCAGTTCAAACGTCTATCTTGTCTGTGAATCAAAGACTGACTTGGACATCAGCGCAACTTTAGCAAGGTTATGCGAAGCGCGCTGGCTCTGATTCGGGGCCGTGTGC : 3861
 GGTAAGCCCTATCAGTTCAAACGTCTATCTTGTCTGTGAATCAAAGACTGACTTGGACATCAGCGCAACTTTAGCAAGGTTATGCGAAGCGCGCTGGCTCTGATTCGGGGCCGTGTGC

QM6a_XYR1 : * 3880 * 3900 * 3920 * 3940 * 3960 * 3980 : 3978
 CL847_XYR1 : CGGAAGATTTAGCTGAGCAGCAGCAGCGACGACGCGAGCTTCTTGCACTATAACCGATGGACTGGTAACGGAACCGGCTGGCCCTCTAAGGAGGCCACTCAATCGTATGACGTTGGA : 3978
 CGGAAGATTTAGCTGAGCAGCAGCAGCGACGACGCGAGCTTCTTGCACTATAACCGATGGACTGGTAACGGAACCGGCTGGCCCTCTAAGGAGGCCACTCAATCGTATGACGTTGGA

QM6a_XYR1 : 80 * 4000 * 4020 * 4040 * 4060 * 4080 * 4100 : 4095
 CL847_XYR1 : TTGGGGACTACACAACACGAAGGCGACCAACATAGGGGGCCGCTCTGCTGCGATATTTCAACATTGTGGCAAATATGAATATCCTTTTCATTTGTTCGGCAAGGGTGTGTTTTGGT : 4095
 TTGGGGACTACACAACACGAAGGCGACCAACATAGGGGGCCGCTCTGCTGCGATATTTCAACATTGTGGCAAATATGAATATCCTTTTCATTTGTTCGGCAAGGGTGTGTTTTGGT

QM6a_XYR1 : 4100 * 4120 * 4140 * 4160 * 4180 * 4200 * 4220 : 4212
 CL847_XYR1 : TTTGATTTGTTTACGGTCTTGGAGGCTATCTTAATACTTTGGGATGTCTTGAAGAATGGTCTAGGTGGGCTGAGGCGCCGGGCAAGGCTGGTAGGATCATGAGCGACTTTATGGTTA : 4212
 TTTGATTTGTTTACGGTCTTGGAGGCTATCTTAATACTTTGGGATGTCTTGAAGAATGGTCTAGGTGGGCTGAGGCGCCGGGCAAGGCTGGTAGGATCATGAGCGACTTTATGGTTA

QM6a_XYR1 : 4220 * 4240 * 4260 * 4280 * 4300 * 4320 : 4329
 CL847_XYR1 : TGACGAAAAAGATATCCCTTGATTATGTGTACGGCAGGCACCTGGCTCGGACGACATGTTTTGTATATTTGGTTGGGACTGCGGGAATCTCTTTGTTCGCGATGATGGGTTGGGCTATG : 4329
 TGACGAAAAAGATATCCCTTGATTATGTGTACGGCAGGCACCTGGCTCGGACGACATGTTTTGTATATTTGGTTGGGACTGCGGGAATCTCTTTGTTCGCGATGATGGGTTGGGCTATG

QM6a_XYR1 : * 4340 * 4360 * 4380 * 4400 * 4420 * 4440 : 4446
 CL847_XYR1 : TTCGGTTTTGAGGATACGATGTCAAATGCTGTATGCCTAGGTAATATGAACTTTTATGAAGAGAAAACAAAAGTGACTTGTTCGAAAGGTAGCTTCCAAGTGCACGATGAGCATGG : 4446
 TTCGGTTTTGAGGATACGATGTCAAATGCTGTATGCCTAGGTAATATGAACTTTTATGAAGAGAAAACAAAAGTGACTTGTTCGAAAGGTAGCTTCCAAGTGCACGATGAGCATGG

QM6a_XYR1 : * 4460 * 4480 * 4500 * 4520 * 4540 * 4560 : 4563
 CL847_XYR1 : CTGTAGAAATTGAGGGTGACCGGTGTCCGGTGGCTCGCCAGCTGTTCCGTCACCTTGCCCTGGCAGCTGGTTTGTCTCATTTCAGGCGCGCAGGCTCGGGGGTTCGTCACCCATCAGATAA : 4563
 CTGTAGAAATTGAGGGTGACCGGTGTCCGGTGGCTCGCCAGCTGTTCCGTCACCTTGCCCTGGCAGCTGGTTTGTCTCATTTCAGGCGCGCAGGCTCGGGGGTTCGTCACCCATCAGATAA

QM6a_XYR1 : * 4580 * 4600 * 4620 * 4640 * 4660 * 4680 : 4680
 CL847_XYR1 : CCACGGTAAATGGGTAGAGCGAGGACAAGAATGAAGGAGATGGTAAAAAGAGAATTTGCTGAAAGTGAAGAGGGCTTGATTTCACAGAATGATTCAATGATGGGGGCAAGAATATCATG : 4680
 CCACGGTAAATGGGTAGAGCGAGGACAAGAATGAAGGAGATGGTAAAAAGAGAATTTGCTGAAAGTGAAGAGGGCTTGATTTCACAGAATGATTCAATGATGGGGGCAAGAATATCATG

```

      *      4700      *      4720      *      4740      *      4760      *      4780      *      4
QM6a_XYR1 : TCTAAGGTACTATATGCTCGGAGCAACGAAGGTGGGAAACGACAAGTACGAGAGCCAGTGTGACGAGTGGCCCCAAAGAGAGCAGAGCGGAGATTTTGTCTTCTCATTGTCTGAAAT : 4797
CL847_XYR1 : TCTAAGGTACTATATGCTCGGAGCAACGAAGGTGGGAAACGACAAGTACGAGAGCCAGTGTGACGAGTGGCCCCAAAGAGAGCAGAGCGGAGATTTTGTCTTCTCATTGTCTGAAAT : 4797
      TCTAAGGTACTATATGCTCGGAGCAACGAAGGTGGGAAACGACAAGTACGAGAGCCAGTGTGACGAGTGGCCCCAAAGAGAGCAGAGCGGAGATTTTGTCTTCTCATTGTCTGAAAT

      800      *      4820      *      4840      *      4860      *      4880      *      4900      *
QM6a_XYR1 : GCTTTGGCCATGGGAATTGTGTTTGTGCGTTTGGCAGTGATTCACGGTCTGCATCTGTGATTTGCGCGTTCTCGCGTGGGAGGTGGAGGGTTTGGAGATGGGAGGATGTTTCGTTGC : 4914
CL847_XYR1 : GCTTTGGCCATGGGAATTGTGTTTGTGCGTTTGGCAGTGATTCACGGTCTGCATCTGTGATTTGCGCGTTCTCGCGTGGGAGGTGGAGGGTTTGGAGATGGGAGGATGTTTCGTTGC : 4914
      GCTTTGGCCATGGGAATTGTGTTTGTGCGTTTGGCAGTGATTCACGGTCTGCATCTGTGATTTGCGCGTTCTCGCGTGGGAGGTGGAGGGTTTGGAGATGGGAGGATGTTTCGTTGC

      4920      *      4940      *
QM6a_XYR1 : TGTGTTTGCACTCGTGAGATTGTTGATATATGGCT : 4950
CL847_XYR1 : TGTGTTTGCACTCGTGAGATTGTTGATATATGGCT : 4950
      TGTGTTTGCACTCGTGAGATTGTTGATATATGGCT

```

ace1

QM6a_ACE1 : CGCTTACGCCTCAGCCCACCCCGCGCTATGCCGAAATGGGCCGCGGGTTAGGCAGCAGCCTGCGATGGGCTGAGCGACCTGGTTTCTAATCAATCAGCCTTCGCTCCACTCCTC : 117
CL847_ACE1 : CGCTTACGCCTCAGCCCACCCCGCGCTATGCCGAAATGGGCCGCGGGTTAGGCAGCAGCCTGCGATGGGCTGAGCGACCTGGTTTCTAATCAATCAGCCTTCGCTCCACTCCTC : 117
CGCTTACGCCTCAGCCCACCCCGCGCTATGCCGAAATGGGCCGCGGGTTAGGCAGCAGCCTGCGATGGGCTGAGCGACCTGGTTTCTAATCAATCAGCCTTCGCTCCACTCCTC

120 * 140 * 160 * 180 * 200 * 220 *
QM6a_ACE1 : GACATTTTCGTTTCCCTCCTCCTCGACAACCTCGATCCGTTGCCCTCTTCTCCTCGTCCAGCGGCCTCTCGAAGTGGTACGAGAGTACAGATACCGCAGCATCGTCCGGTTTGTCTGTTTGT : 234
CL847_ACE1 : GACATTTTCGTTTCCCTCCTCCTCGACAACCTCGATCCGTTGCCCTCTTCTCCTCGTCCAGCGGCCTCTCGAAGTGGTACGAGAGTACAGATACCGCAGCATCGTCCGGTTTGTCTGTTTGT : 234
GACATTTTCGTTTCCCTCCTCCTCGACAACCTCGATCCGTTGCCCTCTTCTCCTCGTCCAGCGGCCTCTCGAAGTGGTACGAGAGTACAGATACCGCAGCATCGTCCGGTTTGTCTGTTTGT

240 * 260 * 280 * 300 * 320 * 340 *
QM6a_ACE1 : CCTCTTCGTTTTCCCTGCTCCTCCACCACCGCACAGACAGGTAATACTCCGTAACATAGTTTCCTGTGCTGCACACGCCCCGGCTCTCCGCTTCTCTTTTCCACTCCC : 351
CL847_ACE1 : CCTCTTCGTTTTCCCTGCTCCTCCACCACCGCACAGACAGGTAATACTCCGTAACATAGTTTCCTGTGCTGCACACGCCCCGGCTCTCCGCTTCTCTTTTCCACTCCC : 351
CCTCTTCGTTTTCCCTGCTCCTCCACCACCGCACAGACAGGTAATACTCCGTAACATAGTTTCCTGTGCTGCACACGCCCCGGCTCTCCGCTTCTCTTTTCCACTCCC

360 * 380 * 400 * 420 * 440 * 460
QM6a_ACE1 : TCGAGTGTCTCAGTTGCTCGTCTGGCTCTCGTCTGCCACCCCGTCGGTGCCGAGTCCTTGGCTTCATATATCCAATTTACGCCCGTCTTCTGTCAAGGCAGTTTTTTCCACACA : 468
CL847_ACE1 : TCGAGTGTCTCAGTTGCTCGTCTGGCTCTCGTCTGCCACCCCGTCGGTGCCGAGTCCTTGGCTTCATATATCCAATTTACGCCCGTCTTCTGTCAAGGCAGTTTTTTCCACACA : 468
TCGAGTGTCTCAGTTGCTCGTCTGGCTCTCGTCTGCCACCCCGTCGGTGCCGAGTCCTTGGCTTCATATATCCAATTTACGCCCGTCTTCTGTCAAGGCAGTTTTTTCCACACA

* 480 * 500 * 520 * 540 * 560 * 580
QM6a_ACE1 : TCTCTCCGCCTCCTTGGCATCTCCTCCTCCTCGTTCAGATTCCATCTCGAGACGATCCTCCTCCGGCACCAGCGTCGTTTCGTTGAGCGTCTGCTTGCCTGCTGCCACTGTCGTCTTCGTC : 585
CL847_ACE1 : TCTCTCCGCCTCCTTGGCATCTCCTCCTCCTCGTTCAGATTCCATCTCGAGACGATCCTCCTCCGGCACCAGCGTCGTTTCGTTGAGCGTCTGCTTGCCTGCTGCCACTGTCGTCTTCGTC : 585
TCTCTCCGCCTCCTTGGCATCTCCTCCTCCTCGTTCAGATTCCATCTCGAGACGATCCTCCTCCGGCACCAGCGTCGTTTCGTTGAGCGTCTGCTTGCCTGCTGCCACTGTCGTCTTCGTC

* 600 * 620 * 640 * 660 * 680 * 700
QM6a_ACE1 : CATCGCGATACACAATTGGTGGTGCCTTCTGTTCGAGATCGATTGTCTGGACCCTGTGACGTTTGCATTGAGAGCCGCCACTGGTTTGACCGCCGTCGCCGTTGAGTCAGCTAATT : 702
CL847_ACE1 : CATCGCGATACACAATTGGTGGTGCCTTCTGTTCGAGATCGATTGTCTGGACCCTGTGACGTTTGCATTGAGAGCCGCCACTGGTTTGACCGCCGTCGCCGTTGAGTCAGCTAATT : 702
CATCGCGATACACAATTGGTGGTGCCTTCTGTTCGAGATCGATTGTCTGGACCCTGTGACGTTTGCATTGAGAGCCGCCACTGGTTTGACCGCCGTCGCCGTTGAGTCAGCTAATT

* 720 * 740 * 760 * 780 * 800 * 82
QM6a_ACE1 : GGTATGGGTTGCGCTTCTCTTTATTCATATATCTGTCTGTCTCTTCCCTGCTGGCCTCAGCGCCTCGGTTTCTGCGGTTTCTCTTCGGTGTTCGATAGCCTTCGTTCTATCTCG : 819
CL847_ACE1 : GGTATGGGTTGCGCTTCTCTTTATTCATATATCTGTCTGTCTCTTCCCTGCTGGCCTCAGCGCCTCGGTTTCTGCGGTTTCTCTTCGGTGTTCGATAGCCTTCGTTCTATCTCG : 819
GGTATGGGTTGCGCTTCTCTTTATTCATATATCTGTCTGTCTCTTCCCTGCTGGCCTCAGCGCCTCGGTTTCTGCGGTTTCTCTTCGGTGTTCGATAGCCTTCGTTCTATCTCG

0 * 840 * 860 * 880 * 900 * 920 *
QM6a_ACE1 : GGCTTGTGACTCACACCTGTGACACCCCGCCCGCCCGCCCTTCAAAGAGCCTCATTTGCAACATCTCTTCTATCGACAAGGAAGCTAACCGCTTGCCGCTGCAGCTGTAGGACTC : 936
CL847_ACE1 : GGCTTGTGACTCACACCTGTGACACCCCGCCCGCCCGCCCTTCAAAGAGCCTCATTTGCAACATCTCTTCTATCGACAAGGAAGCTAACCGCTTGCCGCTGCAGCTGTAGGACTC : 936
GGCTTGTGACTCACACCTGTGACACCCCGCCCGCCCGCCCTTCAAAGAGCCTCATTTGCAACATCTCTTCTATCGACAAGGAAGCTAACCGCTTGCCGCTGCAGCTGTAGGACTC

940 * 960 * 980 * 1000 * 1020 * 1040 *
QM6a_ACE1 : GTGTGGCCATATCTCATACATCTTCCCAACCCACGGAAGGAAAGAAACAGATCTCGGCCGCCATGTCTTTTCGAACCCCGCAGAAAGGACACCGGTGACTCGTCCCGGAACCGA : 1053
CL847_ACE1 : GTGTGGCCATATCTCATACATCTTCCCAACCCACGGAAGGAAAGAAACAGATCTCGGCCGCCATGTCTTTTCGAACCCCGCAGAAAGGACACCGGTGACTCGTCCCGGAACCGA : 1053
GTGTGGCCATATCTCATACATCTTCCCAACCCACGGAAGGAAAGAAACAGATCTCGGCCGCCATGTCTTTTCGAACCCCGCAGAAAGGACACCGGTGACTCGTCCCGGAACCGA

1060 * 1080 * 1100 * 1120 * 1140 * 1160 *
QM6a_ACE1 : CTGCGAACATGGCCTGTCTCTTAAGACTACCATGACCCTCCGCAAGGGTGCCACCTTTCACCTCTCCTACATCTCCCAGCGCTTCATCTGCTGCGGGGCACTTCGTTCCCTCCTACTCT : 1170
CL847_ACE1 : CTGCGAACATGGCCTGTCTCTTAAGACTACCATGACCCTCCGCAAGGGTGCCACCTTTCACCTCTCCTACATCTCCCAGCGCTTCATCTGCTGCGGGGCACTTCGTTCCCTCCTACTCT : 1170
CTGCGAACATGGCCTGTCTCTTAAGACTACCATGACCCTCCGCAAGGGTGCCACCTTTCACCTCTCCTACATCTCCCAGCGCTTCATCTGCTGCGGGGCACTTCGTTCCCTCCTACTCT

QM6a ACE1 : CACGAGGTCCTCAATCGGCTTTTGTATGATGTCGTCGACGCAAGCCGTCGTCGATTTGCCATGACTCTGAACGATATCGACGAGGCCCTCTCCAAGCCCTCGCTCTCCGACAAGAGCCC : 1287
 CL847 ACE1 : CACGAGGTCCTCAATCGGCTTTTGTATGATGTCGTCGACGCAAGCCGTCGTCGATTTGCCATGACTCTGAACGATATCGACGAGGCCCTCTCCAAGCCCTCGCTCTCCGACAAGAGCCC : 1287
 CACGAGGTCCTCAATCGGCTTTTGTATGATGTCGTCGACGCAAGCCGTCGTCGATTTGCCATGACTCTGAACGATATCGACGAGGCCCTCTCCAAGCCCTCGCTCTCCGACAAGAGCCC

QM6a ACE1 : TCGGCCGAAGCCCTTTCGCGACACCAGCCTGCCCGTCCCTCGCGGCTTCCCTCGAACCTCCCGTCGTCGACCCCGCCATGAACAAACAAGAGCCTGAGCGAAGGGTCTCGCCCTCG : 1404
 CL847 ACE1 : TCGGCCGAAGCCCTTTCGCGACACCAGCCTGCCCGTCCCTCGCGGCTTCCCTCGAACCTCCCGTCGTCGACCCCGCCATGAACAAACAAGAGCCTGAGCGAAGGGTCTCGCCCTCG : 1404
 TCGGCCGAAGCCCTTTCGCGACACCAGCCTGCCCGTCCCTCGCGGCTTCCCTCGAACCTCCCGTCGTCGACCCCGCCATGAACAAACAAGAGCCTGAGCGAAGGGTCTCGCCCTCG

QM6a ACE1 : CTCTGTTTCGACGCACCAGAAACCACGCTCCGACAGCGGCATTGGCAGCTCAGTCGTCGACAAACGACAAGGCTGGCGCCGCGACTCTACAAAGAAGCCCTCCGCCCT : 1521
 CL847 ACE1 : CTCTGTTTCGACGCACCAGAAACCACGCTCCGACAGCGGCATTGGCAGCTCAGTCGTCGACAAACGACAAGGCTGGCGCCGCGACTCTACAAAGAAGCCCTCCGCCCT : 1521
 CTCTGTTTCGACGCACCAGAAACCACGCTCCGACAGCGGCATTGGCAGCTCAGTCGTCGACAAACGACAAGGCTGGCGCCGCGACTCTACAAAGAAGCCCTCCGCCCT

QM6a ACE1 : GACAAGGTCGGCCGCTCGAGCACCACCGCGATGCTTCCAGCCTCAGCCACCGCGCTGTCAACCGCATCCGCGAACACACTCTCCGCCCTCTGCTGGAGAAGCCACGTTGAAGGA : 1638
 CL847 ACE1 : GACAAGGTCGGCCGCTCGAGCACCACCGCGATGCTTCCAGCCTCAGCCACCGCGCTGTCAACCGCATCCGCGAACACACTCTCCGCCCTCTGCTGGAGAAGCCACGTTGAAGGA : 1638
 GACAAGGTCGGCCGCTCGAGCACCACCGCGATGCTTCCAGCCTCAGCCACCGCGCTGTCAACCGCATCCGCGAACACACTCTCCGCCCTCTGCTGGAGAAGCCACGTTGAAGGA

QM6a ACE1 : ATTTCGAGCCATCGTCTAGACGTGCCCGGGCGCATCCGATCCAAGGAAATCATTTGCTTGGCAGATCTCGAGAAGACCCCTGATCTTTCATGGCACCGGTAAGTCGACTTCTGACTGA : 1755
 CL847 ACE1 : ATTTCGAGCCATCGTCTAGACGTGCCCGGGCGCATCCGATCCAAGGAAATCATTTGCTTGGCAGATCTCGAGAAGACCCCTGATCTTTCATGGCACCGGTAAGTCGACTTCTGACTGA : 1755
 ATTTCGAGCCATCGTCTAGACGTGCCCGGGCGCATCCGATCCAAGGAAATCATTTGCTTGGCAGATCTCGAGAAGACCCCTGATCTTTCATGGCACCGGTAAGTCGACTTCTGACTGA

QM6a ACE1 : TGACGGTGTTTGGGGTGATGCTTATCGGATGTTACGTCAAAGGAAAAGGCCAAGTCCGCCGCTTATACCTTGATTTCTGCCTCAGTCCGTCGGATGCATTCAAGCGACAGTCTGA : 1872
 CL847 ACE1 : TGACGGTGTTTGGGGTGATGCTTATCGGATGTTACGTCAAAGGAAAAGGCCAAGTCCGCCGCTTATACCTTGATTTCTGCCTCAGTCCGTCGGATGCATTCAAGCGACAGTCTGA : 1872
 TGACGGTGTTTGGGGTGATGCTTATCGGATGTTACGTCAAAGGAAAAGGCCAAGTCCGCCGCTTATACCTTGATTTCTGCCTCAGTCCGTCGGATGCATTCAAGCGACAGTCTGA

QM6a ACE1 : ATATCTCACCACCGCGAACAAGTTCGCCCGGGCGACCGGCCTTACACTAACGGATACTTTATCGACCTGAAGGAGCAAATCTACCAGTACGGCAAGCAACTGGCCGCCATCAAGGA : 1989
 CL847 ACE1 : ATATCTCACCACCGCGAACAAGTTCGCCCGGGCGACCGGCCTTACACTAACGGATACTTTATCGACCTGAAGGAGCAAATCTACCAGTACGGCAAGCAACTGGCCGCCATCAAGGA : 1989
 ATATCTCACCACCGCGAACAAGTTCGCCCGGGCGACCGGCCTTACACTAACGGATACTTTATCGACCTGAAGGAGCAAATCTACCAGTACGGCAAGCAACTGGCCGCCATCAAGGA

QM6a ACE1 : AAAGGGAAGCCTTGCCGACGACATGGACATTGACCCGTACGTAACCTGATACCTCTCGTAGACGTCGAGACCACCCCTACTAACGCCCGTCAAGATCTGACGAGGTTCC : 2106
 CL847 ACE1 : AAAGGGAAGCCTTGCCGACGACATGGACATTGACCCGTACGTAACCTGATACCTCTCGTAGACGTCGAGACCACCCCTACTAACGCCCGTCAAGATCTGACGAGGTTCC : 2106
 AAAGGGAAGCCTTGCCGACGACATGGACATTGACCCGTACGTAACCTGATACCTCTCGTAGACGTCGAGACCACCCCTACTAACGCCCGTCAAGATCTGACGAGGTTCC

QM6a ACE1 : CCTCTATGGCGCGTCGCTGAGAACGGCCGCCCTGCTGAGCTCATCCGCGTCAAGAAGGACGGCACTGCCTACTCAATGGCCACCGGAAAGATTGTGACATGACCGAATCCCTAC : 2223
 CL847 ACE1 : CCTCTATGGCGCGTCGCTGAGAACGGCCGCCCTGCTGAGCTCATCCGCGTCAAGAAGGACGGCACTGCCTACTCAATGGCCACCGGAAAGATTGTGACATGACCGAATCCCTAC : 2223
 CCTCTATGGCGCGTCGCTGAGAACGGCCGCCCTGCTGAGCTCATCCGCGTCAAGAAGGACGGCACTGCCTACTCAATGGCCACCGGAAAGATTGTGACATGACCGAATCCCTAC

QM6a ACE1 : GCCGCTCAAGCGCTCCCTCAGCGAGCAGCGTGAGGACGAGGAGGAGATTATGCGGTCATGGCCCGCGGAAGAAGAAGCCACCCCGAGGAGCTGGCGCCCAAGAAGTGC CGCGGA : 2340
 CL847 ACE1 : GCCGCTCAAGCGCTCCCTCAGCGAGCAGCGTGAGGACGAGGAGGAGATTATGCGGTCATGGCCCGCGGAAGAAGAAGCCACCCCGAGGAGCTGGCGCCCAAGAAGTGC CGCGGA : 2340
 GCCGCTCAAGCGCTCCCTCAGCGAGCAGCGTGAGGACGAGGAGGAGATTATGCGGTCATGGCCCGCGGAAGAAGAAGCCACCCCGAGGAGCTGGCGCCCAAGAAGTGC CGCGGA

QM6a ACE1 : G C C C G G T G C A C C A A G G A G T T C A A G C G C C C T T G C G A C C T C A C C A A G C A G A G A G A C T C A C T C T C G T C C C T G G A A G T G C C C C A T C C C C A C T T G C A A G T A C C A G A G T A C G G C T G G C C : 2457
CL847 ACE1 : G C C C G G T G C A C C A A G G A G T T C A A G C G C C C T T G C G A C C T C A C C A A G C A G A A G A C T C A C T C T C G T C C C T G G A A G T G C C C C A T C C C C A C T T G C A A G T A C C A G A G T A C G G C T G G C C : 2457
G C C C G G T G C A C C A A G G A G T T C A A G C G C C C T T G C G A C C T C A C C A A G C A G A G A G A C T C A C T C T C G T C C C T G G A A G T G C C C C A T C C C C A C T T G C A A G T A C C A G A G T A C G G C T G G C C

460 * 2480 * 2500 * 2520 * 2540 * 2560 *
QM6a ACE1 : C A C C G A G A A G G A A A T G G A C C G C C A C A T C A A C G A C A A G C A C T C G G A C G C C C C G G C C A T G T A C G A A T G C C T C T T C A A G C C C T G C C C G T A C A A G T C G A A G C G T G A G T C G A A C T G C A A G C A : 2574
CL847 ACE1 : C A C C G A G A A G G A A A T G G A C C G C C A C A T C A A C G A C A A G C A C T C G G A C G C C C C G G C C A T G T A C G A A T G C C T C T T C A A G C C C T G C C C G T A C A A G T C G A A G C G T G A G T C G A A C T G C A A G C A : 2574
C A C C G A G A A G G A A A T G G A C C G C C A C A T C A A C G A C A A G C A C T C G G A C G C C C C G G C C A T G T A C G A A T G C C T C T T C A A G C C C T G C C C G T A C A A G T C G A A G C G T G A G T C G A A C T G C A A G C A

2580 * 2600 * 2620 * 2640 * 2660 * 2680 *
QM6a ACE1 : G C A C A T G G A A A A G G C C C A C G G T G G A C C T A T G T C C G A C C A A G A C C A A C G G C A A G A A G G C A C C G A G C C A G A A T G G C T C C A C T G C C C A G C A G A C C C C C C T C T C G C C A A C G T G T C T A C : 2691
CL847 ACE1 : G C A C A T G G A A A A G G C C C A C G G T G G A C C T A T G T C C G A C C A A G A C C A A C G G C A A G A A G G C A C C G A G C C A G A A T G G C T C C A C T G C C C A G C A G A C C C C C C T C T C G C C A A C G T G T C T A C : 2691
G C A C A T G G A A A A G G C C C A C G G T G G A C C T A T G T C C G A C C A A G A C C A A C G G C A A G A A G G C A C C G A G C C A G A A T G G C T C C A C T G C C C A G C A G A C C C C C C T C T C G C C A A C G T G T C T A C

2700 * 2720 * 2740 * 2760 * 2780 * 2800
QM6a ACE1 : G C C T T C C T C C A C G C C C A G C T A C A G C G T T C C C A C G C C T C C C C A A G A C C A G G T C A T G T C C A C C G A C T T C C C C A T G T A T C C G G C T G A T G A C G A T T G G C T C G C T A C C T A C G G C G C G A G C C : 2808
CL847 ACE1 : G C C T T C C T C C A C G C C C A G C T A C A G C G T T C C C A C G C C T C C C C A A G A C C A G G T C A T G T C C A C C G A C T T C C C C A T G T A T C C G G C T G A T G A C G A T T G G C T C G C T A C C T A C G G C G C G A G C C : 2808
G C C T T C C T C C A C G C C C A G C T A C A G C G T T C C C A C G C C T C C C C A A G A C C A G G T C A T G T C C A C C G A C T T C C C C A T G T A T C C G G C T G A T G A C G A T T G G C T C G C T A C C T A C G G C G C G A G C C

* 2820 * 2840 * 2860 * 2880 * 2900 * 2920
QM6a ACE1 : C A A C A C C A T C G A C G C C A T G G A C C T G G G T C T C G A G A A C C T T T C C C C T G C C T C T G C A G C T T C C T C G T A C G A G C A G T A C C C T C C C T A C C A G A A C G G T T C C A C T T C A T C A T C A A C G A T G A : 2925
CL847 ACE1 : C A A C A C C A T C G A C G C C A T G G A C C T G G G T C T C G A G A A C C T T T C C C C T G C C T C T G C A G C T T C C T C G T A C G A G C A G T A C C C T C C C T A C C A G A A C G G T T C C A C T T C A T C A T C A A C G A T G A : 2925
C A A C A C C A T C G A C G C C A T G G A C C T G G G T C T C G A G A A C C T T T C C C C T G C C T C T G C A G C T T C C T C G T A C G A G C A G T A C C C T C C C T A C C A G A A C G G T T C C A C T T C A T C A T C A A C G A T G A

* 2940 * 2960 * 2980 * 3000 * 3020 * 3040
QM6a ACE1 : G G A C A T C T A C G C C G C C C A T G T T C A G A T T C C T G C C A G C T G C C C A C T C C T G A G C A G G T G T A C A C C A A G A T G A T G C C C C A G C A A A T G C C G G T C T A C C A C G T C C A G C A G G A G C C A T G C A C : 3042
CL847 ACE1 : G G A C A T C T A C G C C G C C C A T G T T C A G A T T C C T G C C C A G C T G C C C A C T C C T G A G C A G G T G T A C A C C A A G A T G A T G C C C C A G C A A A T G C C G G T C T A C C A C G T C C A G C A G G A G C C A T G C A C : 3042
G G A C A T C T A C G C C G C C C A T G T T C A G A T T C C T G C C C A G C T G C C C A C T C C T G A G C A G G T G T A C A C C A A G A T G A T G C C C C A G C A A A T G C C G G T C T A C C A C G T C C A G C A G G A G C C A T G C A C

* 3060 * 3080 * 3100 * 3120 * 3140 * 316
QM6a ACE1 : C A C C G T T C C C A T C C T G G G C G A G C C T C A A T T C T C C C C A A T G C T C A G C A G A A T G C A G T T C T G T A C A C T C C G A C C T C G C T G C G C G A G G T T G A T G A A G G C T T T G A C G A G T C G T A C G C C G C : 3159
CL847 ACE1 : C A C C G T T C C C A T C C T G G G C G A G C C T C A A T T C T C C C C A A T G C T C A G C A G A A T G C A G T T C T G T A C A C T C C G A C C T C G C T G C G C G A G G T T G A T G A A G G C T T T G A C G A G T C G T A C G C C G C : 3159
C A C C G T T C C C A T C C T G G G C G A G C C T C A A T T C T C C C C A A T G C T C A G C A G A A T G C A G T T C T G T A C A C T C C G A C C T C G C T G C G C G A G G T T G A T G A A G G C T T T G A C G A G T C G T A C G C C G C

0 * 3180 * 3200 * 3220 * 3240 * 3260 *
QM6a ACE1 : A G A C G G C G C G A C T T T C A G C T G T T C C C G G C A G C G G T C G A C A A G A C G G A T G T G T T C C A G T C A T T G T T T A C C G A T A T G C C A A G T G C C A A C C T C G G C T T C T C C C A G A C C A C A C A G C C C G A : 3276
CL847 ACE1 : A G A C G G C G C G A C T T T C A G C T G T T C C C G G C A G C G G T C G A C A A G A C G G A T G T G T T C C A G T C A T T G T T T A C C G A T A T G C C A A G T G C C A A C C T C G G C T T C T C C C A G A C C A C A C A G C C C G A : 3276
A G A C G G C G C G A C T T T C A G C T G T T C C C G G C A G C G G T C G A C A A G A C G G A T G T G T T C C A G T C A T T G T T T A C C G A T A T G C C A A G T G C C A A C C T C G G C T T C T C C C A G A C C A C A C A G C C C G A

3280 * 3300 * 3320 * 3340 * 3360 * 3380 *
QM6a ACE1 : C A T C T T C A A C C A A A T A G A T T G G A G C A A C C T C G A C T A C C A G G G T T T C A A G A G T A A A A A G A T T G C G A C A C A T A C A A T G A C T A C T G C A C A A G A T G C T G C A A A C G C T T A T C C A C T C G C C G : 3393
CL847 ACE1 : C A T C T T C A A C C A A A T A G A T T G G A G C A A C C T C G A C T A C C A G G G T T T C A A G A G T A A A A A G A T T G C G A C A C A T A C A A T G A C T A C T G C A C A A G A T G C T G C A A A C G C T T A T C C A C T C G C C G : 3393
C A T C T T C A A C C A A A T A G A T T G G A G C A A C C T C G A C T A C C A G G G T T T C A A G A G T A A A A A G A T T G C G A C A C A T A C A A T G A C T A C T G C A C A A G A T G C T G C A A A C G C T T A T C C A C T C G C C G

3400 * 3420 * 3440 * 3460 * 3480 * 3500 *
QM6a ACE1 : C T T A C C A C A C T T G T T C T T T T A A C G A T T T C A T G A A G A G G T T T C C G T T G G T C A T T G A A A A A G A T T G C C T T T G T G T A T T A A A G A G T T T T G T T T C C T C A T T T T C A C C T C A T T T T C T T : 3510
CL847 ACE1 : C T T A C C A C A C T T G T T C T T T T A A C G A T T T C A T G A A G A G G T T T C C G T T G G T C A T T G A A A A A G A T T G C C T T T G T G T A T T A A A G A G T T T T G T T T C C T C A T T T T C A C C T C A T T T T C T T : 3510
C T T A C C A C A C T T G T T C T T T T A A C G A T T T C A T G A A G A G G T T T C C G T T G G T C A T T G A A A A A G A T T G C C T T T G T G T A T T A A A G A G T T T T G T T T C C T C A T T T T C A C C T C A T T T T C T T

3520 * 3540 * 3560 * 3580 * 3600 * 3620
QM6a_ACE1 : CTCTTCTCATCTGCACGAGAACGGATGCTTCATTTGCATCGAAACAAGCGGATTAGTTGGCCGTATGGGCCCTGGAGGAGGAACAGCAGCAGGCTTGAGTTGAGGCCATGGGCCAGC : 3627
CL847_ACE1 : CTCTTCTCATCTGCACGAGAACGGATGCTTCATTTGCATCGAAACAAGCGGATTAGTTGGCCGTATGGGCCCTGGAGGAGGAACAGCAGCAGGCTTGAGTTGAGGCCATGGGCCAGC : 3627
CTCTTCTCATCTGCACGAGAACGGATGCTTCATTTGCATCGAAACAAGCGGATTAGTTGGCCGTATGGGCCCTGGAGGAGGAACAGCAGCAGGCTTGAGTTGAGGCCATGGGCCAGC

* 3640 * 3660 * 3680 * 3700 * 3720 * 3740
QM6a_ACE1 : AGCCGTCTTTCAAGATACTGACAGCTTGCTGGTGGGGTAAAGGTTTACTTTTATTACACTCTTGAAGAATGATAGAAGATTTATCCCACCTTCTGGTGACCTTGTGACATGCTGT : 3744
CL847_ACE1 : AGCCGTCTTTCAAGATACTGACAGCTTGCTGGTGGGGTAAAGGTTTACTTTTATTACACTCTTGAAGAATGATAGAAGATTTATCCCACCTTCTGGTGACCTTGTGACATGCTGT : 3744
AGCCGTCTTTCAAGATACTGACAGCTTGCTGGTGGGGTAAAGGTTTACTTTTATTACACTCTTGAAGAATGATAGAAGATTTATCCCACCTTCTGGTGACCTTGTGACATGCTGT

* 3760 * 3780 * 3800 * 3820 * 3840 * 3860
QM6a_ACE1 : GAAAGTAAGGTAGGTAGTACCTTGGTACCTAGTAAGTAAGGTAAGGTAAGGTAAGGTAGGTAGGTAGTAAAGTAATTAGGTCGGTCTAATGCAAACCTGAAGGTAAGGTAAGGTAAGG : 3861
CL847_ACE1 : GAAAGTAAGGTAGGTAGTACCTTGGTACCTAGTAAGTAAGGTAAGGTAAGGTAAGGTAGGTAGGTAGTAAAGTAATTAGGTCGGTCTAATGCAAACCTGAAGGTAAGGTAAGGTAAGG : 3861
GAAAGTAAGGTAGGTAGTACCTTGGTACCTAGTAAGTAAGGTAAGGTAAGGTAAGGTAGGTAGGTAGTAAAGTAATTAGGTCGGTCTAATGCAAACCTGAAGGTAAGGTAAGGTAAGG

* 3880 * 3900 * 3920 * 3940 * 3960 * 3980
QM6a_ACE1 : TTCGTTTGCACGCCTACACGCAAAAAGAATGAATGTGTGCCACACTTCAAAAAGGTGAAGACGATGTTCTCCGATACTGGGAATCGAACCCAGGGCTGCTCGGATCTTCAACTGAA : 3978
CL847_ACE1 : TTCGTTTGCACGCCTACACGCAAAAAGAATGAATGTGTGCCACACTTCAAAAAGGTGAAGACGATGTTCTCCGATACTGGGAATCGAACCCAGGGCTGCTCGGATCTTCAACTGAA : 3978
TTCGTTTGCACGCCTACACGCAAAAAGAATGAATGTGTGCCACACTTCAAAAAGGTGAAGACGATGTTCTCCGATACTGGGAATCGAACCCAGGGCTGCTCGGATCTTCAACTGAA

80 * 4000 * 4020 * 4040 * 4060 * 4080 *
QM6a_ACE1 : GGTGAGAGCGAGCGATGTTAGCCACTACACCATATCGGATGTTGAGATTAAGCAGATGGGGATGTGCTTGACTCGTTGGAGGGTGGCGACGTGAAGTGGTTGATAACCTGTCGTGG : 4095
CL847_ACE1 : GGTGAGAGCGAGCGATGTTAGCCACTACACCATATCGGATGTTGAGATTAAGCAGATGGGGATGTGCTTGACTCGTTGGAGGGTGGCGACGTGAAGTGGTTGATAACCTGTCGTGG : 4095
GGTGAGAGCGAGCGATGTTAGCCACTACACCATATCGGATGTTGAGATTAAGCAGATGGGGATGTGCTTGACTCGTTGGAGGGTGGCGACGTGAAGTGGTTGATAACCTGTCGTGG

4100 * 4120 * 4140 * 4160 * 4180 * 4200 *
QM6a_ACE1 : GCGAGGCTTTATTGCGCTAGTACCTAGCTACTGTAGGTGACTAAAGTAAGAATGCTTTCAACTTGGCAATAGCAGAATATATGGCTAATTTGATTTCCTTTCAAGAAACAACATGTAT : 4212
CL847_ACE1 : GCGAGGCTTTATTGCGCTAGTACCTAGCTACTGTAGGTGACTAAAGTAAGAATGCTTTCAACTTGGCAATAGCAGAATATATGGCTAATTTGATTTCCTTTCAAGAAACAACATGTAT : 4212
GCGAGGCTTTATTGCGCTAGTACCTAGCTACTGTAGGTGACTAAAGTAAGAATGCTTTCAACTTGGCAATAGCAGAATATATGGCTAATTTGATTTCCTTTCAAGAAACAACATGTAT

4220 * 4240 * 4260 * 4280 * 4300 * 4320
QM6a_ACE1 : TTATATAGGCTGGCCAAACTAAAAATCGGGTTGCCTGAGTGTCTAGAGGGGGAAAAAAGAGAAGAATCTTTCATACTGGGCTGTCAAGGGAAAGAGGATCTTAATGGTGATCATGGTG : 4329
CL847_ACE1 : TTATATAGGCTGGCCAAACTAAAAATCGGGTTGCCTGAGTGTCTAGAGGGGGAAAAAAGAGAAGAATCTTTCATACTGGGCTGTCAAGGGAAAGAGGATCTTAATGGTGATCATGGTG : 4329
TTATATAGGCTGGCCAAACTAAAAATCGGGTTGCCTGAGTGTCTAGAGGGGGAAAAAAGAGAAGAATCTTTCATACTGGGCTGTCAAGGGAAAGAGGATCTTAATGGTGATCATGGTG

*
QM6a_ACE1 : GT : 4331
CL847_ACE1 : GT : 4331
GT

ace2

QM6a_ACE2 : GACGGCGACCCAGCGGGCGATGGCATGTCGGGCATTTCGGGATGGCCACGGATCTCCAGCCGCTGGGAGCTGGTGTTCGGCCTACAACGGCATTGCGATGTCGCCTGGTGTGTTGTC : 117
CL847_ACE2 : GACGGCGACCCAGCGGGCGATGGCATGTCGGGCATTTCGGGATGGCCACGGATCTCCAGCCGCTGGGAGCTGGTGTTCGGCCTACAACGGCATTGCGATGTCGCCTGGTGTGTTGTC : 117
GACGGCGACCCAGCGGGCGATGGCATGTCGGGCATTTCGGGATGGCCACGGATCTCCAGCCGCTGGGAGCTGGTGTTCGGCCTACAACGGCATTGCGATGTCGCCTGGTGTGTTGTC

120 * 140 * 160 * 180 * 200 * 220 *
QM6a_ACE2 : GCTGTTAGCGACCCGGTCTTCTCTCTTCCCGGGGAAATGGAGCGACTCACAGTACAACCTCGCGGCCCTCTGAGGCTGGTATGGGATCAACGACATAGCGACAGGCTGCCAATG : 234
CL847_ACE2 : GCTGTTAGCGACCCGGTCTTCTCTCTTCCCGGGGAAATGGAGCGACTCACAGTACAACCTCGCGGCCCTCTGAGGCTGGTATGGGATCAACGACATAGCGACAGGCTGCCAATG : 234
GCTGTTAGCGACCCGGTCTTCTCTCTTCCCGGGGAAATGGAGCGACTCACAGTACAACCTCGCGGCCCTCTGAGGCTGGTATGGGATCAACGACATAGCGACAGGCTGCCAATG

240 * 260 * 280 * 300 * 320 * 340 *
QM6a_ACE2 : TCCGTCGTGAGGAAGAGCGCCAGCATGGCGAAGCTGCAGGGAGGGCAGATTTGTAGCAAAAGCAATTGAAGGCGATGCAGGCGGATGGTGAAGATTGGAGAAGCTTCTGGGATCGT : 351
CL847_ACE2 : TCCGTCGTGAGGAAGAGCGCCAGCATGGCGAAGCTGCAGGGAGGGCAGATTTGTAGCAAAAGCAATTGAAGGCGATGCAGGCGGATGGTGAAGATTGGAGAAGCTTCTGGGATCGT : 351
TCCGTCGTGAGGAAGAGCGCCAGCATGGCGAAGCTGCAGGGAGGGCAGATTTGTAGCAAAAGCAATTGAAGGCGATGCAGGCGGATGGTGAAGATTGGAGAAGCTTCTGGGATCGT

360 * 380 * 400 * 420 * 440 * 460
QM6a_ACE2 : GGGATGTGATGAAAAAAAAAAGTCAAGGTGGCGTGGTGTGCTTCTTACTCCGTACATGTACAGGGGAATTGAGATCTATGCCTCAGTGCCGTAAGAGCACAGGTCGAGGGCAGTT : 468
CL847_ACE2 : GGGATGTGATGAAAAAAAAAAGTCAAGGTGGCGTGGTGTGCTTCTTACTCCGTACATGTACAGGGGAATTGAGATCTATGCCTCAGTGCCGTAAGAGCACAGGTCGAGGGCAGTT : 468
GGGATGTGATGAAAAAAAAAAGTCAAGGTGGCGTGGTGTGCTTCTTACTCCGTACATGTACAGGGGAATTGAGATCTATGCCTCAGTGCCGTAAGAGCACAGGTCGAGGGCAGTT

* 480 * 500 * 520 * 540 * 560 * 580
QM6a_ACE2 : CAATGAGCTCCGGTGGCGACGTAATTGATGGCGGGGCCCCGACAGCTGGCTGGGCCGTTCTAGCGGTGCTTTAGCGGGCGCTCCACTGAACCTGACAGGCTGCAGCAGCCCCGCATCG : 585
CL847_ACE2 : CAATGAGCTCCGGTGGCGACGTAATTGATGGCGGGGCCCCGACAGCTGGCTGGGCCGTTCTAGCGGTGCTTTAGCGGGCGCTCCACTGAACCTGACAGGCTGCAGCAGCCCCGCATCG : 585
CAATGAGCTCCGGTGGCGACGTAATTGATGGCGGGGCCCCGACAGCTGGCTGGGCCGTTCTAGCGGTGCTTTAGCGGGCGCTCCACTGAACCTGACAGGCTGCAGCAGCCCCGCATCG

* 600 * 620 * 640 * 660 * 680 * 700
QM6a_ACE2 : CTAATAACGGCGATGCAGATGCTCACATCAGTGAACCTGCCACTGTACCTAGGTATTATTTCTCAATAGGTGGATGTACATACGTGGACCGTGGAGGCGATGGAGGCGATCAGCATAAA : 702
CL847_ACE2 : CTAATAACGGCGATGCAGATGCTCACATCAGTGAACCTGCCACTGTACCTAGGTATTATTTCTCAATAGGTGGATGTACATACGTGGACCGTGGAGGCGATGGAGGCGATCAGCATAAA : 702
CTAATAACGGCGATGCAGATGCTCACATCAGTGAACCTGCCACTGTACCTAGGTATTATTTCTCAATAGGTGGATGTACATACGTGGACCGTGGAGGCGATGGAGGCGATCAGCATAAA

* 720 * 740 * 760 * 780 * 800 * 82
QM6a_ACE2 : TGGGCAGATCCAGTATTGGAGGCGAATGCCTACATGTATCTTGCATTTCCCATCGGTCTCTGAAAGCCAGACGCTCTTGCATATTCTGAGTCTCGCCCTGGCATCCCACCGATGGAA : 819
CL847_ACE2 : TGGGCAGATCCAGTATTGGAGGCGAATGCCTACATGTATCTTGCATTTCCCATCGGTCTCTGAAAGCCAGACGCTCTTGCATATTCTGAGTCTCGCCCTGGCATCCCACCGATGGAA : 819
TGGGCAGATCCAGTATTGGAGGCGAATGCCTACATGTATCTTGCATTTCCCATCGGTCTCTGAAAGCCAGACGCTCTTGCATATTCTGAGTCTCGCCCTGGCATCCCACCGATGGAA

0 * 840 * 860 * 880 * 900 * 920 *
QM6a_ACE2 : CTGTTTAGGCTAGGTACGCAAGTGCAAAGAACATCGCCGTCATGGCGTGATATCGGTCCATGTTGCCAGCATGGCCCTACCGTATCCGAGACTCCCGCGGGCTGCCCTGCCTGGGG : 936
CL847_ACE2 : CTGTTTAGGCTAGGTACGCAAGTGCAAAGAACATCGCCGTCATGGCGTGATATCGGTCCATGTTGCCAGCATGGCCCTACCGTATCCGAGACTCCCGCGGGCTGCCCTGCCTGGGG : 936
CTGTTTAGGCTAGGTACGCAAGTGCAAAGAACATCGCCGTCATGGCGTGATATCGGTCCATGTTGCCAGCATGGCCCTACCGTATCCGAGACTCCCGCGGGCTGCCCTGCCTGGGG

940 * 960 * 980 * 1000 * 1020 * 1040 *
QM6a_ACE2 : CTTGTTCGGACGACATCTGCTCTCTCTCAACGCTGTGCATCATCGCGATTCCCTTCTGATTTCGCCATGGACCTCCGGCAAGCATGTGATCGATGCCACGACAAGAAGCTCAGGTGTCC : 1053
CL847_ACE2 : CTTGTTCGGACGACATCTGCTCTCTCTCAACGCTGTGCATCATCGCGATTCCCTTCTGATTTCGCCATGGACCTCCGGCAAGCATGTGATCGATGCCACGACAAGAAGCTCAGGTGTCC : 1053
CTTGTTCGGACGACATCTGCTCTCTCTCAACGCTGTGCATCATCGCGATTCCCTTCTGATTTCGCCATGGACCTCCGGCAAGCATGTGATCGATGCCACGACAAGAAGCTCAGGTGTCC

1060 * 1080 * 1100 * 1120 * 1140 * 1160 *
QM6a_ACE2 : CAGGATTTCCGGCTCGCCCTGCTGCAGCCGCTGCGCAAGGGCCAATGTCGCCTGCGTCTTTAGTCCGCCATCGAGGCCATTTCCGCCCTCACGAGCCCTTGAACCCACAGCCATGAACA : 1170
CL847_ACE2 : CAGGATTTCCGGCTCGCCCTGCTGCAGCCGCTGCGCAAGGGCCAATGTCGCCTGCGTCTTTAGTCCGCCATCGAGGCCATTTCCGCCCTCACGAGCCCTTGAACCCACAGCCATGAACA : 1170
CAGGATTTCCGGCTCGCCCTGCTGCAGCCGCTGCGCAAGGGCCAATGTCGCCTGCGTCTTTAGTCCGCCATCGAGGCCATTTCCGCCCTCACGAGCCCTTGAACCCACAGCCATGAACA

QM6a ACE2 : CAGTCATAGTCACAGTCACAATCATAACGGGGTAGGGCTCAGCTTTGACTGGCTCGATCTCATGAGCTTGGAGCAGCAGCAGGAGCAGCAACAAGGCCAGCCTCAACATCCTCCACC : 1287
 CL847 ACE2 : CAGTCATAGTCACAGTCACAATCATAACGGGGTAGGGCTCAGCTTTGACTGGCTCGATCTCATGAGCTTGGAGCAGCAGCAGGAGCAGCAACAAGGCCAGCCTCAACATCCTCCACC : 1287
 CAGTCATAGTCACAGTCACAATCATAACGGGGTAGGGCTCAGCTTTGACTGGCTCGATCTCATGAGCTTGGAGCAGCAGCAGGAGCAGCAACAAGGCCAGCCTCAACATCCTCCACC

QM6a ACE2 : ACCAGTCCA AACCCCTCTCCGAACGGCTGGCAGCTCTTCTGTGTGCCCTGGACCGCATGCTCCAAGCCGTACCCTCATCCCTCGACATGCATCAGCTCTCAAGGCAGCAGCTGAGAGA : 1404
 CL847 ACE2 : ACCAGTCCA AACCCCTCTCCGAACGGCTGGCAGCTCTTCTGTGTGCCCTGGACCGCATGCTCCAAGCCGTACCCTCATCCCTCGACATGCATCAGCTCTCAAGGCAGCAGCTGAGAGA : 1404
 ACCAGTCCA AACCCCTCTCCGAACGGCTGGCAGCTCTTCTGTGTGCCCTGGACCGCATGCTCCAAGCCGTACCCTCATCCCTCGACATGCATCAGCTCTCAAGGCAGCAGCTGAGAGA

QM6a ACE2 : GTACGCCGACACCGTGGGAACCGGCTTCGACCTGCAGTCCACCCTCGACAGCCTCCACCACGCCAGGATCTCGCCTCCCTCTATTCCGAGGCCGTACCCGCCTCGTTCAACAA : 1521
 CL847 ACE2 : GTACGCCGACACCGTGGGAACCGGCTTCGACCTGCAGTCCACCCTCGACAGCCTCCACCACGCCAGGATCTCGCCTCCCTCTATTCCGAGGCCGTACCCGCCTCGTTCAACAA : 1521
 GTACGCCGACACCGTGGGAACCGGCTTCGACCTGCAGTCCACCCTCGACAGCCTCCACCACGCCAGGATCTCGCCTCCCTCTATTCCGAGGCCGTACCCGCCTCGTTCAACAA

QM6a ACE2 : GCGCACAACCGCTGCCGAGGCCGACGCCCTCTGTGCCGTTCGGACTGCGTACACCAGGACCGCACCTCGTTGCACACGACGCGGTTGCCCAAGCTGGACCACGCCCTGTTGAACCT : 1638
 CL847 ACE2 : GCGCACAACCGCTGCCGAGGCCGACGCCCTCTGTGCCGTTCGGACTGCGTACACCAGGACCGCACCTCGTTGCACACGACGCGGTTGCCCAAGCTGGACCACGCCCTGTTGAACCT : 1638
 GCGCACAACCGCTGCCGAGGCCGACGCCCTCTGTGCCGTTCGGACTGCGTACACCAGGACCGCACCTCGTTGCACACGACGCGGTTGCCCAAGCTGGACCACGCCCTGTTGAACCT

QM6a ACE2 : CGTCATGGCGTGCCACATCCGTTCTGCTCGATGTCATGGACACTCTCGCAGAGCACGGGCGGATGTGCCGCTTTCATGGTGGCCACTCTGCCGCCGACTACGACCCCAAGTTTGGCGT : 1755
 CL847 ACE2 : CGTCATGGCGTGCCACATCCGTTCTGCTCGATGTCATGGACACTCTCGCAGAGCACGGGCGGATGTGCCGCTTTCATGGTGGCCACTCTGCCGCCGACTACGACCCCAAGTTTGGCGT : 1755
 CGTCATGGCGTGCCACATCCGTTCTGCTCGATGTCATGGACACTCTCGCAGAGCACGGGCGGATGTGCCGCTTTCATGGTGGCCACTCTGCCGCCGACTACGACCCCAAGTTTGGCGT

QM6a ACE2 : TCCTGAGATACGGGTGGGCACTTTTGTGCGCCCTACCGATACGGCTGCCTCAATGCTGCTCTCTGTTGTGCTGGAGCTTCAAACCGTGTGGTGGCGAGGGTCAAGGACTTGGTGGC : 1872
 CL847 ACE2 : TCCTGAGATACGGGTGGGCACTTTTGTGCGCCCTACCGATACGGCTGCCTCAATGCTGCTCTCTGTTGTGCTGGAGCTTCAAACCGTGTGGTGGCGAGGGTCAAGGACTTGGTGGC : 1872
 TCCTGAGATACGGGTGGGCACTTTTGTGCGCCCTACCGATACGGCTGCCTCAATGCTGCTCTCTGTTGTGCTGGAGCTTCAAACCGTGTGGTGGCGAGGGTCAAGGACTTGGTGGC

QM6a ACE2 : CATGGTTGACCAGGTGAAGGATGATGCAAGAGCGGGCAGAGAAGCAAAGGTCGTTCCGCTGTCAGTGTGGGATTTTACTGGAACGAGCCGAGTCGACGCTTGGAGAGTGGTCCAGGTT : 1989
 CL847 ACE2 : CATGGTTGACCAGGTGAAGGATGATGCAAGAGCGGGCAGAGAAGCAAAGGTCGTTCCGCTGTCAGTGTGGGATTTTACTGGAACGAGCCGAGTCGACGCTTGGAGAGTGGTCCAGGTT : 1989
 CATGGTTGACCAGGTGAAGGATGATGCAAGAGCGGGCAGAGAAGCAAAGGTCGTTCCGCTGTCAGTGTGGGATTTTACTGGAACGAGCCGAGTCGACGCTTGGAGAGTGGTCCAGGTT

QM6a ACE2 : CAAGGACGGGCTGGTCAAGTCCAGACTGCTGAAGTGAGTCTCTCATATGGACGGCAACGGTGAGGGATTTGGCAGTTCCTACAGTGAACAGTGAATGCCGGGACATGTTCTGCACGA : 2106
 CL847 ACE2 : CAAGGACGGGCTGGTCAAGTCCAGACTGCTGAAGTGAGTCTCTCATATGGACGGCAACGGTGAGGGATTTGGCAGTTCCTACAGTGAACAGTGAATGCCGGGACATGTTCTGCACGA : 2106
 CAAGGACGGGCTGGTCAAGTCCAGACTGCTGAAGTGAGTCTCTCATATGGACGGCAACGGTGAGGGATTTGGCAGTTCCTACAGTGAACAGTGAATGCCGGGACATGTTCTGCACGA

QM6a ACE2 : CTAGAACTCGGATATGATGTCTGATCAGCCGGAAACGAGGGGCGACAGCTCTAACGCAGTACGGGGCTGATGCTCAAGGAGCAGCCGAGCCTGCATGTACTTCCGACGTTACCT : 2223
 CL847 ACE2 : CTAGAACTCGGATATGATGTCTGATCAGCCGGAAACGAGGGGCGACAGCTCTAACGCAGTACGGGGCTGATGCTCAAGGAGCAGCCGAGCCTGCATGTACTTCCGACGTTACCT : 2223
 CTAGAACTCGGATATGATGTCTGATCAGCCGGAAACGAGGGGCGACAGCTCTAACGCAGTACGGGGCTGATGCTCAAGGAGCAGCCGAGCCTGCATGTACTTCCGACGTTACCT

QM6a ACE2 : TACCTACATGCAGTGATTGGCCGGGCTCTTCGATCGTCTCTGGGGCAGCATCCTCTTCTCTAACCCCTCTTGCGAAGCACTTACAGCTCATGCTTGGATGTGGATATCGGAATCTGA : 2340
 CL847 ACE2 : TACCTACATGCAGTGATTGGCCGGGCTCTTCGATCGTCTCTGGGGCAGCATCCTCTTCTCTAACCCCTCTTGCGAAGCACTTACAGCTCATGCTTGGATGTGGATATCGGAATCTGA : 2340
 TACCTACATGCAGTGATTGGCCGGGCTCTTCGATCGTCTCTGGGGCAGCATCCTCTTCTCTAACCCCTCTTGCGAAGCACTTACAGCTCATGCTTGGATGTGGATATCGGAATCTGA

QM6a_ACE2 : ACAGAACTTGTGCCGTCATGGGCTCGCGAACCAGCCACCTGCATGTCCTTGAAGTCATCAAAAACAGACGAGCTCGACACGGCTCCACCCTCCCTATGCTTCGGCGACTTACACAA : 2457
 CL847_ACE2 : ACAGAACTTGTGCCGTCATGGGCTCGCGAACCAGCCACCTGCATGTCCTTGAAGTCATCAAAAACAGACGAGCTCGACACGGCTCCACCCTCCCTATGCTTCGGCGACTTACACAA : 2457
 ACAGAACTTGTGCCGTCATGGGCTCGCGAACCAGCCACCTGCATGTCCTTGAAGTCATCAAAAACAGACGAGCTCGACACGGCTCCACCCTCCCTATGCTTCGGCGACTTACACAA

QM6a_ACE2 : GGAGATAACATCCACTATCGCCGTCCTCAACAGCAAATCAAGGCCGCCCGAGGCTCCATATCGAAGAGATCAAGATTTGAAGCTGTGCCACCCTCTTTCGCACTTGTAGTTGCTCCGGT : 2574
 CL847_ACE2 : GGAGATAACATCCACTATCGCCGTCCTCAACAGCAAATCAAGGCCGCCCGAGGCTCCATATCGAAGAGATCAAGATTTGAAGCTGTGCCACCCTCTTTCGCACTTGTAGTTGCTCCGGT : 2574
 GGAGATAACATCCACTATCGCCGTCCTCAACAGCAAATCAAGGCCGCCCGAGGCTCCATATCGAAGAGATCAAGATTTGAAGCTGTGCCACCCTCTTTCGCACTTGTAGTTGCTCCGGT

QM6a_ACE2 : GAAGATCTCCACCTCATGCAACGGAGACGCAGCTTTGGCGAGGTTACGAATAGGGAGGATCTGTGCGTGGTGTGTGTACCAACTCTGTCCTCCACAGAGACGATACGTGTAACACA : 2691
 CL847_ACE2 : GAAGATCTCCACCTCATGCAACGGAGACGCAGCTTTGGCGAGGTTACGAATAGGGAGGATCTGTGCGTGGTGTGTGTGTACCAACTCTGTCCTCCACAGAGACGATACGTGTAACACA : 2691
 GAAGATCTCCACCTCATGCAACGGAGACGCAGCTTTGGCGAGGTTACGAATAGGGAGGATCTGTGCGTGGTGTGTGTGTGTACCAACTCTGTCCTCCACAGAGACGATACGTGTAACACA

QM6a_ACE2 : GCTGCTATTCGGCCGTCGCTGTTGTCCGTTTTTGTCAAACCTCCGCGGTGCTGCTGTGACTGGTGACCATATTGGTTGTTGTTTCGAAGAGAAGCCTCGTTTTTGTCTCGTGTACTATTA : 2808
 CL847_ACE2 : GCTGCTATTCGGCCGTCGCTGTTGTCCGTTTTTGTCAAACCTCCGCGGTGCTGCTGTGACTGGTGACCATATTGGTTGTTGTTTCGAAGAGAAGCCTCGTTTTTGTCTCGTGTACTATTA : 2808
 GCTGCTATTCGGCCGTCGCTGTTGTCCGTTTTTGTCAAACCTCCGCGGTGCTGCTGTGACTGGTGACCATATTGGTTGTTGTTTCGAAGAGAAGCCTCGTTTTTGTCTCGTGTACTATTA

QM6a_ACE2 : GTAGTAGTATGTCCGAAGAGATGTTGAATGGCACGGACAGGCATCATCGTACAGAGCACAAAGTAATACGGAGCACAAACGTGAACGTACATGTATTTATTCGGTAACAAGCGTTTT : 2925
 CL847_ACE2 : GTAGTAGTATGTCCGAAGAGATGTTGAATGGCACGGACAGGCATCATCGTACAGAGCACAAAGTAATACGGAGCACAAACGTGAACGTACATGTATTTATTCGGTAACAAGCGTTTT : 2925
 GTAGTAGTATGTCCGAAGAGATGTTGAATGGCACGGACAGGCATCATCGTACAGAGCACAAAGTAATACGGAGCACAAACGTGAACGTACATGTATTTATTCGGTAACAAGCGTTTT

QM6a_ACE2 : GCTACAATATGTCATGCCAGCATCACGTACCGAGTACTAGTATGCAGTACATGACTGCCACGTCCTCACAAACAGCCGTGTCGAAGCATTCCTCAGGGGAT : 3026
 CL847_ACE2 : GCTACAATATGTCATGCCAGCATCACGTACCGAGTACTAGTATGCAGTACATGACTGCCACGTCCTCACAAACAGCCGTGTCGAAGCATTCCTCAGGGGAT : 3026
 GCTACAATATGTCATGCCAGCATCACGTACCGAGTACTAGTATGCAGTACATGACTGCCACGTCCTCACAAACAGCCGTGTCGAAGCATTCCTCAGGGGAT

7.3 Publications

- Jourdren, L., Duclos, A., Brion, C., **Portnoy, T.**, Mathis, H., Margeot, A., and Le Crom, S. (2010). Teolenn: an efficient and customizable workflow to design high-quality probes for microarray experiments. *Nucleic Acids Res.*
- Portnoy, T.**, Margeot, A., Linke, R., Atanasova, L., Fekete, E., Sandor, E., Hartl, L., Karaffa, L., Druzhinina, I.S., Seiboth, B., *et al.* (2011a). The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. *BMC Genomics* **12**, 269.
- Portnoy, T.**, Margeot, A., Seidl-Seiboth, V., Le Crom, S., Ben Chaabane, F., Linke, R., Seiboth, B., and Kubicek, C.P. (2011b). Differential Regulation of the Cellulase Transcription Factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* Strains Producing High and Low Levels of Cellulase. *Eukaryotic Cell* **10**, 262-271.

Teolenn: an efficient and customizable workflow to design high-quality probes for microarray experiments

Laurent Jourden¹, Aurélie Duclos¹, Christian Brion^{1,2}, Thomas Portnoy^{1,2},
Hugues Mathis², Antoine Margeot² and Stéphane Le Crom^{1,*}

¹Institut de Biologie de l'École normale supérieure (IBENS), Institut National de la Santé et de la Recherche Médicale U1024, Centre National de la Recherche Scientifique UMR8197, 75005 Paris and ²IFP, Département Biotechnologie, 1-4 Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France

Received October 14, 2009; Accepted February 8, 2010

ABSTRACT

Despite the development of new high-throughput sequencing techniques, microarrays are still attractive tools to study small genome organisms, thanks to sample multiplexing and high-feature densities. However, the oligonucleotide design remains a delicate step for most users. A vast array of software is available to deal with this problem, but each program is developed with its own strategy, which makes the choice of the best solution difficult. Here we describe Teolenn, a universal probe design workflow developed with a flexible and customizable module organization allowing fixed or variable length oligonucleotide generation. In addition, our software is able to supply quality scores for each of the designed probes. In order to assess the relevance of these scores, we performed a real hybridization using a tiling array designed against the *Trichoderma reesei* fungus genome. We show that our scoring pipeline correlates with signal quality for 97.2% of all the designed probes, allowing for a posteriori comparisons between quality scores and signal intensities. This result is useful in discarding any bad scoring probes during the design step in order to get high-quality microarrays. Teolenn is available at <http://transcriptome.ens.fr/teolenn/>.

INTRODUCTION

The development of high-throughput sequencing technologies challenges the classical application fields conquered by microarrays during the last decade (1).

To face this new competition, microarray providers concentrate their efforts on the improvement in feature density and sample multiplexing. Thus, biochips became flexible enough to face a range of applications not yet easily available with high-throughput sequencing. These applications are clearly complementary to the ones found with sequencing such as genotyping (2). Microarrays are still attractive to perform comparative genomic hybridization (CGH) or high-resolution transcriptome analyses on small microorganism genomes. Affordable multiplexing arrays are also of great interest for system biology approaches where kinetic experiments are needed to create dynamic models.

However, the oligonucleotide design step is a real bottleneck for all these applications. Despite all the existing probe-design software, this step remains complex and is considered something of a black box for many microarray users. Designing probes for microarrays requires dealing with several parameters, such as sensitivity and specificity. Sensitivity is defined by the strength with which a probe binds to its target sequence. It influences the level of the signal read from the microarray and the relevance of the obtained information. Specificity is defined according to the ability of the probe to bind to non-target sequences in the hybridization sample. Cross-hybridization is usually one major source of non-specificity. The selection of the best oligonucleotide design solution is therefore even more difficult (3) since it is often dependent on the expected application. Several solutions that allow the design of tiling arrays with long oligonucleotides are freely available for academics. First, OligoTiler (4) and Lipson *et al.* (5) algorithms optimize the tiling path (succession of oligonucleotides along the genome) in order to obtain the most even distribution of probes. OligoTiler is able to find probes even in repeated regions but offers a limited control on the sensitivity and the specificity of

*To whom correspondence should be addressed. Tel: +33 1 44 32 23 72; Fax: +33 1 44 32 39 88; Email: lecrom@biologie.ens.fr

the designed oligonucleotides. The Lipson *et al.* solution works on a subset of oligonucleotides with an acceptable quality, and then selects probes that ensure the most evenly distributed tiling path. For both these solutions, probe selection is therefore mainly based on probe position and not on their quality. In contrast, Tileomatic (6) and ArrayDesign (7) focus mainly on probe quality; either by selecting the probe of best quality in each window of the tiling path (ArrayDesign), or by using an implementation of the shortest path algorithm (Tileomatic). The specificity calculation, which is necessary to estimate probe quality, is done using either a suffix array approach (Tileomatic) or a uniqueness score calculation based on minimum unique prefix count for each oligonucleotide (ArrayDesign). Both these programs allow the design of oligonucleotides in repeated regions if a large unique overlapping probe can be found. Among these four solutions, only OligoTiler and ArrayDesign are implemented. Each of these software programs runs according to its own properties. Both lack flexibility and are not able to assign probe quality value to each designed probe.

Here, we describe a new flexible and universal probe design solution (Teolenn) based on an open module system approach. This software answers everyone's needs and remains user-friendly. In addition, our design program assigns a quality score to each designed probe so that a posteriori filtering steps and correlations can be made. We compared the results of several tests we performed with the ones obtained from other available solutions. Finally, we designed a microarray against a fungal genome in order to perform hybridizations; this allowed us to compare the quality scores we calculated with the intensity of the signal obtained from a genomic DNA-hybridization experiment.

MATERIALS AND METHODS

Teolenn probe parameter calculation

After the creation of the library of all possible probes, we used the SOAP software (8) to detect redundant oligonucleotides. Teolenn launched SOAP v1 with a seed size of 12, and a limit of five maximum mismatches on a read. Only probes with a unique match were conserved in the library. Complexity is evaluated using the masked genome by counting the number of masked bases for each probe. T_m values are calculated using the nearest neighbour thermodynamic model (9). The 'uniquesub' function of genome tools (7) is used to calculate the uniqueness of each probe in the library.

Trichoderma reesei probe design

We downloaded the unmasked fasta file of the *T. reesei* genome v.2.0 from the Department of Energy Joint Genome Institute website (JGI): <http://genome.jgi-psf.org/Trire2/Trire2.home.html>. We designed a *T. reesei* tiling array with 60 mer oligonucleotides (oligo length) each 150 bp (oligo distance) using OligoTiler from its web interface (<http://tiling.gersteinlab.org/OligoTiler/oligotiler.cgi>). The advanced parameters

were set up as follows: 'IR region' = 5, 'IR require' = 3, 'repeat region overlap' = 4.

ArrayDesign software was retrieved from the author website (<http://www.ebi.ac.uk/~graef/arraydesign/>). We created sequence windows of 150 bp every 149 bp along the 87 scaffolds using Exonerate tools (10). Minimal unique prefix was computed with the MAX_PREFIX_LENGTH variable set at 15. Finally, we launched the oligonucleotide selection using the following parameters: minimal uniqueness score = 0, offset to shift window over unit for uniqueness score = 1, T_m value range = 60–80°C, G number cut-off = 15, percent palindromic filter = 40%, maximum number of synthesis cycles allowed = 185. No deviation of probe length was allowed.

The *T. reesei* design was done with Teolenn using a MAX_PREFIX_LENGTH set to 15 for the uniqueness calculation made by genome tools. No filters were applied after probe parameter calculations. In order to obtain the probe quality score, the calculated parameters were weighted as follows: 0.4 for T_m , 0.3 for uniqueness, 0.2 for GC content and 0.1 for complexity. To get final oligonucleotide scores, a weighting of 0.75 was assigned to quality scores, and a weighting of 0.25 was assigned to position scores. We also performed a variable probe length design using the same parameters than above and allowing a four-base variation of the total probe length.

Probe design comparison

We used the Unafold suite (11) to compute the free energies of the most probable secondary structures for each oligonucleotide. We used the melt.pl script with a hybridization temperature fixed to 65°C. We set DNA concentration to 0.00001 M, sodium to 1 M and magnesium to 0 M. The interval between oligonucleotides was calculated by measuring the distance between the first position (start) of two successive oligonucleotides. To calculate the number of designed probes per transcript, we first retrieved the 'Filtered Models' transcript file from the JGI website. We selected all 'exon' features from this file, and we calculated the number of oligonucleotides fully included in each of these annotated exons. For each transcript, we merged all exon information, and retrieved the total number of oligonucleotides.

To estimate Kane's parameters, we launched WU-BLAST (12) on each oligonucleotide with the *T. reesei* reference genome using the following parameters: expectation threshold for reporting database hits = 1.2 (E), seed word length for the ungapped BLAST algorithm = 11 (W), negative penalty score for mismatch nucleotides in the BLASTN search mode = -3 (N), penalty for a gap of length one = 3, per-residue penalty for extending a gap = 3. Gapped alignments were not allowed to be created. From the output files, we counted the number of hits where the percentage of identity exceeded 75% with an alignment length of 60 bp.

Comparative genomic hybridization

The microarray data and the related protocols are available at the GEO web site (<http://www.ncbi.nlm.nih>

.gov/geo/) under accession number: GSE17752. Briefly, chromosomal DNA from *T. reesei* QM6a strain was prepared as described previously (13). Two times 4 μ g of genomic DNA were labelled with Alexa 555 or Alexa 647 using the BioPrime Total aCGH-labelling kit (Invitrogen). The two samples were then mixed together and hybridized according to the oligo aCGH/ChIP-on-Chip hybridization kit (Agilent) on a 244k array ordered from Agilent. The array was read using a GenePix 4000B scanner (Molecular Devices) and signal analysis was done using the GenePix Pro 6.1 software. Data pretreatment was applied on each result file to discard GenePix flag and saturating spots. The data were normalized without background subtraction by the global Lowess method performed with the Goulphar software (14).

The GenePix Pro analysis software flagged 'not found' spots when the 'align blocks' algorithm was not able to locate features on the slide. A spot was labelled as 'detectable' when the raw mean intensities were above the background. The background threshold was calculated by adding 2 SDs to the average intensity of all the 'not found' features. The detection threshold was 6.06 in our CGH experiment.

RESULTS

Software implementation

Teolenn works as a four-step workflow (Figure 1). Each step uses the modular organization we set up and is fully customizable. New parameter calculations or new filters can be added to fit the final user needs.

Step 1—Probe library creation. The first step of the workflow consists of creating a library of all the possible probes along the genome sequence of the organism of interest against which the probes have to be designed (called 'reference'). These probes can be designed either with a fixed or variable length. An analysis of probe redundancy is performed after the library creation, in order to discard all the probes that are found several times in the collection. The analysis is set to search for similarity among probes, allowing several mismatch positions.

Step 2—Probe parameter calculation. For all the probes of the created library, the software calculates the parameters needed to estimate oligonucleotide quality. These calculations are implemented as modules. Any developer can add new parameters to the workflow by programming its own module. In its current version, the workflow can calculate the melting temperature (T_m), the GC percent, the complexity and the uniqueness for each probe. T_m is calculated with the nearest neighbour method (9); complexity is measured using the masked reference genome; and uniqueness is estimated based on GenomeTools (7).

Step 3—Probe filtering. This step allows customization of the workflow in order to obtain the final oligoset that best fits the user's needs. This filtering process is also based on modules and can therefore be conducted in several ways.

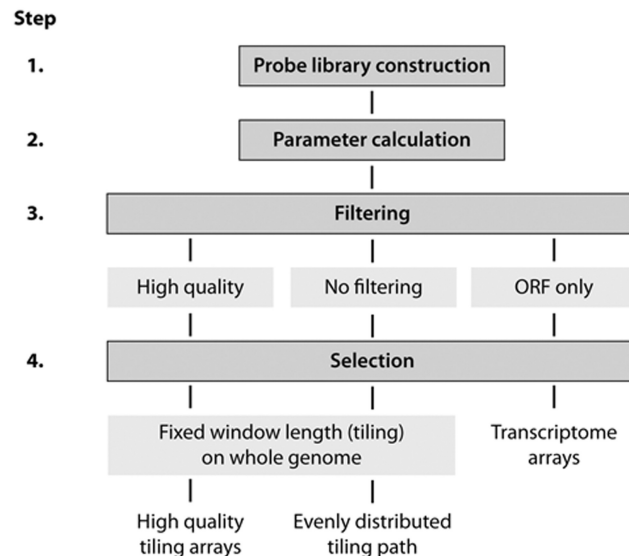


Figure 1. Probe design workflow. The probe design workflow is composed of four steps (see text for details). Here we show several application possibilities depending on the parameters used for the filtering and selection steps. For example, to design a classical transcriptome microarray, the filtering step will keep only the probes from the library that are found in ORFs from genome annotations. Next, the selection step will select the best probes (or several) for each ORF according to user's specifications. At the opposite, one can make a design without any a priori on genome annotation for tiling arrays. On this figure, we show two different ways to design tiling arrays. The first one creates 'high-quality' tiling arrays by filtering low quality probes from the library out (e.g. high or low GC content, non-homogenous melting temperature, etc.). In each window with a fixed length, the best probe will be selected only among the highest quality oligonucleotides of the library. With such a design it is possible that several windows do not get to any corresponding probe. At the opposite, if one wants to favour an even distribution of the tiling path, the constraint on probe filtering can be relaxed in order to keep most of the probes from the library. All the windows will then have a probe selected, though with lower quality parameters.

New filters that give the design workflow new abilities can be set up by developers. For example, the user can apply filtering ranges on the calculated parameters in order to keep only high-quality probes. Conversely, a user who wants to design tiling arrays with a homogenous distribution of probes can keep all the possible probes by using less stringent probe quality filters.

Another possibility may be to set filters that are based on genome annotations in order to select probes that only cover ORFs, small RNAs and so on.

Step 4—Probe selection. The final step of the probe design workflow is dedicated to the selection of the best probe in each window. The idea of a genomic window is very flexible here, as a window can be set up to a fixed or a variable length. A score position is calculated for each oligonucleotide of the selected windows in a way that the most central probe will get the best score.

A final oligonucleotide score is then calculated by combining this position score and the quality score calculated during Step 2. The best probe for each window for microarray design is the one that gets the highest final score. All these score calculations can be

balanced to modify the relative weight of each parameter. Teolenn can output the results in various file formats (plain text, fasta files, GFF or custom output) to provide ready-to-use data. For example, a GFF output in a genome browser like gBrowse (15) exhibits designed probe positions on the genome.

Probe design tests

In order to test our probe selection workflow, we designed a tiling microarray against the genome of *T. reesei* (*Hypocrea jecorina*) (16). Thanks to its ability to degrade plant cell wall polysaccharides efficiently by synthesizing and secreting cellulase enzymes, this fungus is increasingly being investigated in various fields of biotechnology, especially in biofuel production from lignocellulosic biomass. However, it is not available from microarray providers, which makes it a good model for organisms that require the use of probe design software.

We chose to cover the whole genome of *T. reesei* (34 Mb) with a 244 000-feature microarray. This led us to design one oligonucleotide (60 mer each) every 150 bases. The library construction step resulted in 0.35% probes discarded as 33 334 323 probes were produced out of the 33 449 658 possible oligonucleotides. These discarded probes were actually found to be redundant within the genome of *T. reesei* and consequently not suitable for microarrays. We next calculated the parameters needed for the estimation of probe sensitivity and specificity. Figure 2 shows that the distribution of probe parameters was very homogenous within the library. Most of the probes (75.64%) that were found in the library had a melting temperature of between 70 and 80°C; 89.78% of them had a GC percent of between 40 and 70%. This analysis of probe parameter distribution helped to set the thresholds that filter out bad probes from the library.

As we decided to favour a homogenous distribution of probes along the genome, we chose to apply low stringent filters during the filtering process in order to keep the highest possible number of probes. Next, the selection

step was done with the aim of selecting the best probe in all of the 150-base windows along *T. reesei* genome. This phase led to the selection of 222 690 oligonucleotides. Teolenn was able to design probes in 99.87% of all the possible 150-base windows along the *T. reesei* genome.

Comparison with other probe design software

We compared Teolenn results with some other tools available. To this end, the same *T. reesei* tiling design was done using ArrayDesign (7) and OligoTiler (4). Those solutions designed 144 201 (64,67%) and 222 778 (99,91%) probes, respectively. The first and the most striking result from this comparison is that, considering all designed probes with these three solutions, 4333 identical probes were found in two different designs and only 16 probes were common to the three designs.

We then compared the sensitivity of the probes obtained from each design. Figure 3 shows that Teolenn achieved a better combination of homogeneity and high median values regarding GC percent and T_m distribution. Indeed, Teolenn exhibited a median GC percent of $53.3 \pm 0.07\%$ instead of $53.3 \pm 0.1\%$ for OligoTiler and $46.6 \pm 0.07\%$ for ArrayDesign, and a median melting temperature of $74.3 \pm 3.3^\circ\text{C}$ instead of $74.3 \pm 5.0^\circ\text{C}$ for OligoTiler and $71.2 \pm 3.3^\circ\text{C}$ for ArrayDesign. It is noteworthy that no limitation was fixed during this calculation using Teolenn, contrary to ArrayDesign that exhibits no upper outlier (Figure 3A and B). In addition, we compared these results with the ones obtained using a variable probe length design. The T_m distribution of probes (Supplementary Figure S1) shows that we achieved only a slightly better T_m homogeneity median with variable probe length ($74.35 \pm 3.23^\circ\text{C}$) than with the fixed length design ($74.29 \pm 3.30^\circ\text{C}$).

The secondary structure minimum free energy tells us how much a probe is available for interaction. A low free energy creates strong secondary structures, which makes the probes unable to bind their target. As shown in Figure 3C, ArrayDesign obtained the narrowest and

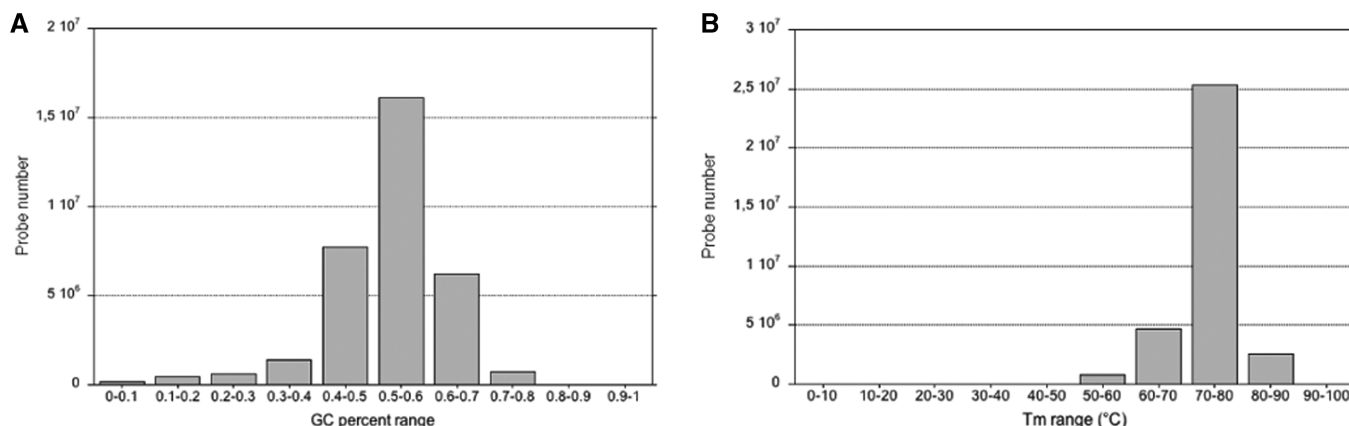


Figure 2. Library probes parameter distribution. Distribution of two parameters calculated for each probe of the oligonucleotide library. (A) Distribution of GC content for all possible probes from the library. The GC percent range is from 0 to 1 on the x-axis and the total number of probes in each category is shown on the y-axis. (B) Distribution of the melting temperature for all possible probes from the library. The T_m range in $^\circ\text{C}$ is displayed from 0 to 100 on the x-axis and the total number of probes in each category is shown on the y-axis.

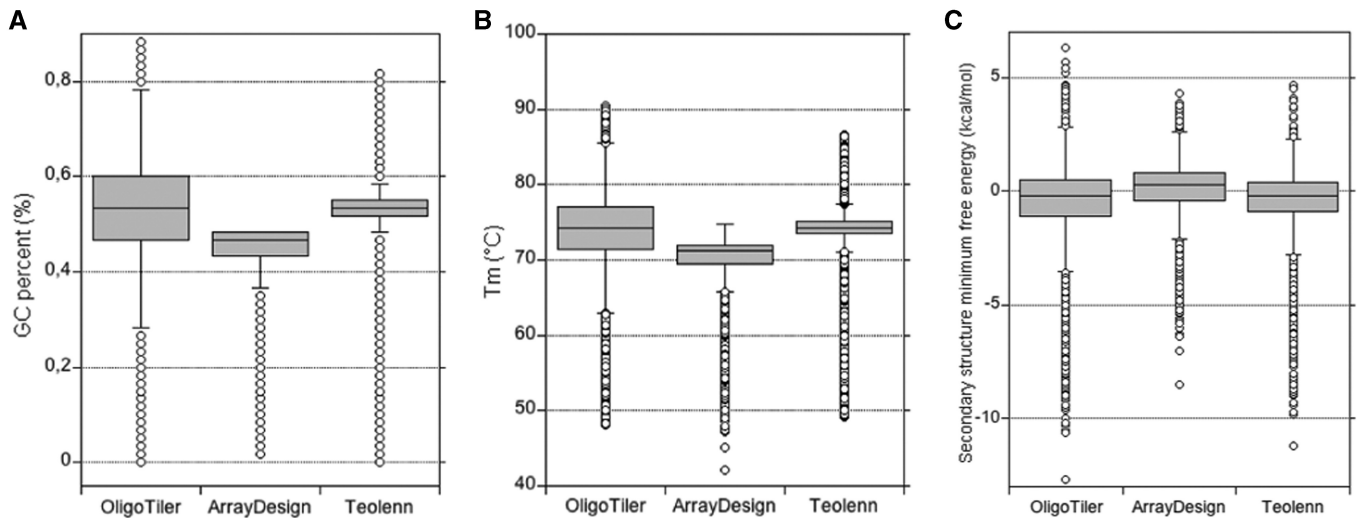


Figure 3. Comparison of the sensitivity of probe sets designed with OligoTiler, ArrayDesign and Teolenn software. For each probe set designed, boxplots show the distribution of (A) the GC percent, (B) the melting temperature (T_m) in °C and (C) the secondary structure free energy in kcal/mol.

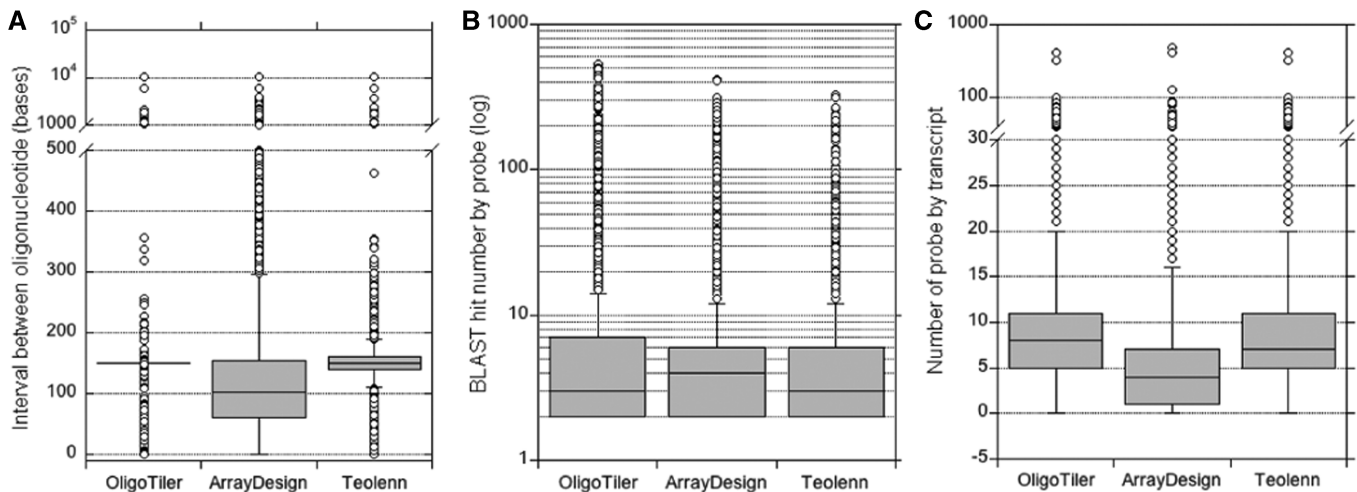


Figure 4. Comparison of the specificity of probe sets designed with OligoTiler, ArrayDesign and Teolenn software. (A) Distribution of the distance in base pairs between two consecutive oligonucleotides. (B) Distribution of the number of BLAST hits by oligonucleotide using the first Kane parameter (see 'Materials and Methods' section for details). (C) Distribution of the number of designed probes per annotated transcript in the reference genome.

highest free energy distribution of the three designed probe sets with a median of $0.3 \text{ kcal mol}^{-1}$. OligoTiler and Teolenn exhibited very similar distributions (median of $-0.2 \text{ kcal mol}^{-1}$); Teolenn had a lower inter-quartile range ($1.3 \text{ kcal mol}^{-1}$) than OligoTiler ($1.6 \text{ kcal mol}^{-1}$).

We compared the respective distribution of the probes along the genome obtained with each design software (Figure 4A). Unsurprisingly, OligoTiler achieved the best even distribution of probes, since this program was created to optimize this parameter; the obtained median interval between two consecutive probes was 150 ± 27.8 bases. Teolenn almost reached the same value with a median interval of 150 ± 36.1 bases. This slight difference in standard deviation can be explained by the fact that Teolenn also optimizes probe specificity. Finally, ArrayDesign exhibited a 165 ± 221.3 bases median interval. This result is actually not surprising since ArrayDesign focuses on specificity optimization. This

constraint leads to several windows with no probes selected, which artificially increases the interval between some probes.

The specificity of each probe set was investigated using the first Kane parameter (17). Using wuBLAST (12), we computed the number of hits for each probe that exhibited a full size alignment (60 bp) against the *T. reesei* genome with more than 75% of identity. The number of oligonucleotides with only one hit was slightly greater (97.1%) with Teolenn than with OligoTiler (96.8%) or ArrayDesign (95.8%). Considering only probes with more than one hit (Figure 4B), the median hit number by oligonucleotide was 4 with ArrayDesign and 3 with OligoTiler and Teolenn.

Finally, in order to estimate the coverage of potential biological entities, we measured the number of probes that cover transcript units for each design (Figure 4C). The annotation of *T. reesei* genome (16) lists a number of

9108 transcripts. We found that OligoTiler covered 98.9% of the annotated transcripts with 9.18 probes per transcript on average. Teolenn covered 98.6% of the annotated CDS with a mean of 9.10 probes per transcript, whereas ArrayDesign covered only 78.6% of the known transcripts (3.40 probes per transcript in average).

Analysis of a real hybridization

In order to assess whether our probe scoring strategy is able to reflect oligonucleotide quality, we wanted to evaluate the strength of the interaction between our designed probes and their targets within the context of a real hybridization. Indeed, some approximations (e.g. thermodynamic estimations) are made in order to calculate some of the scoring parameters. For example, the calculation of the free energy of the probe–target duplex refers to free molecules in solutions, whereas probes are linked to a slide surface with microarrays. To this end, we ordered the slide designed using Teolenn, and we performed a CGH using a self-hybridization of the *T. reesei* wild type reference strain (QM6a) against itself. In this context, no biological modifications are supposed to bias the observed results; this experiment is therefore very informative since the signal obtained for each probe is directly linked to the intensity of its interaction with its target. Image analysis and normalization were performed on a slide scan that exhibited no saturating spots, in order to get the highest number of exploitable measures.

We got a detectable signal (see ‘Materials and Methods’ section) for 97.4% of the spots on the array. We plotted the average intensity of each spot against its probe quality score. The linear correlation shown on Figure 5A clearly indicates a direct relationship between intensities and scores, even if the scatter of points is spread for high-scoring oligonucleotides ($R = 0.41$). This chart

shows that, overall, our design strategy works properly, and that Teolenn is able to design high-quality probes. The intensity of scatter points found for oligonucleotide score over 0.6 (97.2% of detected probes) is representative of the classical bell shape distribution of signal intensities on microarrays; most intensities are found between 8 and 12. In addition, 95.8% of the detected features had good quality scores (≥ 0.6) and intensities that were higher than background levels (≥ 8). Oligonucleotides with low scores (< 0.6 , 2.8% of detected probes) mostly exhibited low intensities. Among them, only 901 had an intensity equal or greater than 10, which represents 13% of bad quality scores (< 0.6) and 0.38% of all the detectable probes.

When looking at the distribution of the parameters that were calculated for these ‘bad’ probes, we found that all of them (i.e. T_m , GC percent, uniqueness and complexity) were lower than the ones associated with good oligonucleotides (Figure 5B). Finally, our score calculation was well adapted for 97.2% of the probes. Scoring calculation is also useful, because no linear correlation is detectable if the parameters are taken independently (Supplementary Figure S2).

DISCUSSION

Here we describe Teolenn, a new probe design workflow that was created to be flexible and customizable. To this end, we developed all the calculations, filtering and selection processes as independent modules. This organization is very useful for several reasons. First, the organization in modules allows for activating or disabling each function according to the available resources or to the needs of the user. For example, the complexity calculation can be deactivated if the masked reference genome of the organism is not available; in this case, this parameter would not be taken into account for

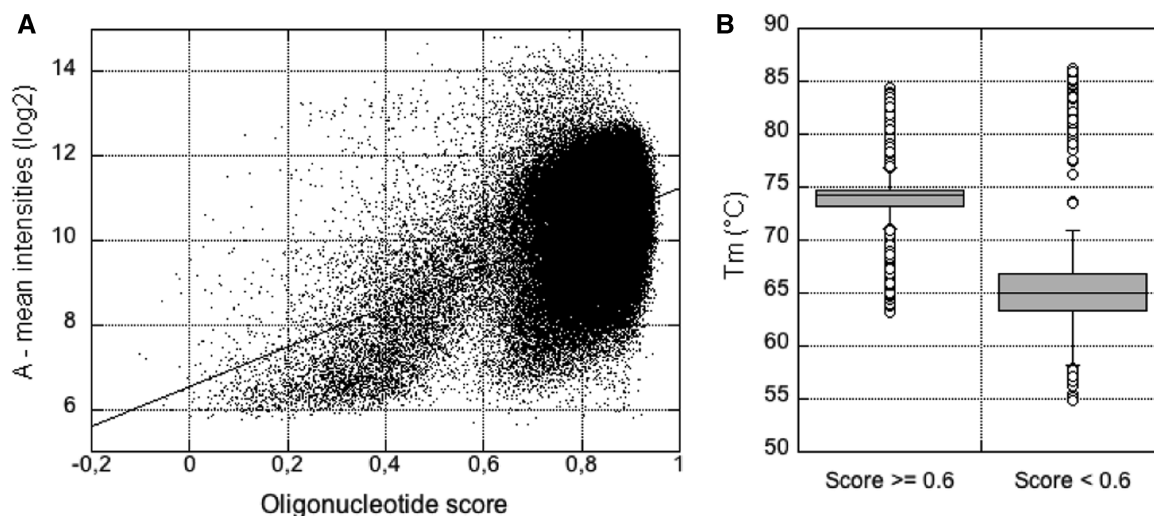


Figure 5. Correlation between oligonucleotide scores and spot intensities. (A) The graph displays for each probe the average of \log_2 intensities (A value) for the corresponding spot as a function of the oligonucleotide score calculated with Teolenn software. The straight line is the linear correlation between the two axes. (B) Distribution of T_m values for two sets of probes with intensities > 10 (see A). Boxplots display the melting temperatures (T_m values) of high-quality scores (≥ 0.6 , left) and low quality ones (< 0.6 , right).

the calculation of the final probe quality score. Second, it is possible for developers to write new modules if needed. For example, the specificity calculation based on uniqueness can be replaced by another cross-hybridization calculation method like BLAST or suffix arrays (18). It is also possible to add modules for calculating the evaluation of secondary structures. Such calculation is time consuming; it is not implemented in the current version of Teolenn.

The flexible construction of Teolenn software fits every level of the probe design pipeline, particularly for the filtering and selection phases. The filtering process is done to influence the output of the probe design directly. Oligonucleotides can be filtered using probe quality. The probe library built at the beginning of the workflow contains all the possible non-redundant oligonucleotides. Without any filtering, a selection of the best probe is done in each window even in low-quality regions (e.g. repeats). This strategy is interesting for the users who want to get the most even distribution of probes along the genome. Conversely, it is possible to filter bad probes of the oligonucleotide library out. This ensures better results after hybridization but induces gaps in the tiling path, since no probes will be designed in AT rich or repeated regions, for example. Furthermore, the quality score and the calculated parameters are available for every probe designed by Teolenn. It is therefore possible to link bad detected signals with oligonucleotide scoring a posteriori, using the results obtained with control experiments (e.g. self-hybridizations). Here, we tested the influence of probe filtering on our *T. reesei* design. If we set up thresholds on T_m , GC percent and uniqueness in order to discard 15% of the worst probes, we were able to design probes in 98.45% of all possible windows compared with 99.87% without any filtering on probe quality. The probe-filtering step can also be used to filter probes according to genome annotations. Such filtering helps to create custom functional genomic microarrays in order to detect mRNA (transcriptome), miRNA, splicing events and so on. In these cases, the filtration step is done according to the coordinates of annotations along the genome. This filtration process can be combined, of course, with probe quality filtering and therefore allows for designing high-quality transcriptome arrays.

Customization can also be done with the probe selection process. Scores for each oligonucleotide are calculated during this last step, using the parameters obtained from Step2 of the design workflow (Figure 1). The major advantage of customization at this stage is that users can set priorities among parameters. Indeed, a weight is assigned to each parameter in order to calculate the global oligonucleotide score. Changing these weights therefore modifies the way the best probe is selected in each window. For example, favouring the position parameter better than probe quality achieves a more even distribution of probes, but induces a wider distribution of sensitivity and specificity values. Setting the priority on melting temperature also makes it possible to obtain an almost isothermal distribution of the T_m in the probe design without modifying the length of each probe. In our example, introducing length flexibility in the design only

led to marginal improvements, allowing an increase of the T_m median of 0.06°C and a decrease of 0.07 of the standard deviation.

Since it depends on the chosen window size, the position score is also calculated during the selection step. Here, the idea of ‘window’ can be interpreted in several ways. For a tiling microarray design, the window has to be a region with a fixed length all along the reference genome. In this case, the best position score is given to the most central oligonucleotide in each selected region. In contrast, for a custom array based on genome annotation (e.g. transcriptome arrays), the window size can be variable and dependant on coding sequences. The user may want to design a limited number of probes per transcript. In this case, the probe orientation is important, and selecting the most central one is not of interest. If only one probe is supposed to be designed, its position score is calculated in reference to either 5'- or 3'-end of the transcript, according to the reverse transcription method used (19). Teolenn is able to deal with all these cases for the design of probe sets whereas usual transcriptome probe design software only offer some of these choices (20–24).

All the parameters set by the user are gathered into one xml file, which allows convenient specification of the user properties during the process. However, developers have the possibility of directly modifying or adding new modules to the Teolenn workflow. Another worthwhile feature of Teolenn is that each step can be launched independently. Consequently, if users want to test the effects of different parameter weights on probe selection, Teolenn can launch only the selection step using results from the previous steps of the workflow, since they were already done. The calculation time for each test is therefore very short. It is noteworthy that the most time-consuming step, probe library construction, has to be done only once for each reference genome. Thus, Teolenn allows for testing parameters more easily than any other available probe design software. Furthermore, this way of working can also be useful for a large functional genomic project. With the same probe library, it is possible to first design a whole genome tiling array against a reference genome and then, when all transcript events have been detected without any a priori on gene annotation, to design a smaller and less expensive custom transcriptome array by filtering probes according to coding sequences. Thus, the flexibility of the Teolenn solution allows for quick and efficient construction of an infinite number of possible probe design solutions.

We designed probes against the genome of the fungus *T. reesei* with OligoTiler, ArrayDesign and Teolenn. Melting temperatures obtained with Teolenn were close to isothermal conditions with high median T_m , even when the oligonucleotide length was fixed. This was obtained without any cut-off on T_m values, contrary to ArrayDesign. Probes with a high T_m may cause saturating signals after scanning, whereas low T_m probes may lead to weak undetectable signals from the background. Since hybridization occurs at the same temperature for all probes during a microarray experiment, obtaining the narrowest T_m distribution helps avoiding low or high signal detection. In addition, Teolenn is able to control

simultaneously an even distribution of the tiling path. Since OligoTiler has been developed for tiling path optimization, Teolenn was not able to reach the same homogeneity level as this software, but it optimizes all sensitivity parameters at the same time, and without any loss of specificity. Indeed, Teolenn got the best result on specificity using the first Kane parameter; close to ArrayDesign, which was developed with the aim of getting the most specific probes along a tiling design.

With Teolenn, the user has permanent access to all the parameters and scores calculated during the probe design process. A posteriori correlations between probe quality score and signal strength on microarrays are therefore possible. Using a self-hybridization control array, we were able to detect a clear correlation between quality and signal intensities. This result validates our probe selection strategy. When investigating the properties of bad quality probes that exhibited a high signal level on our array, we were not able to understand this unexpected behaviour, since their parameters show low sensitivity and specificity for all parameters. This means that other properties that we were not able to evaluate may influence hybridization. However, these probes represented only 0.38% of all the detectable probes on the array.

Scores calculated by Teolenn may also be useful to validate some results after hybridization. Indeed, when working with transcriptome arrays, on which several probes are dedicated to the same transcript, discrepancies between probe results can sometimes be observed for the same transcript. This results could be explained either by a lack of information on transcript annotation (e.g. smaller RNA in the analysed condition, alternate splicing, incomplete reverse transcription resulting in 3' truncated cDNAs), or by the fact that hybridization with low quality probes leads to bad signals. With Teolenn scoring values, it is now possible to differentiate between these two hypotheses.

In conclusion, we developed new flexible and customizable probe design software. Teolenn integrates a probe quality score calculation that is useful for correlations with signal strength after hybridization. The program is based on open modules in order to allow easy access to parameter setting and the development of new abilities. Teolenn probe design software is available at <http://transcriptome.ens.fr/teolenn>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the ENS transcriptome platform staff Corinne Blugeon, Fanny Couplier and Véronique Tanty for their support during hybridization experiments. They also thank Bernd Jagla (Pasteur Institute, Paris) for Teolenn testing on Mac OS.

FUNDING

Réseau National Génopôle (RNG); Infrastructures en Biologie Santé et Agronomie (IBISA). Foundation Tuck 'Enerbio' fund doctorate fellowship to TP. Funding for open access charge: Work Institution.

Conflict of interest statement. None declared.

REFERENCES

- Wold, B. and Myers, R.M. (2008) Sequence census methods for functional genomics. *Nat. Meth.*, **5**, 19–21.
- Shendure, J. (2008) The beginning of the end for microarrays? *Nat. Meth.*, **5**, 585–587.
- Lemoine, S., Combes, F. and Le Crom, S. (2009) An evaluation of custom microarray applications: the oligonucleotide design challenge. *Nucleic Acids Res.*, **37**, 1726–1739.
- Bertone, P., Trifonov, V., Rozowsky, J.S., Schubert, F., Emanuelsson, O., Karro, J., Kao, M.Y., Snyder, M. and Gerstein, M. (2006) Design optimization methods for genomic DNA tiling arrays. *Genome Res.*, **16**, 271–281.
- Lipson, D., Yakhini, Z. and Aumann, Y. (2007) Optimization of probe coverage for high-resolution oligonucleotide aCGH. *Bioinformatics*, **23**, e77–e83.
- Schliep, A. and Krause, R. (2007) *Algorithms in Bioinformatics*, Vol. 4645. Springer Berlin, Heidelberg, Heidelberg, pp. 383–394.
- Graf, S., Nielsen, F.G., Kurtz, S., Huynen, M.A., Birney, E., Stunnenberg, H. and Flicek, P. (2007) Optimized design and assessment of whole genome tiling arrays. *Bioinformatics*, **23**, i195–i204.
- Li, R., Li, Y., Kristiansen, K. and Wang, J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics*, **24**, 713–714.
- SantaLucia, J. Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
- Slater, G.S. and Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.
- Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3–31.
- Chao, K.M., Pearson, W.R. and Miller, W. (1992) Aligning 2 sequences within a specified diagonal band. *Comp. Applications Biosci.*, **8**, 481–487.
- Seidl, V., Gamauf, C., Druzhinina, I.S., Seiboth, B., Hartl, L. and Kubicek, C.P. (2008) The *Hypocrea jecorina* (Trichoderma reesei) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome. *BMC Genomics*, **9**, 327.
- Lemoine, S., Combes, F., Servant, N. and Le Crom, S. (2006) Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics*, **7**, 467.
- Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
- Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D. *et al.* (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.*, **26**, 553–560.
- Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D. and Madore, S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res.*, **28**, 4552–4557.
- Manber, U. and Myers, G. (1993) Suffix arrays – a new method for online string searches. *SIAM J. Computing*, **22**, 935–948.
- Tomiuk, S. and Hofmann, K. (2001) Microarray probe selection strategies. *Brief Bioinform.*, **2**, 329–340.
- Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B. and DeRisi, J.L. (2003) Expression profiling of the schizont and

- trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol.*, **4**, R9.
21. Li, F. and Stormo, G.D. (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics*, **17**, 1067–1076.
 22. Nordberg, E.K. (2005) YODA: selecting signature oligonucleotides. *Bioinformatics*, **21**, 1365–1370.
 23. Rouillard, J.M., Zuker, M. and Gulari, E. (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.*, **31**, 3057–3062.
 24. Wernersson, R. and Nielsen, H.B. (2005) OligoWiz 2.0—integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res.*, **33**, W611–W615.

RESEARCH ARTICLE

Open Access

The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation

Thomas Portnoy^{1,2}, Antoine Margeot¹, Rita Linke³, Lea Atanasova⁴, Erzsébet Fekete⁵, Erzsébet Sándor⁶, Lukas Hartl⁴, Levente Karaffa⁵, Irina S Druzhinina⁴, Bernhard Seiboth⁴, Stéphane Le Crom² and Christian P Kubicek^{4*}

Abstract

Background: The identification and characterization of the transcriptional regulatory networks governing the physiology and adaptation of microbial cells is a key step in understanding their behaviour. One such wide-domain regulatory circuit, essential to all cells, is carbon catabolite repression (CCR): it allows the cell to prefer some carbon sources, whose assimilation is of high nutritional value, over less profitable ones. In lower multicellular fungi, the C2H2 zinc finger CreA/CRE1 protein has been shown to act as the transcriptional repressor in this process. However, the complete list of its gene targets is not known.

Results: Here, we deciphered the CRE1 regulatory range in the model cellulose and hemicellulose-degrading fungus *Trichoderma reesei* (anamorph of *Hypocrea jecorina*) by profiling transcription in a wild-type and a delta-cre1 mutant strain on glucose at constant growth rates known to repress and de-repress CCR-affected genes. Analysis of genome-wide microarrays reveals 2.8% of transcripts whose expression was regulated in at least one of the four experimental conditions: 47.3% of which were repressed by CRE1, whereas 29.0% were actually induced by CRE1, and 17.2% only affected by the growth rate but CRE1 independent. Among CRE1 repressed transcripts, genes encoding unknown proteins and transport proteins were overrepresented. In addition, we found CRE1-repression of nitrogenous substances uptake, components of chromatin remodeling and the transcriptional mediator complex, as well as developmental processes.

Conclusions: Our study provides the first global insight into the molecular physiological response of a multicellular fungus to carbon catabolite regulation and identifies several not yet known targets in a growth-controlled environment.

Background

Many filamentous fungi have developed a predominantly saprobic lifestyle, in which successful competition with other microorganisms for the limited resources present in the environment is the key for survival. To this end mechanisms evolved that allow a rapid adaptation to changing nutrient conditions. One such wide-domain regulatory circuit is carbon catabolite repression (CCR): it allows the preferred assimilation of carbon sources of high nutritional value over others [1-4]. This is usually achieved

through inhibition of gene expression of enzymes involved in the catabolism of other carbon sources than the preferred ones. In multicellular ascomycetes, the C2H2 type transcription factor CreA/CRE1*, which is related to Mig1/Mig2/Mig3 proteins that mediate glucose repression in *Saccharomyces cerevisiae* [5] and to the mammalian Krox20/Egr and Wilm's tumour proteins [6], has been shown to act as a repressor mediating CCR [7,8]. CreA/CRE1 binds to the promoters of the respective target genes via the consensus motif 5'-SYGGRG-3', whose function *in vivo* has been shown both in *Aspergillus nidulans* and *Trichoderma reesei* [9-12]. Functional CreA/CRE1 binding sites frequently consist of two closely spaced 5'-SYGGRG-3' motifs, and it has been suggested that direct repression would only occur through such double binding

* Correspondence: ckubicek@mail.zserv.tuwien.ac.at

⁴Research Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Technische Universität Wien, Getreidemarkt 9/166, A-1060 Vienna, Austria

Full list of author information is available at the end of the article

sites [10,11]. In addition, phosphorylation of a serine in a conserved short stretch within an acidic domain of *T. reesei* CRE1 has been demonstrated to regulate its DNA binding [13].

Today, a plethora of genes have been shown to be under control of CreA or CRE1 (reviewed in [7]) but the mechanisms triggering regulation by CreA/CRE1 are less well understood. In *A. nidulans*, regulation by CreA can be initiated by several so called “repressing” hexoses, requires their phosphorylation, and is affected by the growth rate [14-16]. Most studies on CCR in fungi have been made with gene model systems where CCR functions in the counteraction of gene induction [7]. In contrast, little information is available about which genes directly respond to a relief from CCR. Since *creA/cre1*-knock out mutants display severe phenotypic changes such as reduced growth, abnormal hyphal morphology and sporulation [17,18], such studies are only possible under carefully controlled conditions. Here we chose to use chemostat cultures on D-glucose as a carbon source at two different growth rates (one repressing and one derepressing [16]) to investigate the genome-wide changes in gene expression in relation to CRE1 function, using a $\Delta cre1$ recombinant mutant strain of *T. reesei* and corresponding control strain.

* Footnote: we accept the gene/protein nomenclature of Sordariomycetes and therefore name the *Trichoderma* CreA orthologue CRE1.

Results

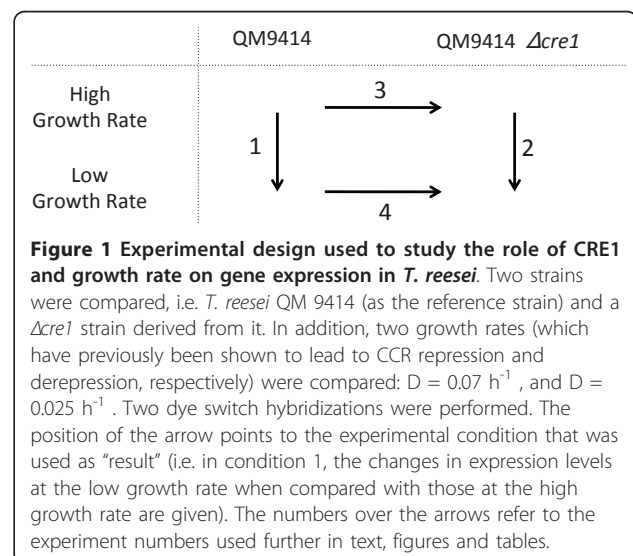
Construction and phenotypic characteristics of a *cre1* knock-out strain of *T. reesei*

We constructed a $\Delta cre1$ recombinant strain of *T. reesei* QM 9414 by replacing its ORF and part of its 5'- and 3'-nt regions by a hygromycin B resistance gene [19] under constitutive expression signals. In accordance with Nakari-Setälä *et al.* [18], the corresponding knock-out strains exhibited a reduced radial growth rate on plates, and formed smaller colonies, fewer aerial hyphae and less spores. In addition, the $\Delta cre1$ strain displayed shorter but more robust hyphae, that contained a considerably thickened cell wall and less septa (Additional File 1: Figure S1 and Table S1). All these mutant phenotypes were eliminated by retransforming the $\Delta cre1$ strain with the *cre1* gene, thus proving that they are *cre1* specific (data not shown). We used phenotype microarrays to analyze whether the above noted reduction in the growth rate is general or specific for some carbon sources only. The results showed that - in contrast to the radial growth on plates - the $\Delta cre1$ strain grew significantly slower on only 5 of 95 carbon sources (Glycogen -37 [\pm 4], $p = 0.004$; arbutin -27 [\pm 4], $p = 0.0006$; adenosine -42 [\pm 5], $p = 0.022$; salicin -34 [\pm 4], $p = 0.007$; and amygdalin -42 [\pm 6], $p = 0.013$), but on the other hand was unaffected on the

majority of them (within $\pm 25\%$ of variation, Chi square test $p > 0.05$; see Additional File 1, Figure S2). However, increased growth of $> 30\%$ of the control, which would be expected if CRE1 represses growth on a given carbon source, was observed for 9 carbon sources shown in Additional File 1, Figure S3 (D-galactose +38 [\pm 3], $p = 0.0008$; L-sorbose +34 [\pm 3], $p = 0.027$; D-xylose +48 [\pm 5], $p = 0.002$; palatinose +31 [\pm 6], $p = 0.033$; maltose +58 [\pm 6], $p = 0.002$; stachyose +45 [\pm 4], $p = 0.03$; xylitol +37 [\pm 4], $p = 0.017$; adonitol +46 [\pm 6], $p = 0.0008$); and glucuronamide +73 [\pm 3], $p = 0.044$).

Wide domain regulation by CRE1: experimental design and properties

We have used whole-genome DNA microarrays with the goal of identifying the main genes in *T. reesei* that are controlled by CRE1. Since carbon catabolite repression is known to be dependent on the growth rate, the microarray experiments were therefore performed in chemostat cultures at two constant growth rates that were earlier shown [16] to be carbon repressing (0.07 h^{-1}) and potentially carbon derepressing (0.025 h^{-1}), respectively (Figure 1). After data pretreatment and normalization, we applied the linear modeling approach and the Bayes statistics implemented in the limma R package [20] to our biological replicates as described in the Methods part. Using these criteria we retrieved a list of 251 genes whose expression was regulated in at least one of the four experimental conditions. One of them was *cre1* itself (in the parent strain), which is trivial and was omitted from all further investigations. The (in part overlapping) occurrence of the truly regulated 250 genes under the four different experimental conditions is shown in Additional File 1, Figure S4. We used clustering algorithms (see Methods) to divide them into 9 different clusters (Figure 2). The effect of CRE1 and its interplay



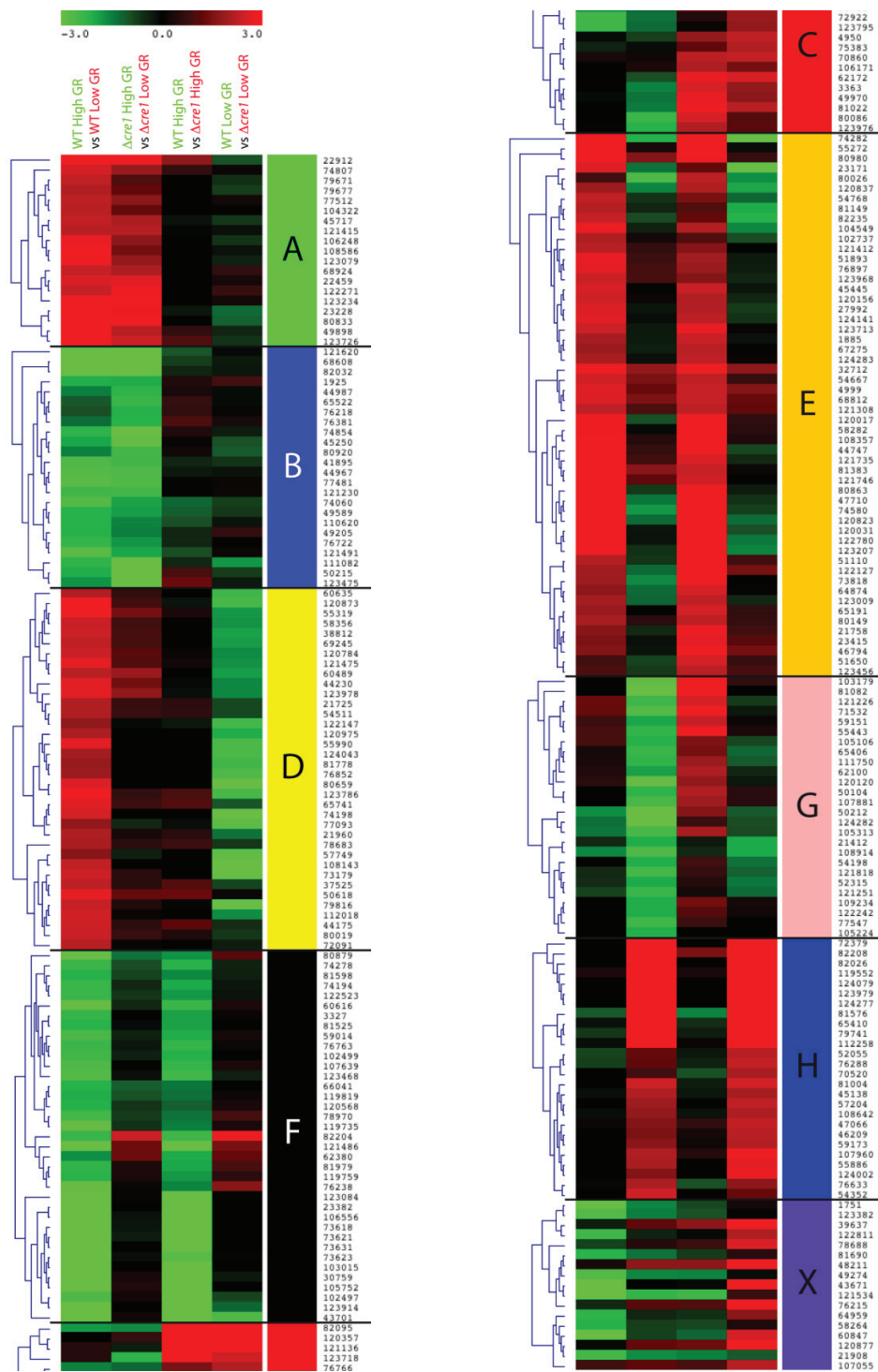


Figure 2 Heat map of strongly expressed genes. Heat map displaying the result from hierarchical clustering of the strongly regulated genes. The colored vertical bars give the letter identifying the respective clusters. Genes contained in the clusters are given by the ID of the encoded proteins. GR: growth rate; wt: QM 9414 wild-type strain.

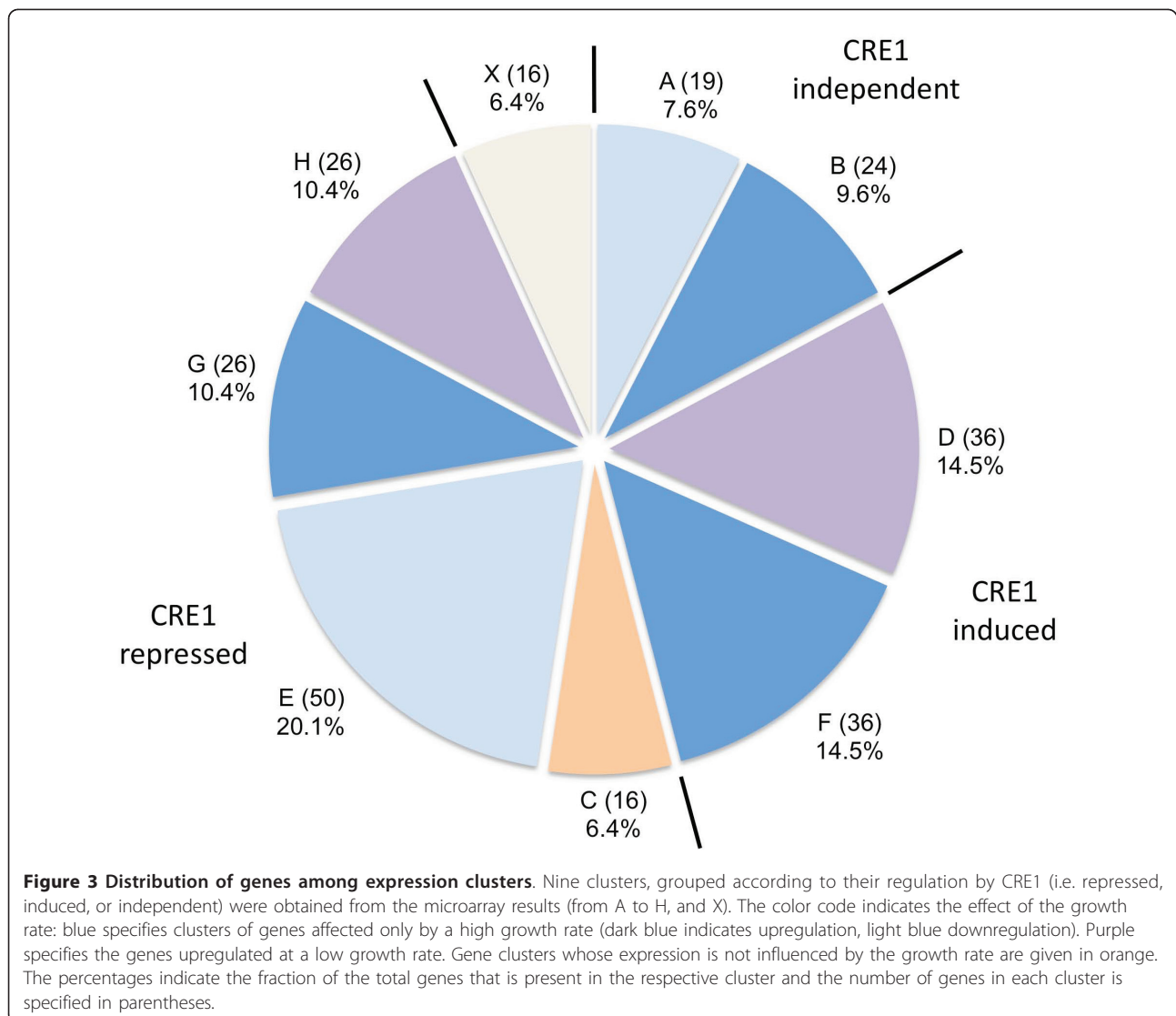
with the growth rate is summarized in Figure 3: 47.3% of the identified genes are in fact repressed by CRE1, but 29.0% are CRE1 induced and 17.2% are CRE1 independent. In addition, 62.2% of the genes are influenced by the high growth rate (34.5% induced and 27.7% repressed), 24.9% are derepressed at the low growth rate and 6.4% are growth rate independent. A complete list of all genes is given in Additional File 1, Table S2. To confirm these microarray results, quantitative Real-Time-PCR (qRT-PCR) was performed on a subset of the genes belonging to different clusters. The gene for isocitrate lyase was included in this set, as its regulation by CRE1-mediated carbon catabolite repression is known ([16], and references therein) and its appearance in our dataset confirms that the growth rates were chosen appropriately. As shown in Additional File 1, Table S3, these genes that showed differential expression between two or more of the used

conditions in the microarray study were also differentially expressed in the same direction upon Real Time-PCR analysis. We therefore conclude that the microarray expression ratios indeed reflect differences in the expression of these genes.

Up and down regulated genes were annotated and categorized as described in the Methods section. The enrichment for the major FunCat categories was assessed in each cluster, and a Gene Ontology (GO) enrichment analysis was also done on the whole dataset as further control (see Additional File 1, Tables S4 and S5).

Genes upregulated in the absence of CRE1 function in *T. reesei*

As shown in Figure 3, genes which were upregulated in the $\Delta cre1$ mutant strain could be grouped into four classes: genes that were upregulated in the CRE1



knockout independently of the growth rate (cluster C, 16 genes); genes upregulated in the CRE1 knockout at high growth rate only (cluster E, 50 genes); genes for which CRE1 function counteracted an induction at high growth rate (cluster G, 26 genes); and genes for which CRE1 function counteracted an induction at low growth rate (cluster H, 26 genes).

Figure 4a shows the distribution of functional categories (FunCat) within these 4 gene clusters. Genes which have orthologues in other fungi but for which no function can be predicted were most abundant in clusters E and G. Genes for carbohydrate degradation were most abundant in the cluster that groups together the genes upregulated in the CRE1 knockout at high growth rate (cluster E). In this cluster, we detected a significant enrichment of genes involved in cellular transport (see Table S4). This category was also significantly enriched in cluster C, which comprises the genes upregulated in the CRE1 knockout independently of the growth rate. The nature of cellular transport proteins was diverse in cluster E, but was dominated by permeases for transport of nitrogenous compounds in cluster C.

The cluster of the genes upregulated in the CRE1 knockout at high growth rate (E) also exhibited genes linked to transcription regulation (e.g. the helicase SNF2 involved in chromatin remodeling and the transcriptional regulator MedA involved in sporulation), as well as PhiA and RAS1, both participating in hyphal development.

This latter function was also represented with the *bys1* transcript in the cluster of genes for which CRE1 function counteracted an induction at high growth rate (cluster G). This cluster comprised different genes linked to carbohydrate metabolism (e.g. glycoside

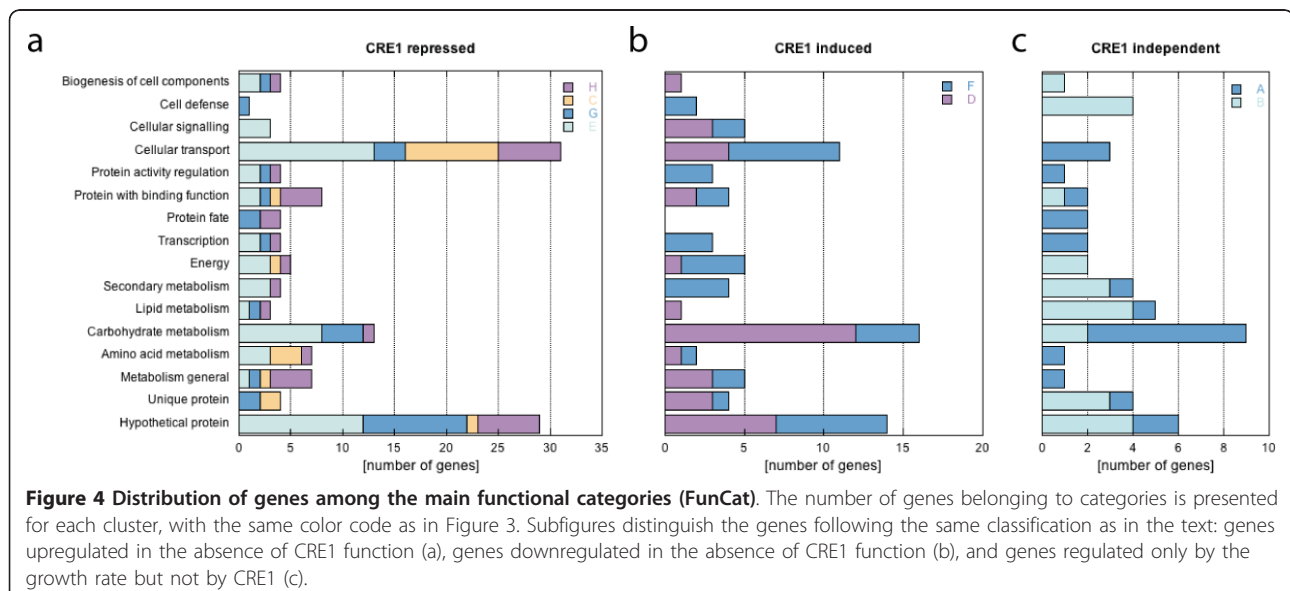
hydrolases and glycoside transferases), as well as *hsp30*, a single integral plasma membrane heat shock protein that is induced by several stresses, including glucose limitation [21].

As found for cluster E, the cluster that groups together genes for which CRE1 function counteracted an induction at low growth rate (cluster H) exhibited a majority of genes involved in cellular transport. A membrane protein linked to sporulation, CMP1, has also been found in this cluster. The significance of the enrichment of these functional categories has been verified using Gene Ontology annotation (Additional File 1, Table S5).

Genes contained in cluster X exhibited a heterogenous pattern of gene expression, but their expression was consistently highest in the $\Delta cre1$ strain at low growth rates, and lowest in the parental strain at high growth rates. We therefore conclude that genes of this cluster are also subject to CRE1-dependent CCR. Gene with putative functions did not reveal any common characteristics. However, we detected the orthologue of *A. nidulans creD* (*cre4*), which encodes a protein with arrestin and PY motifs known to interact with ubiquitin ligase [22]. Disruption of *creD* in *A. nidulans* confers partial relief from CCR [23].

Genes downregulated in the absence of CRE1 function in *T. reesei*

The expression of a smaller number of genes was decreased in the $\Delta cre1$ strain, either only at the low (cluster D, 36 genes) or the high growth rate (cluster F, 36 genes), suggesting that CRE1 acts - either directly or indirectly - positively on these genes. The highest number of genes from the cluster that groups together the



genes downregulated in the CRE1 knockout at high growth rate (F) corresponds to genes that encode proteins involved in cellular transport or proteins for which no function has been characterized (Figure 4b). This cluster also comprised 4 genes for carbohydrate metabolism, 2 synthases and 2 oxidoreductases involved in secondary metabolite formation, as well as 3 transcription factors with Zn2Cys6 and bZIP DNA binding domains. Cluster D, which comprise genes downregulated in the CRE1 knockout at low growth rate, exhibited a significant enrichment of genes encoding carbohydrate active enzymes. This cluster also exhibited a broad composition of other various gene categories, those encoding unknown proteins and compounds involved in cellular transport being most abundant.

Genes regulated only by the growth rate but not by CRE1
 17.2% of the genes detected in this study displayed an expression pattern that was independent of CRE1 function, i.e. their expression was different only between the two growth rates, but not in any direct comparison between the parent and the $\Delta cre1$ mutant strain. 24 of them were activated (cluster B), whereas 19 (cluster A) were actually repressed by a high growth rate. A significant enrichment of genes for carbohydrate metabolism has been detected in the cluster of genes repressed by the high growth rate (cluster A), whereas the group of genes induced at high growth rate (cluster B) comprised the whole set of genes needed to catabolize N-acetylglucosamine to fructose-6-phosphate and exhibited five genes dedicated to fatty acid metabolism (Figure 4c; see also Additional File 1, Tables S4 and S5 for enrichment statistics).

Presence of CRE1-binding sites in the 5'-region of the identified genes

In order to link the diverse regulatory effects of CRE1 identified in this study to a direct gene target, we looked for the CRE1 binding pattern 5'-SYGGRG-3' [10,11] in sequences 1 kb upstream of the coding region of the genes that were up- or down-regulated in the microarray experiments. For each cluster, an average number of sites per gene was calculated and normalized to the average number of sites per gene found in the whole *T. reesei* genome. Since binding of CreA/CRE1 has been described to require two adjacent binding sites, or may be clustered in cis-regulatory element enriched regions (CRERs, [24]), we also looked for the occurrence of direct or inverted repeats (motif pairs). The total number of these "paired motifs" was calculated for each gene, and an average per gene was again measured for each cluster, and then normalized to the average number in the genome. Two spacing distances between the two motifs were tested, which were maximally 20 bp or 50 bp. For a value of 50 bp, the average number of complex motifs per gene for the whole genome was 0.9 while it was 0.5 for a 20 bp distance value. For each motif pair alone, values ranged from 0.1 to 0.5. Significance of the differences in average per cluster compared to the average on the whole genome was assayed with bi-directional Student t-test. The relatively high frequency of the motifs, combined with the possible presence of genuine primary and secondary CreA/CRE1 target, resulted in often high standard deviation, making statistical validation difficult. However, the results (Table 1) show that the CRE1 binding motif, was significantly more abundant in the CRE1 repressed cluster E, and in the CRE1 activated cluster D. This

Table 1 Results from the promoter analysis

		SYGGRG		SAGGGRRGR	
		20 bp distance		50 bp distance	
Average number per gene in whole genome		4.13	0.47	0.9	0.57
	A	-8%	- 66%**	-24%	-17%
	B	- 3%	- 2%	- 7%	17%
	C	26%	33%	11%	34%
Enrichment in clusters	D	+23%**	36%	+48%**	+105%**
	E	+16%**	+58%*	26%	+72%**
	F	- 3%	9%	11%	-15%
	G	1%	- 43%*	- 6%	-26%
	H	1%	+23%*	3%	8%
	X	-16%	- 50%**	-28%	24%

Enrichments compared to whole genome values are shown.

Results of statistical tests (Student t-test) are indicated as follow:

* p-value between 0.05 and 0.1;

** p-value < 0.05. Values in bold are significant positive enrichments, which are predominantly discussed in the text.

observation was less clear for cluster H (genes for which CRE1 function counteracted an induction at low growth rate), while no enrichment was detected for cluster F and G (genes downregulated in the CRE1 knockout at high growth rate, and genes for which CRE1 function counteracted an induction at high growth rate, respectively). While we were not able to statistically validate this, a trend in enrichment is visible for cluster C, which groups together genes that were upregulated in the CRE1 knockout independently of the growth rate. These results suggest that clusters D, E and maybe C are enriched in direct CRE1 targets while clusters F and H are indirect results of the CRE1 regulation. In support of this conclusion, the values obtained for the CRE1 independent genes contained in clusters A and B were in the range of (or even below) the genomic average.

We also looked for the presence of nucleotide motifs that are shared by genes regulated by CRE1, as described in the Methods section. One of the detected motif using the RSA-Tools software (5'-SAGGGRG-3') was indeed significantly enriched in clusters D and E (i.e. the clusters most enriched in CRE1 motifs). This motif is otherwise found at least once at a 57% frequency in all *T. reesei* promoters.

Discussion

The carbon catabolite regulator CreA/CRE1 is one of a few wide-domain master regulators identified in multicellular fungi, and has been shown to govern the repression of genes involved in polysaccharide degradation and in the utilization of ethanol and amino acids as carbon sources [7,8]. However, the major drawback inherent to these studies is that the growth rate was not controlled. Using chemostat cultures, we show here that - depending of the gene considered - the function of *T. reesei* CRE1 can be dependent on the growth rate, thus implying that the physiology of the fungal cell does affect the function of CRE1 in different ways for different genes. In addition, we show that in 29.0% of the detected genes, CRE1 actually activates (rather than represses) their expression. This number is similar to CreA-activated genes found in *A. nidulans*, although the latter investigated only a subset of the genome [25]. Our data therefore illustrate that the function of CRE1 is more complex than previously thought and clearly goes beyond CCR alone. CRE1 should therefore rather be considered a master regulator of carbon metabolism that adjusts gene regulation in relation to the rate of glucose assimilation.

Since the genome of *T. reesei* contains 9,129 predicted genes [26], the 250 highly differentially regulated genes we detected account for 2.8%, which may be considered only a low percentage. However, this study did not detect those genes where CRE1 interferes with the

induction but not with basal transcription (e.g. cellulases and hemicellulases, ethanol catabolism [7], see below); the number of genes actually controlled by CRE1 is therefore probably larger and in many cases carbon-source-dependent. In this context, it is interesting to note that an orthologue of Med2, a component of the mediator complex that is conserved in all eukaryotes [27], was found to be repressed by CRE1. This protein complex mediates signals between enhancer-bound factors (activators) and the core transcriptional machinery. In *S. cerevisiae*, Med2 - together with Gal11 - has been shown to mediate the strongest activations [28], thus making them being logical targets for regulation. The identification of *T. reesei med2* as a CCR-repressed gene may hence be interpreted as an effective means by which the cell controls expression of a broad set of genes when a drop-down in the growth rate renders their induced expression uneconomic.

207 of the genes identified in this study were in fact regulated by CRE1, whereas the other 43 only influenced by the growth rate independently of CRE1 function. 118 of the CRE1-regulated genes appeared to be repressed by CRE1. Besides unknown proteins, genes encoding membrane permeases represent the highest portion indicating that *T. reesei* carbon catabolite repression acts preferentially at the entry of substrates into the cell. Carbon catabolite repression of a high-affinity hexose permease is known in *A. nidulans* [29], and constitutes a mechanism by which fungi can retrieve even traces of high value carbon sources at a high rate. Interestingly, most of the permeases transporting nitrogenous compounds were regulated by CRE1 in a growth-rate independent manner, suggesting that proteins and their degradation products are among the preferred substitutes for fast metabolizable carbohydrates. A similar increase in amino acid uptake upon CCR has been reported for *S. cerevisiae* [30], but the effect was MIG1 independent in this case. Since the natural habitat of *T. reesei* (decaying wood) is poor in nitrogen but also in repressing carbon sources, this mechanism may enable the fungus to recruit available nitrogenous compounds at an enhanced rate.

There is a general believe that the transcription of genes encoding extracellular hydrolases (CAZome) are repressed by CRE1. Although this was in fact shown for several such genes in this study, they constituted only a very minor portion of the *T. reesei* CAZome (200 genes). Particularly, although the regulation of e.g. cellulolytic and xylanolytic enzymes by CRE1 has been demonstrated earlier [10,18], none of these genes was strongly upregulated in the $\Delta cre1$ mutant. These data are in agreement with the findings that CRE1 mediated carbon catabolite repression mainly affects their induction but not their basal expression [31].

We and others have previously shown that loss-of-function of CRE1 or CreA leads to an alteration in nucleosome repositioning upon addition of glucose [32-34]. Our detection of *snf2* as a gene subject to repression by CRE1 at a high growth rate offers an explanation for this finding: Snf2 is a component of the yeast Swi/Snf multisubunit chromatin remodelling complex [35], one of the cellular mechanisms altering chromatin structure by modulating DNA-histone interactions [36]. Swi/Snf is required for the transcriptional regulation of about 5% of the total yeast genome [37] and Snf2 plays an essential role in it by associating with nucleosomes two helical turns from the dyad axis [38].

It has also previously been reported that a loss-of-function mutation in *creA/cre1* leads to an altered morphology and an impairment in sporulation. We identified several genes that could be responsible for the alterations in morphology in the $\Delta cre1$ strain: a central position may be played by RAS1, a small GTPase that was repressed by CRE1 only at high growth rates but not by the growth rate itself. In the fission yeast *Schizosaccharomyces pombe*, Ras1 regulates two distinct pathways: one that controls mating through the Byr2-mitogen-activated protein kinase cascade and one that signals through Scd1-Cdc42 to maintain elongated cell morphology. In the filamentous fungi *Neurospora crassa* and *Aspergillus fumigatus*, the RAS1/RasA orthologue has been reported to regulate morphology, asexual development and cell wall integrity [39,40]. Also, *S. cerevisiae* RAS1 is regulated by glucose [41]. It is therefore possible that RAS1 is responsible for the phenotypic consequences of a *cre1* loss-of-function, and the other genes may only indirectly be affected via *ras1*. In fact, three of the genes associated with morphology or sporulation (encoding PAG1, MedA and a phosphoproteoglycan, respectively) were repressed by CRE1 also at the high growth rate and the effect of CRE1 could therefore be via RAS1. Pag1 encodes a protein associated with protein kinase Cbk1p and that is required for cell morphogenesis and proliferation in *S. cerevisiae* [42]. MedA encodes an orthologue of *Fusarium oxysporum* REN1, *Aspergillus nidulans* MedA and *Magnaporthe grisea* ACR1 [43] that are transcription regulators involved in conidogenesis, and whose loss-of-function leads to abnormal conidiophores and rod-shaped, conidium-like cells. Two other genes (encoding orthologues of the *Blastomyces* yeast phase specific extracellular protein BYS1, and the *A. fumigatus* cell wall protein PhiA) were affected by *cre1* also at the high growth rate. This regulation under conditions where *ras1* is repressed implies that their expression is independent of the CRE1 effects on *ras1* and thus likely direct. Both genes contain a paired CRE1 binding motif in their promoter, which would support this assumption. Our data suggest that the interplay between CRE1 and

RAS1 may be an important factor regulating developmental processes in *T. reesei*.

An investigation of the occurrence of the established CRE1 target sequence 5'-SYGGRG-3' in the regulated gene set confirmed its relevance to CCR, as 2 clusters out of the 6 that were regulated by CRE1 were also enriched in this CRE1 consensus site. In addition, an even higher enrichment was found for motif pairs biased toward CRERs (Cis Regulated Enriched Regions). These findings are in accordance with previous results showing that only a double CRE1 target is functionally *in vivo* [10,11]. In addition, we also identified an additional GC-rich (5'-SAGGGRGR-3') consensus to be overrepresented in CRE1 regulated promoters. This motif has not been previously described in filamentous fungi or any other organism, and we therefore do not know which proteins, if any, bind to it. However, a study on motifs in Mig1/Mig2/Mig3 regulated promoters in *S. cerevisiae* [5] detected a 5'-GGGAGG-3' motif, which is completely covered by the 5'-SAGGGRGR-3' sequence. The authors reasoned that the 5'-GGGAGG-3' motif may bind a transcription factor that regulates genes involved in phosphate metabolism. However, none of the genes of our study that show an enrichment of the 5'-SAGGGRGR-3' sequence in their 5' nontranslated sequences, are involved in phosphate metabolism. Notably, the presence or absence of the 5'-SYGGRG-3' motifs (or any other motif) in the promoters of regulated genes did not correlate with the way they are regulated by CRE1 (i.e. with a up- or a down-regulation).

Conclusions

Carbon catabolite repression by CreA/CRE1 in filamentous fungi has mostly been studied in relation to the utilization of alternative carbon sources of either industrial interest (e.g. plant biomass components), or genetic model systems such as ethanol and proline catabolism [7,8,10,11], but these studies have so far not revealed the impact of this regulator on the physiology of the fungus in its natural environment. Here, we have identified new targets for *T. reesei* CRE1 and also dissected the dependence of their regulation on the rate of growth (equivalent to the nutritional condition) of the fungus. Noteworthy, a predominant effect of CCR seems to act at the transporter level and on the use of nitrogen substrates. Several genes that could explain morphological changes and sporulation behavior were also affected. Additionally, these growth-controlled environments also reflect some industrial conditions that can be used with fungi. The data and genes obtained will be a valuable basis for future attempts towards understanding the role of carbon nutrition for saprobic fungi. Finally, the strategy used in this paper may be useful also in further studies of other wide domain regulators in fungi.

Methods

Fungal strains and cultivation conditions

H. jecorina QM 9414 and the $\Delta cre1$ strain derived from it (see below) were maintained on malt extract agar. Constant-mass, chemostat-type continuous cultivations were performed in a 2.5 l glass bioreactor with a working volume of 2 l, essentially as described earlier [31]. The feeding medium contained 3 g l⁻¹ glucose, a concentration low enough to make the culture carbon-limited. Steady-state of the cultures was established when no changes in biomass dry weight were observed in three successive samples taken over a period of three residence times (= the reciprocal value of the dilution rate). In D-glucose limited cultures, the steady-state biomass concentration was 1.46 ± 0.21 g l⁻¹, irrespective of the dilution rate. The residual steady-state concentrations of D-glucose in the medium were 0.08-0.10 mM. The calculated growth yield (grams of biomass formed per gram of carbon source consumed) was between 46 and 49% for all cultures, which correlates well with our previous studies [34]. Two subsequently achieved, independent steady-states were sampled and analysed for each dilution rate and fungal strain.

Construction of a *cre1* knock-out strain of *T. reesei* QM 9414

A *cre1* deletion vector was constructed by using the double joint PCR technique [44]. Primers used are given in Additional File 1, Table S6. Oligonucleotides cre5'F and Cre5'Rtailhph were used for the promoter region, cre3'R and Cre3'Ftailhph for the terminator region, and M1LHhph and M2LHhph for the hygromycin B expression cassette [45] amplification. The outside primers (cre5'F and cre3'R) were used for the amplification of the whole deletion fragment in a fusion PCR assay. The resulting fragment was subsequently cloned in pGEMT-Easy resulting in p $\Delta cre1$ hph. The *cre1* deletion fragment was released from plasmid p $\Delta cre1$ hph by a *NotI* restriction digest. The respective 5.3 kb fragment was eluted with a QIAGEN Gel Extraction Kit (QIAGEN GmbH) and was used to transform protoplasts of the strain QM 9414 as described by Gruber *et al.* [46]. Transformants were selected on malt extract medium containing hygromycin (50 µg/ml). 40 transformants were selected from the transformation plates and transferred to small malt extract plates containing hygromycin (50 µg/ml). 25 transformants exhibited a stable phenotype and their respective spores were plated on malt extract medium containing hygromycin (50 µg/ml) and Triton X-100 (0.1%) to isolate single spore colonies. The *cre1* loci of the six transformants and the parental strain QM 9414 as a control were amplified with the primers cre5'F and cre3'R using the following PCR conditions: 2 min of

denaturation (94°C), were followed by 32 cycles of 45 s denaturation (94°C), 45 s of annealing (55°C), and 6 min of elongation (72°C), concluded by 7 min at 72°C. Genomic DNA of strain QM 9414 yielded a fragment of 4.3 kb, while homologous insertion of the deletion vector lead to an increase of the fragment size to 5.3 kb fragment in *cre1* deleted strains. For the retransformation of the $\Delta cre1$ strain the *cre1* gene was amplified with oligonucleotides cre1retrafo1 and cre1retrafo2, which are located about 1500 bp upstream and 700 bp downstream of the coding region, respectively. The fragment was cloned into pGEM-T Easy (Promega). Transformation was performed with the *A. nidulans amdS* as marker [47]. Successful retransformation of the $\Delta cre1$ strain with plasmid pcre1amdS was verified by PCR amplification of the *cre1* gene and resulted in colonies, which showed a QM 9414 growth phenotype.

Nucleic acid isolation and microarray hybridizations

Fungal mycelia were harvested by filtration, washed with distilled cold water, frozen and ground under liquid nitrogen. For extraction of genomic DNA, plasmid DNA and RNA, purification kits (Wizard Genomic DNA Purification Kit, PureYield Plasmid Midiprep System and SV Total RNA Isolation System, respectively, all from Promega) were used according to the manufacturer's protocol. Standard methods were used for electrophoresis, blotting and hybridization of nucleic acids. The microarray data and the related protocols are available at the GEO web site (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number: GSE21072. Briefly, the fungal RNAs of each experiment were reverse-transcribed and labelled with Cy3 or Cy5 dye using the indirect labelling procedure and dye-switch on two biological replicates. We then hybridized 1 µg of labelled cDNA with the 1 × 244k *T. reesei* DNA chip manufactured by Agilent and designed using the Teloenn software as described previously [48]. Arrays were read using a GenePix 4000B scanner (Molecular Devices) and signals analysed by the GenePix Pro 6.1 software. Spots when the "align blocks" algorithm was not able to locate features on the slide were flagged "not found". A spot was labelled as "detectable" when the raw mean intensities were above the background. Data pretreatment was applied on each result file to discard GenePix flag and saturating spots. The data were normalized without background subtraction by the global Lowess method performed with the Goulphar software [49]. The background threshold was calculated by adding two standard deviations to the average intensity of all the "not found" features. For each probe the log₂ hybridization ratio was linked to genome annotation coming from the JGI website. The final log₂ ratio for each transcript was obtained by

averaging the “detectable” hybridization values from all probes located inside the coding sequence on the matching strand. Transcripts with no or only one probe marked as “detectable” were discarded from further analysis. For the two biological replicates on each four experiments we applied on the pretreated results the linear modeling approach implemented by *lmFit* (using each replicate as independent variable) and the empirical Bayes statistics implemented by *eBayes* both from the *limma* R package [20]. We selected the list of statistically regulated genes with an adjusted p-value, using the Benjamini-Hochberg multiple test correction, lower than 0.05. Finally we kept as the most highly regulated targets only transcripts with a final log₂ hybridization ratio greater than 2 or lower than -2.

Real Time PCR quantification of *T. reesei* transcripts

DNase treated (DNase I, RNase free; Fermentas) RNA (5 µg), obtained from *T. reesei* QM 9414 and the delta-*cre1* strain grown in chemostat cultures (*vide supra*), was reverse transcribed with the RevertAid™ First Strand cDNA Kit (Fermentas) according to the manufacturer’s protocol with a combination of the provided oligo-dT and random hexamer primers. All real-time RT-PCR experiments were performed on a Bio-Rad (Hercules, CA) iCycler IQ. For the reaction the IQ SYBR Green Supermix (Bio-Rad, Hercules, CA) was prepared for 25 µl assays with standard MgCl₂ concentration (3 mM) and a final primer concentration of 100 nM each. Primers used are given in Additional File 1, Table S7. All assays were carried out in 96-well plates, which were covered with optical tape. The amplification protocol consisted of an initial denaturation step (3 min at 95°C) followed by 40 cycles of denaturation (15 sec at 95°C), annealing (20 sec at 57°C) and elongation (10 sec at 72°C). Determination of the PCR efficiency was performed using triplicate reactions from a dilution series of cDNA (1, 0.1, 10⁻² 10⁻³). Amplification efficiency was then calculated from the given slopes in the IQ5 Optical system Software v2.0. Expression ratios were calculated using REST[®] Software [50].

Biolog phenotype microarray experiments

Growth of *T. reesei* QM 9414 and two *Δcre1* independent transformed clones from it on 95 carbon sources was investigated using Biolog[®] phenotype microarrays essentially as described by Druzhinina *et al.* [51]. Biomass concentration (OD₇₅₀) was measured after 96 h of growth. One-way or main-effect analyses of variance (ANOVAs) were used to compare the growth of selected strains on individual carbon sources. Tukey’s honest significant difference (Tukey HSD) test as implemented in STATISTICA 6.1 was used for post hoc comparisons to detect the contribution of each variable to the main effect of the F test resulting from the ANOVA. The

significance of increased growth on selected carbon sources was tested by the Chi-square test. Only p-values < 0.05 were considered as significant.

Cluster analysis of microarray results

Clustering analysis was done using the MultiExperiment Viewer software [52]. An expression matrix was built from all the genes sorted as strongly regulated for the four conditions we hybridized. From this matrix we selected the expression profiles corresponding to the main behaviour categories we expected which are: CRE1 repressed genes at high growth rate (GR), CRE1 induced genes at high GR, genes induced by CRE1 at low GR, genes repressed by CRE1 at low GR, genes repressed or induced by GR alone, CRE1 repressed genes alone and CRE1 induced genes alone. Then each of these gene expression groups was extended using the Pavlidis Template Matching method [53] fixing the absolute R threshold when all gene included in the group used as template are retrieved. Genes not classified in one of the groups were clustered using the Clustering Affinity Search Technique [54] procedure using the Pearson correlation coefficient and a 0.8 threshold. From the clusters obtained, two of them could be associated to already detected groups. Finally a hierarchical clustering was performed on each expression groups using Euclidian distance and the average linkage method. From this clustering each gene expression group was separated in two subclusters using the two more external nodes of each tree, resulting in clusters A to H and X.

Gene identification and functional prediction

Genes were first identified according to their ID number. In the case of genes that were either poorly or not yet annotated, they were subjected to BLASTX at NCBI, and all hits with E < -50 retrieved and aligned. Proteins with >75% identity and over >90% of the amino acid sequence in other fungal taxa were considered to be their potential orthologues, and this hypothesis tested by phylogenetic analysis. Proteins that fell outside these criteria were termed “unknown proteins”. Proteins with < 25% of identity and less than 75% of the size of the best hits were considered “unique”. Identified proteins were categorized according to the Functional Catalogue (FunCat [55]) using the *T. reesei* genome database implemented in the Pedant server of the Munich Information Center for Protein Sequences [56]. For each of the FunCat categories selected we calculated for each gene cluster the enrichment ratio compared to the whole FunCat annotation and we used the hypergeometric distribution to compute the statistical p-value associated to this enrichment.

We performed a whole enrichment analysis for all the 171 annotated genes (among 250) detected as highly regulated targets using the Gene Ontology annotation of *T. reesei* from the JGI web site. We calculated the significance of

the enrichment ratio using the hypergeometric distribution and the p-value where adjusted for multiple test using the Benjamini-Hochberg FDR correction. There are 4,977 genes annotated with at least one Gene Ontology term among the 9,129 genes from the *T. reesei* genome.

Promoter sequence analysis

All analyses were performed with the RSAT software suite [24]. Promoter sequences were obtained using the “retrieve sequence” algorithm, with -1000 to -1 coordinates input and the “noORF” option unchecked. Motifs were searched using the “DNA Pattern Matching” algorithm, with the “prevent overlapping matches” parameter checked. For a given set of genes (any cluster, or whole genome), the total number of motifs found was collected and an average number of sites per gene was calculated. Finally this number was normalized to the average number of sites in the whole genome. Significance in mean values between cluster and whole genomes groups were assessed using a Student t-test with Microsoft Excel “ttest” function, using bidirectional and heteroscedastic (unequal variances) options. P-values of 0.1 and 0.05 were considered (and indicated in Table 1). For search of new regulation motifs, the “oligo analysis” and “Dyad Analysis” algorithms were used for each cluster or group of clusters. To ensure the selection of statistically relevant motifs, for each condition tested (Dyad, 5, 6, 7 or 8 words length), three different random gene groups of the same size were assessed in parallel (using the “random gene selection” algorithm). Only motifs with scores higher than the highest score obtained with the corresponding control set +0.5 were considered. This usually resulted in a 1.5 score selection threshold. Selected motifs were validated with enrichment in clusters and statistical tests described above.

Additional material

Additional file 1: The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. additional file 1 contains 4 figures (Figure S1 - S4) and 7 tables (Table S1-S7) that complement the results of the main paper: Figure S1: Morphological changes in *Trichoderma reesei* $\Delta cre1$. Figure S2: Effect of *Trichoderma reesei* $cre1$ -knock out on biomass formation on different carbon sources. Figure S3: Growth of *T. reesei* wild-type and $cre1$ -knock out on carbon sources whose utilization is CRE1-repressed. Figure S4: Distribution of gene profiles among experiments. Table S1: Average hyphal and cell wall width of *Trichoderma reesei* QM9414 and $\Delta cre1$ strains. Table S2: Transcripts and encoded proteins identified in this study. Table S3: Quantitative expression patterns determined by qRT-PCR of selected genes. Table S4: Enrichment analysis on FunCat categories. Table S5: Complete enrichment analysis with Gene Ontologies Table S6: Primers used for construction of the *T. reesei* $\Delta cre1$ strain. Table S7: Primers for Real Time quantification of selected genes.

Acknowledgements

This work was supported by grants from the Austrian Science Foundation (FWF P-19143 and FWF P-19421) to CPK and BS, respectively. Work by SLC was supported by the Infrastructures en Biologie Santé et Agronomie (IBISA).

TP is recipient of a Fondation Tuck Enerbio PhD. Fellowship. Work by LK, EF and ES was supported by the Hungarian Scientific Research Fund (OTKA; K67667) and the National Office for Research and Technology (NKTH; A2-2006-0017). EF is supported by a Bolyai János Research Scholarship (BO/00519/09/8).

Author details

¹IFP Energies nouvelles, Département Biotechnologie, 1-4 Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France. ²École normale supérieure, Institut de Biologie de l'ENS, IBENS, Paris, F-75005 France. Inserm, U1024, Paris, F-75005 France. CNRS, UMR 8197, Paris, F-75005 France. ³Austrian Center of Industrial Biotechnology, c/o Institute of Chemical Engineering, Technische Universität Wien, Getreidemarkt 9/166, A-1060 Vienna, Austria. ⁴Research Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Technische Universität Wien, Getreidemarkt 9/166, A-1060 Vienna, Austria. ⁵Department of Biochemical Engineering, Faculty of Science and Technology, University of Debrecen, H-4010, P.O.Box 56, Debrecen, Hungary. ⁶Department of Plant Protection, Faculty of Agriculture and Food Sciences and Environmental Management, University of Debrecen, H-4032 Böszörményi út 138., Debrecen, Hungary.

Authors' contributions

SLC, AM and TP designed and analysed the DNA arrays, EF and LK performed the chemostat cultivations, EK performed the morphological analysis, RL and BS did the Real Time PCR experiments, LA and ISD performed the Biolog Phenotype Microarray experiments. CPK designed the study, annotated the genes found and - together with TP, SLC and AM - wrote the paper. All authors read and approved the final manuscript.

Received: 17 January 2011 Accepted: 27 May 2011

Published: 27 May 2011

References

- Görke B, Stülke J: Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* 2008, **6**:613-624.
- Deutscher J: The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* 2008, **11**:87-93.
- Gancedo JM: Carbon catabolite repression in yeast. *Microbiol Mol Biol Rev* 1998, **62**:334-361.
- Ronne H: Glucose repression in fungi. *Trends Genet* 1995, **11**:12-17.
- Westholm JO, Nordberg N, Murén E, Ameer A, Komorowski J, Ronne H: Combinatorial control of gene expression by the three yeast repressors Mig1, Mig2 and Mig3. *BMC Genomics* 2008, **9**:601.
- Dowzer CEA, Kelly JM: Analysis of the *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. *Mol Cell Biol* 1991, **11**:5701-5709.
- Ruijter GJ, Visser J: Carbon repression in *Aspergilli*. *FEMS Microbiol Letts* 1997, **151**:103-114.
- Felenbok B, Flippin M, Nikolaev I: Ethanol catabolism in *Aspergillus nidulans*: a model system for studying gene regulation. *Progress Nucleic Acid Res Mol Biol* 2001, **69**:149-204.
- Panozzo C, Cornillot E, Felenbok B: The CreA repressor is the sole DNA-binding protein responsible for carbon catabolite repression of the *alcA* gene in *Aspergillus nidulans* via its binding to a couple of specific sites. *J Biol Chem* 1998, **273**:6367-6372.
- Mach RL, Strauss J, Zeilinger S, Schindler M, Kubicek CP: Carbon catabolite repression of *xyn1* (xylanase I-encoding) gene expression in *Trichoderma reesei*. *Mol Microbiol* 1996, **21**:1273-1281.
- Cubero B, Scazzocchio C: Two different, adjacent and divergent zinc finger binding sites are necessary for CRE-mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans*. *EMBO J* 1994, **13**:407-415.
- Takashima S, Iikura H, Nakamura A, Masaki H, Uozumi T: Analysis of Cre1 binding sites in the *Trichoderma reesei* *cbh1* upstream region. *FEMS Microbiol Letts* 1996, **145**:361-366.
- Cziferszky A, Mach RL, Kubicek CP: Phosphorylation positively regulates DNA binding of the carbon catabolite repressor Cre1 of *Hypocrea jecorina* (*Trichoderma reesei*). *J Biol Chem* 2002, **277**:14688-14699.
- Bailey C, Arst HN Jr: Carbon catabolite repression in *Aspergillus nidulans*. *Eur J Biochem* 1975, **51**:573-577.
- Strauss J, Horvath HK, Abdallah BM, Kindermann J, Mach RL, Kubicek CP: The function of CreA, the carbon catabolite repressor of *Aspergillus*

- nidulans*, is regulated at the transcriptional and post-transcriptional level. *Mol Microbiol* 1999, **32**:169-178.
16. Ilyés H, Fekete E, Karaffa L, Fekete E, Sándor E, Szentirmai A, Kubicek CP: **CreA-mediated carbon catabolite repression of β -galactosidase formation in *Aspergillus nidulans* is growth rate dependent.** *FEMS Microbiol Letts* 2004, **235**:147-151.
 17. Shroff RA, O'Connor SM, Hynes MJ, Lockington RA, Kelly JM: **Null alleles of *creA*, the regulator of carbon catabolite repression in *Aspergillus nidulans*.** *Fungal Genet Biol* 1997, **22**:28-38.
 18. Nakari-Setälä T, Paloheimo M, Kallio J, Vehmaanperä J, Penttilä M, Saloheimo M: **Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production.** *Appl Environ Microbiol* 2009, **75**:4853-4860.
 19. Mach RL, Schindler M, Kubicek CP: **Transformation of *Trichoderma reesei* based on hygromycin resistance using homologous expression signals.** *Curr Genet* 1994, **25**:567-570.
 20. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Statistical Applications in Genetics and Molecular Biology* 2004, **3**.
 21. Piper PW, Ortiz-Calderon C, Holyoak C, Cooté P, Cole M: **Hsp30, the integral plasma membrane heat shock protein of *Saccharomyces cerevisiae*, is a stress-inducible regulator of plasma membrane H (+)-ATPase.** *Cell Stress Chaperones* 1997, **2**:12-24.
 22. Kubicek CP, Mikus M, Schuster , Schmolli M, Seiboth B: **Metabolic engineering strategies for improvement of cellulase production by *Hypocrea jecorina*.** *Biotechnol Biofuels* 2009, **2**:19.
 23. Boase NA, Kelly JM: **A role for *creD*, a carbon catabolite repression gene from *Aspergillus nidulans*, in ubiquitination.** *Mol Microbiol* 2004, **53**:929-940.
 24. Thomas-Chollier M, Sand O, Turatsinze JV, Janky R, Defrance M, Vervisch E, Brohée S, van Helden J: **RSAT: regulatory sequence analysis tools.** *Nucleic Acids Res* 2008, **1**:W119-W127.
 25. Mogensen J, Nielsen HB, Hofmann G, Nielsen J: **Transcription analysis using high-density micro-arrays of *Aspergillus nidulans* wild-type and *creA* mutant during growth on glucose or ethanol.** *Fungal Genet Biol* 2006, **43**:593-603.
 26. Martinez D, Berka RM, Henriksen B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmolli M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS: **Genome sequence analysis of the cellulolytic fungus *Trichoderma reesei* (syn *Hypocrea jecorina*) reveals a surprisingly limited inventory of carbohydrate active enzymes.** *Nature Biotechnol* 2008, **26**:553-560.
 27. Taatjes DJ, Tjian R: **Structure and function of CRSP/Med2; a promoter-selective transcriptional coactivator complex.** *Mol Cell* 2004, **14**:675-683.
 28. Balciunas D, Hallberg M, Björklund S, Ronne H: **Functional interactions within yeast mediator and evidence of differential subunit modifications.** *J Biol Chem* 2003, **278**:3831-3839.
 29. MacCabe AP, Miró P, Ventura L, Ramón D: **Glucose uptake in germinating *Aspergillus nidulans* conidia: involvement of the *creA* and *sorA* genes.** *Microbiology* 2003, **149**:2129-2136.
 30. Peter GJ, Düring L, Ahmed A: **Carbon catabolite repression regulates amino acid permeases in *Saccharomyces cerevisiae* via the TOR signaling pathway.** *J Biol Chem* 2006, **281**:5546-5552.
 31. Karaffa L, Fekete E, Gamauf C, Szentirmai A, Kubicek CP, Seiboth B: **D-galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates.** *Microbiology* 2006, **152**:1507-1514.
 32. Zeilinger S, Schmolli M, Pail M, Mach RL, Kubicek CP: **Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction.** *Mol Genet Genomics* 2003, **270**:46-55.
 33. García I, Gonzalez R, Gómez D, Scazzocchio C: **Chromatin rearrangements in the *prnD-prnB* bidirectional promoter: dependence on transcription factors.** *Eukaryot Cell* 2004, **3**:144-56.
 34. Reyes-Dominguez Y, Narendja F, Berger H, Gallmetzer A, Fernandez-Martin R, García I, Scazzocchio C, Strauss J: **Nucleosome positioning and histone H3 acetylation are independent processes in the *Aspergillus nidulans prnD-prnB* bidirectional promoter.** *Eukaryot Cell* 2008, **7**:656-63.
 35. Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA: **Dissecting the regulatory circuitry of a eukaryotic genome.** *Cell* 1998, **95**:717-728.
 36. Becker PB, Horz W: **ATP-dependent nucleosome remodeling.** *Annu Rev Biochem* 2002, **71**:247-273.
 37. Sudarsanam P, Iyer VR, Brown PO, Winston F: **Whole-genome expression analysis of *snf/swi* mutants of *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2000, **97**:3364-3369.
 38. Dechassa ML, Zhang B, Horowitz-Scherer R, Persinger J, Woodcock CL, Peterson CL, Bartholomew B: **Architecture of the SWI/SNF-nucleosome complex.** *Mol Cell Biol* 2008, **28**:6010-6021.
 39. Kanauchi A, Yamashiro CT, Tanabe S, Murayama T: **A ras homologue of *Neurospora crassa* regulates morphology.** *Mol Gen Genet* 1997, **254**:427-432.
 40. Fortwendel JR, Fuller KK, Stephens TJ, Bacon WC, Askew DS, Rhodes JC: ***Aspergillus fumigatus* RasA regulates asexual development and cell wall integrity.** *Eukaryot Cell* 2008, **7**:1530-1539.
 41. Breviaro D, Hinnebusch AG, Dhar R: **Multiple regulatory mechanisms control the expression of the RAS1 and RAS2 genes of *Saccharomyces cerevisiae*.** *EMBO J* 1988, **7**:1805-1813.
 42. Du LL, Novick P: **Pag1p, a novel protein associated with protein kinase Cbk1p, is required for cell morphogenesis and proliferation in *Saccharomyces cerevisiae*.** *Mol Biol Cell* 2002, **13**:503-514.
 43. Ohara T, Inoue I, Namiki F, Kunoh H, Tsuge T: **REN1 is required for development of microconidia and macroconidia, but not of chlamydospores, in the plant pathogenic fungus *Fusarium oxysporum*.** *Genetics* 2004, **166**:113-24.
 44. Yu JH, Hamari Z, Han KH, Seo JA, Reyes-Dominguez Y, Scazzocchio C: **Double-joint PCR: a PCR-based molecular tool for gene manipulations in filamentous fungi.** *Fungal Genet Biol* 2004, **41**:973-81.
 45. Hartl L, Kubicek CP, Seiboth B: **Induction of the *gal* pathway and cellulase genes involves no transcriptional inducer function of the galactokinase in *Hypocrea jecorina*.** *J Biol Chem* 2008, **282**:18654-18659.
 46. Gruber F, Visser J, Kubicek CP, De Graaff L: **Cloning of the *Trichoderma reesei pyrG* gene and its use as a homologous marker for a high-frequency transformation system.** *Curr Genet* 1990, **18**:451-456.
 47. Kelly JM, Hynes MJ: **Transformation of *Aspergillus niger* by the *amdS* gene of *Aspergillus nidulans*.** *EMBO J* 1985, **4**:475-479.
 48. Jourden L, Duclos A, Brion C, Portnoy T, Mathis H, Margeot A, Le Crom S: **Teolenn: an efficient and customizable workflow to design high-quality probes for microarray experiments.** *Nucleic Acids Res* 2010, **38**:e117.
 49. Lemoine S, Combes F, Servant N, Le Crom S: **Goulphar: Rapid access and expertise for standard two-color microarray normalization methods.** *BMC Bioinformatics* 2006, **7**:467.
 50. Pfaffl MW, Horgan GW, Dempfle L: **Relative expression software tool (REST[®]) for group-wise comparison and statistical analysis of relative expression results in real-time PCR.** *Nucleic Acid Research* 2002, **30**:e36.
 51. Druzhinina IS, Schmolli M, Seiboth B, Kubicek CP: **Global carbon utilization profiles of wild-type, mutant, and transformant strains of *Hypocrea jecorina*.** *Appl Environ Microbiol* 2006, **72**:2126-2133.
 52. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovskiy I, Liu Z, Vinsavich A, Trush V, Quackenbush J: **TM4: a free, open-source system for microarray data management and analysis.** *Biotechniques* 2003, **34**:374-378.
 53. Pavlidis P, Noble WS: **Analysis of strain and regional variation in gene expression in mouse brain.** *Genome Biol* 2001, **2**:RESEARCH0042.
 54. Ben-Dor A, Shamir R, Yakhini Z: **Clustering gene expression patterns.** *J Comput Biol* 1999, **6**:281-97.
 55. Walter MC, Rattei T, Arnold R, Güldener U, Münsterkötter M, Nenova K, Kastenmüller G, Tischler P, Wölling A, Volz A, Pongratz N, Jost R, Mewes HW, Frishman D: **PEDANT covers all complete RefSeq genomes.** *Nucleic Acids Res* 2009, **37** Database: D408-411.
 56. Institute of Bioinformatic and Systems Biology: **Helmholtz Zentrum München.** [http://pedant.helmholtz-muenchen.de/].

doi:10.1186/1471-2164-12-269

Cite this article as: Portnoy et al.: The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. *BMC Genomics* 2011 **12**:269.

Differential Regulation of the Cellulase Transcription Factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* Strains Producing High and Low Levels of Cellulase^{∇†}

Thomas Portnoy,^{1,3‡} Antoine Margeot,^{1‡*} Verena Seidl-Seiboth,² Stéphane Le Crom,³
Fadhel Ben Chaabane,¹ Rita Linke,² Bernhard Seiboth,² and Christian P. Kubicek^{2*}

IFP Energies nouvelles, Département Biotechnologie, 1-4 Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex,¹ and Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS, Inserm, U1024, and CNRS, UMR 8197, Paris F-75005,³ France, and Research Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorferstrasse 1a, A-1060 Vienna, Austria²

Received 30 August 2010/Accepted 7 December 2010

Due to its capacity to produce large amounts of cellulases, *Trichoderma reesei* is increasingly being investigated for second-generation biofuel production from lignocellulosic biomass. The induction mechanisms of *T. reesei* cellulases have been described recently, but the regulation of the genes involved in their transcription has not been studied thoroughly. Here we report the regulation of expression of the two activator genes *xyl1* and *ace2*, and the corepressor gene *ace1*, during the induction of cellulase biosynthesis by the inducer lactose in *T. reesei* QM 9414, a strain producing low levels of cellulase (low producer). We show that all three genes are induced by lactose. *xyl1* was also induced by D-galactose, but this induction was independent of D-galactose metabolism. Moreover, *ace1* was carbon catabolite repressed, whereas full induction of *xyl1* and *ace2* in fact required CRE1. Significant differences in these regulatory patterns were observed in the high-producer strain RUT C30 and the hyperproducer strain *T. reesei* CL847. These observations suggest that a strongly elevated basal transcription level of *xyl1* and reduced upregulation of *ace1* by lactose may have been important for generating the hyperproducer strain and that thus, these genes are major control elements of cellulase production.

Recent demands for the production of biofuels from renewable carbon sources have led to increased interest in the hydrolysis of cellulose, a β -(1,4)-linked glucose polymer which is the product of the utilization of solar energy and carbon dioxide by plants and exhibits an annual production of about 7.2×10^{10} tons. Its degradation is a key transformation step in the biological carbon cycle in nature and also in the production of second-generation biofuels (8). *Trichoderma reesei*, the anamorph of the tropical ascomycete *Hypocrea jecorina*, is one of many saprobic fungi that are capable of efficiently degrading plant cell wall polysaccharides, such as cellulose or hemicelluloses. Possibly due to an intriguing clustering of cellulase genes in its genome (21), *T. reesei* is superior in its ability to form and secrete cellulases and is the major fungus for industrial cellulase production (2, 17, 25).

The *T. reesei* genome encodes 10 cellulases and 16 hemicellulases (21). From a biological perspective, the expression of

these genes must be tightly regulated, because the synthesis of the proteins encoded and their secretion requires a great deal of energy. They are therefore synthesized only in the presence of an inducer, which can be cellulose itself or disaccharides generated by its degradation (such as sophorose) (4, 16). Also, the galactosyl- β -1,4-glucoside lactose is a strong inducer and virtually the only soluble inducing carbon source for commercial cellulase production (16).

Most of the cellulase genes are regulated in a consistent manner (11), suggesting a fine-tuned cooperation of the respective transcription factors (TFs). At present, five such TFs have been identified: the positive regulators XYR1, ACE2, and the HAP2/3/5 complex, as well as the repressor ACE1 and the carbon catabolite repressor CRE1 (16). XYR1 (xylanase regulator 1) is considered the main activator of cellulase and hemicellulase gene expression (35, 36), because its deletion eliminates cellulase induction by all inducers and also impairs the induction of different hemicellulase genes involved in xylan and arabinan degradation (1, 35). It is an orthologue of the *xlnR* gene of *Aspergillus niger*, which controls the transcription of more than 20 genes encoding hemicellulases and cellulases (36). Deletion of the second cellulase activator-encoding gene characterized, *ace2*, lowered the transcript levels of the major cellulases and reduced cellulase activity to 30 to 70% when the fungus was grown on cellulose but had no effect on cellulase induction by sophorose (5). Interestingly, XYR1 and ACE2 are able to bind to the same promoter motif [GGC(T/A)₄] (12). Stricker et al. (38) suggested that phosphorylation and dimerization are prerequisites for the binding of ACE2 to its

* Corresponding author. Mailing address for Antoine Margeot: IFP Energies nouvelles, Département Biotechnologie, 1-4 Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France. Phone: (33) 1 47527196. Fax: (33) 1 47527001. E-mail: antoine.margeot@ifpenergiesnouvelles.fr. Mailing address for Christian P. Kubicek: Research Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorferstrasse 1a, A-1060 Vienna, Austria. Phone: (43) 1 58801-166550. Fax: (43) 1 58801-17299. E-mail: ckubicek@mail.zserv.tuwien.ac.at.

† Supplemental material for this article may be found at <http://ec.asm.org/>.

‡ T.P. and A.M. contributed equally to this work.

∇ Published ahead of print on 17 December 2010.

TABLE 1. *Trichoderma reesei* strains

Name or genotype	Reference
QM 9414.....	Mandels and Andreotti, 1978 (20)
<i>Δcre1</i>	This study
<i>Δxyl1</i>	Stricker et al., 2006 (35)
<i>Δgall</i>	Seiboth et al., 2004 (30)
<i>Δxyl1</i>	Seiboth et al., 2007 (29)
<i>Δgall Δxyl1</i>	Seiboth et al., 2007 (29)
RUT C30.....	Eveleigh and Montencourt, 1979 (10)
CL847.....	Durand et al., 1988 (9)

target promoter. The HAP2/3/5 complex is believed to be necessary for the generation of an open chromatin structure, necessary for full transcriptional activation (41). The repressor role of ACE1 has been suggested by studies demonstrating that deletion of *ace1* results in an increase in the expression of all the main cellulase and hemicellulase genes in sophorose- and cellulose-induced cultures (3). ACE1 also represses *xyl1* expression during growth on D-xylose (19). CRE1 is the main transcription factor mediating carbon catabolite repression (CCR) (14, 34), a mechanism that favors the assimilation of high-energy-yielding carbon sources over that of sources yielding less energy. CRE1 is impaired in strain RUT C30 (14), which produces very high levels of cellulase (hyperproducer), and the importance of this gene for the improvement of enzyme production in *T. reesei* has been emphasized recently (24).

TFs are often present in cells only in small amounts, and they function as pacemakers for the processes in which they are involved (27). Many of them, therefore, are further induced by the conditions for which they are needed and are degraded once they are no longer needed (40). In contrast, *xyl1* expression has been reported to be regulated solely by CRE1-dependent CCR and by repression by the specific transcription factor ACE1, not by induction (19, 37). Such a mechanism, if correct, would tightly link *xyl1* expression to the flux through carbohydrate catabolism but would preclude the possibility of responding to the presence of a potential inducer.

The objective of this paper was therefore to reassess how *xyl1* and two other putatively specific cellulase-related TFs (i.e., ACE1 and ACE2) are regulated under conditions of cellulase induction and how this regulation relates to carbon catabolite repression. Lactose was chosen as the cellulase-inducing carbon source, because growth on lactose is independent of cellulase activity. In addition, we compared the expression patterns of these TFs with those observed in a high-producer strain in the public domain (RUT C30) and an industrial hyperproducer strain (CL847). Our findings offer new insights into the mechanism of cellulase induction, as well as targets for the improvement of *T. reesei* strains.

MATERIALS AND METHODS

Fungal strains. The *T. reesei* strains used in this study and their references are given in Table 1. All strains were maintained on malt extract agar. Spores were resuspended in a sterile NaCl (9 g · liter⁻¹)–20% (wt/vol) glycerol solution and were stored at –80°C. A *cre1* deletion vector was constructed by using the double joint PCR technique (39). The oligonucleotides used for the construction of the deletion fragment and the *Δcre1* strain are described in the supplemental material.

Cultivation conditions. Frozen spores were used to inoculate a Fernbach flask containing 250 ml of the culture medium described previously (13). Cultivation

was carried out at 30°C with stirring at 150 rpm. After 72 h, the medium containing mycelia was used as an inoculum for bioreactor culture.

Cellulase was produced in a 4-liter bioreactor with a two-step cultivation procedure. Strains were first grown at 28°C in 2 liters of a medium containing 30 g · liter⁻¹ of glucose as a carbon source and pH regulated at 4.8 with 5.5 M NH₃. The airflow was adjusted at 0.5 volume per volume per min (VVM), and initial stirring was set at 400 rpm. This parameter was gradually increased to maintain partial O₂ pressure (pO₂) above 40% oxygen saturation. When the initial glucose was close to depletion (<20% of initial glucose content), the fed-batch phase was initiated. During this second step, a 250-g · liter⁻¹ carbon source solution (either glucose or lactose) was injected at a rate of 0.98 g · h⁻¹. Samples were collected periodically to determine the biomass, carbon, and protein concentrations.

D-Galactose pathway deletion strains were cultivated in shake flasks as described previously (29). For transfer cultures, strains were pregrown for 24 h in 1-liter flasks on a rotary shaker (250 rpm) at 30°C in 250 ml of the medium described by Mandels and Andreotti (20) with glycerol (1% [wt/vol]) as a carbon source. They were then harvested, washed with a medium without a carbon source, and transferred to a fresh medium containing D-galactose or D-glucose as a carbon source (1% [wt/vol]), as specified in Results. Samples for mRNA extraction were then harvested after 2, 4, and 6 h.

Analytical methods. Biomass concentrations were determined by gravimetric analysis. Ten milliliters of the collected culture medium containing mycelia was filtrated on 1.2-μm-pore-size GF/C glass microfiber membranes. Dry biomass was measured 24 h after incubation of the membrane at 105°C.

The glucose concentration during the batch phase was assessed by enzymatic reaction using an Analox GM10 glucose analyzer (Imlab) to predict the convenient time to start the fed-batch phase.

The concentrations of all carbon sources were quantified *a posteriori* by high-performance liquid chromatography (HPLC) using a HPX-87P column (Bio-Rad) maintained at 85°C. He-degassed distilled water was used as an eluant at a flow rate of 0.4 ml · liter⁻¹.

Concentrations of extracellular protein were quantified by the Quick Start Bradford protein assay kit (Bio-Rad) with bovine serum albumin (BSA) as a standard.

RNA sample preparation. RNA samples for analyses of gene expression were prepared from mycelium powder obtained by grinding the filtrated biomass from bioreactor cultivations—or from shake flasks for D-galactose pathway deletion strains and transfer cultures—in liquid nitrogen. The powder was subjected to a phenol treatment using TRI reagent solution (Applied Biosystems). The extracted total RNA was then isolated with bromochloropropane, precipitated with isopropanol, washed with ethanol, and solubilized in nuclease-free water according to the manufacturer's instructions. Samples were cleaned up by following the Qiagen RNeasy procedure and were subjected to on-column DNase digestion with the RNase-Free DNase set (Qiagen).

Transcript analysis by quantitative PCR (qPCR) analysis. Five hundred nanograms of RNA was reverse transcribed (RT) using the iScript cDNA synthesis kit (Bio-Rad) containing a blend of oligo(dT) and random hexamer primers.

qPCRs were performed in an iCycler iQ real-time detection system (Bio-Rad). Each reaction mixture contained 1 μl of the template (1/10 of the previous RT reaction product), 12.5 μl of iQ SYBR green Supermix 2× (Bio-Rad), 0.8 pmol · μl⁻¹ forward and reverse primers, and nuclease-free water to a final volume of 25 μl. The PCR protocol consisted of 3 min of initial denaturation at 95°C, followed by 40 cycles of 10 s at 95°C and 30 s at 60°C. A melting curve was performed after each run to check PCR product specificity.

The oligonucleotide primers used are given in Table 2. To ensure the absence of genomic DNA, control samples were subjected to the procedure described above, but no reverse transcriptase was added, and PCRs without a template were set up to rule out contamination of other PCR components. All PCRs were carried out in triplicate within a plate, and every condition was tested in at least two different plates. Identical samples were present on every plate; this calibrator was used to counteract the effects of potential interrun variations for all other samples.

Data were analyzed with iQ 5 Optical System software (Bio-Rad). Using the PCR baseline-subtracted mode, the detection threshold level was set automatically by the software, and the amplification efficiency for each gene was determined. All threshold cycle (C_T) values of a plate were adjusted so that C_T value differences between interrun calibrators were eliminated. In order to compare different samples, the threshold cycles for all genes investigated were corrected with a factor for *gpd1* amplification, as described previously (33).

Expression ratios were analyzed on samples collected just before the beginning of the fed-batch phase (BFB) or 1 h, 6 h, and 24 h after the beginning of extracellular protein accumulation during the fed-batch phase. Since strain QM 9414 started its protein production 18 h after the beginning of the fed-batch

TABLE 2. qPCR oligonucleotides used in this study

Transcript studied	Probe type	Sequence
<i>gpd1</i>	Forward	GGAGCTCTTTGAAGAGGA
<i>gpd1</i>	Reverse	GGCAGGTACTTGACGTTTTC
<i>xyl1</i>	Forward	CCATCAACCTTCTAGACGAC
<i>xyl1</i>	Reverse	AACCCTGCAGGAGATAGAC
<i>cbh1</i>	Forward	CCGAGCTTGGTAGTTACTCTG
<i>cbh1</i>	Reverse	GGTAGCCTTCTTGAACCTGAGT
<i>bga1</i>	Forward	GTACAAGGTTGACTCCCTTC
<i>bga1</i>	Reverse	CCATAGTCAGAGCCGTAGAG
<i>ace1</i>	Forward	GGACGAGGAGGAGATTATG
<i>ace1</i>	Reverse	GTGAGTCTTCTCGTGCTT
<i>ace2</i>	Forward	GACAAGAAGCTCAGGTGTC
<i>ace2</i>	Reverse	ACTGTGTTTCATGGCTGTG

phase (as can be seen in Fig. 1Ai), which corresponds to the time needed to release CCR, transcript expression levels during the fed-batch phase were assessed 1 h, 6 h, and 24 h after the beginning of the fed-batch phase for the $\Delta cre1$, CL847, and RUT C30 strains and 19 h, 24 h, and 42 h after the beginning of the fed-batch phase for the parental strain QM 9414. The means \pm standard deviations (SD) of replicates are shown in the figures. For each transcript, the lowest ratio measured among strains and conditions was set to 1, and all other ratios were related to it, in order to allow better comparisons.

Statistical analyses. Student *t* tests were performed on all sets of data. All the differences mentioned fulfill the criterion of a *P* value of 0.05.

RESULTS

Accumulation of *cbh1* and *bga1* transcripts during the growth of *T. reesei* QM 9414 on lactose and D-glucose. In order to investigate the transcriptional regulation of the cellulase regulatory genes by lactose, we set up a fed-batch system in which mycelia were first cultivated on D-glucose (batch phase) and then—when the concentration of D-glucose had decreased below 20% of the initial content—were fed with lactose at a rate of 0.98 g · h⁻¹ (fed-batch phase). The suitability of this system for the investigation of the secretome in *T. reesei* has been shown recently (13). A D-glucose feed at the same rate was used as a control.

Figure 1Ai shows the pattern of growth, the carbohydrate consumption, and the specific protein production rate in the low cellulase producer *T. reesei* QM 9414. D-Glucose was consumed within 30 h of cultivation, and secretion of cellulases and hemicellulases started around 18 h after the beginning of the fed-batch phase. With this protocol, more than 80% of the extracellular protein comprised cellulase and hemicellulase proteins (13). The concentrations of dry biomass reached 8 to 12 g · liter⁻¹ during the batch phase and stabilized after the beginning of protein production.

In contrast, almost no protein was secreted with a D-glucose feed (Fig. 1Bi), although dry mass concentrations and D-glucose consumption were similar during the batch phase.

The accumulation of the *cbh1* and *bga1* transcripts, encoding the cellobiohydrolase 1 CEL7A and the extracellular β -galactosidase BGA1, respectively, correlated with the utilization of the fed lactose (Fig. 1Aii and iii). The expression of *cbh1* and *bga1* at the onset of lactose feeding was very low (before the beginning of the fed-batch phase [BFB]) but increased considerably during the fed-batch phase (1 h, 6 h, and 24 h). In contrast, *cbh1* and *bga1* expression was much weaker with D-glucose feeding (Fig. 1Bii and iii). These data demonstrate

that the experimental system chosen is well suited to document the induction of *cbh1* and *bga1* by lactose and thus is appropriate for the study of the transcriptional regulation of the cellulase regulatory proteins by lactose.

***xyl1* expression is induced by lactose in *T. reesei* QM 9414.** Having set up an appropriate cultivation system for cellulase induction, we tested the expression of *xyl1*, encoding the major transcriptional regulator of cellulase and hemicellulase formation. As can be seen in Fig. 2, the abundance of the *xyl1* transcript in strain QM 9414 was low at the end of the batch phase but increased about 250-fold during lactose feeding. During feeding with D-glucose, however, only a small increase in *xyl1* expression was observed and its final level was only about 5% of that observed on lactose. These data are in accordance with the assumption of a specific role for lactose in *xyl1* induction.

***xyl1* is positively regulated by induction and by the carbon catabolite repressor CRE1.** Stricker et al. have recently postulated that *xyl1* is not regulated by induction but solely by carbon catabolite (de)repression (37). In order to rule out the possibility that the enhanced expression of *xyl1* on lactose is due to relief of carbon catabolite repression on this carbon source, we used a $\Delta cre1$ strain of *T. reesei* QM 9414, cultivating it under the same conditions as those described above for QM 9414. The carbon catabolite-derepressed strain indeed accumulated higher *cbh1* mRNA levels at the onset of lactose feeding (see Table S2 in the supplemental material), a finding consistent with previous data showing stimulation of *cbh1* gene expression in a carbon catabolite-derepressed background.

xyl1 transcript accumulation, however, was lower in the $\Delta cre1$ strain: only at the end of the batch phase was the abundance higher (about 10-fold) than that in QM 9414 (Fig. 2), challenging the repressing role of CRE1/CCR on *xyl1* (37). Feeding of D-glucose to the $\Delta cre1$ strain increased the relative *xyl1* transcript abundance to the level seen in the parent strain QM 9414, whereas feeding of lactose gradually increased *xyl1* transcript abundance to a final level of 6-fold relative to the BFB level. However, this final transcript abundance was only one-third of that in the parent strain. We therefore conclude that *xyl1* is induced by lactose and that full induction requires the positive action of CRE1.

Lactose induction of *xyl1* is independent of D-galactose metabolism. The induction of cellulases by lactose has previously been shown to require metabolism of lactose through the Leloir pathway and through the alternative reducing pathway of D-galactose catabolism (29, 32). We therefore tested whether *xyl1* induction by lactose in strain QM 9414 would also require this metabolism. To this end, we used isogenic QM 9414 strains bearing knockouts in *gal1* (galactokinase; blocking the Leloir pathway [30]), *xyl1* (aldose reductase; blocking the alternative pathway [29]), and both genes (*gal1 xyl1* [29]). The results (Fig. 3A) show that none of these pathway mutations impaired *xyl1* gene expression on lactose, which apparently occurs even in the absence of D-galactose metabolism.

Since lactose metabolism in *T. reesei* occurs by extracellular hydrolysis of lactose to D-glucose and D-galactose (31, 32), the findings described above raise the question of whether the hydrolysis product D-galactose would also act as an inducer of *xyl1*. Since D-galactose is a repressing carbon source and is able to induce cellulases only at low growth rates (15), we compared

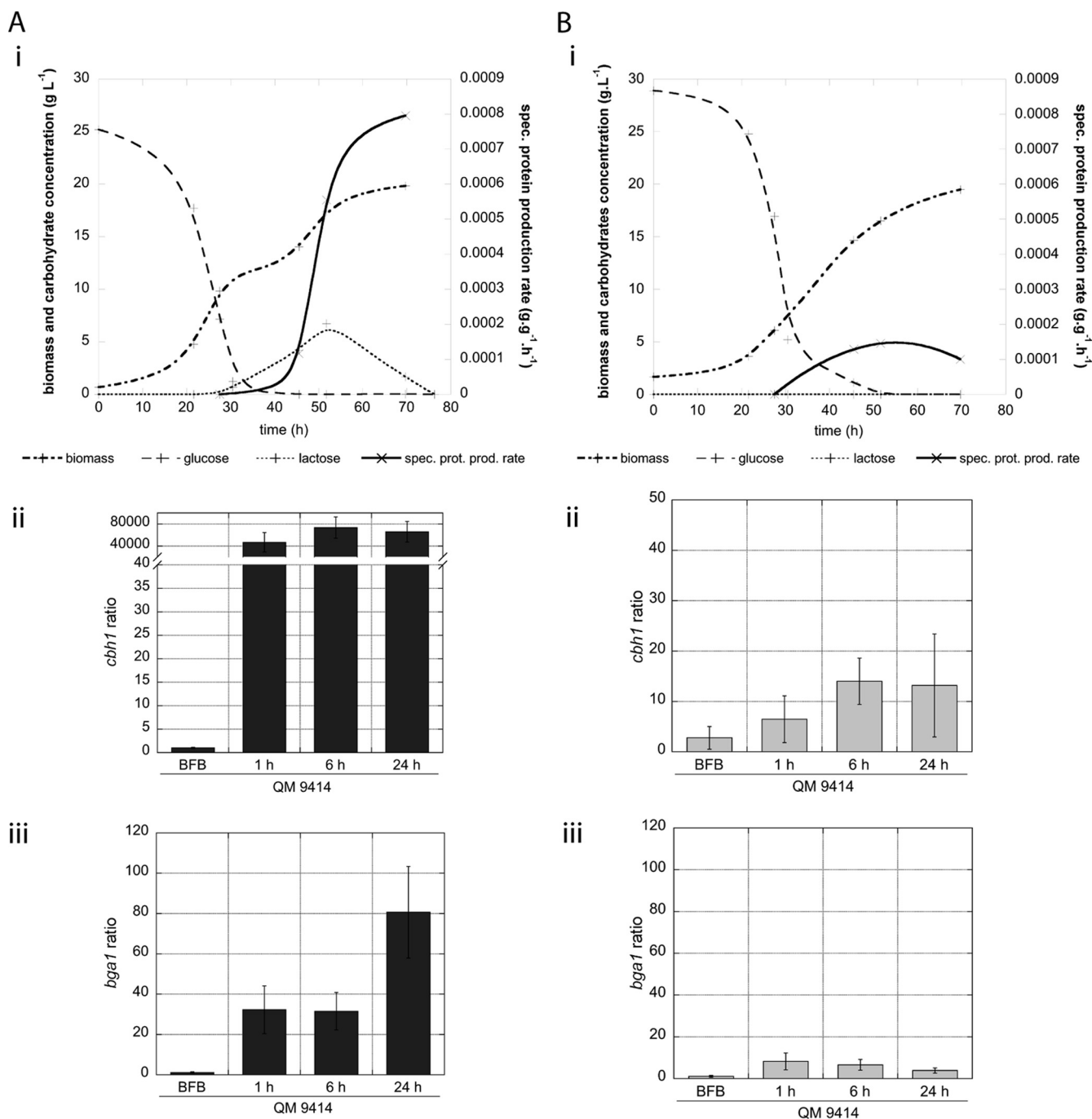


FIG. 1. *T. reesei* QM 9414 accumulates CEL7A and BGA1 when cultivated under a lactose-feeding regimen. After a phase of growth on glucose, *T. reesei* QM 9414 was subjected to a lactose (A) or a glucose (B) feed. (i) The dry biomass concentration, carbohydrate consumption, and specific (spec.) protein production rate were assessed throughout the cultivations. (ii and iii) The ratios of the expression of the *cbh1* (ii) and *bga1* (iii) transcripts to that of the steady-state *gpd1* reference gene were calculated just before (BFB) and 1 h, 6 h, and 24 h after the beginning of the fed-batch phase.

xyl1 expression in shake flasks on D-glucose and D-galactose as carbon sources. Indeed, we found 5- to 10-fold-higher expression of *xyl1* with D-galactose than with D-glucose and glycerol cultures (Fig. 3B). The degree of induction by D-galactose was actually lower than that by lactose, but this is consistent with its action as a carbon catabolite-repressing carbon source, al-

though some of the quantitative differences in transcript abundance may also be due to the different culture conditions used.

ace2 and ace1 are also induced by lactose. Having established that *xyl1* is regulated by lactose, we investigated whether the coactivator gene *ace2* and the repressor gene *ace1* would also be induced by lactose. As shown in Fig. 4A, the *ace2*

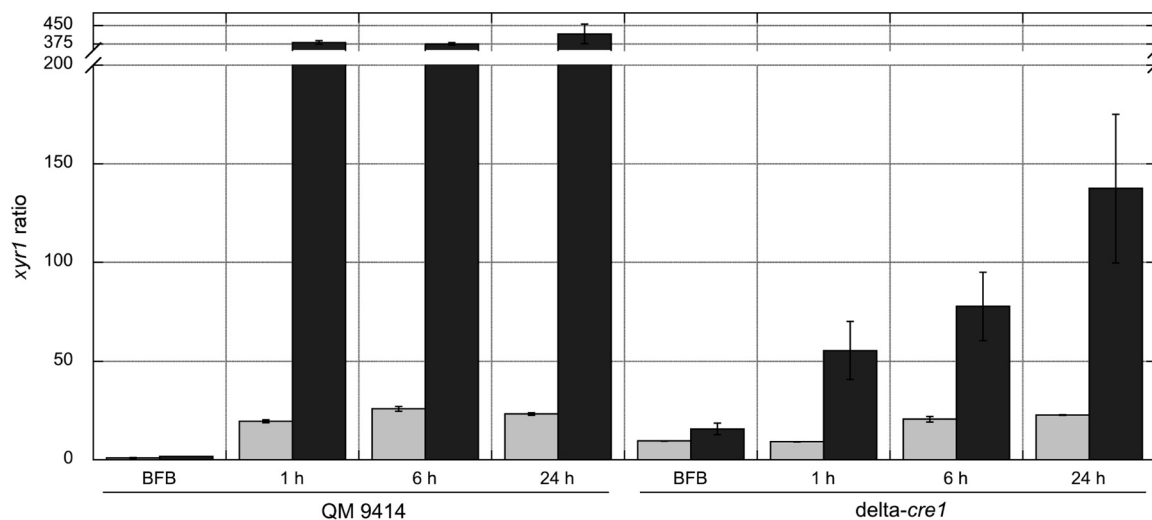


FIG. 2. *xyr1* expression is induced by lactose independently of the release of carbon catabolite repression. The ratio of the expression of the *xyr1* transcript to that of the steady-state *gpd1* reference gene was calculated in *T. reesei* QM 9414 or in a $\Delta cre1$ strain of QM 9414 fed with glucose (shaded bars) or with lactose (filled bars) after a 25-h growth phase on glucose. Ratios before (BFB) and at different time points during the fed-batch phase are shown.

transcript was present during growth on D-glucose, suggesting partial constitutive expression. Lactose feeding led to a further increase (about 3-fold) in *ace2* transcript abundance. While this is much lower than lactose induction of *xyr1*, it nevertheless indicates lactose induction of *ace2*. Expression of *ace2* was not subject to carbon catabolite repression, as evidenced by the fact that no increased expression levels were observed in the $\Delta cre1$ strain. In contrast, both the transcript abundance on D-glucose and the transcript abundance upon induction by lactose were in fact lower in the $\Delta cre1$ strain, suggesting that CRE1 is also required—either directly or indirectly—for full *ace2* expression.

The *ace1* gene, encoding a corepressor of cellulase gene

transcription, was also expressed during batch growth on glucose. Its expression remained at the same level during glucose feeding. However, expression also increased moderately upon the shift to lactose feeding (Fig. 4B). Both the basal and the inducible expression levels of *ace1* were about doubled in the $\Delta cre1$ strain, proving that they are subject to CCR.

Expression of *xyr1* and *ace1* is altered in the *T. reesei* hyper-producer strain CL847. Having assessed the regulatory patterns of the three specific transcriptional regulators, XYR1, ACE2, and ACE1, in the low cellulase producer mutant QM 9414, we wondered whether their expression and/or regulation would have become altered in the course of mutagenesis for high cellulase production. To explore this question, we chose

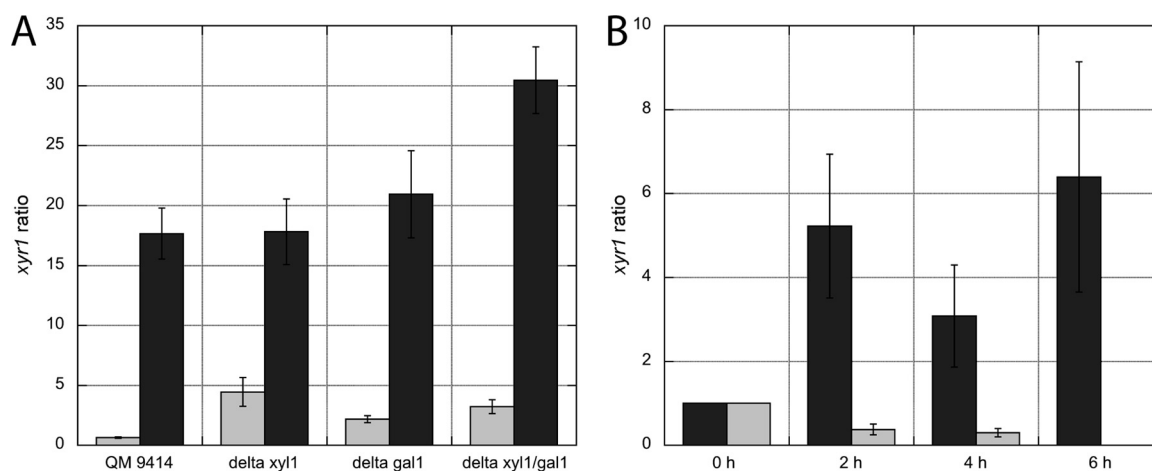


FIG. 3. Induction of *xyr1* by lactose is not dependent on D-galactose metabolism. The ratio of the expression of the *xyr1* transcript to that of the steady-state *gpd1* reference gene was calculated. (A) *xyr1* transcript expression ratios in QM 9414 and galactose pathway mutant strains upon incubation for 20 h (shaded bars) or 30 h (filled bars) on lactose. (B) *xyr1* transcript expression ratios in strain QM 9414, pregrown in shake flasks with glycerol as a carbon source and then transferred to D-galactose (filled bars) or D-glucose (shaded bars) for 2, 4, or 6 h as indicated (only the ratio for galactose was measured at 6 h). The ratio obtained on glycerol (24 h of precultivation) was set to 1.0, and all other ratios are given relative to this value (*n*-fold).

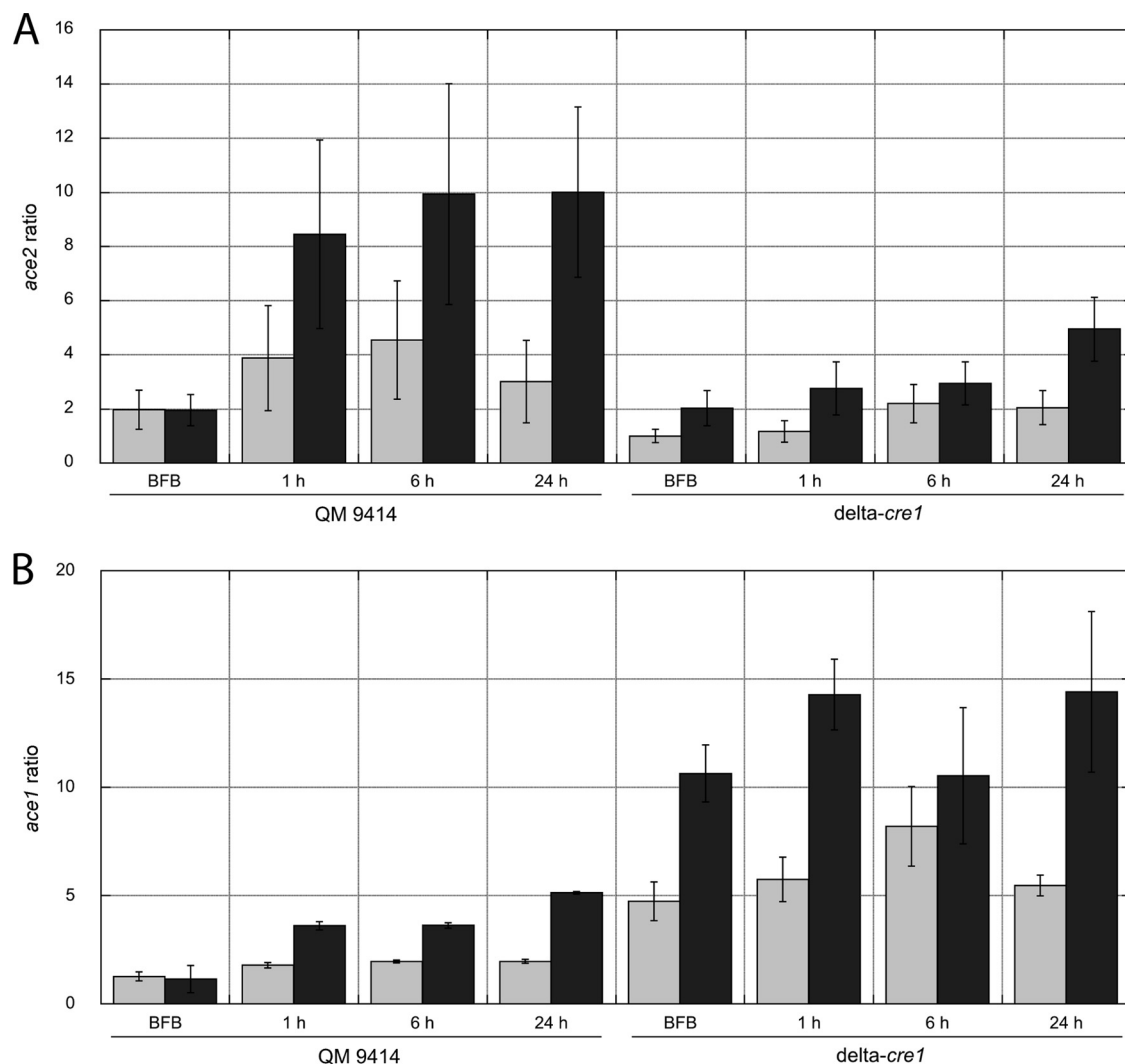


FIG. 4. *ace2* and *ace1* are both regulated by lactose, but only *ace1* is subject to CCR. The ratios of the expression of the *ace2* (A) and *ace1* (B) transcripts to that of the steady-state *gpd1* reference gene were calculated in *T. reesei* QM 9414 and in a $\Delta cre1$ strain of QM 9414 fed with glucose (shaded bars) or lactose (filled bars) after a 25-h growth phase on glucose. Ratios before (BFB) and at different time points during the fed-batch phase are shown.

the mutant strain CL847 (9), which is the highest cellulase producer strain published so far, and—for analysis of *xyl1*—also the high-producer strain RUT C30 (10, 23). Strain RUT C30 is known to bear a truncated version of the *cre1* gene and is thus carbon catabolite derepressed. Since the mutagenesis scheme published by Durand et al. (9) also includes a screen for 2-deoxyglucose resistance to eliminate CCR, we tested whether CL847 still contains an intact *cre1* locus. To our surprise, it contains the same truncated *cre1* locus as strain RUT C30 (see Fig. S1 in the supplemental material), suggesting that it is a mutant derived from RUT C30. This assumption is further supported by genome sequencing of CL847 (A. Margeot and S. Le Crom, unpublished data). We also sequenced the loci of *xyl1*, *ace1*, and *ace2* in CL847 and found that neither their promoter regions nor their open reading frames (ORF) bear any mutations (see Fig. S2 to S4 in the supplemental material). We also have evidence that all three genes occur as single copies in the CL847 genome (data not shown).

We cultivated CL847 and RUT C30 under the same conditions as those used for strain QM 9414. As shown in Fig. 5, D-glucose consumption and biomass accumulation by CL847 during the batch phase were similar to those for QM 9414. The concentrations of protein excreted into the culture medium were 5- to 15-fold higher than those for QM 9414 at comparable time points (e.g., $0.08 \text{ g} \cdot \text{liter}^{-1}$ in QM 9414 versus $1.33 \text{ g} \cdot \text{liter}^{-1}$ in CL847 after 24 h of induction; $0.58 \text{ g} \cdot \text{liter}^{-1}$ in QM 9414 versus $4.23 \text{ g} \cdot \text{liter}^{-1}$ in CL847 after 50 h of induction).

The relative transcript abundance of *cbh1* in strain CL847 was already very high before the beginning of lactose feeding. After 24 h, the relative transcript abundance of *cbh1* in CL847 was 10-fold higher than that in QM 9414 (Fig. 6A). The expression of *xyl1* was also already significantly higher at the onset of the feed, and fed-batch levels were 6- to 7-fold higher than those in QM 9414 after 24 h of lactose feeding. During cultivation on D-glucose, the abundance of *xyl1* mRNA in

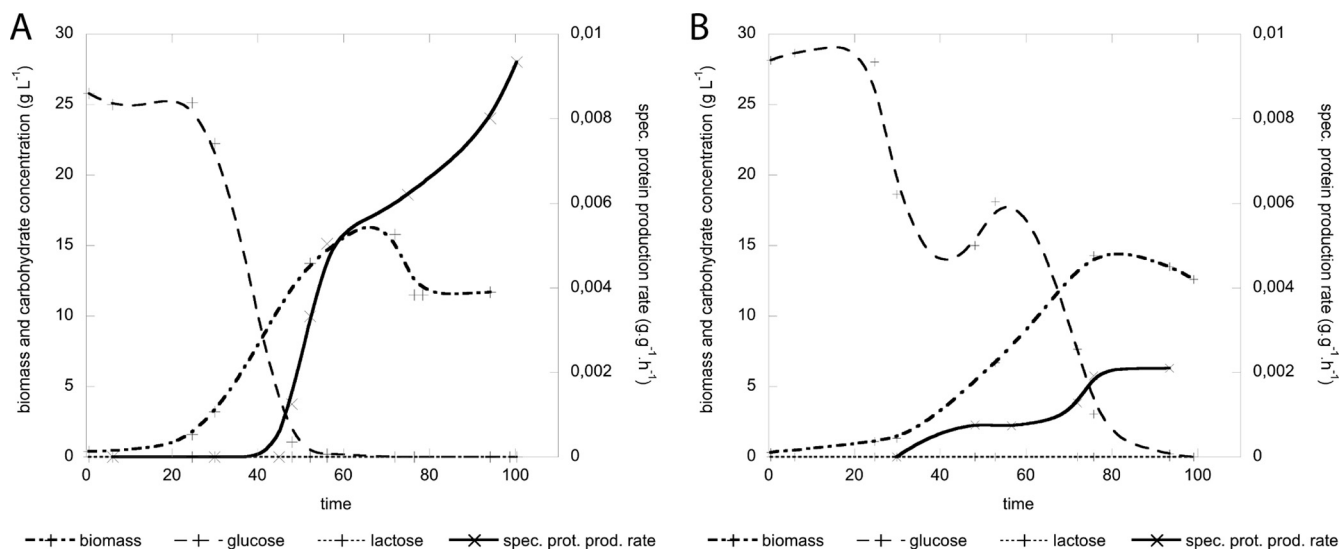


FIG. 5. The *T. reesei* hyperproducer strain CL847 produces cellulases when cultivated under a lactose-feeding regimen. After a phase of growth on glucose, *T. reesei* CL847 was subjected to a lactose (A) or a glucose (B) feed. The dry biomass concentration, carbohydrate consumption, and specific protein production rate were assessed throughout the cultivations.

CL847 was 8-fold higher than that in the $\Delta cre1$ strain at the beginning, and the *xyl1* mRNA in the CL847 strain remained at about this level during the D-glucose fed-batch process (Fig. 6B). In contrast, *xyl1* transcript abundance in strain CL847 after 24 h was almost 15-fold higher than that at the onset of the lactose feed.

Having identified the *T. reesei* high-producer mutant RUT C30 as an ancestor of CL847 (see above), we were interested in whether *xyl1* transcription had already become elevated in this strain, too. As shown in Fig. 6B, the transcript abundance of *xyl1* in RUT C30 was at about the same level as that in QM 9414, both for basal and for lactose-induced gene expression. The high upregulation of *xyl1* is thus specific for strain CL847.

The expression of *ace2* was less affected in CL847: the basal expression level resembled that observed in the $\Delta cre1$ strain and thus probably is due to the nonfunctional CRE1 protein in this mutant. Induction by lactose, however, led to about the same expression level as that in strain QM 9414, indicating that the loss of CRE1 does not affect inducible transcription of *xyl1* (Fig. 6C). The basal expression of *ace1* in strain CL847 also resembled that in the $\Delta cre1$ strain and can be explained by the relief from CCR. Interestingly, *ace1* expression decreased during glucose feeding. Expression on lactose was also higher than that for QM 9414, but well below the levels seen in the $\Delta cre1$ strain and—notably—was not further induced by lactose during feeding (Fig. 6D).

DISCUSSION

XYR1 is the only transcription factor identified so far in *T. reesei* whose knockout results in the loss of the formation of all cellulases, xylanases, and two arabinofuranosidases (1), as well as in that of some of the enzymes required for the metabolism of the hemicellulose monomers D-xylose, L-arabinose, and D-galactose (29, 35). XYR1 is therefore currently considered to be the main regulator of the cellulase/hemicellulase system of

T. reesei. Consequently, regulation of *xyl1* expression has a significant impact on the ability of *T. reesei* to produce the various carbohydrate depolymerases and to grow on plant cell wall polysaccharides. In agreement with previous data (19), we found that only a very low level of *xyl1* transcript was formed during growth on D-glucose. However, in contrast with the claim that *xyl1* is not regulated by induction, we found strong induction of *xyl1* by lactose and its hydrolysis product D-galactose. Interestingly, Mach-Aigner et al. also reported an increase in the abundance of the *xyl1* transcript during incubation with another hemicellulose monomer, D-xylose, but interpreted this increase as due to carbon catabolite derepression (19). We consider this interpretation flawed, because D-xylose has been shown to be a carbon catabolite-repressing carbon source in *Saccharomyces cerevisiae* (6) and *Aspergillus nidulans* (26). We admit that proof for D-xylose as a carbon catabolite-repressing carbon source in *T. reesei* has not yet been published, but its operation can be deduced from the fact that cultivation of *T. reesei* on D-xylose does not even lead to the low cellulase levels that are formed during carbon catabolite derepression (28). We therefore speculate that the data of Mach-Aigner et al. are in fact consistent with our conclusion that *xyl1* is regulated by a feed-forward activating circuit similar to that seen with *bga1* and *xyl1* (29, 31). In summary, our results indicate that *xyl1* expression is increased by lactose induction. In addition, our data show that the carbon catabolite repressor CRE1 is—in contrast—required for full induction by lactose, rather than repressing its expression. A positive effect of *A. nidulans* CreA on gene expression has been reported previously (22) and is also observed for several other *T. reesei* genes (T. Portnoy et al., submitted for publication).

The transcriptional regulation of the other two putative specific regulators also revealed induction by lactose, although to a lower degree than for *xyl1*. The expression of the coactivator gene *ace2* resembles that of *xyl1* in being inducible by lactose and also requiring CRE1 for full induction. The regulation of

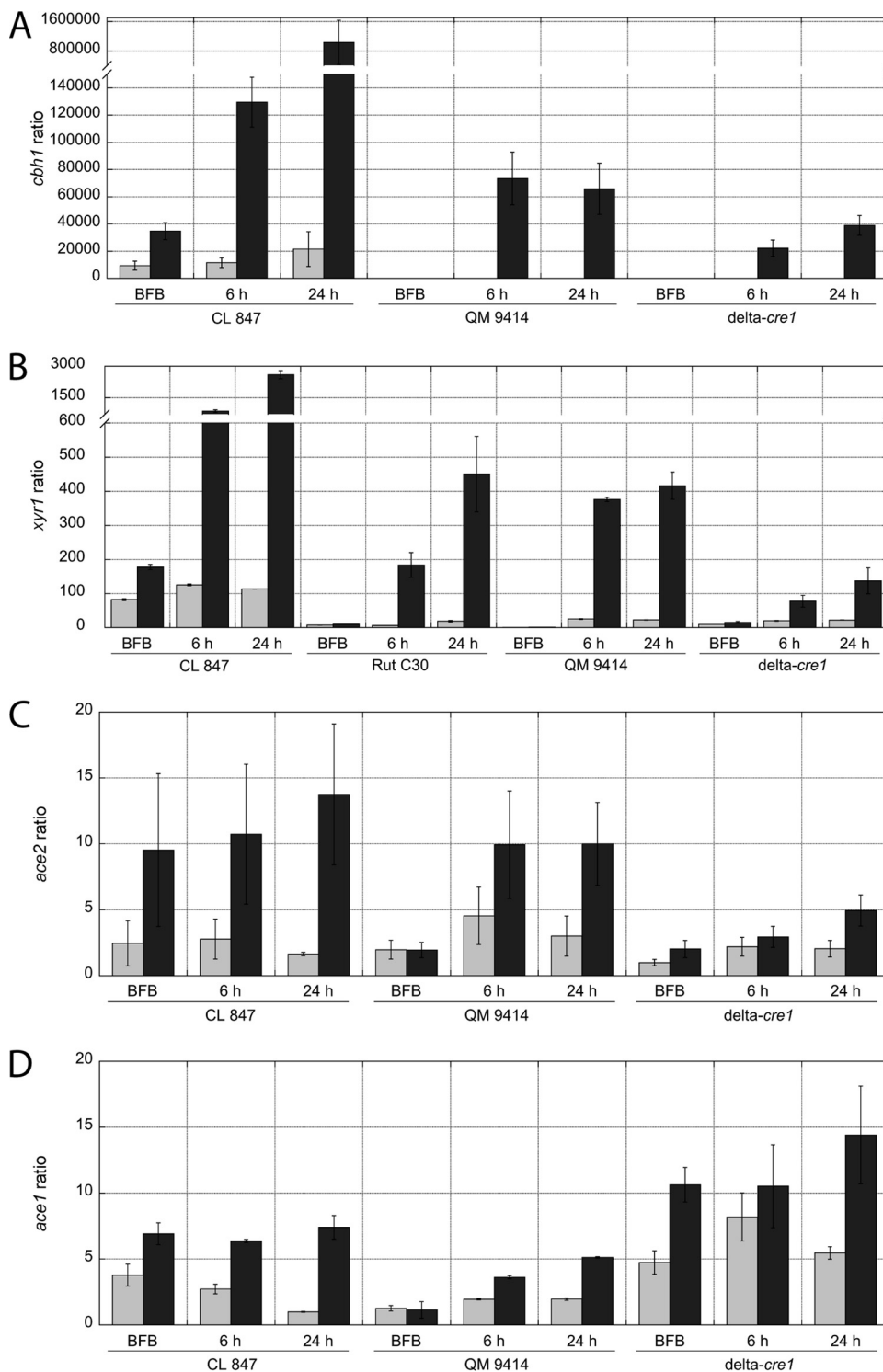


FIG. 6. The expression of cellulase transcriptional regulator genes is altered in the hyperproducing strain CL847. The ratios of the expression of *cbh1* (A), *xyr1* (B), *ace2* (C), and *ace1* (D) transcripts to that of the steady-state *gpd1* reference gene were calculated in *T. reesei* CL847 (and in *T. reesei* RUT C30 [B]) fed with glucose (shaded bars) or with lactose (filled bars) after a 25-h growth phase on glucose. Ratios before (BFB) and 6 h and 24 h after the beginning of the fed-batch phase are shown. Relevant data from Table S2 in the supplemental material and from Fig. 1, 2, and 4 are replotted here for better comparisons.

ace1 expression has been reported previously to be slightly enhanced in the presence of cellulose or sophorose (3), whereas contradictory data on its possible regulation by CCR have been published (3, 19). Here we show—in support of the

latter authors—that both basal expression of *ace1* and its induction by lactose are indeed strongly repressed by CRE1. All these findings, as well as data about the effect of galactose on *xyr1* induction (see below), are summarized in Fig. 7.

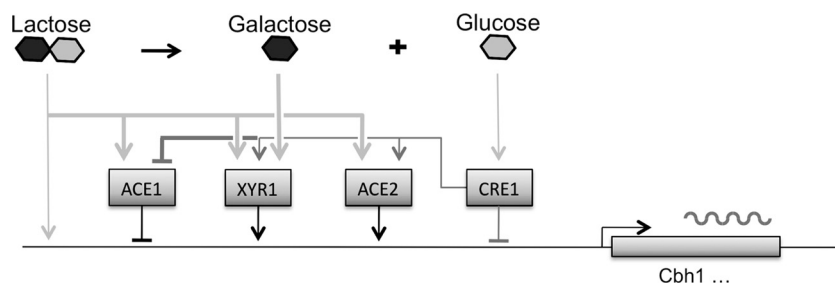


FIG. 7. Regulatory network of cellulase gene expression on lactose. New regulatory relationships described in this work are indicated by bold lines, while the network described previously is indicated by thin lines.

The induction of cellulase gene expression by lactose has recently been shown to be dependent both on the operation of the Leloir pathway and on that of the alternative reductive pathway of D-galactose metabolism (16, 29, 30). Interestingly, the present study shows that lactose induction of *xyl1* was completely independent of galactose metabolism and could also be achieved by D-galactose. The mechanism by which *xyl1* responds to lactose and D-galactose remains unclear. Lactose is not taken up by *T. reesei* (31, 32), and thus, its effect, even in the absence of D-galactose metabolism, must be due to a signaling process. Brunner et al. have recently identified a gene encoding a G protein-coupled receptor from *Trichoderma atroviride* whose silencing resulted—among other features—in a drastic decrease in the ability of this fungus to grow on β -linked galactooligosaccharides (7). An orthologue of this gene is present in *T. reesei* (B. Seiboth and S. Zeilinger, unpublished data) and is expressed under cellulase-forming conditions (expressed sequence tag [EST] clone CB901961) (11). The possible role of this receptor in the lactose induction of *xyl1* is currently under investigation in our laboratory. From a biological perspective, these data show that the presence of lactose (or, in the natural habitat of *T. reesei*, more likely β -galactosides derived from plant cell wall polymers) is a major physiological elicitor for *T. reesei* to shift its transcriptional response toward the degradation of cellulose and hemicelluloses. In support of this hypothesis, we have recently found that XYR1 also regulates L-arabinose and arabinan metabolism in *T. reesei* (1).

There is currently huge interest in the further improvement of cellulase production in existing *T. reesei* mutant strains for second-generation biofuel production (16, 18). We hypothesized that the higher-producing mutants may have also undergone alterations in their transcriptional regulation of the TFs that govern cellulase gene expression. Therefore, we have compared the regulation of *xyl1* in the highest producer strain for which data have been published, *T. reesei* CL847 (9), with that in its obvious parent strain, RUT C30 (10, 23). Under the conditions used in this work, the formation of extracellular protein by strain CL847 on lactose occurred at a \sim 15-fold-higher rate than that for strain QM 9414. This correlates well with the \sim 15-fold-increased abundance of the *cbh1* transcript. These data are consistent with the assumption that gene expression may indeed be a major limiting step for cellulase biosynthesis in the parental strain. Consistent with its role as the major transcriptional regulator of cellulase gene expression, we found strongly increased basal expression of *xyl1* in

strain CL847, which was further induced by lactose to the approximately 6-fold amplitude similar to the one observed in strain QM 9414. Since the *xyl1* locus in CL847 is perfectly wild type, this increase in gene expression must be due to improved function of the transcriptional machinery required for *xyl1* expression.

The basal expression of *ace2* was not significantly altered in strain CL847, and the inducible level was the same as that in strain QM 9414. This indicates that the lack of CRE1 function, which seems to be required for *ace2* gene expression, as indicated by the lower expression levels in the $\Delta cre1$ mutant, has been overcome during the breeding of CL847. While these data suggest that *ace2* expression is not limiting for cellulase induction on lactose, they nevertheless show that wild-type expression levels appear to be necessary for the formation of high levels of cellulase.

Expression of *ace1*—even though it is a repressor of cellulase formation—was also increased in the mutant strain CL847. However, *ace1* is subject to CRE1-dependent CCR, and these data therefore must be viewed against the expression levels in the $\Delta cre1$ strain. The comparison reveals that the basal expression level of *ace1* in CL847 is lower than that in the $\Delta cre1$ strain and decreases during the glucose feed. The approximate doubling of this level during the lactose feed is conserved, however. We therefore conclude that carbon catabolite derepression of *ace1* has partially reverted in CL847, leading to a lower concentration of this repressor under cellulase-producing conditions.

Taken together, our data strongly suggest that lactose signaling is a trigger for the cellulase transcriptional machinery and that the limiting points in this response reside in the transcription components regulating cellulase gene expression. These data are also in agreement with the recently published genome sequence of a cellulase-producing mutant lineage of *T. reesei*, which revealed that the majority of the affected loci encode formerly unknown transcription factors, proteins acting on mRNA stability, and components of nuclear traffic (18). Mach-Aigner et al. have recently hypothesized that posttranslational modification of XYR1 would be the major mechanism governing cellulase gene expression (19). While we have not specifically looked into this, the present findings of reduced *xyl1* but increased *cbh1* transcription in the $\Delta cre1$ strain would be consistent with the operation of posttranslational modification of XYR1. Nevertheless, our data show clearly that the expression of *xyl1*, *ace1*, and *ace2* has been significantly altered in the hyperproducer CL847, suggesting that their wild-type expression was insufficient for hyperproduction. Identification

of the proteins and genes responsible for the mechanisms observed in this paper may result in a major breakthrough in the understanding of cellulase formation and may offer a straightforward means for its improvement.

ACKNOWLEDGMENTS

This work was supported by grants from the Austrian Science Foundation (FWF P-19143 and FWF P-19421, to C.P.K. and B.S., respectively). T.P. is the recipient of a Fondation Tuck Enerbio Ph.D. fellowship.

We thank Sarah Martinez, Pauline Sacareau, Hélène Ipas, Sabine Prigent, Bernard Chaussepied, and Hugues Mathis for excellent technical assistance and help in the analysis of early results.

REFERENCES

1. Akel, E., B. Metz, B. Seiboth, and C. P. Kubicek. 2009. Molecular regulation of arabinan and L-arabinose metabolism in *Hypocrea jecorina* (*Trichoderma reesei*). *Eukaryot. Cell* **8**:1837–1844.
2. Antoni, D., V. V. Zverlov, and W. H. Schwarz. 2007. Biofuels from microbes. *Appl. Microbiol. Biotechnol.* **77**:23–35.
3. Aro, N., M. Ilmen, A. Saloheimo, and M. Penttila. 2003. ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Appl. Environ. Microbiol.* **69**:56–65.
4. Aro, N., T. Pakula, and M. Penttila. 2005. Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiol. Rev.* **29**:719–739.
5. Aro, N., A. Saloheimo, M. Ilmen, and M. Penttila. 2001. ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. *J. Biol. Chem.* **276**:24309–24314.
6. Belinchón, M. M., and J. M. Gancedo. 2003. Xylose and some non-sugar carbon sources cause catabolite repression in *Saccharomyces cerevisiae*. *Arch. Microbiol.* **180**:293–297.
7. Brunner, K., et al. 2008. *Trichoderma* G protein-coupled receptors: functional characterisation of a cAMP receptor-like protein from *Trichoderma atroviride*. *Curr. Genet.* **54**:283–299.
8. Chang, M. C. 2007. Harnessing energy from plant biomass. *Curr. Opin. Chem. Biol.* **11**:677–684.
9. Durand, H., M. Clanet, and G. Tiraby. 1988. Genetic improvement of *Trichoderma reesei* for large scale cellulase production. *Enzyme Microb. Technol.* **10**:341–346.
10. Eveleigh, D. E., and B. S. Montencourt. 1979. Increasing yields of extracellular enzymes. *Adv. Appl. Microbiol.* **25**:57–74.
11. Foreman, P. K., et al. 2003. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *J. Biol. Chem.* **278**:31988–31997.
12. Furukawa, T., et al. 2009. Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. *Fungal Genet. Biol.* **46**:564–574.
13. Herpoël-Gimbert, I., et al. 2008. Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnol. Biofuels* **1**:18.
14. Ilmén, M., C. Thrane, and M. Penttila. 1996. The glucose repressor gene *creI* of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form. *Mol. Gen. Genet.* **251**:451–460.
15. Karaffa, L., et al. 2006. D-Galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates. *Microbiology* **152**:1507–1514.
16. Kubicek, C. P., M. Mikus, A. Schuster, M. Schmoll, and B. Seiboth. 2009. Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnol. Biofuels* **2**:19.
17. Kumar, R., S. Singh, and O. V. Singh. 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J. Ind. Microbiol. Biotechnol.* **35**:377–391.
18. Le Crom, S., et al. 2009. Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **106**:16151–16156.
19. Mach-Aigner, A. R., et al. 2008. Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. *Appl. Environ. Microbiol.* **74**:6554–6562.
20. Mandels, M., and R. E. Andreotti. 1978. Problems and challenges in the cellulose to cellulase fermentation. *Process Biochem.* **13**:6–15.
21. Martinez, D., et al. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.* **26**:553–560.
22. Mogensen, J., H. B. Nielsen, G. Hofmann, and J. Nielsen. 2006. Transcription analysis using high-density micro-arrays of *Aspergillus nidulans* wild-type and *creA* mutant during growth on glucose or ethanol. *Fungal Genet. Biol.* **43**:593–603.
23. Montencourt, B. S., and D. E. Eveleigh. 1977. Preparation of mutants of *Trichoderma reesei* with enhanced cellulase production. *Appl. Environ. Microbiol.* **34**:777–782.
24. Nakari-Setälä, T., et al. 2009. Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production. *Appl. Environ. Microbiol.* **75**:4853–4860.
25. Percival Zhang, Y. H., M. E. Himmel, and J. R. Mielenz. 2006. Outlook for cellulase improvement: screening and selection strategies. *Biotechnol. Adv.* **24**:452–481.
26. Prathumpai, W., M. McIntyre, and J. Nielsen. 2004. The effect of CreA in glucose and xylose catabolism in *Aspergillus nidulans*. *Appl. Microbiol. Biotechnol.* **63**:748–753.
27. Ptashne, M. 1988. How eukaryotic transcriptional activators work. *Nature* **335**:683–689.
28. Schafner, D. W., and R. T. Toledo. 1992. Cellulase production in continuous culture by *Trichoderma reesei* on xylose-based media. *Biotechnol. Bioeng.* **39**:865–869.
29. Seiboth, B., C. Gamauf, M. Pail, L. Hartl, and C. P. Kubicek. 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. *Mol. Microbiol.* **66**:890–900.
30. Seiboth, B., et al. 2004. The galactokinase of *Hypocrea jecorina* is essential for cellulase induction by lactose but dispensable for growth on D-galactose. *Mol. Microbiol.* **51**:1015–1025.
31. Seiboth, B., et al. 2005. Role of the *bga1*-encoded extracellular β -galactosidase of *Hypocrea jecorina* in cellulase induction by lactose. *Appl. Environ. Microbiol.* **71**:851–857.
32. Seiboth, B., B. S. Pakdaman, L. Hartl, and C. P. Kubicek. 2007. Lactose metabolism in filamentous fungi: how to deal with an unknown substrate. *Fungal Biol. Rev.* **21**:42–48.
33. Seidl, V., I. S. Druzhinina, and C. P. Kubicek. 2006. A screening system for carbon sources enhancing β -N-acetylglucosaminidase formation in *Hypocrea atroviridis* (*Trichoderma atroviride*). *Microbiology* **152**:2003–2012.
34. Strauss, J., et al. 1995. Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*. *FEBS Lett.* **376**:103–107.
35. Stricker, A. R., K. Grosstessner-Hain, E. Wurleitner, and R. L. Mach. 2006. Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryot. Cell* **5**:2128–2137.
36. Stricker, A. R., R. L. Mach, and L. H. de Graaff. 2008. Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). *Appl. Microbiol. Biotechnol.* **78**:211–220.
37. Stricker, A. R., M. G. Steiger, and R. L. Mach. 2007. Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. *FEBS Lett.* **581**:3915–3920.
38. Stricker, A. R., P. Trefflinger, N. Aro, M. Penttila, and R. L. Mach. 2008. Role of Ace2 (activator of cellulases 2) within the *xyn2* transcriptosome of *Hypocrea jecorina*. *Fungal Genet. Biol.* **45**:436–445.
39. Szewczyk, E., et al. 2006. Fusion PCR and gene targeting in *Aspergillus nidulans*. *Nat. Protoc.* **1**:3111–3120.
40. Tansey, W. P. 2001. Transcriptional activation: risky business. *Genes Dev.* **15**:1045–1050.
41. Zeilinger, S., M. Schmoll, M. Pail, R. L. Mach, and C. P. Kubicek. 2003. Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. *Mol. Genet. Genomics* **270**:46–55.