UNIVERSITÉ LAVAL

Étude génomique de l'agent de lutte biologique *Pseudozyma flocculosa*

Thèse

François Lefebvre

Doctorat en biologie végétale

Philosophiae doctor (Ph.D.)

Québec, Canada

Résumé

La lutte aux maladies des cultures est une bataille de tous les instants. Au cours du XX^{e} siècle, l'essor de l'industrie chimique a mené à la création de nombreux produits phytosanitaires d'une grande efficacité qui a permis d'accroître de manière substantielle la productivité des cultures. Aujourd'hui, l'utilisation intensive de certains de ces produits est reconnue comme entrainant de fâcheuses conséquences sur la santé humaine ou celle des écosystèmes. À la suite de cette prise de conscience, plusieurs initiatives visant à réduire la dépendance des systèmes agricoles aux pesticides les plus nocifs ont vu le jour.

Une des approches préconisées, celle de la lutte biologique, consiste à tirer profit de l'antagonisme naturel qui caractérise la relation qu'entretiennent certains organismes entre eux. La découverte de l'agent de lutte biologique (ALB) *Pseudozyma flocculosa* et les efforts de recherche qui en ont découlé s'inscrivent directement dans cette mouvance. Depuis 1987, les découvertes concernant ce champignon n'ont cessé d'étonner et d'entretenir de grands espoirs quant à l'élaboration d'un biofongicide efficace et écologique pour la lutte contre la maladie du blanc (*oïdium*) qui affecte plusieurs espèces végétales.

La présente thèse constitue la somme des découvertes les plus récentes concernant *P. flocculosa*. En effet, l'avènement des technologies de séquençage a créé de nouvelles opportunités de recherche qui ont mené à de nombreuses découvertes d'une importance capitale. D'une part, il a été possible de caractériser avec précision la batterie de gènes associée à la synthèse de la flocculosine. Pour ce faire, l'utilisation d'un système d'expression hétérologue basé sur la transformation génétique de l'espèce *Ustilago maydis* a été nécessaire pour déterminer la fonction de certains gènes. D'autre part, une analyse comparative en profondeur du génome de l'ALB et de ceux d'Ustilaginales apparentées a permis de mieux comprendre l'évolution de l'organisme tout en générant un ensemble d'informations de haute qualité et de grande valeur.

Le développement de nouvelles approches basées sur la biologie d'organismes bénéfiques demande beaucoup d'efforts. Cependant, les retombées potentielles associées à la réduction de l'utilisation de produits nocifs pour l'environnement et la santé humaine dépassent bien souvent les investissements qui y sont consacrés.

Abstract

Controlling pests is an important part of crop cultivation. The rise of the chemical industry during the XXth century led to the design of efficient molecules for the control of plant diseases that prompted a substantial increase in crop productivity. However, the intensive use of pesticides in agricultural systems quickly raised doubts about sustainablility. In response to those concerns, many initiatives were put in place to reduce the reliance of farmers on synthetic pesticides.

Biological control, one of these initiatives, is about favoring the growth and the dispersal of natural antagonists in agricultural systems to fight plant diseases as ecological alternatives to pesticides. The discovery and characterization of the biocontrol agent *Pseudozyma flocculosa* was part of this trend. Since 1987, the discoveries regarding the biology of the fungus never ceased to amaze, maintaining hopes for the development of an efficicent biofongicide to control powdery mildews in many crops.

This thesis presents the most recent work that was accomplished to reveal the most interesting properties associated with the biology of the biocontrol agent. Much of this work was made possible, or at least facilitated, by the access to emerging and powerful DNA sequencing and bioinformatics tools that created opportunities to answer pending questions in new ways. First, we characterized with precision the function and structure of the flocculosin gene cluster. Using the genomic sequences of *P. flocculosa*, we localized a putative gene cluster for flocculosin and caraterized the function of the most ambiguous genes using *Ustilago maydis* as a heterologous expression system. Second, we conducted an in-depth comparative analysis of *P. flocculosa* genome with those of other Ustilaginales species. This work shed a new light on the evolution of the species and that generated a data set of high quality that will be available for future projects.

The development of novel alternatives to pesticides based on the biology of beneficial organisms requires time and effort. However, the beneficial impact associated with a reduction in pesticide use often exceeds the investments that lead to such a positive outcome.

Table des matières

Liste des tableaux

Liste des figures

Chapitre 2 : Mode of Action of Biocontrol Agents: all that Glitters Is not Gold

Chapitre 3 : Identification of a Biosynthesis Gene Cluster for Flocculosin a Cellobiose Lipid Produced by the Biocontrol Agent *Pseudozyma flocculosa*

Chapitre 4 : Beta Hydroxylation of Glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-Dependent β-Hydroxylase

Chapitre 5 : The Transition from a Phytopathogenic Smut Ancestor to an Anamorphic Biocontrol Agent Deciphered by Comparative Whole-Genome Analysis

Annexe I : Chapter 5 – Supplemental tables and figures

Remerciements

Une formation doctorale est d'abord le fruit d'une volonté personnelle. Cependant, elle serait bien insuffisante si elle ne pouvait compter sur la générosité de nombreuses personnes dont les divers rôles sont autant d'appuis sur lesquels se bâtit la réussite.

Ma reconnaissance va d'abord à mon directeur de recherche, M. Richard Bélanger, qui m'a fait l'honneur de me proposer un projet de recherche des plus stimulants et grâce auquel j'ai acquis un bagage personnel et professionnel des plus précieux. Que ce soit pour la conduite de projets scientifiques, l'interprétation et la présentation des résultats, la combativité à défendre une idée ou le dynamisme lié aux fonctions académiques, les enseignements dont j'ai bénéficié au cours de mes études sont autant d'outils qui me seront utiles pour la suite des choses.

Je me dois de souligner l'apport essentiel qu'a représenté la présence de M. François Belzile, à titre de codirecteur, dans les divers projets auxquels j'ai participé. Par la finesse de ses analyses et la pertinence de ces recommandations, j'ai pu accroître la qualité de mon travail tout en bénéficiant d'une formation d'une grande qualité.

Mais qu'auraient été ces années d'étude sans le support moral et technique d'une professionnelle de recherche aussi dévouée que Mme Caroline Labbé? Sa grande expérience au laboratoire et son flair scientifique ont été les adjuvants de bon nombre de projets scientifiques, des plus appliqués aux plus fondamentaux.

J'ai aussi une pensée spéciale pour Beate Teichmann et Wilfried Rémus-Borel qui m'ont épaulé à mon arrivée au laboratoire et pour la réalisation des premiers projets dont j'ai été responsable. Merci pour votre patience et votre générosité.

Je souligne l'aide reçue des organismes subventionnaires pour la recherche du Québec et du Canada.

Je remercie aussi mes parents qui ont toujours encouragé ma curiosité et m'ont donné les outils nécessaires pour y répondre.

Je tiens finalement à remercier ma conjointe qui chaque jour adoucit mon existence et dont l'appui indéfectible est une des plus grandes preuves d'amour.

Avant-propos

Les chapitres deux, trois, quatre et cinq de cette thèse sont sous la forme d'articles scientifiques. La contribution des coauteurs est détaillée ci-bas :

Chapitre 2 – Mode of Action of Biocontrol Agents: all that Glitters Is not Gold

Richard R. Bélanger, Caroline Labbé, François Lefebvre et Beate Teichmann.

Cette revue de littérature scientifique a été publiée dans le *Canadian Journal of Phytopathology*. Richard R. Bélanger a rédigé l'ébauche de l'article. Caroline Labbé, François Lefebvre et Beate Teichmann ont participé à la révision et à la rédaction de certaines parties de l'ébauche.

Chapitre 3 – Identification of a Biosynthesis Gene Cluster for Flocculosin a Cellobiose Lipid Produced by the Biocontrol Agent *Pseudozyma flocculosa*

Beate Teichmann, Caroline Labbé, François Lefebvre, Michael Bölker, Uwe Linne et Richard R. Bélanger. (voir chapitre 4)

Chapitre 4 – Beta Hydroxylation of Glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-Dependent β-Hydroxylase

Beate Teichmann, François Lefebvre, Caroline Labbé, Michael Bölker, Uwe Linne et Richard R. Bélanger.

Les articles scientifiques des chapitres trois et quatre ont été publiés dans les revues *Molecular microbiology* et *Applied and Environmental Microbiology*. Beate Teichmann a conçu le dispositif expérimental, avec l'aide de Caroline Labbé, et participé à l'interprétation des données et rédigé l'ébauche de l'article. Caroline Labbé a participé à l'analyse des résultats. François Lefebvre a conduit les expériences pour les gènes *fat2* et *fhd1,* participé à l'analyse des données, à l'interprétation des résultats et à la rédaction de l'ébauche des articles. Michael Bölker et Uwe Linne ont fourni les souches d'*Ustilago maydis* mutées et participé à la rédaction de l'ébauche des articles. Richard R. Bélanger a participé à l'interprétation des données et à la rédaction de l'ébauche des articles.

Chapitre 5 – The Transition from a Phytopathogenic Smut Ancestor to an Anamorphic Biocontrol Agent Deciphered by Comparative Whole-Genome Analysis

François Lefebvre, David L. Joly, Caroline Labbé, Beate Teichmann, Rob Linning, François Belzile, Guus Bakkeren et Richard R. Bélanger.

François Lefebvre a conduit les analyses génomiques (assemblage des génomes et transcriptomes, prédiction et curation des structures géniques, annotation fonctionnelle, et analyse comparative), participé à l'interprétation des données, soumis les séquences à GenBank, et rédigé l'ébauche de l'article. David L. Joly a vérifié la présence de gènes de pathogénicité, de conjugaison (*mating*) et de méiose, amélioré l'annotation des protéines sécrétées, analysé les données génomiques, participé à l'interprétation des résultats et à la rédaction de l'ébauche de l'article. Caroline Labbé a participé à l'annotation génomique et à l'interprétation des données. Beate Teichmann a mené l'expérience des essais de conjugaison des différentes souches d'*U. maydis* et *P. flocculosa*. Rob Linning a conduit l'analyse bio-informatique des éléments répétés. François Belzile a participé à l'interprétation des données et à la rédaction de l'ébauche de l'article. Guus Bakkeren a analysé la structure et la phylogénie des gènes de la conjugaison (*mating-type genes*), participé à l'interprétation des données et à la rédaction de l'ébauche de l'article. Richard R. Bélanger a participé à l'annotation génomique, à l'interprétation des données et à la rédaction de l'ébauche de l'article.

Chapitre 1 : Introduction

1. La lutte biologique

Depuis les débuts de l'agriculture, la protection des cultures représente une préoccupation constante pour l'humain. Bien que les avancées spectaculaires faites dans le domaine depuis plus d'une centaine d'années, et surtout depuis les années 50, aient permis d'augmenter de manière importante le rendement et la qualité des produits qui en découlent, il n'en reste pas moins que certains produits toxiques demeurent abondamment utilisés en l'absence de meilleures alternatives (Enserink *et al.*, 2013).

Au cours des dernières décennies, la recherche incessante de nouvelles approches phytosanitaires plus efficaces et moins dommageables pour l'environnement et la santé humaine a mené au développement de nouveaux outils prometteurs inspirés de processus biologiques et écologiques retrouvés en nature. Par exemple, de nombreux microorganismes possèdent la capacité naturelle d'inhiber la croissance ou même de tuer d'autres espèces de manière à protéger leur niche écologique ou à disposer d'une source de nutriments. Depuis longtemps, consciemment ou non, les hommes ont su tirer avantage de cette réalité afin d'améliorer le rendement des cultures ou la qualité des produits alimentaires disponibles. Cependant, c'est grâce à l'émergence de la science moderne que les causes de tels phénomènes, jusque-là inconnues, ont pu être expliquées. À partir de ce moment, la reconnaissance de l'importance des microorganismes dans de multiples sphères liées aux activités humaines n'a cessé de se renforcer.

En agriculture, la flore microbienne présente dans l'environnement des végétaux cultivés est devenue l'objet d'un intérêt accru. De nouvelles espèces microbiennes ont été décrites et de nouvelles propriétés bénéfiques ou néfastes ont pu leur être associées. De ces connaissances est née une nouvelle manière de concevoir la lutte aux agents nuisibles. Elle allait tirer profit des propriétés de toutes sortes d'organismes afin de promouvoir la santé des végétaux cultivés. Cette nouvelle approche a été baptisée « lutte biologique ». Elle vise à contrôler toute espèce nuisible par le recours à d'autres espèces qui leur sont naturellement antagonistes ou qui promeuvent les réactions de défense des plantes. Elle peut se faire par l'introduction en masse de l'organisme dans le système agricole ou bien par l'aménagement d'environnements le favorisant, s'il y est déjà présent. Elle peut aussi

bien être utilisée dans un contexte de lutte aux espèces végétales indésirables, aux insectes nuisibles ou aux maladies végétales.

2. *Pseudozyma flocculosa*

Les années 1980 ont été particulièrement fertiles dans la recherche d'agents potentiels de lutte biologique contre diverses maladies végétales. C'est dans ce contexte que des efforts ont été entrepris pour colliger et identifier des organismes susceptibles d'être efficaces pour combattre le blanc, une maladie foliaire causée par différentes espèces de champignons Erysiphales (*Oidium lycopersici* (tomate), *Podosphaera fusca* (concombre), *Sphaerotheca pannosa* var. *rosae* (rose), *Blumeria graminis* (céréales)) et qui peut atteindre plusieurs espèces végétales, notamment dans les conditions de culture en serre.

Parmi les différents organismes étudiés, un champignon levuroïde, *Pseudozyma flocculosa*, initialement baptisé *Sporothrix flocculosa* (Traquair *et al.*, 1988), s'est avéré posséder une prédisposition particulière pour attaquer le blanc. Depuis cette découverte, de nombreux efforts ont été déployés afin de révéler les secrets qui se cachent derrière son activité dans le but d'en exploiter les propriétés qui lui permettent d'être si efficace contre le blanc.

Le présent document retrace d'abord, au chapitre 2, l'évolution de nos connaissances concernant *P. flocculosa* depuis sa découverte. On y présente les hypothèses et les résultats qui ont tracé les lignes d'un parcours scientifique singulier jusqu'à l'avènement des technologies de séquençage dites de deuxième génération.

3. Problématiques et découvertes

Au fil des travaux de recherche concernant *P. flocculosa*, de nombreuses hypothèses ont été énoncées quant aux fondements génétiques liés à sa biologie particulière. Toutefois, certains obstacles, comme celui lié à la difficulté de transformer génétiquement le champignon par recombinaison homologue, limitent toujours notre capacité à vérifier certaines d'entre elles. L'avènement des technologies de séquençage hautement parallèle et à haut débit a cependant fait apparaître de nouvelles opportunités de recherche nous permettant d'approfondir nos connaissances au sujet de l'espèce.

Depuis les premiers essais ayant démontré l'activité antifongique de *P. flocculosa*, on soupçonne le champignon d'agir par antibiose. La découverte de la flocculosine, un glycolipide aux propiétés antifongiques pouvant être synthétisé et sécrété en grande quantité par l'ALB, est venue appuyer cette idée.

C'est aussi cette découverte qui a permis d'établir un rapprochement avec le champignon pathogène *Ustilago maydis*, un organisme modèle qui cause la maladie du charbon du maïs, et qui produit, lui aussi, un glycolipide quasi identique à la flocculosine.

Ces deux découvertes expliquent pourquoi il devenait nécessaire d'identifier et de caractériser l'ensemble des gènes permettant la biosynthèse de cette molécule par l'ALB. Il s'agissait donc de vérifier si une batterie de gènes, semblable à celle retrouvée chez *U. maydis*, est présente chez *P. flocculosa* et si les légères différences observées entre les deux structures moléculaires s'expliquent par des différences de composition et d'organisation des gènes y étant associées. Pour ce faire, nous avons dû inactiver un certain nombre de gènes chez *U. maydis* et complémenter chaque souche avec un gène provenant de *P. flocculosa* de manière à vérifier si les gènes possèdent une même fonction en ce qui a trait à la synthèse de leur glycolipide respectif.

Ensuite, à plus grande échelle, la structure et l'organisation du génome de l'ALB suscitaient tout autant de questionnements. Bien que la proximité phylogénétique entre *U. maydis* et *P. flocculosa* ait été démontrée il y a plusieurs années, les détails au sujet du contenu et de la distribution des gènes chez l'ALB demeuraient l'objet de spéculations. En effet, d'un côté, l'analyse du génome de l'agent pathogène *U. maydis,* disponible depuis 2006, a mené à d'importantes découvertes et a généré une masse d'information fort utile pour différents projets (Kämper *et al.*, 2006). D'un autre côté, l'ignorance de ces aspects de la biologie de *P. flocculosa* limitait les possibilités d'avancement quant à la découverte de nouvelles pistes d'explications concernant son pouvoir antagoniste et sa parenté avec les agents pathogènes de l'ordre des Ustilaginales. En caractérisant le génome de l'ALB, nous devenions en mesure de déterminer si certains groupes de gènes démontrent une distribution particulière, signe d'une évolution de l'organisme vers un certain mode de vie qui pourrait nous éclairer dans l'étude de son comportement face au blanc. Pour ce faire, de nombreux outils bio-informatiques ont été utilisés. Un premier travail de reconstruction a

été effectué par l'assemblage du génome, l'identification de l'emplacement physique des gènes et l'identification de leur fonction. Par la suite, différentes analyses ont pu être réalisées, que ce soit sur la base de l'organisation génomique ou sur la base de l'homologie des séquences présentes avec celles disponibles dans le domaine public. Combiné à différentes méthodes de biologie moléculaire utilisées à certaines étapes, ce travail a permis de vérifier plusieurs des hypothèses qui n'avaient toujours pas trouvé de réponses.

D'abord, il a été possible d'identifier et de caractériser la fonction des gènes composant la batterie de gènes responsable de la synthèse de la flocculosine, une molécule glycolipidique aux propriétés antifongiques (chapitre 3). Notre compréhension de la synthèse de cette molécule s'est alors raffinée par l'identification d'un gène dont le produit participe directement à la ß-hydroxylation de la courte chaîne d'acides gras de cette molécule (chapitre 4). Ces travaux ont montré en effet que la structure des molécules étudiées correspondait à l'organisation de la batterie de gènes à laquelle elles étaient associées. Ensuite, une analyse en profondeur de la séquence d'ADN du génome de l'ALB a permis de dresser un portrait global des ressources génétiques disponibles pour cette espèce. En comparant le génome à ceux d'espèces phytopathogènes apparentées, le niveau de conservation de différents éléments génomiques présents chez ces espèces a été déterminé avec précision. Il en résulte un portrait détaillé et une compréhension approfondie de l'évolution des ressources génétiques qui distingue *P. flocculosa* des autres champignons auxquel il a été comparé (chapitre 5). Finalement, le document se termine par une discussion en guise de conclusion (chapitre 6).

4. Références

Enserink, M., Hines, P.J., Vignieri, S.N., Wigginton, N.S. et Yeston, J.S. (2013). The Pesticide Paradox. *Science*, *341*(6147), 728-729.

Kämper, J., Kahmann, R., Bölker, M., Ma, L.J., Brefort, T., Saville, B.J., *et al.* (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*, *444*(7115), 97-101.

Traquair, J.A., Shaw, L.A. et Jarvis, W.R. (1988). New species of *Stephanoascus* with *Sporothrix* anamorphs. *Canadian Journal of Botany*, *66*(5), 926-933.

Chapitre 2 : Mode of Action of Biocontrol Agents: all that Glitters Is not Gold

Bélanger, R.R., Labbé, C., Lefebvre, F. et Teichmann, B. (2012). Mode of action of biocontrol agents: all that glitters is not gold. *Canadian Journal of Plant Pathology*, *34*(4), 469–478

Résumé

Pseudozyma flocculosa (syn.: *Sporothrix flocculosa*) a été découvert et décrit pour la première fois en 1987 comme une levure épiphyte se développant sur les feuilles de trèfle infectées par le blanc. Par la suite, il a été démontré qu'il s'agissait d'un puissant antagoniste des blancs, ce qui a suscité son étude et son développement à titre d'agent de lutte biologique (ALB). La plupart des ALB agissent selon un ou plusieurs des modes suivants: compétition, parasitisme, antibiose et résistance induite. Dans le cas de *P. flocculosa*, des épreuves biologiques in vitro, des études au microscope électronique et des analyses chimiques ont toutes indiqué un seul mode d'action: l'antibiose. Cette conclusion a été renforcée par la caractérisation et la purification d'une molécule active, la flocculosine, ainsi que par la démonstration de sa puissante activité antimicrobienne. Ceci a de plus été appuyé par la découverte d'un groupe de gènes complexe qui régule la synthèse de la flocculosine, une molécule presque identique à l'acide ustilagique – un composé produit par *Ustilago maydis*, déterminé également par un groupe de gènes semblable. Malgré cette solide preuve, il y a de nouveaux indices laissant croire que la flocculosine ne joue qu'un rôle secondaire, voire aucun, dans l'activité antagoniste de *P. flocculosa*. Le processus de lutte biologique semble plutôt relié à une interaction complexe impliquant des nutriments produits par la plante, récoltés par l'agent phytopathogène et utilisés par *P. flocculosa*. Avec la sortie imminente du séquençage du génome de *P. flocculosa*, la publication récente du génome du blanc de l'orge (*Blumeria graminis* f. sp. *hordei*) et la disponibilité du génome de nombreuses plantes, les dernières avancées dans le domaine du traitement des données et des analyses transcriptomiques nous offriront un aperçu sans précédent de l'équilibre complexe et fragile caractérisant cette unique interaction tripartite.

Abstract

Pseudozyma flocculosa (syn: *Sporothrix flocculosa*) was first discovered and described in 1987 as an epiphytic yeast on powdery mildew-infected clover leaves. It was subsequently found to be a powerful antagonist of powdery mildews which prompted its study and development as a biocontrol agent (BCA). Most BCAs exert their activity through the manifestation of one or more of the following modes of action: competition, parasitism, antibiosis and induced resistance. In the case of *P. flocculosa*, in vitro bioassays, electron microscopy studies, and chemical analyses all pointed to a single mode of action: antibiosis. This conclusion was reinforced by the characterization and purification of an active molecule, flocculosin, and by the demonstration of its powerful antimicrobial activity. This was further supported by the discovery of a complex gene cluster regulating the synthesis of flocculosin, a molecule nearly identical to ustilagic acid – a compound produced by *U. maydis* under the control of a similar gene cluster. Despite this strong evidence, there is new evidence to indicate that flocculosin plays a secondary, if any, role in the antagonistic activity of *P. flocculosa*. The biocontrol process instead appears to be mediated by an intricate interaction involving nutrients produced by the plant, harvested by the phytopathogen and exploited by *P. flocculosa*. With the imminent completion of sequencing of the *P. flocculosa* genome, the recent publication of the barley powdery mildew genome (*Blumeria graminis* f. sp. *hordei*), and the availability of many plant genomes, the latest developments in DNA sequencing and transcriptomic analyses will allow unparalleled insight into the complex and delicate balance defining this unique tripartite interaction.

1. Introduction

The development and study of biocontrol agents (BCAs) against plant pathogens has been and continues to remain the focal point of research in many laboratories. While practical implementation of biological control has failed to meet early expectations, many scientific, legal and commercial reasons can explain this situation (Fravel 2005; Paulitz et Bélanger, 2001). In many instances, our inability to properly exploit the intrinsic properties of a BCA accounts for this limited success.

Our laboratory has been working with the yeast-like epiphyte *Pseudozyma flocculosa* (Traquair, L. A. Shaw & Jarvis) Boekhout & Traquair (syn: *Sporothrix flocculosa*) for more than 20 years. From its initial discovery on powdery mildew-infected clover leaves in 1988 (Traquair *et al.*, 1988), its efficacy and ability at controlling members of the Erysiphales have been the driving force for past and current initiatives to develop a biocontrol product with *P. flocculosa* as the main active ingredient. However, from a research perspective, the bulk of our efforts have been devoted to trying to decipher its mode of action in order to improve the expression of its biocontrol properties. From what appeared to be a straightforward case of antibiosis, this endeavour has proven to be significantly more challenging than initially anticipated.

We review in this paper the many milestones that have both enlightened and influenced the study of the mode of action of *P. flocculosa* over the last 20 years. At the same time, we will describe how the path has been filled with fortuitous discoveries, and unexpected phenomena that have led the way to some remarkable and unique research opportunities that could not have been anticipated 20 or even 10 years ago.

2. The mode of action of BCAs

Most BCAs identified to date have been categorized as exerting their activity through the manifestation of one or more of four modes of action: competition, parasitism, antibiosis and induced resistance (Bélanger et Avis, 2002; Whipps 2001). Understanding precisely how BCAs kill their target has always been perceived as the key to maximize the efficacy of BCAs in the field. From a theoretical point of view, if the mode of action can be reduced to a single trait, this greatly facilitates manipulation and over-expression of the trait in key

situations for a BCA to suppress a plant pathogen. As a matter of fact, this strategy has been tried with a few systems with mitigated results. For instance, in cases where parasitism appeared to be the predominant mode of action, several attempts have been made to increase production of lytic enzymes such as chitinases and glucanases (Kubicek *et al.*, 2001; Lorito *et al.*, 2001). The approach targeted either the selection of BCA strains with superior ability to produce such enzymes or the direct cloning and over-expression of relevant genes conferring greater degrading properties. These initiatives, while conceptually sound, failed to deliver BCAs with notable increased activity. In retrospect, it was generally agreed that parasitism (or at least the enzymes associated with this mode of action) is not, as initially proposed (Elad *et al.*, 1982), the key factor in the biocontrol activity of some fungi, notably in the case of *Trichoderma* spp. (Woo *et al.*, 2006).

The mode of action became paramount to the study of every BCA in the early 1990s. At the time, there seemed to have been a necessity to associate a BCA with a specific mode of action. This trend appeared to take its origin in the efforts to register a BCA as a biocontrol product. A rumour quickly circulated that if a BCA acted by antibiosis, i.e. produced antibiotics, regulation agencies around the world would deny or delay registration of the BCA. The source of this misinformation is unknown, but it had a profound impact on the scientific community working in biological control. Scientists started screening bacteria and fungi through bioassays, and systematically rejected organisms that displayed antibiosis in vitro. At the same time, some researchers went to great lengths to demonstrate that their BCAs did not act by antibiosis (Castoria *et al.*, 1997; Elad *et al.*, 1998). While the notion that antibiotic-producing organisms may be more difficult to register still persists, regulation agencies will not automatically reject them. In fact, one of the first biocontrol agents approved for use in Canada, Mycostop (*Streptomyces griseoviridis*) acts primarily through antibiosis. A thorough description of the properties of the active molecules and their fate in the environment will satisfy most requirements.

3. Antibiosis: the case of *Pseudozyma flocculosa*

It became apparent early on that *P. flocculosa* had all the characteristics of acting by antibiosis, as much as we would have liked to describe it as a non-antibiotic producer. Electron microscopic studies showed that the antagonist did not penetrate powdery mildew

cells but induced a rapid cell collapse (Fig. 1 a; Hajlaoui et Bélanger, 1993). In addition, we observed by using gold-labelling of chitin residues that the cell walls of powdery mildew hyphae were not degraded in the presence of *P. flocculosa*. These results eliminated the possibility that the fungus acted through parasitism and production of cell wall-degrading enzymes. The biological fact that powdery mildew fungi are biotrophs de facto ruled out competition for substrate. No evidence of induced resistance was ever observed, so antibiosis appeared to be the sole and logical mode of action by which *P. flocculosa* antagonized fungi from the Erysiphales. This conclusion was reinforced by in vitro bioassays showing clear zones of lysis when *P. flocculosa* was confronted to other fungi (Fig. 1).

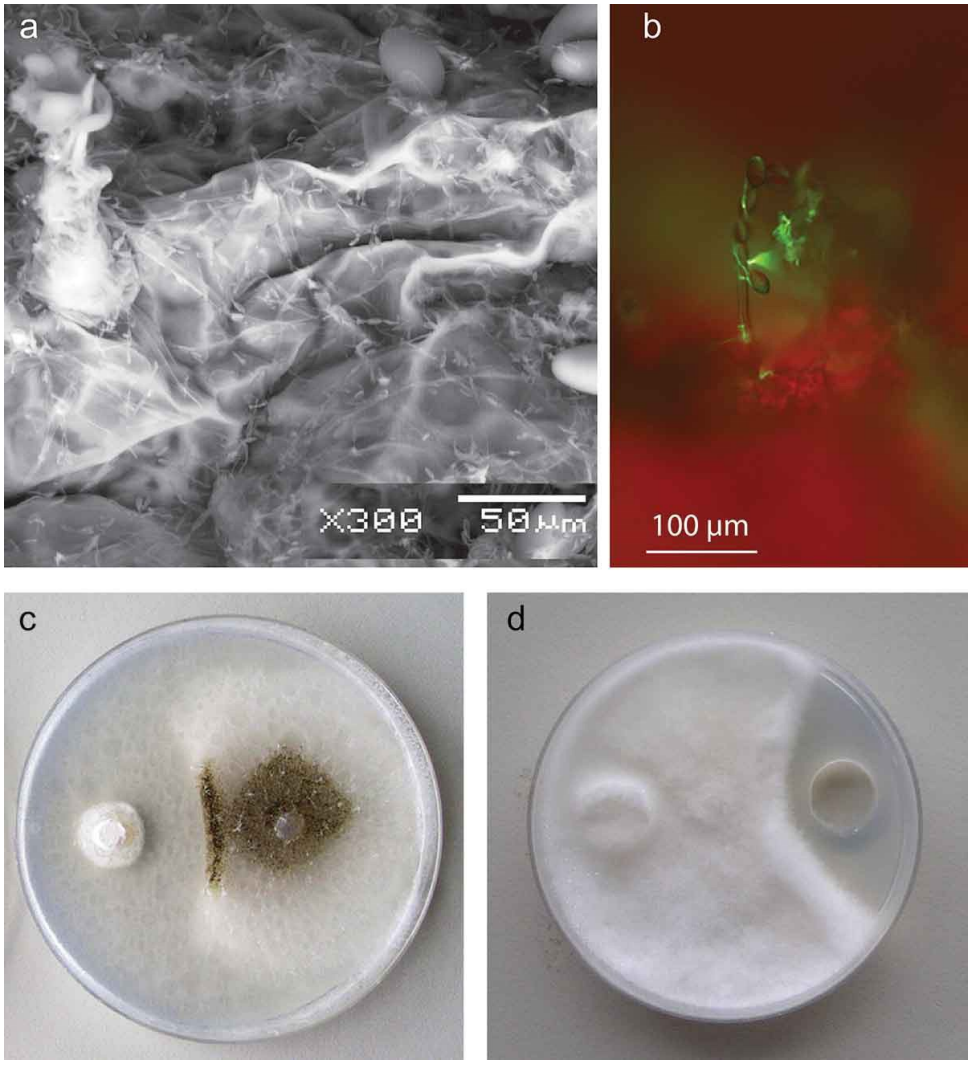


Figure 1 : *Pseudozyma flocculosa*, 24 h post-inoculation, on powdery mildew infected leaves. **a,** under scanning electron microscopy. **b,** under fluorescent microscopy (when using a *P. flocculosa* strain expressing GFP. **c,** *P. flocculosa* against *Phomopsis* sp. on potato dextrose agar (PDA) medium after 2 days. **d,** flocculosin against *Pythium ultimum* grown for 2 days on PDA.

4. The search for active molecules

We began working on the premise that *P. flocculosa* produced molecules with antifungal properties, so we sought to isolate and purify these active compounds. Our initial approach was based on the understanding that secondary metabolites were produced in the later stage of culture. Our work identified three fatty acids displaying strong antimicrobial properties (Choudhury *et al.*, 1994, 1995). Purification and synthesis of the most active molecule, 6-

methyl-9-heptadocenoic acid, provided the opportunity to carry out bioassays aimed at identifying the molecular target of the compound.

The fatty acid induced a rapid plasmolysis of sensitive cells with the plasma membrane being seemingly affected in the presence of the molecule. By testing it against a number of fungi, it appeared that the fatty acid was more deleterious to certain fungi than others. Analysis of the plasma membrane revealed that fungi with a high sterol content were much less sensitive than fungi containing none (e.g. Pythiaceae) or limited sterols (Avis et Bélanger, 2001; Benyagoub *et al.*, 1996). This explained, at least in part, why *P. flocculosa*, with a high sterol content, was not affected by its own secondary metabolites. However, given the fairly large spectrum of fungi affected by the fatty acid, it still raised the question as to why *P. flocculosa* was nearly exclusively antagonistic towards members of the Erysiphales.

5. Toward molecular approaches

A seemingly trivial change occurred at the time we were trying to decipher the mode of action of *P. flocculosa*. Based on new molecular tools to classify fungi, Boekhout (1995) suggested that ascomycetous fungi known hitherto as *Sporothrix* anamorphs, including *Sporothrix flocculosa*, were in fact more closely related to the basidiomycetes and should be reclassified as such under the genus *Pseudozyma*. Further analyses confirmed this finding and Begerow *et al.* (2000) officially renamed *Sporothrix flocculosa* as *Pseudozyma flocculosa* (Fig. 2). We paid little attention to this new classification until we ventured into molecular approaches.

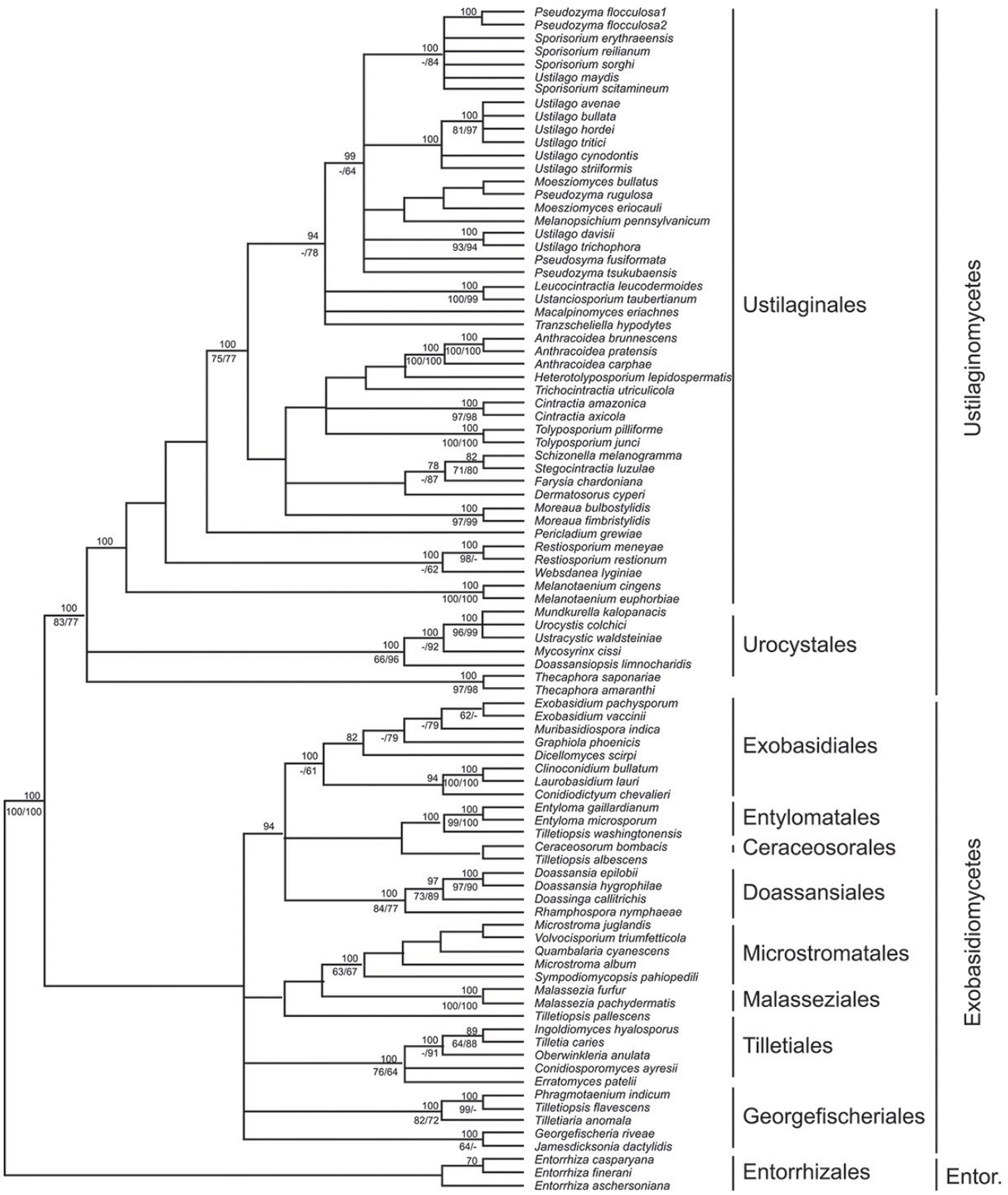


Figure 2 : Phylogenetic relationship among different members of Ustilaginales. Note the close relationship between *Pseudozyma flocculosa* and *Ustilago maydis* in the upper clade. (Adapted from Begerow *et al.*, 2000).

Our initial attempts to transform *P. flocculosa* in the early 2000s were unsuccessful. Based on the premise that the fungus was a yeast-like organism, we tried all the tools available that had been developed with the widely studied yeast *Saccharomyces cerevisiae*, but to no avail (Cheng *et al.*, 2001). Those failures served as incentives to rethink our approach and this was when we considered for the first time the implications of the new classification. What if *P. flocculosa* responded better to tools developed for genetic manipulation of basidiomycete fungi? Upon further analysis, we realized that *P. flocculosa* belonged to the Ustilaginales and was actually closely related to the model fungus *Ustilago maydis* (Fig. 2). This prompted us to try a plasmid commonly used in *U. maydis* transformation. Within a fortnight, we obtained our first *P. flocculosa* transformants (Cheng *et al.*, 2001). Through random insertion of the plasmid, some transformants had seemingly lost their ability to produce antibiotics. By analysing and comparing the metabolites present in the culture media of both the wild-type and the antibiotic-minus transformants, we discovered a compound that was abundant and unique to the wild-type. Although the compound, later named flocculosin, contained fatty acid chains, it appeared to be quite different from the fatty acids identified some 10 years earlier. As a matter of fact, it was so unusual and so complex that the structure we first described turned out to be incorrect (Cheng *et al.*, 2003).

6. Flocculosin and ustilagic acid

The fortuitous discovery of a paper dating back to the early 1950s and describing a glycolipid, ustilagic acid, was arguably the most altering event in our work with *P. flocculosa*. For one, ustilagic acid was nearly similar to flocculosin in its chemical composition and it allowed us to define the proper structure of the latter molecule (Fig. 3; Mimee *et al.*, 2005). It consists of cellobiose, O-glycosidically linked to 3,15,16-trihydroxypalmitic acid. The sugar moiety is acylated with 2-hydroxyoctanoic acid and acetylated at two positions. Secondly and most importantly, ustilagic acid was produced by *U. maydis*, a coincidence reinforcing the phylogenetic link between *P. flocculosa* and *U. maydis*. Furthermore, both molecules were reported to display strong antibiotic properties (Haskins et Thorn, 1951; Mimee *et al.*, 2005, 2009). While this discovery opened up the field of potential research endeavours, it raised the question about the role of ustilagic acid. If flocculosin conferred a biocontrol activity to *P. flocculosa*, what was the

function of ustilagic acid for *U. maydis* considering that the fungus was strictly known as a plant pathogen?

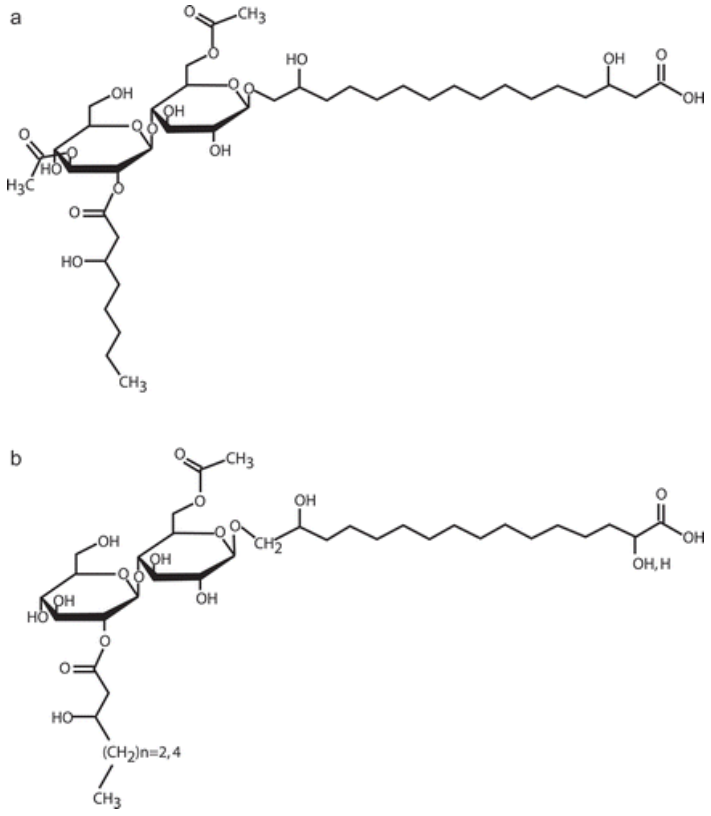


Figure 3 : Chemical structure of flocculosin and ustilagic acid produced by *Pseudozyma flocculosa* and *Ustilago maydis*, respectively. **a,** Flocculosin. **b,** Ustilagic acid.

On the basis of the similarity between the two molecules, M. Bölker's group in Germany took advantage of the annotated genome of *U. maydis* and the particular structure of the long chain fatty acid on the glycolipid to hypothesize that a gene coding for a cytochrome P450 monooxygenase had to be involved in the synthesis of ustilagic acid. Hewald *et al.* (2005) were thus able to identify the gene *cyp1* responsible for the terminal hydroxylation of the long chain fatty acid conferring the link between the fatty acid and the sugar moiety. This discovery was the cornerstone of scientific inferences leading to the description of a unique cluster of 12 genes involved in the complete synthesis of the glycolipid (Teichmann *et al.*, 2007).

We followed in the footsteps of that discovery and the phylogenetic link between *P. flocculosa* and *U. maydis* to hypothesize that *P. flocculosa* must contain a homolog of

cyp1 in order to synthesize flocculosin. Using the *U. maydis cyp1* cDNA as a probe against all known species of *Pseudozyma*, we were indeed able to show that it hybridized specifically with *P. flocculosa* (Marchand *et al.*, 2009) which was the only other strain producing flocculosin. This indicated that *cyp1* had to be involved in flocculosin production. The presence of *cyp1* in *P. flocculosa* raised the obvious possibility of the existence of a cluster similar to the one found in *U. maydis* regulating the production of flocculosin.

7. Genomic sequencing of *P. flocculosa*

It soon became evident that genome sequencing and assembly of *P. flocculosa* were a necessity if we were to make significant headway into the understanding of the implication of flocculosin in the biocontrol activity of *P. flocculosa*. Using Roche 454 Titanium technology, we generated 525 Mb of shotgun data and 167 Mb of 2.6 and 4.5 kb mate-pair sequences for a ca. 30× coverage of the genome. The latest assembly contains 40 scaffolds and has a N50 value of nearly 1 Mb for an estimated genome size of 22 Mb, similar to that of *U. maydis*. On the basis of sequence homology with genes found in *U. maydis*, we were able to find a gene cluster comprising 10 genes that were necessary for the biosynthesis of flocculosin (Teichmann *et al.*, 2010) (Fig. 4). In contrast to the cluster of *U. maydis*, the flocculosin biosynthesis cluster contains an additional gene encoding an acetyl-transferase and is lacking a gene homologous to the α-hydroxylase Ahd1 necessary for UA hydroxylation. The functions of three acyl/acetyl-transferase genes (Fat1, Fat2 and Fat3) including the additional acetyl-transferase were studied by complementing the corresponding *U. maydis* mutants (Teichmann *et al.*, 2011b). This showed that the additional acetyl-transferase is necessary for acetylation of the glucose moiety, explaining the differences between the two molecules.

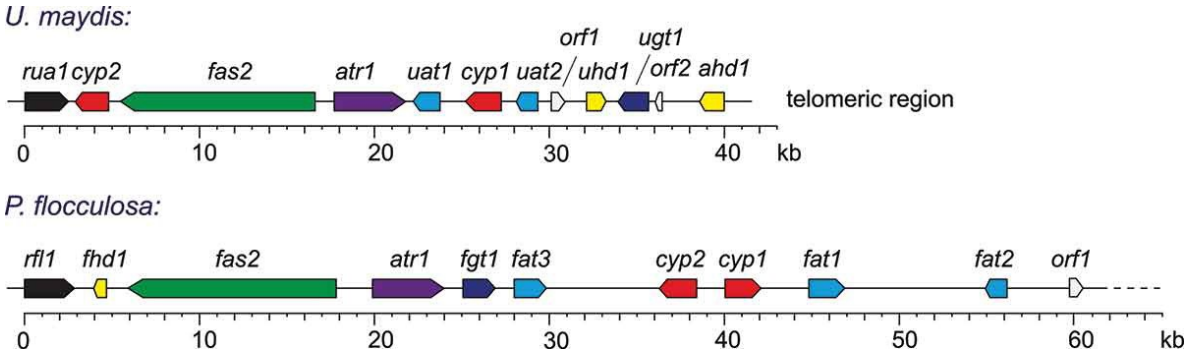


Figure 4 : Comparative alignment of gene clusters found in *Ustilago maydis* and *Pseudozyma flocculosa* controlling the biosynthesis of ustilagic acid (*U. maydis*) and flocculosin (*P. flocculosa*).

8. Is flocculosin involved in the biocontrol activity of *P. flocculosa?*

The fortuitous discovery that two seemingly unrelated fungi with opposite lifestyles would be found to be very similar genetically has created a very unique opportunity to investigate the subtle genetic and molecular phenomena that differentiate the two types of pathogens; a BCA and a plant pathogen. At the same time, it has raised important questions about the role of the glycolipids in the respective ecology of the two fungi.

On the basis of the new phylogeny of Ustilaginales, it is quite clear that *Pseudozyma* spp. are either anamorph vestiges or descendants of *Ustilago* spp. which have lost the ability (or the necessity) for sexual reproduction. However, as the diploid state is essential to initiate the infection process in smut fungi (Bakkeren *et al.*, 2008), *Pseudozyma* spp. have not maintained the pathogenic state of their diploid counterparts. In the same vein, it is quite revealing that *U. maydis* can produce ustilagic acids only in the haploid state (Hewald *et al.*, 2005). Therefore, is the glycolipid-production trait in *U. maydis* really linked to pheromone recognition as suggested by Hewald *et al.* (2005), considering that *P. flocculosa*, the only other known producer of similar glycolipids, has lost its mating capabilities? Is this trait essential for *P. flocculosa* survival or more a measure of its fitness? If it is the latter case, does it have the same ecological role during the anamorph state of *U. maydis* which could explain the conservation of the trait among related species?

We had always assumed that antibiosis was the main mode of action of *P. flocculosa* based on the overwhelming evidence supporting this conclusion. Because it appeared to be

dictated by the ability to produce flocculosin, it would thus stand to reason that any organism capable of synthesizing a similar glycolipid would have biocontrol properties. Bölker's group (Teichmann *et al.*, 2007) had suggested that *U. maydis* had biocontrol potential against *Botrytis cinerea*. However, this suggestion came from the limited observation that ustilagic acid was fungitoxic to the spores of *B. cinerea*.

In an effort to correlate, or refute, the production of flocculosin (or ustilagic acid) with biocontrol properties, we tested and compared *P. flocculosa*, *U. maydis* and other species of *Pseudozyma* which did or did not produce glycolipids. In all our bioassays, only *P. flocculosa* was capable of antagonizing powdery mildews. Using GFP-technology, we observed that none of the other tested fungi could grow over the colonies, while *P. flocculosa* developed abundantly in the presence of the powdery mildew host. Interestingly, this result was confirmed by q-PCR analyses whereby the population of *P. flocculosa* increased dramatically within the first 72 h while that of the other compared fungi declined (Fig. 5; Clément-Mathieu *et al.*, 2008).

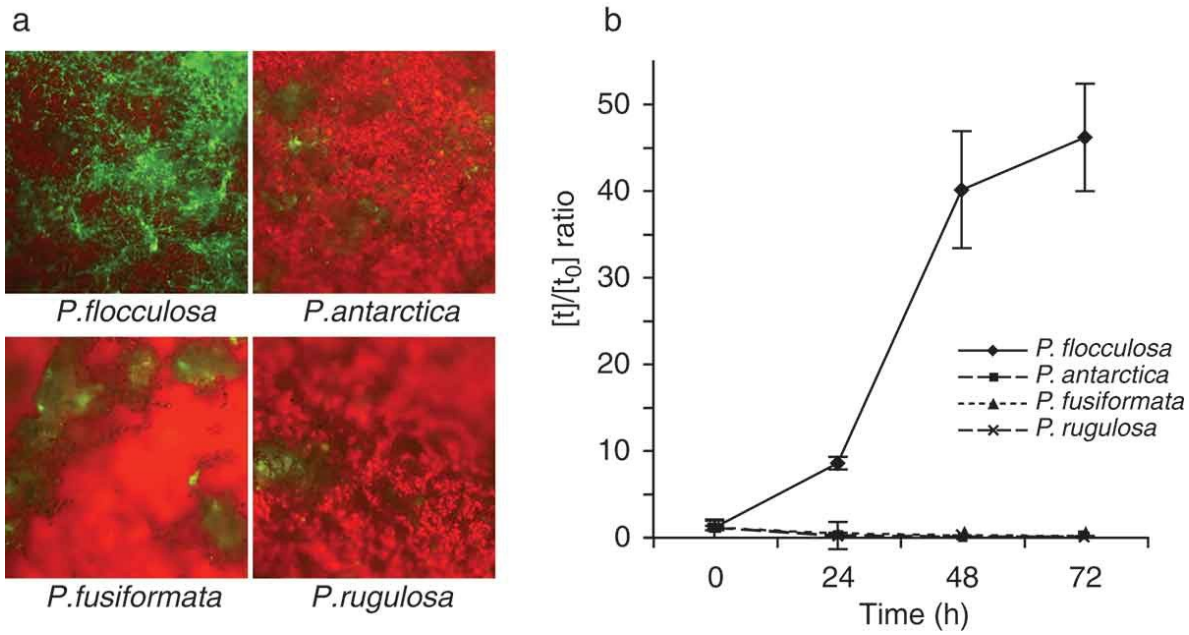


Figure 5 : **a**, Microscope observations of leaf colonization by four GFP-transformed *Pseudozyma* spp. on *Podosphaera xanthii*-infected leaf disks from *Cucumis sativus* L. 'Corona', 24 h after treatment. **b**, Quantification by RT (real time)-PCR of population development over time of four *Pseudozyma* spp. following inundative application of conidial suspensions on *P. xanthii*-infected leaf disks from *Cucumis sativus* L. 'Corona'. Ratios represent the number of cells at a specific time ([t]) divided by the initial number of sprayed cells ([t_0]). Error bars represent the standard error for each treatment.

We concluded from these latter results that the biocontrol specificity of *P. flocculosa* could not be attributed to its ability to produce glycolipids alone, given that other similar organisms producing the same glycolipids were incapable of biocontrol activity. The results rather suggested that the particular properties of *P. flocculosa*, if modulated by flocculosin, were highly dependent on other factors stimulating the growth and development of the yeast-like fungus in presence of powdery mildew colonies.

9. The search for an alternate mode of action

We had to contemplate a new hypothesis after nearly 20 years of working on the assumption that antibiosis was the mode of action of *P. flocculosa*. With access to new genetic tools, we set out to determine if subtle factors came into play to provide *P. flocculosa* with special mechanisms to release flocculosin when in contact with powdery mildews. For this purpose, we followed the expression of *cyp1*, a key gene involved in the synthesis of flocculosin. Initially, in vitro experiments confirmed that the gene was highly

over-expressed when the fungus was grown under conditions conducive to flocculosin release (Fig. 6 a; Marchand *et al.*, 2009). This result confirmed that the gene was a reliable marker to follow flocculosin release in situ. We proceeded to follow the expression of the gene over time following the confrontation between *P. flocculosa* and powdery mildew colonies. We surmised that if flocculosin was important in the process, the fungus would release it in the early stages of antagonism given that it usually overtook colonies within 12 h. This notion was proven incorrect as we could not find any significant increase in *cyp1* expression at any time during the antagonistic process (Fig. 6 b). It thus appears that in spite of all the evidence and the conservation of a complex trait aimed at producing an antifungal compound, the latter bears no direct role in the biocontrol activity of *P. flocculosa*. The ultimate demonstration would obviously come from engineering a flocculosin-minus mutant and further assessment of the phenotype. However, despite repeated attempts and application of different approaches, it has been impossible so far to obtain *P. flocculosa* mutants via homologous recombination. On the other hand, Bölker's group was able to obtain *U. maydis* mutants altered in their ability to produce ustilagic acid. Such mutants were still able to mate and infect, which prompted Hewald *et al.* (2005) to suggest that ustilagic acid acted possibly as a pheromone facilitating recognition between two compatible mating strains. If it is indeed the case, this would mean that the trait serves a different purpose for *P. flocculosa*, given that it is the only other known producer of similar glycolipids, and it has lost its mating capabilities.

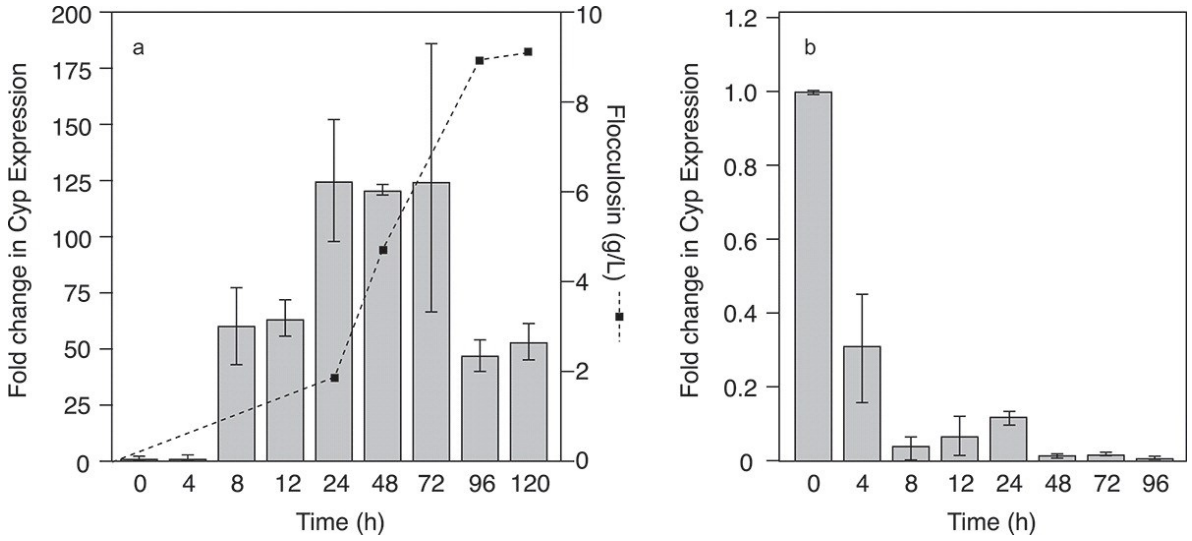


Figure 6 : Reverse transcription RT-PCR analysis of *Pseudozyma flocculosa actin*-normalized *cyp1* expression with **(a)** flocculosin quantification within a growth medium conducive to flocculosin production and **(b)** after inundative applications of conidia on powdery mildew-infected cucumber leaf discs. Error bars represent the standard error for each treatment.

It is now clear that antibiosis, if involved at all, represents a minor component in the overall antagonistic process of *P. flocculosa* against powdery mildews. Considering that BCAs are reported to act by one (or more) of four modes of action – competition, antibiosis, parasitism or resistance inducer – we have seemingly run out of alternatives. Parasitism and induced resistance can be ruled out on the basis of overwhelming evidence against either mode of action (Marchand *et al.*, 2007). This leaves competition, a mechanism highly improbable in the context of the specificity of *P. flocculosa* against biotrophic powdery mildews. Indeed, considering the latter having no saprophytic phase, it is hard to conceive how competition for substrate or ecological niche would take place. Besides, the activity of *P. flocculosa* is manifest when the powdery mildews are anchored into the epidermal cells.

A possible explanation came from the observation of a peculiar phenomenon. *Pseudozyma flocculosa* will display distinct morphologies in culture depending on the nutrients in the media. Incidentally, the morphology resembling the most typical of *P. flocculosa* growth on powdery mildew colonies can be easily replicated by adjusting micro-elements. By systematically adding or subtracting individual micro-elements in the medium, we observed that the complex of Zn/Mn played a key role in the interaction. On the other hand, *U. maydis* did not appear to have the same nutritional requirements and hence lacked the

ability to colonize powdery mildews. Strangely enough, a simple solution of Zn/Mn sprayed on a leaf stimulated the development of *P. flocculosa* in absence of the pathogen (Hammami *et al.*, 2011). The latter interactions indicate that some form of competition or exchange of nutrients takes place between *P. flocculosa* and powdery mildews. The BCA relies on micronutrients for its development and would draw them from the pathogen, which draws them from the plant. It would thus constitute an intricate tripartite interaction. Evidence supporting this sophisticated system of antagonism is threefold : (1) the development of *P. flocculosa* is halted as soon as the powdery mildew colonies have collapsed thus interrupting the flow of nutrients from the plants; (2) *P. flocculosa* will not antagonize powdery mildew spores separated from their host; and (3) this phenomenon can only be observed with the biotrophic powdery mildews that maintain an intimate association with the plant and not with necrotrophs such as *Botrytis cinerea* (Hammami *et al.*, 2011). Interestingly, the only instance under which we have observed the upregulation of genes involved in flocculosin synthesis is when *P. flocculosa* was sprayed on leaves infected with *B. cinerea*. We surmise that sugars released by macerated plant tissues explain this phenomenon since flocculosin will invariably be released in vitro in media rich in sugars and poor in nutrients (Hammami *et al.*, 2008).

10. Transcriptomic analyses

What remains unclear are the factors that would elicit such an unusual interaction? How does *P. flocculosa* recognize powdery mildews as a host? What are the genes involved in this process? What is present (or activated) in *P. flocculosa* and not in *U. maydis* to explain their diametrically opposed lifestyles in spite of their close genetic relatedness? We believe that the latest developments in high-throughput DNA sequencing offers unparalleled opportunities to decipher the minute details of this complex BCA–plant pathogen–plant system. Indeed, the powdery mildew genome of *Blumeria graminis* f. sp. *hordei* (powdery mildew of barley) has recently been sequenced (Spanu *et al.*, 2010) and is now publicly available. We are in the process of completing the annotation of *P. flocculosa* genome, a task that is highly facilitated by relying on the reference fungi *U. maydis*, *U. hordei* and *Sporisorium reilianum*. This means that it is now possible to contemplate the in-depth analysis of the differential transcriptomic response over time of both *P. flocculosa* and

powdery mildew when the former antagonizes the latter. This future project should reveal the salient genes involved in the antagonistic activity of *P. flocculosa*. This can be further supported by using *U. maydis* as a negative control and looking into the expression, or lack thereof, of virulence genes. This should provide invaluable information into the evolutionary separation between plant pathogens and BCAs. At some point, it will also be interesting to incorporate the host plant and conduct a simultaneous analysis of the transcriptome of all actors in this tripartite interaction. A difficult task, but one that may be well worth the effort.

11. Conclusion

Pseudozyma flocculosa is an effective antagonist of powdery mildews but its specific activity toward this particular group of plant pathogens appears to be a lot more intricate and complex than what was hypothesized for nearly 20 years. In light of all the accumulated evidence, it does seem counter-intuitive, however, that a BCA would conserve an elaborate trait such as a potent and unusual antifungal molecule and yet evolve a different mechanism to antagonize other fungi. Recent advancements in DNA sequencing offer powerful tools to investigate these subtle genetic determinants of BCA–pathogen–plant interactions. Exploiting these advancements will reveal fascinating phenomena and useful information to optimize biological control of plant pathogens.

12. References

Avis, T.J. and Bélanger, R.R. (2001). Specificity and mode of action of the antifungal fatty acid *cis*-9-heptadecenoic acid produced by *Pseudozyma flocculosa. Applied and Environmental Microbiology*, *67*(2), 956-60.

Bakkeren, G., Kämper, J. and Schirawski, J. (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. *Fungal Genetics and Biology*, *45*(Suppl. 1), S15-21.

Begerow, D., Bauer, R. and Boekhout, T. (2000). Phylogenetic placements of ustilaginomycetous anamorphs as deduced from nuclear LSU rDNA sequences. *Mycological Research*, *104*(1), 53-60.

Benyagoub, M., Willemot, C. and Bélanger, R.R. (1996). Influence of a subinhibitory dose of antifungal fatty acids from *Sporothrix flocculosa* on cellular lipid composition in fungi. *Lipids*, *31*(10), 1077-82.

Bélanger, R.R. and Avis, T.J. (2002). Ecological processes and interactions occuring in leaf surface fungi. In S. E. Lindow, E. I. Hecht-Poinar, and V. J. Elliott (Eds.), *Phyllosphere microbiology* (pp. 193-208). St. Paul, MN: The American Phytopathological Society.

Castoria, R., De Curtis, F., Lima, G. and De Cicco, V. (1997). β-1,3-glucanase activity of two saprophytic yeasts and possible mode of action as biocontrol agents against postharvest diseases. *Postharvest Biology and Technology*, *12*(3), 293-300.

Cheng, Y., Belzile, F., Tanguay, P., Bernier, L. and Bélanger, R.R. (2001). Establishment of a gene transfer system for *Pseudozyma flocculosa*, an antagonistic fungus of powdery mildew fungi. *Molecular Genetics and Genomics*, *266*(1), 96-102.

Cheng, Y., McNally, D.J., Labbe, C., Voyer, N., Belzile, F. and Bélanger, R.R. (2003). Insertional mutagenesis of a fungal biocontrol agent led to discovery of a rare cellobiose lipid with antifungal activity. *Applied and Environmental Microbiology*, *69*(5), 2595.

Choudhury, S.R., Traquair, J.A. and Jarvis, W.R. (1994). 4-Methyl-7,11-heptadecadienal and 4-methyl-7,11-heptadecadienoic acid: new antibiotics from *Sporothrix flocculosa* and *Sporothrix rugulosa. Journal of Natural Products*, *57*(6), 700-704.

Choudhury, S.R., Traquair, J.A. and Jarvis, W.R. (1995). New extracellular fatty acids in culture filtrates of *Sporothrix flocculosa* and *S. rugulosa. Canadian Journal of Chemistry*, *73*(1), 84-87.

Clément-Mathieu, G., Chain, F., Marchand, G. and Bélanger, R.R. (2008). Leaf and powdery mildew colonization by glycolipid-producing *Pseudozyma* species. *Fungal Ecology*, *1*(2-3), 69-77.

Elad, Y., Chet, I. and Henis, Y. (1982). Degradation of plant pathogenic fungi by *Trichoderma harzianum. Canadian Journal of Microbiology*, *28*(7), 719-725.

Elad, Y., Kirshner, B., Yehuda, N. and Sztejnberg, A. (1998). Management of powdery mildew and gray mold of cucumber by *Trichoderma harzianum* T39 and *Ampelomyces quisqualis* AQ10. *BioControl*, *43*(2), 241-251.

Fravel, D.R. (2005). Commercialization and implementation of biocontrol. *Annual Review of Phytopathology*, *43*, 337-59.

Hajlaoui, M.R. and Bélanger, R.R. (1993). Antagonism of the yeast-like phyllophane fungus *Sporothrix flocculosa* against *Erysiphe graminis* var *tritici. Biocontrol Science and Technology*, *3*, 427-34.

Hammami, W., Castro, C.Q., Remus-Borel, W., Labbe, C. and Bélanger, R.R. (2011). Ecological basis of the interaction between *Pseudozyma flocculosa* and powdery mildew fungi. *Applied and Environmental Microbiology*, *77*(3), 926.

Hammami, W., Labbé, C., Chain, F., Mimee, B. and Bélanger, R.R. (2008). Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa. Applied Microbiology and Biotechnology*, *80*(2), 307-15.

Haskins, R.H. and Thorn, J.A. (1951). Biochemistry of the Ustilaginales VII. Antibiotic activity of ustilagic acid. *Canadian Journal of Botany*, *29*(6), 585-592.

Hewald, S., Josephs, K. and Bolker, M. (2005). Genetic analysis of biosurfactant production in *Ustilago maydis*. *Applied and Environmental Microbiology*, *71*(6), 3033.

Kubicek, C.P., Mach, R.L., Peterbauer, C.K. and Lorito, M. (2001). *Trichoderma*: from genes to biocontrol. *Journal of Plant Pathology*, *83*(2), 11-23.

Lorito, M., Scala, F., Zoina, A. et Woo, S. (2001). Enhancing biocontrol of fungal pests by exploiting the *Trichoderma* genome. In M. Vurro et al. (Eds.), *NATO Science Series I: Life and Behavioural Sciences: Enhancing biocontrol agents and handling risks* (Vol. 339, p. 248). Amsterdam, Netherlands: IOS Press.

Marchand, G., Clément-Mathieu, G., Neveu, B. and Bélanger, R.R. (2007). Enhancing biological control efficacy of yeasts to control fungal diseases through biotechnology. In Z. K. Punja, S. H. De Boer, and H. Sanfaçon (Eds.), *Biotechnology and plant disease management* (pp. 518-531). Wallingford, UK: CABI.

Marchand, G., Rémus-Borel, W., Chain, F., Hammami, W., Belzile, F. and Bélanger, R.R. (2009). Identification of genes potentially involved in the biocontrol activity of *Pseudozyma flocculosa*. *Phytopathology*, *99*(10), 1142-1149.

Mimee, B., Labbé, C., Pelletier, R. and Bélanger, R.R. (2005). Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa*. *Antimicrobial Agents and Chemotherapy*, *49*(4), 1597-9.

Mimee, B., Pelletier, R. and Bélanger, R.R. (2009). In vitro antibacterial activity and antifungal mode of action of flocculosin, a membrane-active cellobiose lipid. *Journal of Applied Microbiology*, *107*(3), 989-96.

Paulitz, T.C. and Bélanger, R.R. (2001). Biological control in greenhouse systems. *Annual Review of Phytopathology*, *39*(1), 103-133.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, *330*(6010), 1543-6.

Teichmann, B., Lefebvre, F., Labbé, C., Bölker, M., Linne, U. and Bélanger, R.R. (2011). Beta hydroxylation of glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-dependent β-hydroxylase. *Applied and Environmental Microbiology*, *77*(21), 7823-7829.

Teichmann, B., Linne, U., Hewald, S., Marahiel, M.A. and Bölker, M. (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Molecular Microbiology*, *66*(2), 525-33.

Teichmann, B., Liu, L., Schink, K.O. and Bölker, M. (2010). Activation of the ustilagic acid biosynthesis gene cluster in Ustilago maydis by the C2H2 zinc finger transcription factor Rua1. *Applied and Environmental Microbiology*, *76*(8), 2633-40.

Traquair, J.A., Shaw, L.A. and Jarvis, W.R. (1988). New species of *Stephanoascus* with *Sporothrix* anamorphs. *Canadian Journal of Botany*, *66*(5), 926-933.

Whipps, J.M. (2001). Microbial interactions and biocontrol in the rhizosphere. *Journal of Experimental Botany*, *52*(suppl 1), 487-511.

Woo, S.L., Scala, F., Ruocco, M. and Lorito, M. (2006). The molecular biology of the interactions between *Trichoderma* spp., phytopathogenic fungi, and plants. *Phytopathology*, *96*(2), 181-5.

Chapitre 3 : Identification of a Biosynthesis Gene Cluster for Flocculosin a Cellobiose Lipid Produced by the Biocontrol Agent *Pseudozyma flocculosa*

Teichmann, B., Labbé, C., Lefebvre, F., Bölker, M., Linne, U. et Bélanger, R.R. (2011). Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa. Molecular Microbiology*, *79*(6), 1483-1495

Résumé

La flocculosine est un glycolipide antifongique produit par le champignon agent de lutte biologique *Pseudozyma flocculosa*. Il s'agit d'un cellobiose attaché par O-glycosylilation à l'acide 3,15,16-trihydroxypalmitique. La fraction glucidique est acétylée avec un acide 2-hydroxy-octanoïque à deux positions. Ici, nous décrivons l'organisation d'une batterie de gènes comprenant 11 gènes nécessaires pour la biosynthèse de la flocculosine. Nous avons comparé cette batterie à celle associée à la synthèse de l'acide ustilagique, un glycolipide quasi identique produit par le champignon phytopathogène *Ustilago maydis*. Contrairement à la batterie retrouvée chez ce dernier, la batterie de gènes de *P. flocculosa* contient un gène additionnel encodant une acétyl-transférase tout en étant dépourvue du gène homologue à l'α-hydrolase Ahd1, nécessaire pour l'hydroxylation de l'acide ustilagique. Les fonctions des trois gènes codant les acyl/acetyl-transférases (*fat1*, *fat2* et *fat3*), incluant l'acétyl-transférase additionnelle, ont été étudiées en complémentant les mutants d'*U. maydis* correspondants. Alors que les séquences des protéines Fat1 et Fat3 chez *P. flocculosa* sont homologues à la protéine Uat1 chez *U. maydis*, la protéine Fat2 partage un niveau d'identité de 64 % avec la protéine Uat2, une protéine impliquée dans la biosynthèse de l'acide ustilagique, mais dont la fonction était inconnue jusqu'à maintenant. Par analyse génétique et spectrométrie de masse, nous montrons qu'Uat2 et Fat2 sont nécessaires pour l'acétylation du glycolipide correspondant. Ces résultats apportent un nouvel éclairage aux propriétés de lutte biologique de *P. flocculosa* et de nouvelles opportunités afin d'accroître son efficacité.

Summary

Flocculosin is an antifungal glycolipid produced by the biocontrol fungus *Pseudozyma flocculosa*. It consists of cellobiose, O-glycosidically linked to 3,15,16-trihydroxypalmitic acid. The sugar moiety is acylated with 2-hydroxy-octanoic acid and acetylated at two positions. Here we describe a gene cluster comprising 11 genes that are necessary for the biosynthesis of flocculosin. We compared the cluster with the biosynthesis gene cluster for the highly similar glycolipid ustilagic acid (UA) produced by the phytopathogenic fungus *Ustilago maydis*. In contrast to the cluster of *U. maydis*, the flocculosin biosynthesis cluster contains an additional gene encoding an acetyl-transferase and is lacking a gene homologous to the α-hydroxylase Ahd1 necessary for UA hydroxylation. The functions of three acyl/acetyl-transferase genes (Fat1, Fat2 and Fat3) including the additional acetyl-transferase were studied by complementing the corresponding *U. maydis* mutants. While *P. flocculosa* Fat1 and Fat3 are homologous to Uat1 in *U. maydis*, Fat2 shares 64% identity to Uat2, a protein involved in UA biosynthesis but with so far unknown function. By genetic and mass spectrometric analysis, we show that Uat2 and Fat2 are necessary for acetylation of the corresponding glycolipid. These results bring unique insights into the biocontrol properties of *P. flocculosa* and opportunities for enhancing its activity.

1. Introduction

Many fungi have the ability to synthesize secondary metabolites, which are produced under certain environmental conditions or at selected times during the life cycle of the organism. These bioactive compounds are not considered essential for viability but will usually help fungal cells cope with environmental stresses or compete with other microorganisms (Hoffmeister et Keller, 2007).

Biocontrol agents (BCA), namely bacteria and fungi, will often rely on the release of secondary metabolites to reduce the activity or destroy the plant pathogens that they antagonize (Shoda 2000). The basidiomycetous fungus *Pseudozyma flocculosa* (Traquair, Shaw & Jarvis) Boekhout & Traquair has been characterized for its antagonistic activity against powdery mildew pathogens (Paulitz et Bélanger, 2001). Indeed, *P. flocculosa* synthesizes large amounts of the extracellular glycolipid flocculosin, which is believed to be responsible for antibiosis since it shows antimicrobial activity against different microorganisms (Mimee *et al.*, 2005, 2009). Flocculosin consists of 3,15,16-trihydroxypalmitic acid O-glycosidically linked at its terminal hydroxyl group to cellobiose. The cellobiose moiety is further composed of two acetyl groups and a short-chain hydroxy fatty acid (Fig. 1A).

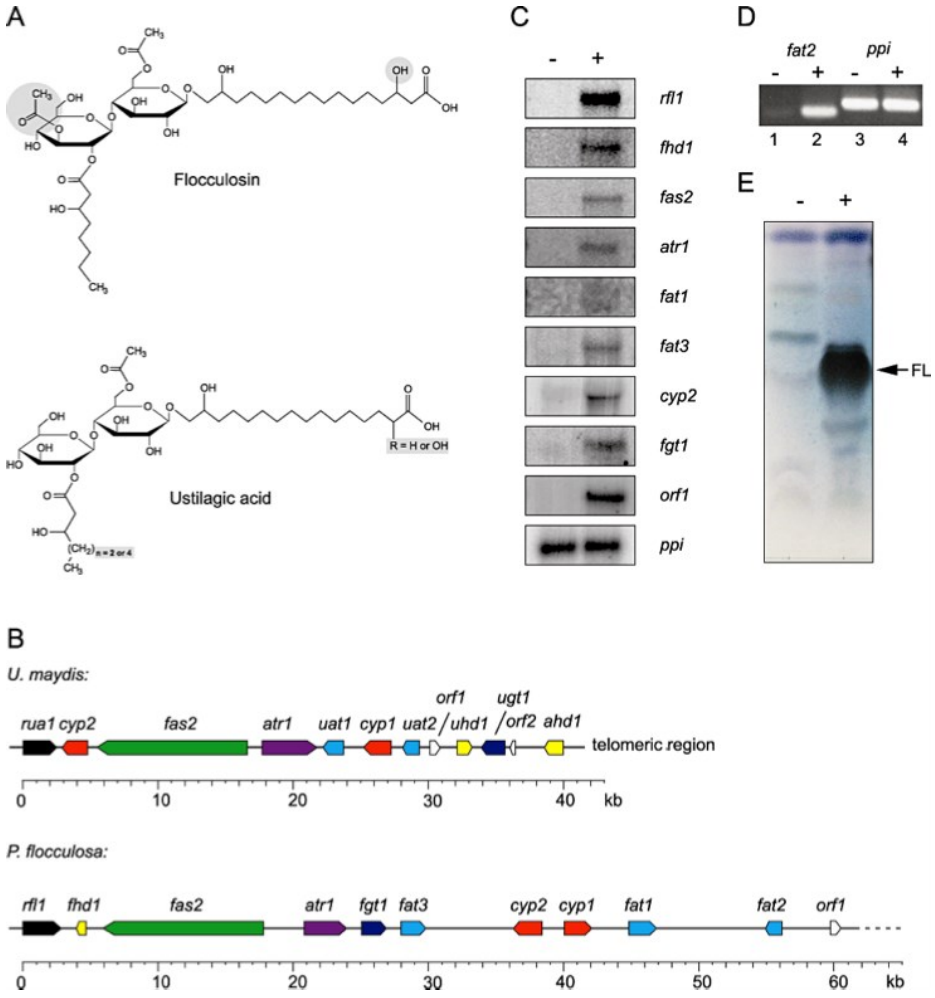


Figure 1 : Molecular structures and gene cluster comparison of flocculosin and ustilagic acid (UA). **A,** Flocculosin as well as UA consists of a cellobiose moiety O-glycosidically linked to a long-chain fatty acid (3,15,16-trihydroxypalmitic acid for flocculosin and 15,16-dihydroxy- or 2,15,16-trihydroxy-palmitic acid for UA). The cellobiose of flocculosin is esterified with 2-hydroxy-octanoic acid and acetylated at two positions. UA in contrast contains a cellobiose esterified with 2-hydroxy-hexanoic or -octanoic acid and is acetylated at only one position. **B,** Genetic organization of the UA and flocculosin biosynthesis gene clusters. For gene designations see Table 1. **C,** Northern blot analysis of cluster genes. Total RNA was prepared from an inducing (+) and non-inducing (−) medium. All cluster genes are strongly upregulated under inducing conditions. The *ppi* gene encoding peptidyl-prolyl *cis-trans* isomerase is constitutively expressed and serves as loading control. **D,** Reverse transcriptase PCR of gene *fat2*. cDNA was prepared from strains under inducing (+) and non-inducing (−) conditions. Amplification with primers specific for *fat2* (lanes 1 and 2) and *ppi* (lanes 3 and 4). E. Thin-layer chromatography (TLC) analysis of extracellular glycolipids produced by *P. flocculosa* under flocculosin (FL) inducing (+) and non-inducing (−) conditions.

At present, besides *P. flocculosa*, only a few fungi, most of them belonging to the Ustilaginales, are known to produce this type of unusual glycolipid. Two other *Pseudozyma*

species, *P. fusiformata* (Kulakovskaya *et al.*, 2005) and *P. graminicola* (Golubev *et al.*, 2008), as well as the causal agent of corn smut, *Ustilago maydis* (de Candolle) Corda (Boothroyd *et al.*, 1956), have been described to produce a similar glycolipid called ustilagic acid (UA) (Fig. 1A), which has also been shown to have antibiotic properties (Haskins et Thorn, 1951; Kulakovskaya *et al.*, 2005).

Recently, the biosynthetic pathway for this glycolipid was identified in *U. maydis* (Teichmann *et al.*, 2007). While flocculosin has only one isomer, UA is a mixture of four different derivatives consisting of 15,16-dihydroxypalmitic or 2,15,16-trihydroxypalmitic acid linked to the cellobiose moiety. In addition the UA molecules contain only one acetyl group and the length of the short-chain hydroxy fatty acid varies between six or eight carbons (Fig. 1A).

The enzymes involved in biosynthesis of UA are encoded by genes which are co-regulated and located in a 45 kb spanning gene cluster encompassing 12 open reading frames (ORFs) (Fig. 1B; Teichmann *et al.*, 2007). The gene cluster is located in the subterminal region of chromosome 23 and contains all genes necessary for the production of this glycolipid including two cytochrome P450 monooxygenases (*cyp1* and *cyp2*), a complete single-chain fatty acid synthase (*fas2*), a glycosyl transferase (*ugt1*), two hydroxylases (*uhd1* and *ahd1*) and a potential export protein of the ABC-transporter family (*atr1*) (Teichmann *et al.*, 2007). The regulation of the gene cluster is mediated by a C2H2 zinc finger transcriptional factor whose gene (*rua1*) is also located within the cluster (Teichmann *et al.*, 2010). In addition, the gene cluster contains three ORFs (*uat1*, *uat2* and *orf1*) whose function has not been confirmed. While *uat1* shows some homology to acetyl-transferases, *uat2* and *orf1* show no homology to known proteins when compared with genetic databases (Teichmann *et al.*, 2007).

In this work, we describe the identification of a biosynthetic gene cluster for flocculosin in *P. flocculosa*. The cluster contains 11 ORFs sharing high similarity to the corresponding *U. maydis* homologues. Interestingly, an additional acetyl-transferase could be identified within the flocculosin gene cluster probably involved in the acetylation of the second acetyl group at the cellobiose moiety. Furthermore, the flocculosin gene cluster lacks a gene homologous to *ahd1*, the α-hydroxylase necessary for α-hydroxylation of the fatty acid.

Notwithstanding these small differences, we propose that the biosynthesis pathway for both glycolipids is highly conserved between these two related fungi in spite of their different lifestyles.

2. Results

Identification of a gene cluster necessary for flocculosin production

In order to identify genes involved in flocculosin biosynthesis, we used the protein sequences of the UA biosynthesis genes in *U. maydis* to scan for homologous genes among the 1281 scaffolds of the recently sequenced genome of *P. flocculosa*. Since both organisms produce nearly identical glycolipids, whose chemical composition is rather unusual, we presumed that the biosynthesis pathways of both fungi could share common steps. Through this approach, we were able to identify specifically 11 homologues of the 12 UA cluster genes in *P. flocculosa*, which lay adjacent to each other also indicating a common cluster organization (Fig. 1B). Because the chromosome organization of *P. flocculosa* has not yet been determined, we were not able to verify the absolute position of the cluster on the chromosome but all these were found on a common scaffold with no indication of partial or complete duplication elsewhere. Interestingly, although the genome size of both organisms is nearly identical [20.5 Mb for *U. maydis* (Kämper *et al.*, 2006) and c. 23.5 Mb for *P. flocculosa* (B. Teichmann, F. Lefebvre and R.R. Bélanger, unpubl. data)], the spanning region of the flocculosin gene cluster covered over 60 kb, a much longer region than the UA gene cluster. Furthermore, the organization of the genes within the flocculosin gene cluster differs from that of the UA biosynthesis cluster (Fig. 1B) and the individual genes contain more introns than the UA biosynthesis genes when compared with homologous sequences of *U. maydis* (Table 1).

Table 1 : Flocculosin biosynthesis genes in *P. flocculosa*

Gene (accession No.)	Size (bp)	Putative function	Number of amino acids	Identity on amino acid level to *U. maydis* orthologue		Intron number[a]	
				Name	E-value	*P. flocculosa*	*U. maydis*
rfl1 (HQ292628)	2205	Regulation of flocculosin gene cluster	734	Rua1	2×10^{-74} (23%)	0	0
fhd1 (HQ292619)	1241	Hydroxylation of short-chain fatty acid	323	Uhd1	6×10^{-46} (95%)	2	1
fas2 (HQ292620)	11461	Synthesis of short and/or long-chain fatty acid	3700	Fas2	0.0 (95%)	3	0
atr1 (HQ292621)	4842	Export of flocculosin	1391	Atr1	0.0 (94%)	6	0
fgt1 (HQ292622)	1747	Glycosylation of fatty acid	549	Ugt1	6×10^{-92} (92%)	1	0
fat3 (HQ292623)	1636	Acetyl-/acylation of cellobiose moiety	511	Uat1	2×10^{-50} (95%)	1	0
cyp2 (HQ292624)	2068	ω-1 Hydroxylation of long-chain fatty acid	654	Cyp2	0.0 (74%)	1	1
cyp1 (ACB59278)	1910	ω-Hydroxylation of long-chain fatty acid	603	Cyp1	0.0 (85%)	1	0
fat1 (HQ292625)	1669	Acetyl-/acylation of cellobiose moiety	519	Uat1	9×10^{-111} (97%)	1	0
fat2 (HQ292626)	1539	Acetyl-/acylation of cellobiose moiety	512	Uat2	8×10^{-14} (64%)	0	2
orf1 (HQ292627)	1155	Unknown	384	Um06465 (orf1)	4×10^{-91} (90%)	0	0

a. *P. flocculosa* introns were predicted from genome sequence based on gene sequence homology to *U. maydis* except for *fat1*, *fat3* and *fhd1* which were determined by cDNA sequencing.

To determine the expression of all cluster genes in relation with flocculosin biosynthesis, we conducted Northern blot analysis after growing *P. flocculosa* under flocculosin-inducing and non-inducing conditions. After 24 h, RNA was prepared and flocculosin production was confirmed by thin-layer chromatography (TLC) (Fig. 1E). All genes were upregulated under flocculosin-inducing conditions (Fig. 1C). In the case of *fat2*, the mRNA was superposed by ribosomal RNA because it has exactly the same size. Therefore, gene expression was confirmed by reverse transcriptase PCR (Fig. 1D).

Sequential analysis of the genes involved in flocculosin biosynthesis

Based on the *U. maydis* homologue *rua1*, the flocculosin gene cluster also seems to be regulated by its own transcription factor located at the left border and designated *rfl1* (regulator of flocculosin biosynthesis) (Fig. 1B, Table 1). It contains a C2H2 zinc finger region, which is highly similar to that of *rua1*. The gene located upstream of *rfl1* is not upregulated under flocculosin-inducing conditions showing that the left border of the gene cluster is defined by *rfl1*. Adjacent to the *rfl1* gene, we identified a gene showing homology to *uhd1*, a gene involved in UA hydroxylation. According to its probable function in flocculosin biosynthesis we called this gene *fhd1* (flocculosin hydroxylase). The next gene within the gene cluster encodes the fatty acid synthase *fas2* likely involved in the synthesis of the long- and/or short-chain fatty acid. Like *fas2* from *U. maydis*, the *P. flocculosa* protein seems to be a fatty acid synthase containing both α- and β-subunit of fatty acid synthases. Interestingly, as in *U. maydis*, the *P. flocculosa fas2* is lacking the C-terminal 4′-phosphopantetheinyl transferase (PPT) subunit. The next gene, *atr1*, encodes a transport protein of the ABC transporter family probably involved in flocculosin export. The glycosylation of the long fatty acid chain is presumably performed by an UDP-glucose-dependent glycosyl-transferase *fgt1*, which shares 92% identity to *ugt1* from *U. maydis*. Flocculosin, as well as UA, is esterified with acetyl and acyl groups. *U. maydis* contains a gene, *uat1*, probably involved in one of these biosynthetic steps. For its part, the flocculosin gene cluster contains two ORFs, *fat1* and *fat3*, with high homology to *uat1* of *U. maydis*. In *U. maydis*, terminal and subterminal hydroxylation of palmitic acid is performed by two cytochrome P450 monooxygenases, Cyp1 and Cyp2 (Teichmann *et al.*, 2007). Homologues of both genes are also present in the biosynthetic gene cluster for flocculosin of which

Cyp1 has already been described to be involved in the biosynthesis of flocculosin (Marchand *et al.*, 2009). Interestingly, the flocculosin cluster is lacking a gene homologous to *ahd1* which was shown to be necessary for the α-hydroxylation of palmitic acid in *U. maydis* (Teichmann *et al.*, 2007). The remaining genes *fat2* and *orf1* showed homology to *uat2* and *orf1* (XP_762612.1) in *U. maydis*, encoding two proteins with unknown function. (Fig. 1B and Table 1).

The genes *uat1*, *fat1* and *fat3* show high similarity to the BAHD superfamily of acetyl-transferases

While UA is acetylated only at one position on the molecule, flocculosin contains an additional acetyl group at the second glucose molecule of the cellobiose moiety (Fig. 1A, shadowed in grey). In order to find proteins involved in the acetylation of the two glycolipids, we analysed the two gene clusters and found that three genes, one in *U. maydis* (*uat1*) and two in *P. flocculosa* (*fat1* and *fat3*), when analysed with the blast algorithm on GenBank, correspond to proteins of the BAHD superfamily of CoA-dependent acetyl-transferases. The deduced amino acid sequences of *uat1*, *fat1* and *fat3* display high similarity with plant proteins frequently involved in the acetylation of secondary metabolites (Fig. 2A). All three proteins share the short sequence motif (Pfam domain 02458) characteristic for the BAHD superfamily of CoA-dependent acetyl-transferases (St-Pierre *et al.*, 1998) (Fig. 2B). Therefore we proposed that *uat1* might be necessary for the acetylation of UA and the *P. flocculosa* homologues *fat1* and *fat3* are the acetyl-transferases transferring the two acetyl groups to flocculosin.

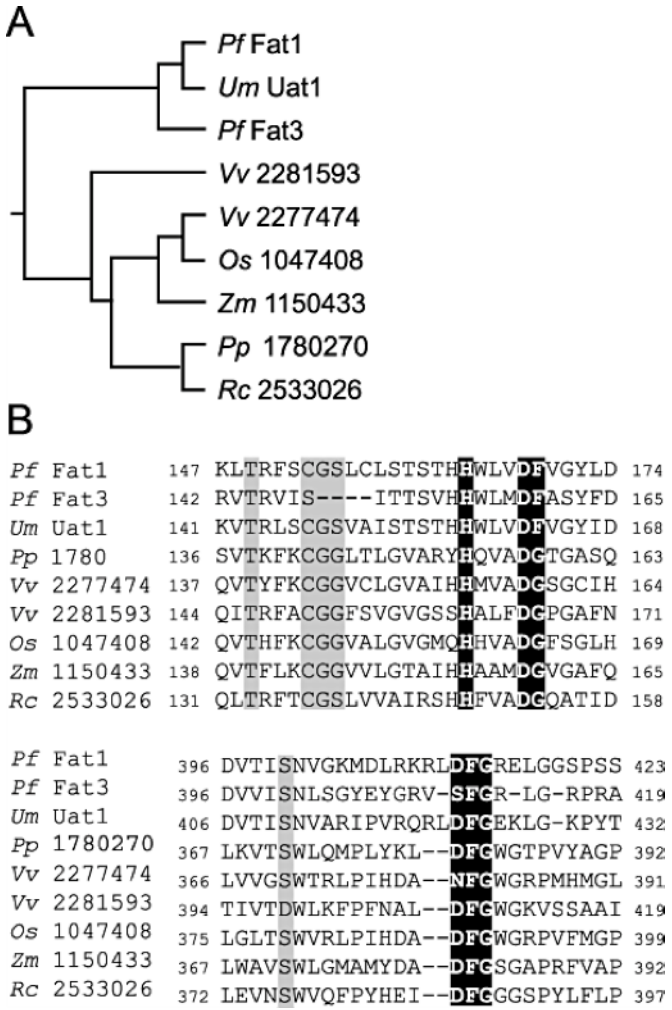


Figure 2 : The *P. flocculosa* proteins Fat1 and Fat3 are members of the BAHD superfamily of acyl-transferases. **A,** Molecular phylogeny of Fat1 and Fat3. The GenBank accession numbers are as follows: Pf Fat1, *P. flocculosa* HQ292625; Pf Fat3, *P. flocculosa* HQ292623; Um Uat1, *U. maydis* XP_762609.1; Pp 1780270, Physcomitrella patens XP_001780270; Vv 2277474, Vitis vinifera XP_002277474; Vv 2281593, V. vinifera XP_002281593; Os 1047408, Oryza sativa NP_001047408; Zm 1150433, Zea mays NP_001150433; Rc 2533026, Ricinus communis XP_002533026. Sequences were aligned using clustalw. **B,** Sequence alignment of the short signature sequences (HXXXDG and DFG) typical for members of the BAHD family of acyltransferases (Pfam domain 02458). The common motif is shaded in black. Highly conserved amino acids are shaded in grey.

Deletion of uat1 in *U. maydis*

In order to confirm the involvement of *uat1* in UA acetylation, we constructed deletion mutants and analysed glycolipid production by mass spectrometry (Fig. 3). LC-MS analysis showed that Δ*uat1* mutant strains had a different UA spectrum when compared with wild-type UAs. Wild-type strains secrete four UA derivatives with masses at 807.3992 *m/z*,

823.3993 *m/z*, 835.4592 *m/z*, 851.4659 *m/z*[M + K]+ (Fig. 3A; Teichmann *et al.*, 2007). As expected, mass spectrometric analysis for Δ*uat1* mutants showed that no wild-type UAs were produced any more. Among the new masses observed, three of them (530.9338 *m/z*, 693.1184 *m/z* and 709.1084 *m/z*[M + K]+) corresponded to altered UAs all lacking the acyl group, instead of the expected acetyl group (Fig. 3B, Table 2). The compound with a mass of 530.9338 *m/z* is a UA derivative lacking the second glucose molecule as well as the acyl group but still carrying the acetyl group. The compounds with the masses of 693.1184 *m/z* and 709.1084 *m/z* are non-acylated UAs; the latter is hydroxylated at the α-position of the long-chain fatty acid (Table 2). The masses 488.8292 *m/z* and 504.8152 *m/z* are UA precursors corresponding to a 2- or 3-hydroxy-fatty acid carrying one glucose molecule. These data indicate that *uat1* encodes an acyl-transferase necessary for the transfer of the shorter fatty acid to the UA precursor, and not the expected acetyl-transferase.

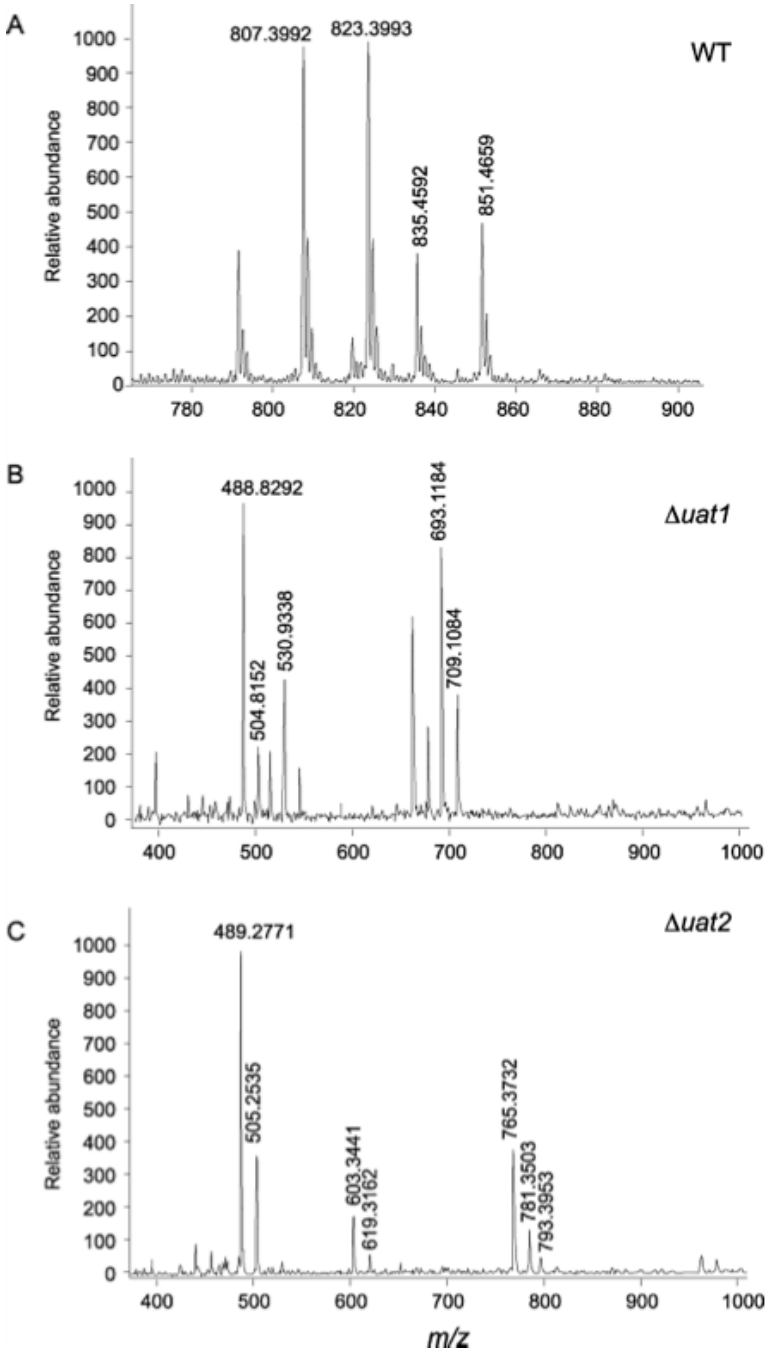
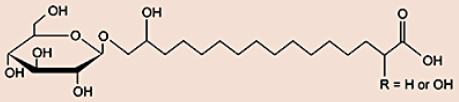
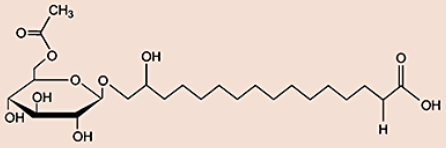
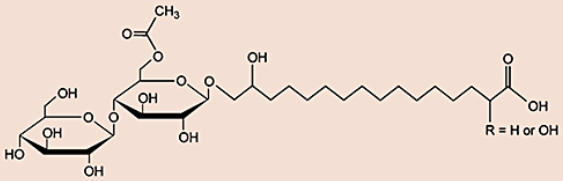
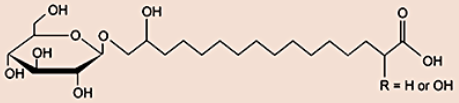
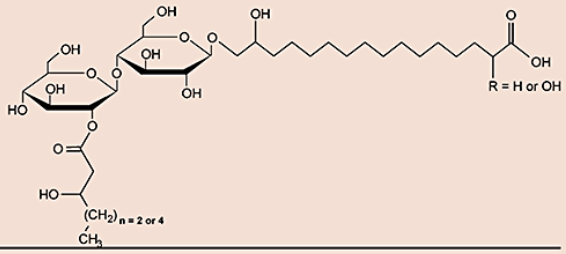


Figure 3 : Analysis of *uat1* and *uat2* deletion strains. LC-MS analysis of glycolipids produced by wild-type, Δ*uat1* and Δ*uat2* mutant strains. **A,** Wild-type strains secrete four UA derivatives with masses at 807.3992 *m/z*, 823.3993 *m/z*, 835.4592 *m/z*, 851.4659 *m/z*$[M + K]^+$. **B**, In FB1Δ*uat1* mutant strains, five masses (488.8292 *m/z*, 504.8152 *m/z*, 530.9338 *m/z*, 693.1184 *m/z* and 709.1084 *m/z*$[M + K]^+$) have been detected and correspond to altered UAs. **C,** FB1Δ*uat2* mutant strains secrete five masses (489.2771 *m/z* and 505.2535 *m/z*, 765.3732 *m/z*, 781.3503 *m/z*, 793.3953 *m/z*, $[M + K]^+$), corresponding to altered UAs.

Table 2 : Structures of the different mass fragments analysed by LC-MS produced by Δ*uat1* and Δ*uat2* mutants

Mutant	Formula	$[M + K]^+$	Structure
Δ*uat1*	$C_{22}H_{42}O_9$/ $C_{22}H_{42}O_{10}$	488.8292/ 504.8152	
	$C_{24}H_{44}O_{10}$	530.9338	
	$C_{30}H_{54}O_{15}$/ $C_{30}H_{54}O_{16}$	693.1184/ 709.1084	
Δ*uat2*	$C_{22}H_{42}O_9$/ $C_{22}H_{42}O_{10}$	489.2771/ 505.2535	
	$C_{34}H_{62}O_{16}$/ $C_{34}H_{62}O_{17}$/ $C_{36}H_{66}O_{16}$	765.3732/ 781.3503/ 793.3953	

Uat2 is required for UA acetylation

In order to identify the acetyl-transferase, we surmised that only *uat2* and *orf1* were possible candidates because they were the only genes for which no homology to proteins with known function was found. Therefore we deleted those genes and analysed the secreted glycolipids with TLC or mass spectrometry. While Δ*orf1* mutants displayed the same phenotype as wild-type strains (data not shown), *uat2* deletion strains showed a different glycolipid pattern compared with the wild-type when analysed with LC-MS (Fig. 3C). Three UA derivatives with masses of 765.3732 *m/z*, 781.3503 *m/z*, 793.3953 *m/z* were produced, corresponding to wild-type UA isoforms each lacking the acetyl group at the sugar moiety (Table 2). The masses 489.2771 *m/z* and 505.2535 *m/z* are, like in Δ*uat1* mutant strains, UA precursors corresponding to a 2- or 3-hydroxy-fatty acid carrying one

glucose molecule. These results clearly confirm that Uat2 is the acetyl-transferase essential for UA acetylation.

Functional analysis of the three acyl/acetyl-transferase genes in *P. flocculosa*

Given that *uat1* was not the expected acetyl-transferase, it was difficult to deduce the function for the two homologous genes *fat1* and *fat3* in *P. flocculosa*, because biosynthesis of flocculosin requires two acetyl-transferases and only one acyl-transferase. To clarify the acyl/acetylation steps during flocculosin biosynthesis, it was therefore necessary to further analyse the function of *fat1* and *fat3* as well as the *uat2* homologue *fat2*. Because *P. flocculosa* does not respond easily to transformation, we were not able to delete the corresponding genes in *P. flocculosa*. Since we had already generated mutant strains in *U. maydis* (Δ*uat1* and Δ*uat2*), we tried to determine the function of *fat1*, *fat2* and *fat3* by complementing these mutants with the corresponding genes from *P. flocculosa*. For this purpose *fat1*, *fat2* and *fat3* were expressed both in the *U. maydis* wild-type strain FB1, and in the mutant strains FB1Δ*uat1* and FB1Δ*uat2*. Glycolipids, produced by the transformants, were analysed by TLC and/or LC-MS.

Figure 4A shows that *fat1* was able to complement the phenotype of a Δ*uat1* mutant strain (lane 3). While introducing *fat2* did not change the phenotype (lane 4), expression of *fat3* did not rescue the phenotype, but showed a new UA derivative on the TLC probably carrying two acetyl groups (lane 5). Expression of *fat2* in the Δ*uat2* strain was able to complement this mutant (lane 9). In contrast, introducing *fat1* and *fat3* did not change the phenotype of the mutant (lanes 8 and 10). Expressing *fat1* and *fat2* in the wild-type strain did not change the phenotype (data not shown). FB1 transformants expressing *fat3* (FB1-Potef::*fat3*) showed additional bands when analysed by TLC (Fig. 4A, lane 12). These putatively new glycolipids were further analysed by LC-MS (Fig. 4B) showing that other than the usual UAs produced by wild-type strains, FB1-Potef::*fat3* mutants secreted four additional derivatives (849.4568 *m/z*, 865.4572 *m/z*, 877.5322 *m/z* and 893.5296 [M + K]+), each of which showed a mass difference of 42.06 {calculated mass for an acetyl group [CH3COOH (60) − H2O (18)] = 42.0377} when compared with the masses of wild-type UAs (Fig. 4B).

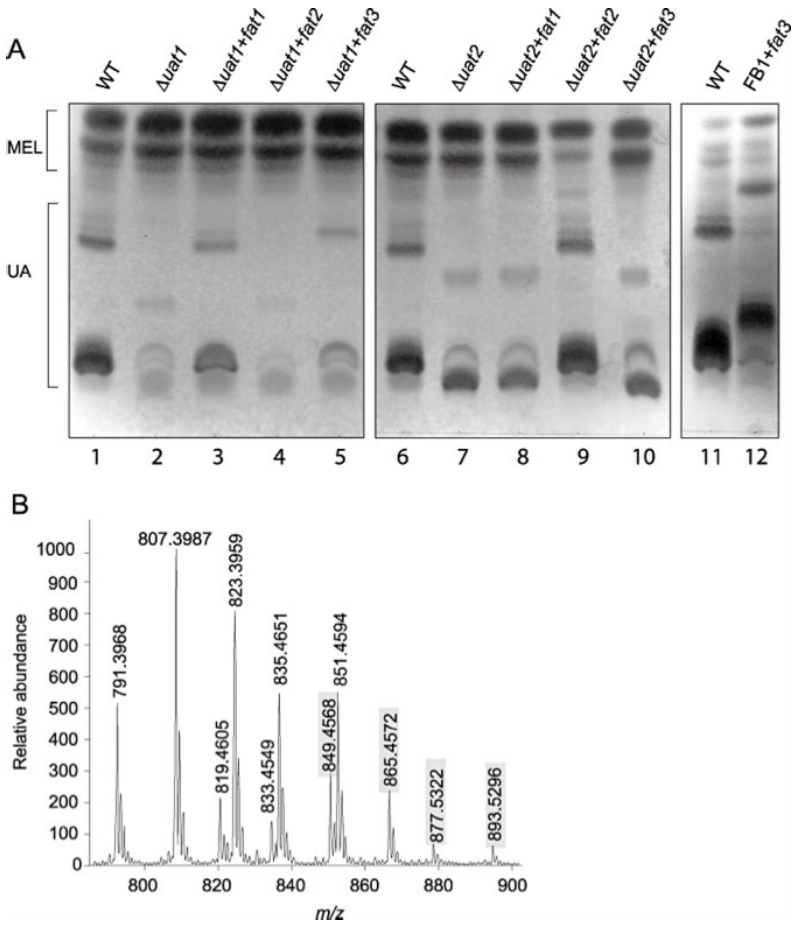


Figure 4 : Proteins involved in flocculosin acyl- and acetylation. **A,** TLC analysis of extracellular glycolipids produced by wild-type strain FB1 (WT), and FB1Δ*uat1* (Δ*uat1*), FB1Δ*uat1*-P$_{otef}$::*fat1* (Δ*uat1* + *fat1*), FB1Δ*uat1*-P$_{otef}$::*fat2* (Δ*uat1* + *fat2*) FB1Δ*uat1*-P$_{otef}$::*fat3* (Δ*uat1* + *fat3*), FB1Δ*uat2* (Δ*uat2*), FB1Δ*uat2*-P$_{otef}$::*fat1* (Δ*uat2* + *fat1*), FB1Δ*uat2*-P$_{otef}$::*fat2* (Δ*uat2* + *fat2*), FB1Δ*uat2*-P$_{otef}$::*fat3* (Δ*uat2* + *fat3*) and FB1-P$_{otef}$::*fat3* (FB1 + *fat3*) mutant strains. While *P. flocculosa* protein Fat1 is able to rescue the FB1Δ*uat1* phenotype (lane 3), expression of *fat2* complements FB1Δ*uat2* (lane 9). Introducing *fat2* into Δ*uat1* did not change the phenotype (lane 4); expression of *fat3* shows a shift of the mutated UA derivatives on the TLC (lane 5). Introducing *fat1* and *fat3* in a *uat2* deletion strain does not change the phenotype of the mutant (lanes 8 and 10). *Fat3* in contrast, when expressed in *U. maydis* wild-type strains FB1 (lane 12), shows additional bands on a TLC. Lanes indicated with WT show the typical glycolipid pattern produced by *U. maydis* wild-type strains. MEL indicates the mannosylerythritol lipids also produced by *U. maydis*. **B,** LC-MS analysis of glycolipids produced by FB1-P$_{otef}$::*fat3* (FB1 + *fat3*) mutant strains. Beside the usual UAs produced by wild-type strains, FB1-P$_{otef}$::*fat3* mutants secreted four additional derivatives (849.4568 *m/z*, 865.4572 *m/z* 877.5322 *m/z* and 893.5296 *m/z*[M + K]$^+$), shaded in grey, each of which showed a mass difference of 42,06 when compared with the masses of wild-type UAs.

These results indicate that *fat1* is the gene homologous to uat1 necessary for the transfer of the shorter fatty acid chain to their corresponding glycolipid. *Fat2* is homologous to *uat2* acetylating the first glucose molecule at the C6 position. *Fat3* has no homologues gene in

U. maydis and encodes an acetyl-transferase essential for acetylation of the C3 position of the second glucose moiety of flocculosin.

3. Discussion

In this work, we have identified a gene cluster comprising 11 genes, involved in the biosynthesis of flocculosin, an antifungal glycolipid produced by the biocontrol fungus *P. flocculosa*. Northern blot analysis showed that all genes within the cluster were simultaneously upregulated under conditions inducing flocculosin production. Because of the peculiar structure of the molecule and its relative rarity within the microbial world, we inferred that its synthesis had to bear very close genetic similarities with that of ustilagic acid, a nearly identical compound produced by the well-studied and close relative of *P. flocculosa*, *U. maydis*. This hypothesis was first raised by Marchand *et al.* (2009) who identified a key gene in the molecule synthesis based on homology with *U. maydis*, and suggested that all other genes must also be arranged within a cluster. This finding gave us the impetus to obtain the full sequence of *P. flocculosa* from which we were able to confirm this hypothesis.

The comparison between the flocculosin gene cluster and the *U. maydis* biosynthesis gene cluster for UA highlighted the conserved similarity between the two clusters in terms of gene presence and function, even though gene arrangement was more random. In fact, only two differences were noted and they accounted for the structural differences between the two molecules. First, the UA biosynthesis gene cluster contains a gene encoding an α-hydroxylase absent in the flocculosin gene cluster and second, the flocculosin gene cluster contains an additional acetyl-transferase gene necessary for the second acetylation on the sugar moiety.

In this work we could characterize two enzymes involved in UA biosynthesis of *U. maydis*, the acyl-transferases Uat1 and the acetyl-transferase Uat2. By TLC and mass spectrometry analysis, we could show that Uat1 is necessary for the transfer of the shorter fatty acid to UA and that Uat2 acetylates the second glucose molecule. The function of Fat1, Fat2 and Fat3 involved in the biosynthesis of flocculosin was demonstrated by expressing the respective genes in the *U. maydis* wild-type, *uat1* and *uat2* deletion strains. While Fat1 was

shown to function as an acyl-transferase transferring 2-hydroxy-octanoic acid to the flocculosin precursor, Fat2 and Fat3 were found to be necessary for flocculosin acetylation, acetylating the molecule at the C6 position at the first glucose molecule and the C3 position at the second glucose molecule respectively. It was even demonstrated that Fat2 and Fat3, the two acetyl-transferases, showed high specificity in acetylating only their position on the molecule. In fact, *fat2* complemented the Δ*uat2* but failed to add another acetyl group to the second glucose molecule while used in the wild type (data not shown). On the other hand, *fat3* was not able to complement a *uat2* deletion strain but was shown to add an additional acetyl group to the wild-type UA. One could even speculate on the order of Fat2 and Fat3 during biosynthesis: Fig. 4A shows that introducing *fat3* in a Δ*uat2* mutant strain does not change the phenotype indicating that Fat3 seems only to be able to acetylate its position if the first glucose molecule is already acetylated.

In fact, we were able to show that *U. maydis* can produce a flocculosin-like molecule, i.e. with two acetyl groups, when expressing *fat3*, a structure never observed before in the culture supernatants of this fungus (Spoeckner *et al.*, 1999; Teichmann *et al.*, 2007).

An interesting observation is the absence of an *ahd1* homologue, necessary for α-hydroxylation of the long-chain fatty acid (Teichmann *et al.*, 2007) in *U. maydis*, in the flocculosin gene cluster, which we could detect neither in the gene cluster nor in the whole genome. Given that the long-chain fatty acid of flocculosin is hydroxylated at the β-position while UA carries this hydroxyl-group at the α-position, we expected to find a gene capable to perform a β-hydroxylation in our cluster. Initial experimental data for the biosynthesis of UA indicated that the shorter fatty acid is β-hydroxylated by Uhd2 a NADPH-dependent oxidoreductase (B. Teichmann and M. Bölker, unpubl. data). Since the flocculosin gene cluster also contains a gene homologous to this β-hydroxylase (*fhd1*), it is equally conceivable that the flocculosin gene cluster does not need an additional gene encoding a β-hydroxylase performing β-hydroxylation of the long-chain fatty acid because *fhd1* is sufficient to hydroxylate both fatty acids at exactly the same position.

Further it is conceivable that *P. flocculosa* uses a similar mechanism described for secondary metabolite biosynthesis of some *Pseudomonas* species. These bacteria have been shown to contain an enzyme, PhaG, which directly utilizes 3-hydroxyacyl-ACP from de

novo fatty acid biosynthesis to synthesize rhamnolipids (Déziel *et al.*, 2003; Rehm *et al.*, 1998). A similar process could explain the absence of a gene encoding a β-hydroxylase.

Although the genome of *P. flocculosa* is still arranged in scaffolds, our results indicate that all genes necessary for flocculosin synthesis appear to be present within the reported cluster, which supports its integral presence on the one studied scaffold. For UA, most steps of the biosynthetic pathway have been described (Teichmann *et al.*, 2007). Based on the similarity between the clusters and the molecules, we propose that the biosynthetic pathway for flocculosin follows an identical pattern. At a first step, 3-hydroxy palmitic acid is hydroxylated by two cytochrome P450 monooxygenases, Cyp1 and Cyp2, at the terminal and subterminal position respectively. The glycosyl-transferase Fft1 would then transfer the cellobiose moiety to 3,15,16-trihydroxy-fatty acid. Initial experimental data for the biosynthesis of UA indicated that Fas2 was necessary for the synthesis of the shorter fatty acid (B. Teichmann and M. Bölker, unpubl. data), which is then hydroxylated by Uhd2. We propose a similar process for the biosynthesis (Fas2) and hydroxylation (Fhd1) of the shorter fatty acid of the flocculosin molecule. We further showed that Fat1 was necessary for the transfer of 3-hydroxy-octanoic acid to flocculosin and Fat2 and Fat3 were essential for the acetylation of flocculosin (Fig. 5).

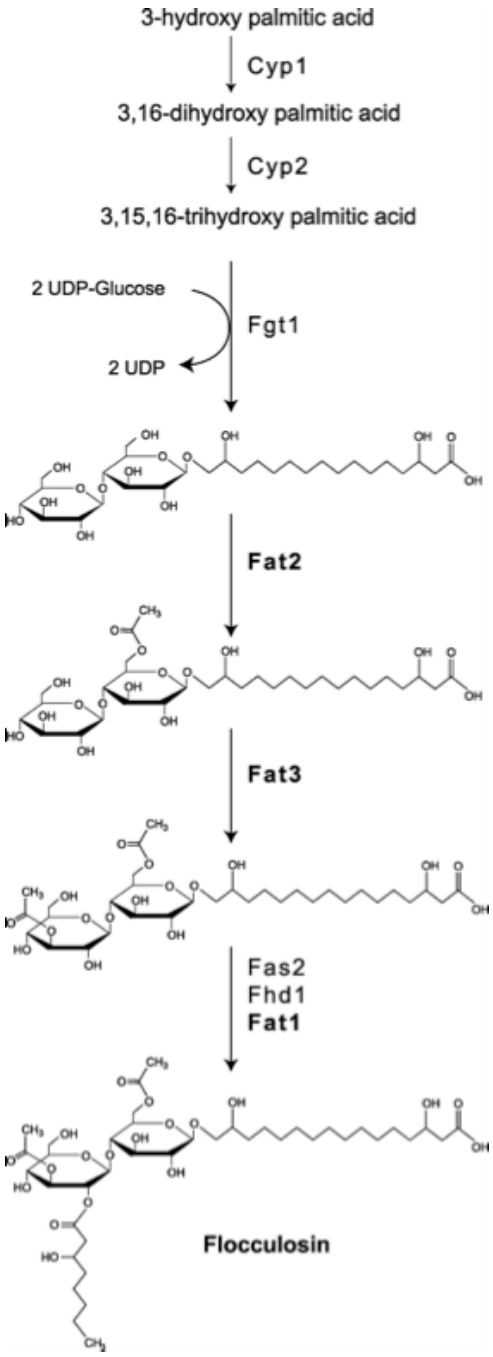


Figure 5 : Proposed biosynthesis pathway for flocculosin. The biosynthesis of flocculosin starts with the hydroxylation of 3-hydroxy-palmitic acid by Cyp1 and Cyp2 at the terminal and subterminal position. The UDP-dependent glycosyl transferase Fgt1 glycosylates the hydroxy fatty acid probably by sequential transfer of two glucose molecules. The cellobiose moiety is acetylated by Fat2 followed by Fat3 and is acylated by Fat1. Biosynthesis and hydroxylation of the short-chain fatty acid is probably performed by Fas2 and Fhd1.

Genes encoding proteins involved in the biosynthesis of secondary metabolites like flocculosin or UA are often found to be clustered. These gene clusters contain generally regulatory genes that control the entire pathway. The advantage of gene clustering is not yet clear, but it has been proposed that it promotes selective transfer of certain genes via homologous gene transfer (Walton 2000). While the gene sequence is relatively highly conserved between the two gene clusters, the order in which the genes are distributed is more random. Furthermore the UA gene cluster is more compact in structure and organization, whereas the flocculosin gene cluster is much larger and the single genes appear to contain more introns. It has already been described that DNA alterations such as recombination processes, gene duplications and gene losses appear during fungal gene cluster evolution (Kroken *et al.*, 2003). This would explain the additional acetyl-transferase gene as well as the missing *ahd1* homologue in the flocculosin cluster. In spite of these differences, both fungi have maintained the ability to produce very similar glycolipids, indicating a common origin for both gene clusters and a selection pressure to maintain this trait.

Interestingly, within the basidiomycetes family, very few fungi have been reported to produce these cellobiose lipids. In fact, only *U. maydis* and three members of the Pseudozyma genus have kept that ability. *Sporisorium reilianum*, which is even more closely related to *U. maydis* than *P. flocculosa* (Begerow *et al.*, 2006), and *Ustilago hordei* are not able to produce these secondary metabolites since they do not contain the appropriate gene cluster (Schirawski *et al.*, 2010; J. Schirawski, R. Kahmann and G. Bakkeren, pers. comm.). *S. reilianum* causes head smut in maize and sorghum (Matyac et Kommedahl, 1985) and seems to have a similar life cycle as *U. maydis*, but infects the plants through the roots causing a systemic infection (Martinez *et al.*, 2000, 1999; Potter 1914). It has been shown that *S. reilianum* produces gibberellins (Matheussen *et al.*, 1991) but until now, no glycolipid production has been reported. The reason why *U. maydis* and *P. flocculosa*, two fungi with different lifestyles, have kept this large gene cluster through their evolution and why two phylogenetically related plant pathogens, *S. reilianum* and *U. hordei*, have not remains to be elucidated. Biosynthesis of secondary metabolites is usually subject to natural selection for the benefits that it confers to the fungus producing it. It has been hypothesized that *P. flocculosa* uses it in its antagonistic activity against

powdery mildews (Cheng *et al.*, 2003), a function obviously not applicable to *U. maydis*. Whether or not the glycolipids play a similar role in both fungi, its intricate synthesis through an elaborate gene cluster conserved between two related fungi with different lifestyles certainly supports the fact that it has an important function in their overall fitness.

4. Experimental procedures

Strains, plasmids and culture conditions

Escherichia coli strain TOP10 was used for all DNA manipulations. Construction of plasmids was performed using standard procedures (Sambrook *et al.*, 1989). *P. flocculosa* (DAOM 196992) and *U. maydis* strains FB1 (a1 b1) (Banuett et Herskowitz, 1989) were maintained on potato dextrose agar (PDA, BD, Mississauga, Canada) at 4°C. *U. maydis* deletion strain FB1Δ*uat1* and FB1Δ*uat2* were generated according to the published protocol for PCR-based generation of gene replacement mutants (Kämper 2004) with minor modifications. Briefly, the 1 kb flanking regions of uat1 and uat2 were amplified by PCR and subjected to agarose gel electrophoresis. Primer sequences are listed in Table 3. After purification from the gel (Gel/PCR DNA Fragments Extraction kit, Avegene, Hamburg, Germany) the PCR products were digested with SfiI (New England Biolabs, Frankfurt, Germany) and ligated to a SfiI-digested hph cassette using 1 U of T4 ligase (Roche Diagnostics, Mannheim, Germany) in a standard 10 µl reaction at 14°C overnight. The ligation product was subjected to agarose gel electrophoresis and, after purification from the gel, the 4 kb fragment was cloned into pCR2.1-TOPO (TOPO® TA Cloning® Kit, Invitrogen, Darmstadt, Germany). Prior to transformation into *U. maydis*, the plasmid was digested with PstI (*uat1*) or NotI (*uat2*) to dissolve the construct from the vector. Transformation of *U. maydis* was performed exactly as described in Brachmann *et al.* (2004). For selection of transformants, PDA plates containing 200 mg ml^{-1} hygromycin or 2 µg ml^{-1} carboxin were used. All mutants were confirmed by Southern analysis and deletion phenotypes were verified by complementation.

Table 3 : Primers used within this article

Name	Sequence (5′→ 3′)[a]	Description
MD469	gcgcctgcaggtcacgaatgcttttgc	Amplification of the 5′ region of *uat1* for deletion construct
MC325	cacggcctgagtggccctgatattgcagggtggtgtc	
MC326	gtgggccatctaggccgccgtcgctaccgacgtaaaag	Amplification of the 3′ region of *uat1* for deletion construct
MD470	gcgcctgcaggcgaatgacagagacgtg	
MC846	gtatgcggccgcttctccgggcgagagcac	Amplification of the 5′ region of *uat2* for deletion construct
MC847	cacggcctgagtggccgatcaggagtgtctgcttgatg	
MC848	gtgggccatctaggccccacattgaaaaattttggtcgacatgg	Amplification of the 3′ region of *uat2* for deletion construct
MC849	cacagcggccgccctgcggcctgaagtgttg	
RB17	atgtctgagccgctctctgg	Amplification of internal probe for gene *fat2* (RT-PCR)
RB18	agaggagtgaggcggtttgg	
RB27	cctcctgctgctgctgctgc	Amplification of internal probe for gene *uhd1*
RB28	gccattcgaagtgtacatgg	
RB29	gcccgcgctggcaaacacgg	Amplification of internal probe for gene *fas2*
RB30	ctctcgctcgggctcgaacg	
RB31	gccggctgacggacaacgcg	Amplification of internal probe for gene *atr1*
RB32	gacgactgacatgaggacgg	
RB33	cgtggccagccatcgggcac	Amplification of internal probe for gene *ugt1*
RB34	ccctcggatcgacggccaag	
RB35	ccctccacgccctgctgtgg	Amplification of internal probe for gene *uat1*
RB36	ctccccgtccggcgctttctg	
RB37	cggctcgaggttcttgagg	Amplification of internal probe for gene *cyp2*
RB38	tgttcaatgccttcctctcc	
RB39	ggatctcgaccaacgacgcc	Amplification of internal probe for gene *fat1*
RB40	ggcccatctgcaccgcaggg	
RB41	cgccatcaaggcgcggtacc	Amplification of internal probe for gene um01428
RB42	ggagcgcaaggtcgtgtag	
RB45	gtctacgtcgcctgccacgc	Amplification of internal probe for gene *orf1*
RB46	gtgggcagcgccagggcgtc	
RB47	cacacggacatggccagcac	Amplification of internal probe for gene *mfc1*
RB48	gtgaggcgcttgggtccgag	
RB51	ccaggccaccgaagcgatgg	Amplification of internal probe for gene *rfl1*
RB52	gcgtcggcgcgtggaacacc	
RB72	ccaacgtcttcttcgacatc	Amplification of internal probe for gene *ppi*
RB73	gcgccgtagatcgacttgcc	

RB65	gcgccatggccatgaccatcgacgagccagc	Amplification of *fat1* ORF
RB56	gcggcggccgcctacaggcagtggcagaac	
RB67	gcgacgcgtcatgtcatcggccgcctctgc	Amplification of *fat3* ORF
RB68	gcggcggccgcctagtacagcacctcacccc	
RB83	gcgggatccatgacgatctcgaccactcta	Amplification of *fat2* ORF
RB84	agtgcggccgccgctttacgccgaatctaa	

[a.] Underlined letters indicate introduced restriction sites.

To induce glycolipid production in *U. maydis*, strains were grown at 28°C in liquid YEPS medium (10 g l^{-1} yeast extract, 20 g l^{-1} peptone, 20 g l^{-1} sucrose) to logarithmic phase and then transferred into a nitrogen starvation medium containing 1.7 g l^{-1} yeast nitrogen base (YNB; BD, Mississauga, Canada) and 5% glucose as a carbon source. Glycolipids were isolated after cultivation of cells for 4 days at 28°C on a rotary shaker set at 200 r.p.m.

Northern analysis

Total RNA from *P. flocculosa* cultures was isolated for Northern blot analysis with Trizol (TRIzol® Reagent, Invitrogen, Burlington, ON, Canada) following the manufacturer's protocol. *P. flocculosa* was grown at 28°C to logarithmic phase in YMPD medium (6 g l^{-1} yeast extract, 3 g l^{-1} malt extract, 5 g l^{-1} peptone water, 10 g l^{-1} dextrose) on a rotary shaker set at 150 r.p.m. Cells were transferred to either a flocculosin-inducing medium (MOD) (Hammami *et al.*, 2008) or a flocculosin-repressing medium (YMPD) and grown for 24 h at 28°C. RNA was prepared after 24 h. Northern blots were prepared using standard procedures (Sambrook *et al.*, 1989). For hybridization, internal probes of the respective genes that were amplified by PCR were used. Primer sequences are listed in Table 3.

Reverse transcriptase PCR

Total RNA from *P. flocculosa* was transcribed into cDNA with SuperScript II Reverse Transcriptase (Invitrogen, Burlington, ON, Canada) following the manufacturer's protocol. A standard PCR using primers RB17 and RB18 (*fat2*) as well as RB72 and RB73 (*ppi*) was performed (Table 3).

Generation of *U. maydis* strains expressing *P. flocculosa* genes

For expression of *fat1*, *fat2* and *fat3* in *U. maydis*, plasmid p123 carrying an enhanced GFP (eGFP) reporter under control of the strong constitutive P_{otef} promoter and a carboxin

resistance (ip) cassette as selectable marker in *U. maydis* was used (Spellig *et al.*, 1996). The three acyl-/acetyl-transferase genes were amplified with primers RB67 and RB68 (*fat3*); RB65 and RB56 (*fat1*) or RB83 and RB84 (*fat2*) digested with MluI/NotI (*fat3*); NcoI/NotI (*fat1*); BamHI/NotI (*fat2*) and introduced into p123 behind the P_{otef} promoter. Prior to integrative transformation into the ip locus of *U. maydis* strains FB1, FB1Δ*uat1* or FB1Δ*uat2* plasmids were linearized with SspI.

Isolation of glycolipids

Extracellular glycolipids were extracted from suspension cultures (0.5 ml) with one volume of ethyl acetate. The ethyl acetate phase was evaporated, and glycolipids were dissolved in methanol. Glycolipids were analysed by TLC on silica plates (Silica gel 60; EMD, Mississauga, ON, Canada) with a solvent system consisting of chloroform–methanol–water (65:25:4, v/v). The plates were dried thoroughly, and sugar-containing compounds were visualized by spraying them with a mixture of glacial acetic acid–sulphuric acid–p-anisaldehyde (50:1:0.5, v/v) and heating them at 150°C for 3 min.

P. flocculosa genome sequencing

Using 454 sequencing Titanium technology (454 Life Sciences, Roche, Branford, CT, USA), the DNA genome of *P. flocculosa* was sequenced at Genome Quebec Innovation Center (Montreal, Quebec, Canada). Briefly, 1.4 million reads were generated and 80% of them were fully assembled into 3410 contigs. Due to the large size of this genome (22 Mb), a long paired end sequencing approach was used to scaffold the contigs (order/orientation) into 1281 scaffolds. The sequences contained on these scaffolds were used for protein identification based on homology with *U. maydis* sequences. The protein sequences of *U. maydis* were used to identify the predicted ORF in the database of the sequenced genome of *P. flocculosa* with the CLC Genomics Workbench 3.6.1 software (CLC bio USA, Cambridge, MA).

These ORFs were then used to find homologous genes and corresponding proteins by submitting the sequences to the National Center for Biotechnology Information (NCBI) and to the *U. maydis* database of the Broad Institute (http://www.broadinstitute.org/annotation/genome/ustilago_maydis/Home.html).

Mass spectrometry of glycolipids

Ustilago maydis strains were grown under glycolipid inducing conditions (see above). Briefly, 4 ml of culture wAS washed with one volume of chloroform to eliminate mannosylerythritol lipids (MELs). After centrifugation (1 min, 13 000 r.p.m.), the aqueous phase (upper phase) was passed through a C18 silica gel cartridge (Sep-Pak® Plus Environmental C18 cartridge, Waters, Mississauga, ON, Canada) previously conditioned with three volumes of MeOH followed by three volumes of water. After sample loading, three volumes of water were used to rinse and the cellobiose lipids contained in the cartridge were finally eluted with three volumes of MeOH. The methanolic fraction was then diluted 10 times in 80% MeOH prior to analysis by LC-MS. MS analyses were performed on a LC-MS (model 'Unique' from LECO Corporation, St. Joseph, MI) system combining electrospray ionization with a time of flight analyser (ESI-TOF MS). Samples were injected by direct infusion at a flow rate of 50 μl min^{-1} and analysed in the positive ionization mode.

5. Acknowledgements

We thank Tyler Avis for critical reading of the manuscript. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada Research Chairs Program to R.R. Bélanger and by the Grant BO 2094/3-1 from the Deutsche Forschungsgemeinschaft (DFG) to M. Bölker. B. Teichmann is grateful to the DFG for financial support (Project: TE 815/1-1).

6. References

Banuett, F. and Herskowitz, I. (1989). Different a alleles of Ustilago maydis are necessary for maintenance of filamentous growth but not for meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(15), 5878-5882.

Begerow, D., Stoll, M. and Bauer, R. (2006). A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data. *Mycologia*, *98*(6), 906-916.

Boothroyd, B., Thorn, J.A. and Haskins, R.H. (1956). Biochemistry of the Ustilaginales. XII. Characterization of extracellular glycolipids produced by *Ustilago* sp. *Canadian Journal of Biochemistry and Physiology*, *34*(1), 10-14.

Brachmann, A., König, J., Julius, C. and Feldbrügge, M. (2004). A reverse genetic approach for generating gene replacement mutants in *Ustilago maydis. Molecular Genetics and Genomics*, *272*(2), 216-226.

Cheng, Y., McNally, D.J., Labbe, C., Voyer, N., Belzile, F. and Bélanger, R.R. (2003). Insertional mutagenesis of a fungal biocontrol agent led to discovery of a rare cellobiose lipid with antifungal activity. *Applied and Environmental Microbiology*, *69*(5), 2595-2602.

Déziel, E., Lépine, F., Milot, S. and Villemur, R. (2003). *rhlA* is required for the production of a novel biosurfactant promoting swarming motility in *Pseudomonas aeruginosa*: 3-(3-hydroxyalkanoyloxy)alkanoic acids (HAAs), the precursors of rhamnolipids. *Microbiology*, *149*(8), 2005-2013.

Golubev, V.I., Kulakovskaia, T.V., Shashkov, A.S., Kulakovskaia, E.V. and Golubev, N.V. (2008). [Antifungal cellobiose lipid secreted by the epiphytic yeast *Pseudozyma graminicola*]. *Mikrobiologiia*, *77*(2), 201-206.

Hammami, W., Labbé, C., Chain, F., Mimee, B. and Bélanger, R.R. (2008). Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa. Applied Microbiology and Biotechnology*, *80*(2), 307-315.

Haskins, R.H. and Thorn, J.A. (1951). Biochemistry of the Ustilaginales VII. Antibiotic activity of ustilagic acid. *Canadian Journal of Botany*, *29*(6), 585-592.

Hoffmeister, D. and Keller, N.P. (2007). Natural products of filamentous fungi: enzymes, genes, and their regulation. *Natural Product Reports*, *24*(2), 393-416.

Kämper, J. (2004). A PCR-based system for highly efficient generation of gene replacement mutants in *Ustilago maydis. Molecular Genetics and Genomics*, *271*(1), 103-110.

Kämper, J., Kahmann, R., Bölker, M., Ma, L.J., Brefort, T., Saville, B.J., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis. Nature*, *444*(7115), 97-101.

Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C. and Turgeon, B.G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15670-15675.

Kulakovskaya, T.V., Shashkov, A.S., Kulakovskaya, E.V. and Golubev, W.I. (2005). Ustilagic acid secretion by *Pseudozyma fusiformata* strains. *FEMS Yeast Research*, *5*(10), 919-923.

Marchand, G., Rémus-Borel, W., Chain, F., Hammami, W., Belzile, F. and Bélanger, R.R. (2009). Identification of genes potentially involved in the biocontrol activity of *Pseudozyma flocculosa. Phytopathology*, *99*(10), 1142-1149.

Martinez, C., Jauneau, A., Roux, C., Savy, C. and Dargent, R. (2000). Early infection of maize roots by *Sporisorium reilianum* f. sp. *zeae. Protoplasma*, *213*(1-2), 83-92.

Martinez, C., Roux, C. and Dargent, R. (1999). Biotrophic development of *Sporisorium reilianum* f. sp. *zeae* in vegetative shoot apex of maize. *Phytopathology*, *89*(3), 247-253.

Matheussen, A.M., Morgan, P.W. and Frederiksen, R.A. (1991). Implication of gibberellins in head smut (*Sporisorium reilianum*) of *Sorghum bicolor*. *Plant Physiology*, *96*(2), 537-544.

Matyac, C.A. and Kommedahl, T. (1985). Factors affecting the development of head smut caused by *Sphacelotheca reiliana* on corn. *Phytopathology*, *75*(5), 577-581.

Mimee, B., Labbé, C., Pelletier, R. and Bélanger, R.R. (2005). Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa*. *Antimicrobial Agents and Chemotherapy*, *49*(4), 1597-1599.

Mimee, B., Pelletier, R. and Bélanger, R.R. (2009). In vitro antibacterial activity and antifungal mode of action of flocculosin, a membrane-active cellobiose lipid. *Journal of Applied Microbiology*, *107*(3), 989-996.

Paulitz, T.C. and Bélanger, R.R. (2001). Biological control in greenhouse systems. *Annual Review of Phytopathology*, *39*(1), 103-133.

Potter, A.A. (1914). Head smut of sorghum and maize. *Journal of Agricultural Research*, *2*(5), 339-372.

Rehm, B.H.A., Krüger, N. and Steinbüchel, A. (1998). A new metabolic link between fatty acid *de novo* synthesis and polyhydroxyalkanoic acid synthesis. *The Journal of Biological Chemistry*, *273*(37), 24044-24051.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning* (Vol. 2). New York, NY, USA: Cold Spring Harbor Laboratory Press.

Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*, *330*(6010), 1546-1548.

Shoda, M. (2000). Bacterial control of plant diseases. *Journal of Bioscience and Bioengineering*, *89*(6), 515-521.

Spellig, T., Bottin, A. and Kahmann, R. (1996). Green fluorescent protein (GFP) as a new vital marker in the phytopathogenic fungus *Ustilago maydis*. *Molecular & General Genetics*, *252*(5), 503-509.

Spoeckner, S., Wray, V., Nimtz, M. and Lang, S. (1999). Glycolipids of the smut fungus *Ustilago maydis* from cultivation on renewable resources. *Applied Microbiology and Biotechnology*, *51*(1), 33-39.

St-Pierre, B., Laflamme, P., Alarco, A.-M., D, V. and Luca, E. (1998). The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. *The Plant Journal*, *14*(6), 703-713.

Teichmann, B., Linne, U., Hewald, S., Marahiel, M.A. and Bölker, M. (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Molecular Microbiology*, *66*(2), 525-533.

Teichmann, B., Liu, L., Schink, K.O. and Bölker, M. (2010). Activation of the ustilagic acid biosynthesis gene cluster in *Ustilago maydis* by the C2H2 zinc finger transcription factor Rua1. *Applied and Environmental Microbiology*, *76*(8), 2633-2640.

Walton, J.D. (2000). Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genetics and Biology*, *30*(3), 167-171.

Chapitre 4 : Beta Hydroxylation of Glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-Dependent β-Hydroxylase

Teichmann, B., Lefebvre, F., Labbé, C., Bölker, M., Linne, U. et Bélanger, R.R. (2011). Beta hydroxylation of glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-dependent β-hydroxylase. *Applied and Environmental Microbiology*, *77*(21), 7823-7829

Résumé

La flocculosine et l'acide ustilagique, deux cellobioses-lipides très similaires, sont respectivement produits par *Pseudozyma flocculosa*, un agent de lutte biologique, et *Ustilago maydis*, un agent pathogène des plantes. Les deux glycolipides contiennent une petite chaîne d'acide gras hydroxylée à la position β, mais diffèrent pour ce qui est de leur longue chaîne d'acide gras, hydroxylée en position α pour l'acide ustilagique et en position β pour la flocculosine. Chez les deux organismes, les gènes de la biosynthèse sont rassemblés sous la forme d'une batterie. La fonction de la plupart des gènes a déjà été identifiée, mais celles des gènes *fhd1,* chez *P. flocculosa,* et *uhd1,* chez *U. maydis,* sont toujours inconnues. La séquence d'acides aminés déduite de ces gènes montre une homologie envers celle de la déshydrogénase/réductase à chaîne courte (DRCC). Ici, nous avons d'abord désactivé le gène *uhd1* chez *U. maydis* et analysé l'acide ustilagique sécrétée. Les souches chez qui le gène *uhd1* est inactivé ont produit des acides ustilagiques sans groupe β-hydroxyl sur la courte chaîne d'acide gras. Pour analyser la fonction de Fhd1 chez *P. flocculosa*, le gène correspondant a été utilisé pour complémenter les mutants Δ*uhd1* d'*U. maydis.* Fhd1 a été en mesure de restaurer la production d'acide ustilagique de type sauvage, indiquant que Fhd1 est responsable de la β-hydroxylation de la courte chaîne d'acide gras. Ensuite, nous nous sommes intéressés à un homologue de *P. flocculosa* pour l'α-hydroxylase de la longue chaîne d'acide gras d'*U. maydis*, Ahd1. Le gène *ahd1* de *P. flocculosa*, qui ne réside pas dans la batterie de gènes de la flocculosine, a été introduit dans les souches mutantes Δ*ahd1* d'*U. maydis*. La synthèse d'Ahd1 de *P. flocculosa* chez *U. maydis* n'a pu complémenter le phénotype Δ*ahd1* d'*U. maydis* ou permettre la production d'acide ustilagique β-hydroxylé. Ces résultats suggèrent que Ahd1, chez *P. flocculosa,* n'est pas impliqué dans l'hydroxylation de la flocculosine.

Abstract

Flocculosin and ustilagic acid (UA), two highly similar antifungal cellobiose lipids, are respectively produced by *Pseudozyma flocculosa*, a biocontrol agent, and *Ustilago maydis*, a plant pathogen. Both glycolipids contain a short-chain fatty acid hydroxylated at the β position but differ in the long fatty acid, which is hydroxylated at the α position in UA and at the β position in flocculosin. In both organisms, the biosynthesis genes are arranged in large clusters. The functions of most genes have already been characterized, but those of the *P. flocculosa fhd1* gene and its homolog from *U. maydis*, *uhd1*, have remained undefined. The deduced amino acid sequences of these genes show homology to those of short-chain dehydrogenases and reductases (SDR). We disrupted the *uhd1* gene in *U. maydis* and analyzed the secreted UA. *uhd1* deletion strains produced UA lacking the β-hydroxyl group of the short-chain fatty acid. To analyze the function of *P. flocculosa* Fhd1, the corresponding gene was used to complement *U. maydis* Δ*uhd1* mutants. Fhd1 was able to restore wild-type UA production, indicating that Fhd1 is responsible for β hydroxylation of the flocculosin short-chain fatty acid. We also investigated a *P. flocculosa* homolog of the *U. maydis* long-chain fatty-acid alpha hydroxylase Ahd1. The *P. flocculosa ahd1* gene, which does not reside in the flocculosin gene cluster, was introduced into *U. maydis* Δ*ahd1* mutant strains. *P. flocculosa* Ahd1 neither complemented the *U. maydis* Δ*ahd1* phenotype nor resulted in the production of β-hydroxylated UA. This suggests that *P. flocculosa* Ahd1 is not involved in flocculosin hydroxylation.

1. Introduction

Extracellular glycolipids are produced by several microorganisms and are composed of different mono- or disaccharides that are either acylated or glycosidically linked to long-chain fatty acids or hydroxy fatty acids. Because of their amphipathic character, they act as biosurfactants; many glycolipids also have antibiotic properties (Ron et Rosenberg, 2001). Because of their unique chemical properties, they have gained importance in the chemical and food industries and are also being exploited for environmental protection (Cameotra et Makkar, 2004).

Ustilago maydis (de Candolle) Corda and *Pseudozyma flocculosa* (Traquair, Shaw & Jarvis) Boekhout & Traquair are two basidiomycete fungi that produce very similar glycolipids with antibiotic activities (Haskins et Thorn, 1951; Mimee *et al.*, 2005, 2009). While *U. maydis* is a phytopathogenic fungus that infects corn plants and causes smut diseases, *P. flocculosa* is a natural inhabitant of the phyllosphere that has been described as a biocontrol agent with particular activity against powdery-mildew pathogens (Avis et Bélanger, 2002; Paulitz et Bélanger, 2001). Ustilagic acid (UA), the glycolipid produced by *U. maydis*, is a mixture of four different derivatives consisting of cellobiose o-glycosidically linked to 15,16-dihydroxypalmitic or 2,15,16-trihydroxypalmitic acid. In addition, the UA molecule is acetylated and esterified with a short-chain β-hydroxy fatty acid with a length of either C6 or C8. In contrast to UA, flocculosin, which is produced by *P. flocculosa*, has only one known isomer consisting of 3,15,16-trihydroxypalmitic acid linked to the cellobiose moiety. In addition, the molecule is diacetylated and decorated with β-hydroxyoctanoic acid.

It has been shown that UA biosynthesis is under the control of a gene cluster that includes all genes necessary for its production (Teichmann *et al.*, 2007). The biosynthesis genes responsible for flocculosin production are also seemingly arranged within a common gene cluster, but at least one gene, the *U. maydis ahd1* homolog, is missing or outside the borders currently defined for the cluster (Fig. 1; Teichmann *et al.*, 2011a). Several genes from these clusters have been cloned and functionally analyzed. Among the proteins encoded by the genes, two cytochrome P450 monooxygenases (Cyp1 and Cyp2) (Hewald *et al.*, 2005; Teichmann *et al.*, 2007), an acyltransferase (Uat1 for *U. maydis*; Fat1 for

P. flocculosa), several acetyltransferases (Uat2 for *U. maydis*; Fat2 and Fat3 for *P. flocculosa*) (Teichmann *et al.*, 2011a), and an enzyme involved in α hydroxylation of the long-chain fatty acid during UA biosynthesis (Ahd1) (Teichmann *et al.*, 2007) have been identified. From this data, a hypothetical biosynthetic pathway for both glycolipids has been postulated (Fig. 2). During UA or flocculosin biosynthesis, several hydroxylation steps are required. The long-chain fatty acids of both glycolipids are hydroxylated at three different positions, two of which, the terminal and subterminal positions, are under the control of Cyp1 and Cyp2, respectively. While α hydroxylation of UA is catalyzed by Ahd1, it is still unknown which enzyme hydroxylates the β position of the long-chain fatty acid of flocculosin. The short-chain fatty acid requires another hydroxylation step to hydroxylate the β position, but it is not yet known which enzyme catalyzes this transfer.

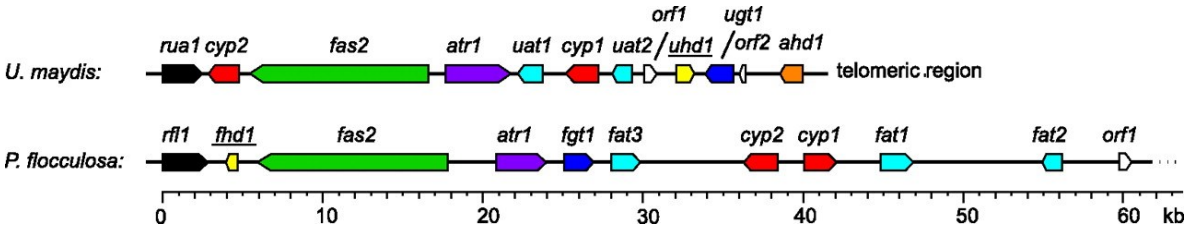


Figure 1 : UA and flocculosin biosynthesis gene clusters. The arrows indicate the positions of genes within the clusters and the orientations of transcription. The designations of genes analyzed within this work are underlined. Homologous genes are shown in the same color. For gene designations, see Teichmann *et al.*, 2011a.

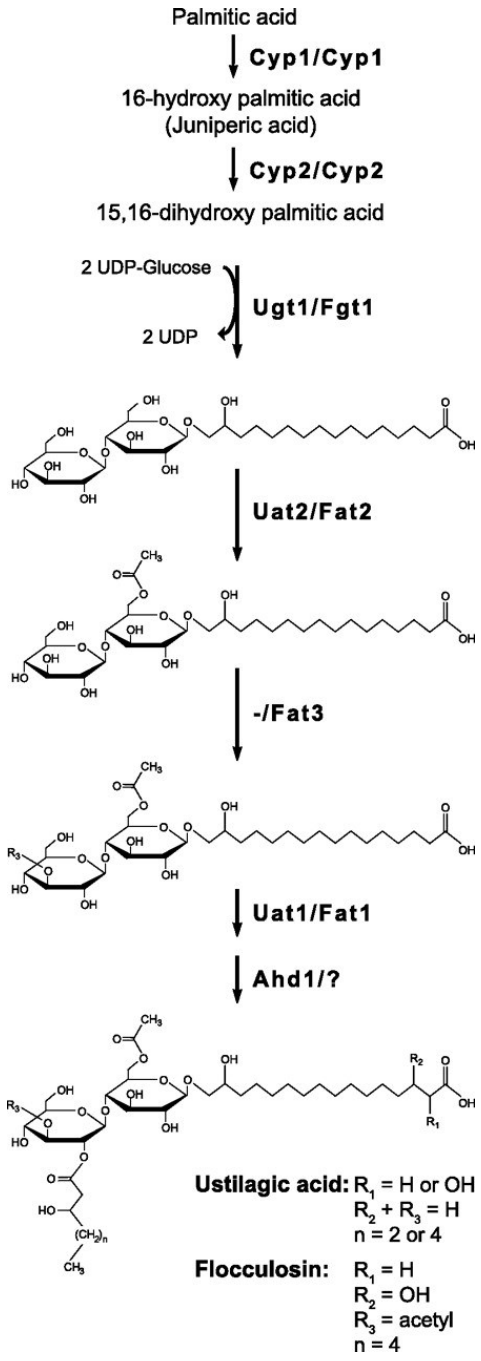


Figure 2 : Biosynthetic pathway for UA and flocculosin. Enzyme names are provided to the right of the arrows (left, *U. maydis* enzyme; right, *P. flocculosa* enzyme). Cyp1 and Cyp2, cytochrome P450 monooxygenases; Ugt1/Fgt1, UDP glucose-dependent glycosyl transferases; Uat2/Fat2 and Fat3, acetyltransferases; Uat1/Fat1, acyltransferases; Ahd1, α-hydroxylase.

Here we report that Uhd1 from *U. maydis* and its homolog Fhd1 from *P. flocculosa* are necessary for β hydroxylation of the short-chain fatty acid. Both proteins show homology

to members of the family of NADPH-dependent oxidoreductases. By deletion studies and heterologous expression of these genes, we were able to analyze their enzymatic functions during flocculosin and UA biosynthesis. We further demonstrate that Ahd1 from *U. maydis* has no functional homolog in *P. flocculosa*, thereby partly explaining the differences between the two organisms in hydroxylation of the long fatty-acid chain.

2. Materials and methods

Strains, plasmids, and culture conditions

Escherichia coli strain TOP10 was used for all DNA manipulations. Construction of plasmids was performed using standard procedures (Sambrook *et al.*, 1989). *P. flocculosa* (DAOM 196992) and *U. maydis* strains FB1 (Banuett et Herskowitz, 1989) and FB1Δ*ahd1* (Teichmann *et al.*, 2007) were maintained on potato dextrose agar (PDA; BD, Mississauga, Canada) at 4°C. To induce glycolipid production in *U. maydis*, strains were grown at 28°C in liquid YEPS medium (10 g/liter yeast extract, 20 g/liter peptone, 20 g/liter sucrose) to the logarithmic phase and then transferred into a nitrogen starvation medium containing 1.7 g/liter yeast nitrogen base (YNB; BD, Mississauga, Canada) and 5% glucose as a carbon source. Glycolipids were isolated after cultivation of cells for 4 days at 28°C on a rotary shaker set at 200 rpm.

Generation of *U. maydis* deletion strains

U. maydis deletion strain FB1Δ*uhd1* (Mips *Ustilago maydis* database [MUMDB] entry number for Uhd1, Um06466) (http://mips.helmholtz-muenchen.de/genre/proj/ustilago/) was generated according to the published protocol for PCR-based generation of gene replacement mutants (Kämper 2004) with minor modifications. Briefly, the 1-kb flanking regions of *uhd1* were amplified by PCR and subjected to agarose gel electrophoresis. Primer sequences are listed in Table 1. After purification from the gel (Gel/PCR DNA fragments extraction kit; Avegene, Hamburg, Germany), the PCR products were digested with SfiI (New England BioLabs, Frankfurt, Germany) and ligated to the hygromycin phosphotransferase gene (*hph*), obtained by digesting plasmid pBS-hhn (Kämper 2004) with SfiI. Ligation was done using 1 U of T4 ligase (Roche Diagnostics, Mannheim, Germany) in a standard 10-μl reaction mixture at 14°C overnight. The ligation product was

subjected to agarose gel electrophoresis, and, after purification from the gel, the 4-kb fragment was cloned into pCR2.1-TOPO (TOPO TA cloning kit; Invitrogen, Darmstadt, Germany). Prior to transformation into *U. maydis*, the plasmid was digested with NotI to separate the construct from the vector. Transformation of *U. maydis* was performed exactly as described by Brachmann *et al.* (2004). For selection of transformants, PDA plates containing 200 mg/ml hygromycin were used. The mutants were confirmed by Southern analysis and verified by complementation.

Table 1 : Oligonucleotides used in the study

Oligo. name	Sequence (5′ → 3′)[a]	Description
MC840	gtatgcggccgctccacacacagtggaggg	Amplification of the 5′ region of *uhd1* for deletion construct
MC841	cacggcctgagtggcctgtgagcagttgtacatgg	
MC842	gtgggccatctaggcctcgtatagcaacggatgaaattcgtg	Amplification of the 3′ region of *uhd1* for deletion construct
MC843	cacagcggccgccaagaatcgatgcttgccgg	
RB89	gagccatggtaaaaatgacaaaggaagccg	Amplification of *fhd1* ORF
RB90	gaggcggccgcgcccgcccgacgatagttat	
RB125a	ccatggccacccagactctatc	Amplification of *ahd1* ORF
RB125b	gcggccgctagccgtctggcccgaatac	
RB18	agaggagtgaggcggtttgg	Amplification of internal probe for gene *fat2* (RT-PCR)
RB19	caagagcccaacgctgaacg	
RB27	cctcctgctgctgctgctgc	Internal control primers for *fhd1*
RB28	gccattcgaagtgtacatgg	
RB72	ccaacgtcttcttcgacatc	Amplification of internal probe for gene *ppi* (RT-PCR)
RB73	gcgccgtagatcgacttgcc	
RB112	gcctcctcgccaccgaggtc	Amplification of internal probe for gene *ahd1* (RT-PCR)
RB113	ggtgccgatgaggaggccgg	

[a] Underlined letters indicate introduced restriction sites.

Generation of *U. maydis* strains carrying *P. flocculosa* genes

For expression of *P. flocculosa fhd1* and *ahd1* (GenBank accession no. HQ292619 [Fhd1] and JN039370 [Ahd1]) in *U. maydis*, plasmid p123, carrying an enhanced green fluorescent

protein (eGFP) reporter under the control of the strong constitutive P_{otef} promoter and a carboxin resistance (ip) cassette as a selectable marker in *U. maydis*, was used (Spellig *et al.*, 1996). Genes *fhd1* and *ahd1* were amplified with primers RB89 and RB90 (*fhd1*) and primers RB125a and RB125b (*ahd1*) digested with NcoI-NotI and introduced into p123 under the control of the P_{otef} promoter. Prior to integrative transformation into the ip locus of *U. maydis* strain FB1, FB1Δ*uhd1*, or FB1Δ*ahd1*, plasmids were linearized with SspI. For selection of transformants, PDA plates containing 2 μg/ml carboxin were used. Mutants were confirmed by PCR amplification using primers RB27 and RB28 (*fhd1*) and primers RB112 and RB113 (*ahd1*) (Table 1).

Isolation of glycolipids

Extracellular glycolipids were extracted from supernatants of suspension cultures with ethyl acetate and analyzed by thin-layer chromatography (TLC) as described previously (Hewald *et al.*, 2005).

Mass spectrometry of glycolipids

Extracellular glycolipids were extracted from supernatants of suspension cultures with ethyl acetate. High-performance liquid chromatography (HPLC) separation of extracted cellobiose lipids (100 ml) was performed with an 1100-HPLC system (Agilent, Germany) equipped with a 3-mm-diameter Nucleosil 250/3 C8 column (Macherey-Nagel, Germany). The gradient applied at a flow rate of 0.4 ml min^{-1} and a column temperature of 45°C was as follows (buffer A was water with 0.05% formic acid; buffer B was methanol with 0.045% formic acid): a linear gradient from 60% buffer B to 95% buffer B within 30 min and then holding of 95% buffer B for 10 min. Electrospray ionization-mass spectrometry (ESI-MSn) for structural elucidation of the compounds was performed with a Finnigan LTQ FT mass spectrometer (Thermo Electron, Germany) equipped with a static nanospray source. The ionization voltage was optimized in the range of 800 to 1,800 V to obtain a stable signal. The capillary temperature was set to 200°C, capillary voltage to 37 V, and lens tube voltage to 120 V. Accurate values for masses (and therefore chemical formulas) were obtained by using an FT mass analyzer, which was operated with a resolution of 100,000. For structural elucidations, compounds of interest were subjected to MSn fragmentation using the LTQ mass analyzer and an isolation width of 3 *m/z* at a normalized

collision energy of 35 V, an activation Q of 0.25, and an activation time of 30 ms. Fragment ions were alternatively analyzed in the LTQ or FT mass analyzer.

Reverse transcriptase PCR (RT-PCR)

Total RNA from *P. flocculosa* cultures was isolated with TRIzol reagent (Invitrogen, Burlington, Ontario, Canada) following the manufacturer's protocol. *P. flocculosa* was grown at 28°C to the logarithmic phase in YMPD medium (6 g/liter yeast extract, 3 g/liter malt extract, 5 g/liter peptone water, 10 g/liter dextrose) on a rotary shaker. Cells were transferred to either a flocculosin-inducing medium (MOD) (Hammami *et al.*, 2008) or a flocculosin-repressing medium (YMPD) and grown for 24 h at 28°C. RNA was prepared after 24 h.

Total RNA was transcribed into cDNA with SuperScript II reverse transcriptase (Invitrogen, Burlington, Ontario, Canada) following the manufacturer's protocol. A standard PCR using primers RB112 and RB113 (*ahd1*), primers RB18 and RB19 (*fat2*), and primers RB72 and RB73 (*ppi*) was performed (Table 1).

P. flocculosa genome

The complete genome sequence of *P. flocculosa*, assembled into 1,281 scaffolds, was used for protein identification based on homology with *U. maydis* sequences (Teichmann *et al.*, 2011a). The protein sequences of *U. maydis* were used to identify the predicted open reading frame (ORF) in the database of the sequenced genome of *P. flocculosa* by the use of CLC Genomics Workbench version 3.6.1 software (CLC bio USA, Cambridge, MA).

Nucleotide sequence accession number.The sequence for the *ahd1* gene has been deposited in GenBank under accession number JN039370.

3. Results

The *uhd1* and *fhd1* genes show homology to the superfamily of extended SDR.The UA and flocculosin biosynthesis gene clusters contain two homologous genes, *uhd1* (UA hydroxylase) and *fhd1* (flocculosin hydroxylase), that have not been characterized so far. The deduced nucleic acid sequence of *uhd1* encodes a protein of 300 amino acid residues and contains one intron, as confirmed by a query of the MUMDB. The coding region of

fhd1 is composed of three exons interrupted by two introns encoding a protein of 323 amino acids. The introns have been confirmed by cDNA sequencing and alignment with the genome sequence of *P. flocculosa*. To elucidate the function of the two proteins, the deduced amino acid sequences were analyzed with the GenBank BLAST algorithm. This analysis revealed that Uhd1 and Fhd1 are members of the superfamily of the extended short-chain dehydrogenase-reductases (SDR) and show homology to proteins of different plants (*Arabidopsis thaliana* and *Ricinus communis*) and fungi (*Aspergillus flavus* and *Aspergillus oryzae*) (Fig. 3). All proteins shared the typical TGxxGxxG and HxAS patterns of extended SDR family proteins necessary for cofactor binding, as well as the active site pattern YxxxK (Kallberg *et al.*, 2002a; Kallberg *et al.*, 2002b; Fig. 3).

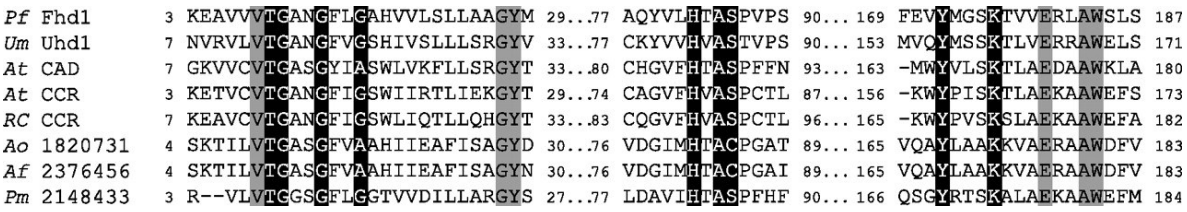


Figure 3 : The genes *uhd1* and *fhd1* show homology to genes encoding the extended short-chain dehydrogenase/reductase (SDR) superfamily. Sequence alignment of the short signature sequences (TGxxGxxG, HxAS, and YxxxK) typical of members of the extended SDR superfamily is shown. The abbreviations representing species and GenBank accession numbers are as follows: Pf Fhd1, *P. flocculosa* ADN97212.1; Um Uhd1, *U. maydis* XP_762613.1; At CAD, *Arabidopsis thaliana* NP_197445.1 At CCR, *A. thaliana* NP_178345.1; Rc CCR, *Ricinus communis* XP_002519377.1; Ao 1820731, *Aspergillus oryzae* XP_001820731.1; Af 2376456, *Aspergillus flavus* XP_002376456.1; Pm 2148433, *Penicillium marneffei* XP_002148433.1. Sequences were aligned using ClustalW. The common motifs are shaded in black. Highly conserved amino acids are shaded in gray.

The *U. maydis* protein Uhd1 hydroxylates the short-chain fatty acid.To determine the function of Uhd1 in *U. maydis*, deletion mutants were generated and the secreted UAs of the mutant strains were analyzed by mass spectrometry. *U. maydis* FB1 wild-type strains produce four different UA derivatives with measured masses of 791.4078 *m/z*, 807.4028 *m/z*, 819.4392 *m/z*, and 835.4345 *m/z* [M + Na]+ (Fig. 4A; Teichmann *et al.*, 2007). The FB1Δ*uhd1* deletion strains secreted four derivatives with measured masses of *m/z* 775.4140, *m/z* 791.4092, *m/z* 803.4455, and *m/z* 819.4383, each of which showed an *m/z* mass difference of 15.994 compared to the masses of the UAs produced by the wild-type strain (Fig. 4C). In each case, this indicated that a hydroxyl group was missing from the

altered UAs produced by Δ*uhd1* mutants, since the calculated mass for one oxygen atom is 15.999 *m/z*. It has already been shown that during UA biosynthesis, the two cytochrome P450 monooxygenases Cyp1 and Cyp2 are necessary for the terminal and subterminal hydroxylation of the long-chain fatty acid while the α-hydroxylase Ahd1 hydroxylates its α position (Hewald *et al.*, 2005; Teichmann *et al.*, 2007). This led to the assumption that Uhd1 might be required for the hydroxylation of the remaining hydroxyl group on the short-chain fatty acid. To confirm this hypothesis, the secreted glycolipids were analyzed by tandem mass spectrometry (MS/MS). Fragmentation of the wild-type UA with the mass of *m/z* 791.4078 showed that the molecule was cleaved into the two glucose molecules, from which the part containing the long-chain fatty acid (*m/z* 515 [M + Na]+) was captured (Fig. 4B; Teichmann *et al.*, 2007). Figure 4D shows the fragmentation spectrum of the smallest compound (*m/z* 775.4140) secreted by *uhd1* deletion strains. The measured fragment showed the same *m/z* value as the wild-type fragment, indicating that the long-chain fatty acid was not affected by the mutation. This confirms the assumption that Uhd1 hydroxylates the β position of the short-chain fatty acid (Fig. 5C).

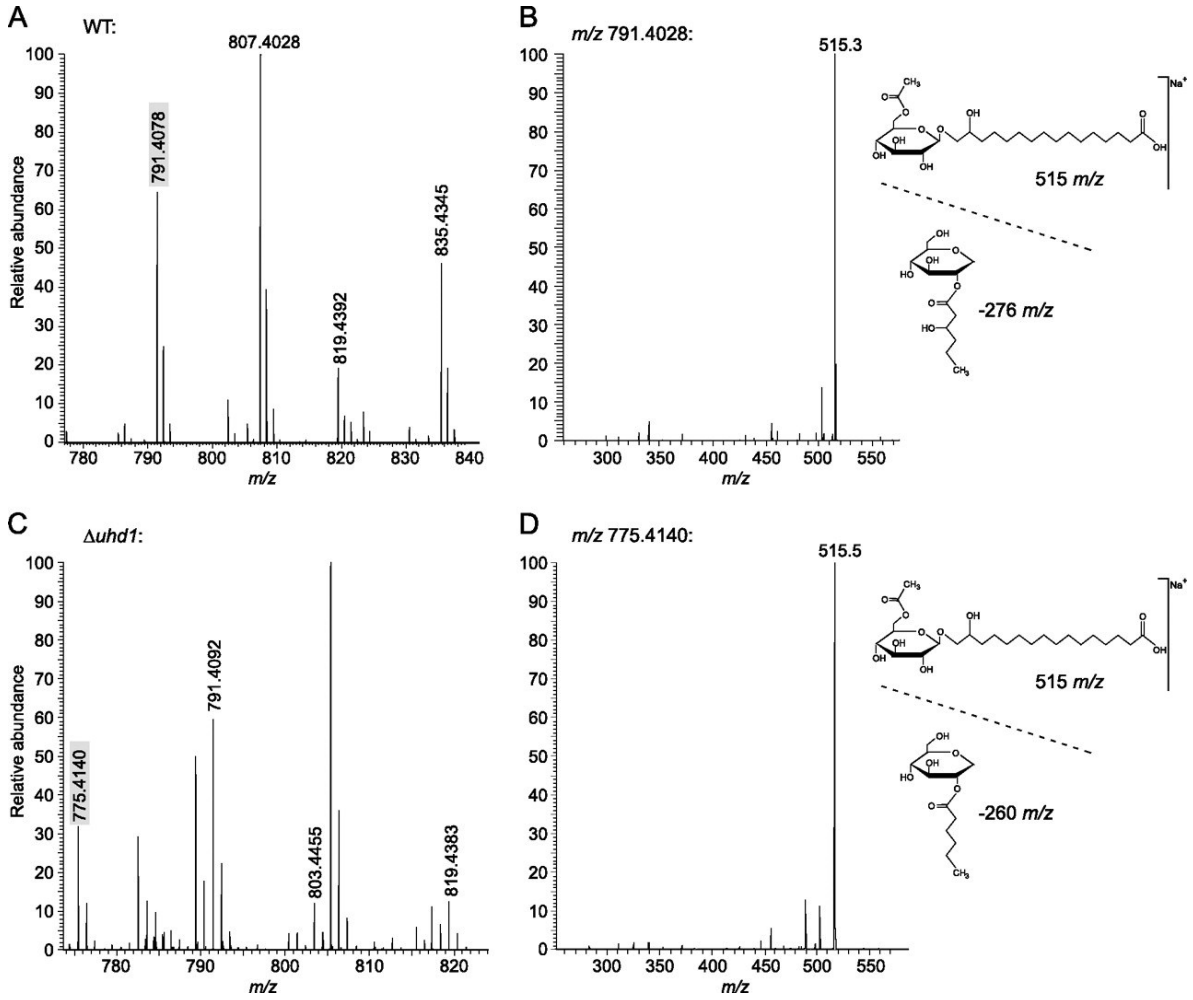


Figure 4 : Mass spectrometry of *uhd1* deletion strains. Glycolipid extracts were subjected to high-resolution FTMS analysis coupled with tandem MS^n. **A,** *U. maydis* wild-type strains secrete four UA derivatives with masses of *m/z* 791.4078, *m/z* 807.4028, *m/z* 819.4392, and *m/z* 835.4345 $[M + Na]^+$. WT, wild type. **B,** Mass fragmentation of the smallest compound (*m/z* 791.4078) showed that the molecule was situated between the two glucose molecules from which the part containing the long-chain fatty acid (*m/z* 515 $[M + Na]^+$) was captured. **C,** Uhd1 deletion strains also synthesize four UA derivatives with masses of *m/z* 775.4140, *m/z* 791.4092, *m/z* 803.4455, and *m/z* 819.4383 $[M + Na]^+$, each differing by exactly 15.994 Da from its corresponding wild-type compound (calculated mass of an oxygen atom, 15.999), suggesting a missing hydroxyl group on the molecule. **D,** Mass fragmentation of the smallest compound (*m/z* 775.4140) secreted by *uhd1* deletion strains. The captured fragment showed the same mass as the wild-type fragment, indicating that the long-chain fatty acid was not affected. This suggests that Uhd1 hydroxylates the short-chain fatty acid.

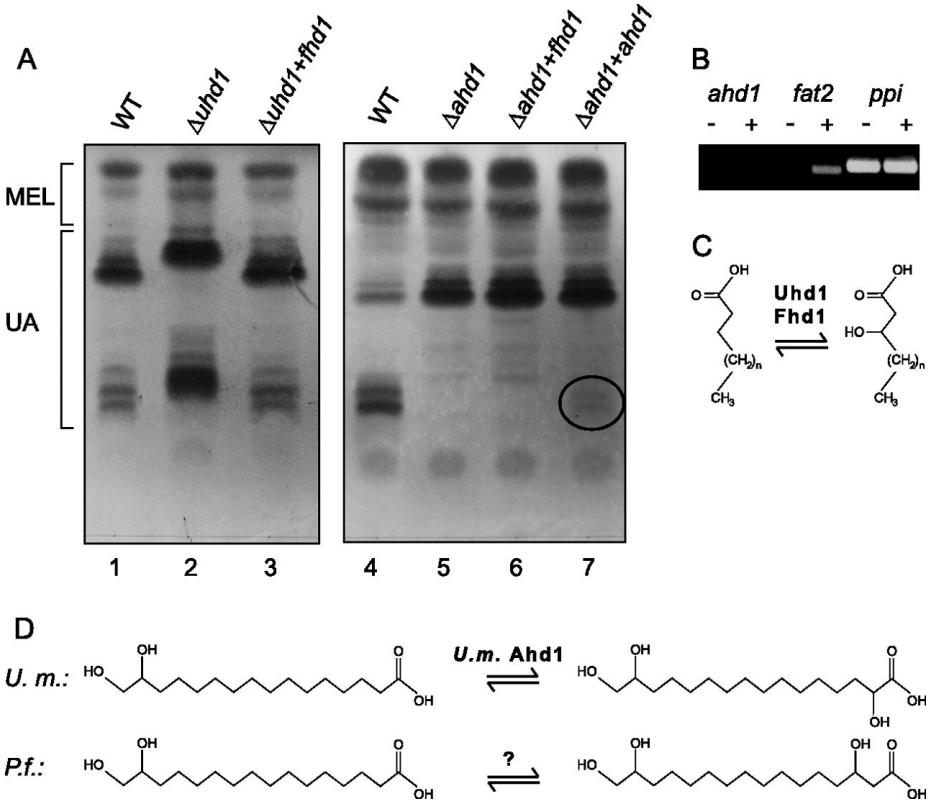


Figure 5 : Functional analysis of Fhd1. **A,** TLC analysis of secreted glycolipids produced by the *U. maydis* FB1 wild-type strain (WT) and mutant strains FB1Δ*uhd1* (Δ*uhd1*), FB1Δ*uhd1*-P_{otef}::*fhd1* (Δ*uhd1+fhd1*), FB1Δ*ahd1* (Δ*ahd1*), FB1Δ*ahd1*-P_{otef}::*fhd1* (Δ*ahd1+fhd1*), and FB1Δ*ahd1*-P_{otef}::P.f.*ahd1* (Δ*ahd1+ahd1*). Deletion of *uhd1* in *U. maydis* resulted in strains producing UAs that showed different UA patterns when analyzed by TLC (lane 2). Expressing the *P. flocculosa* protein Fhd1 in FB1Δ*uhd1* strains rescued the phenotype (lane 3). Introducing *fhd1* in *U. maydis ahd1* deletion strains did not change the phenotype of the mutant (lane 6). *P. flocculosa* Ahd1, when expressed in *U. maydis ahd1* deletion strains, was able to hydroxylate UA, but the reaction was very weak compared to that seen with the wild type (marked with a circle in lane 7). Lanes WT show the typical glycolipid pattern produced by *U. maydis* wild-type strains. "MEL" indicates the mannosylerythritol lipids also produced by *U. maydis*. **B,** Reverse transcriptase PCR of gene *ahd1*. cDNA was prepared from strains under inducing (+) and noninducing (−) conditions. Amplification was performed with primers specific for *P. flocculosa ahd1* (lanes 1 and 2), *fat2* (lanes 3 and 4), and ppi (lanes 5 and 6). While the *fat2* control gene was upregulated under flocculosin-inducing conditions, *ahd1* showed the same expression characteristics under inducing and noninducing conditions. The ppi gene, which encodes the peptidyl-prolyl cis-trans isomerases, is constitutively expressed. **C,** Hypothetical reaction mechanism for Uhd1 and Fhd1. **D,** While the α-hydroxylase Ahd1 hydroxylates UA (Teichmann *et al.*, 2007), it remains unclear which enzyme hydroxylates flocculosin.

Functional analysis of Fhd1 from *P. flocculosa*

To analyze whether the Fhd1 homologous protein from *P. flocculosa* catalyzed the same reaction as Uhd1, we attempted to complement the Δ*uhd1* mutants with *fhd1*, since we

were unable to generate *P. flocculosa* deletion strains (Teichmann *et al.*, 2011a). For this purpose, *fhd1* was constitutively expressed in the FB1Δ*uhd1* mutant strain. Glycolipids produced by *U. maydis* wild-type strains and transformants were analyzed by TLC (Fig. 5A); in each case, all four wild-type UA derivatives could be detected (Fig. 5A, lane 1). However, UA secreted by Δ*uhd1* mutants showed a different glycolipid pattern (lane 2). Figure 5A shows that *fhd1* was able to complement the phenotype of a Δ*uhd1* mutant (lane 3), suggesting that Fhd1 had the same function during flocculosin biosynthesis as Uhd1 during UA biosynthesis (Fig. 5C).

Hydroxylation of the β position of the long-chain fatty acid

One of the major differences between UA and flocculosin is the position of the hydroxyl groups on the long-chain fatty acid: whereas UA is hydroxylated at the α position, a reaction catalyzed by the α-hydroxylase Ahd1 (Teichmann *et al.*, 2007), flocculosin carries a hydroxyl group at the β position. Since the flocculosin gene cluster does not contain an *ahd1* homolog, we wondered whether Fhd1 might also be necessary for this reaction. We therefore expressed the corresponding gene in FB1Δ*ahd1* strains, which are mutant strains producing UAs lacking the α-hydroxyl group on the long-chain fatty acid (Teichmann *et al.*, 2007). TLC analysis of the UAs produced by these mutants showed that Fhd1 was not able to hydroxylate the β position of the long-chain fatty acid (Fig. 5A, lanes 5 and 6), indicating the specificity of Fhd1 for hydroxylation of shorter fatty acids.

To detect a gene that would catalyze hydroxylation of the β position of the long-chain fatty acid, we inspected the genome sequence for candidate genes and found one with 63% homology to *ahd1* (E-value, 2×10^{-101}). TLC analysis of Δ*ahd1* mutant strains expressing *ahd1* of *P. flocculosa* showed that the latter was able to hydroxylate UA but that the reaction was very weak compared to that seen with the wild type (Fig. 5A, lane 7). This led us to question whether *ahd1* was upregulated under flocculosin-producing conditions. Therefore, we analyzed expression of *ahd1* under flocculosin-repressing (-) and -inducing (+) conditions by RT-PCR. Figure 5B shows that the *fat2* control gene, which is part of the cluster, is strongly upregulated under flocculosin-inducing conditions. In contrast, Ahd1 shows the same expression characteristics under either set of conditions, indicating that *ahd1* is not involved in flocculosin biosynthesis (Fig. 5D).

4. Discussion

In this work, we have identified two β-hydroxylase genes, *uhd1* from *U. maydis* and *fhd1* from *P. flocculosa*, that play a role in the biosynthesis of UA and flocculosin, respectively. Both genes encode proteins belong to the extended SDR superfamily involved in different oxidoreductase reactions. By mutational analysis, mass spectrometry, and complementation assays, we have showed that Uhd1 and Fhd1 are necessary for the hydroxylation of the β position of the short-chain fatty acid of their corresponding glycolipids. Deletion of *uhd1* in *U. maydis* resulted in mutant strains producing UA that lacked the β-hydroxyl group on the short-chain fatty acid. Complementation of Δ*uhd1* strains with *fhd1* from *P. flocculosa* restored the Δ*uhd1* phenotype, demonstrating the same function for Fhd1 during flocculosin biosynthesis.

The exact reaction mechanism(s) of Uhd1 and Fhd1 is still unknown. Both proteins show homology to proteins of the extended SDR superfamily. Extended SDRs form a diverse collection of proteins, including isomerases, epimerases, oxidoreductases, and lyases (Bray *et al.*, 2009; Kallberg *et al.*, 2002b). They are similar to the "classical" SDRs but have a less conserved C-terminal extension of approximately 100 amino acids (Kallberg *et al.*, 2002a). All SDRs are characterized by a typical α/β folding pattern, the Rossmann fold, consisting of a central β sheet flanked by α helices required for NAD(P) (H)-binding (Rossmann *et al.*, 1974). Uhd1 and Fhd1in particular showed sequence homology to plant proteins of the cinnamoyl-coenzyme A (cinnamoyl-CoA) reductase (CCR) family involved in lignin biosynthesis and to the dihydroflavonol-4-reductase (DFR) necessary during anthocyanin biosynthesis (Holton et Cornish, 1995; Lacombe *et al.*, 1997; Sparvoli *et al.*, 1994). In both cases, the enzyme catalyzes the reduction of its appropriate substrate, thereby yielding a hydroxylated product. The DFR reaction mechanism has been previously described for Gerbera hybrids. DFR catalyzes the stereospecific reduction of dihydroflavonols to the respective leucoanthocyanidins, which are precursorss for the synthesis of anthocyanins, the major water-soluble pigments in flowers and fruits (Martens *et al.*, 2002). Through this reaction, a ketone is reduced to an alcohol. It has not been shown that the shorter fatty acid carries a ketone group in the β position during flocculosin or UA biosynthesis. If Uhd1 and Fhd1 were necessary for the reduction of a ketone group, deletion of *uhd1* should result in the production of cellobiose lipids carrying a β-ketone

group on the short-chain fatty acid. This indicates that Uhd1 and Fhd1 do not reduce a ketone group but directly transfer a hydroxyl group to the substrate (see Fig. 5C). They might have a reaction mechanism similar to that of cytochrome P450 monooxygenases, which are able to transfer one oxygen atom from molecular oxygen directly to various substrates by the use of NAD(P)H as an electron donor.

Interestingly, Uhd1 and Fhd1 seem to exhibit substrate specificity for short-chain fatty acids. Expressing *fhd1* in *U. maydis* FB1Δ*ahd1* deletion strains did not change the *ahd1* deletion phenotype, indicating that Fhd1 was not able to hydroxylate long fatty acid chains. This specificity might therefore make the enzyme suitable for practical applications in metabolic engineering where one could, for instance, use the enzyme to hydroxylate short-chain fatty acids specifically. Hydroxylated fatty acids have been reported to be valuable compounds in the chemical and medical industry (Bitto *et al.*, 2009).

Which enzyme catalyzes the transfer of the β-hydroxyl group to the long-chain fatty acid of flocculosin is still an open question. We demonstrated that neither Fhd1 nor Ahd1 was responsible for this reaction. For *U. maydis*, it has been shown that hydroxylation of the α position is done by Ahd1 (Teichmann *et al.*, 2007). Expression of a *P. flocculosa* homolog (located outside the flocculosin biosynthesis cluster) in *U. maydis* strains lacking *ahd1* showed that this protein had very low specificity with respect to hydroxylation of UA. Based on this result, the fact that it is not expressed under flocculosin-producing conditions, and its exclusion from the cluster, we conclude that *P. flocculosa* Ahd1 is not significantly involved in flocculosin biosynthesis. As a matter of fact, one could even question the necessity of *ahd1* in the *U. maydis* cluster, given that only half of the molecules are hydroxylated. Preliminary data show that UA produced by Δ*ahd1* strains still show antibiotic activities (B. Teichmann and M. Bölker, unpublished data), but we do not know whether they are lacking another characteristic. The fact that *P. flocculosa* found another mechanism to systematically hydroxylate the long-chain fatty acid seems to indicate that this hydroxyl group is essential for the activity of the molecule or, at the very least, that it confers a fitness advantage to the organism.

In conclusion, we have found that the production of unusual glycolipids by two related yet disparate organisms is the result of an intricate and well-conserved enzymatic process

exclusive to the two studied fungi. From a biological or evolutionary point of view, one has to assume that conservation of this cluster serves a distinct purpose, even though evidence to that effect is still lacking. Two closely related plant pathogens, *Sporisorium reilianum* and *Ustilago hordei*, lack such a cluster (Schirawski *et al.*, 2010; J. Schirawski, R. Kahmann, and G. Bakkeren, personal communication) but are much less prevalent than *U. maydis* (Schirawski *et al.*, 2010). This could indicate that production of glycolipids with antibiotic activity does extend the ecological niche and enhance the ability to reproduce.

Acknowledgments

The work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs Program to R.R.B. and by grant BO 2094/3-1 from the Deutsche Forschungsgemeinschaft (DFG) to M.B. B.T. is grateful to the DFG for financial support (Project TE 815/1-1).

5. References

Avis, T.J. and Bélanger, R.R. (2002). Mechanisms and means of detection of biocontrol activity of *Pseudozyma* yeasts against plant-pathogenic fungi. *FEMS Yeast Research*, *2*(1), 5-8.

Banuett, F. and Herskowitz, I. (1989). Different a alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(15), 5878-5882.

Bitto, N.J., Graichen, F.H. and Monahan, B.J. (2009). Functionality at the end of a fatty acid chain chemical and biological routes to ω-hydroxylated fatty acids. *Lipid Technology*, *21*(10), 216-219.

Brachmann, A., König, J., Julius, C. and Feldbrügge, M. (2004). A reverse genetic approach for generating gene replacement mutants in *Ustilago maydis*. *Molecular Genetics and Genomics*, *272*(2), 216-226.

Bray, J.E., Marsden, B.D. and Oppermann, U. (2009). The human short-chain dehydrogenase/reductase (SDR) superfamily: a bioinformatics summary. *Chemico-biological Interactions*, *178*(1), 99-109.

Cameotra, S.S. and Makkar, R.S. (2004). Recent applications of biosurfactants as biological and immunological molecules. *Current Opinion in Microbiology*, *7*(3), 262-266.

Hammami, W., Labbé, C., Chain, F., Mimee, B. and Bélanger, R.R. (2008). Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa*. *Applied Microbiology and Biotechnology*, *80*(2), 307-315.

Haskins, R.H. and Thorn, J.A. (1951). Biochemistry of the Ustilaginales VII. Antibiotic activity of ustilagic acid. *Canadian Journal of Botany*, *29*(6), 585-592.

Hewald, S., Josephs, K. and Bolker, M. (2005). Genetic analysis of biosurfactant production in *Ustilago maydis*. *Applied and Environmental Microbiology*, *71*(6), 3033-3040.

Holton, T.A. and Cornish, E.C. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *The Plant Cell*, *7*(7), 1071-1083.

Kallberg, Y., Oppermann, U., Jörnvall, H. and Persson, B. (2002a). Short-chain dehydrogenase/reductase (SDR) relationships: A large family with eight clusters common to human, animal, and plant genomes. *Protein Science*, *11*(3), 636-641.

Kallberg, Y., Oppermann, U., Jörnvall, H. and Persson, B. (2002b). Short-chain dehydrogenases/reductases (SDRs). *European Journal of Biochemistry*, *269*(18), 4409-4417.

Kämper, J. (2004). A PCR-based system for highly efficient generation of gene replacement mutants in *Ustilago maydis*. *Molecular Genetics and Genomics*, *271*(1), 103-110.

Lacombe, E., Hawkins, S., Doorsselaere, J., Piquemal, J., Goffner, D., Poeydomenge, O., et al. (1997). Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *The Plant Journal*, *11*(3), 429-441.

Martens, S., Teeri, T. and Forkmann, G. (2002). Heterologous expression of dihydroflavonol 4-reductases from various plants. *FEBS Letters*, *531*(3), 453-458.

Mimee, B., Labbé, C., Pelletier, R. and Bélanger, R.R. (2005). Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa*. *Antimicrobial Agents and Chemotherapy*, *49*(4), 1597-1599.

Mimee, B., Pelletier, R. and Bélanger, R.R. (2009). In vitro antibacterial activity and antifungal mode of action of flocculosin, a membrane-active cellobiose lipid. *Journal of Applied Microbiology*, *107*(3), 989-996.

Paulitz, T.C. and Bélanger, R.R. (2001). Biological control in greenhouse systems. *Annual Review of Phytopathology*, *39*(1), 103-133.

Ron, E.Z. and Rosenberg, E. (2001). Natural roles of biosurfactants. *Environmental Microbiology*, *3*(4), 229-236.

Rossmann, M.G., Moras, D. and Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature*, *250*(5463), 194-199.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning* (Vol. 2). New York, NY, USA: Cold Spring Harbor Laboratory Press.

Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*, *330*(6010), 1546-1548.

Sparvoli, F., Martin, C., Scienza, A., Gavazzi, G. and Tonelli, C. (1994). Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). *Plant Molecular Biology*, *24*(5), 743-755.

Spellig, T., Bottin, A. and Kahmann, R. (1996). Green fluorescent protein (GFP) as a new vital marker in the phytopathogenic fungus *Ustilago maydis*. *Molecular & General Genetics*, *252*(5), 503-509.

Teichmann, B., Labbé, C., Lefebvre, F., Bölker, M., Linne, U. and Bélanger, R.R. (2011). Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa*. *Molecular Microbiology*, *79*(6), 1483-1495.

Teichmann, B., Linne, U., Hewald, S., Marahiel, M.A. and Bölker, M. (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Molecular Microbiology*, *66*(2), 525-533.

Chapitre 5 : The Transition from a Phytopathogenic Smut Ancestor to an Anamorphic Biocontrol Agent Deciphered by Comparative Whole-Genome Analysis

Lefebvre, F., Joly, D.L., Labbé, C., Teichmann, B., Linning, R., Belzile, F., et al. (2013). The transition from a phytopathogenic smut ancestor to an anamorphic biocontrol agent deciphered by comparative whole-genome analysis. *The Plant Cell*, *25*(6), 1946-1959

Résumé

Pseudozyma flocculosa est une espèce fongique apparentée à l'agent pathogène modèle *Ustilago maydis*. Cependant, plutôt qu'être phytopathogène, le premier s'avère être un redoutable agent de lutte biologique efficace contre la maladie du blanc (*oïdium*). Ce lien inusité en fait un organisme unique pour l'étude de l'évolution des facteurs de pathogénicité des agents phytopathogènes. Le génome de *P. flocculosa,* d'une taille approximative de 23 Mb, contient 6 877 gènes prédits et codant pour des protéines. La composition de ce génome, incluant les piliers de la pathogénicité, est très similaire entre *P. flocculosa*, *U. maydis, Sporisorium reilianum* et *Ustilago hordei*. De plus, *P. flocculosa*, se reproduisant uniquement de façon asexuée, a conservé les gènes associés à la conjugaison (*mating-type*) et à la méiose typiques des Ustilaginales. Par opposition, nous avons observé que la perte d'un ensemble de protéines-effecteurs sécrétées candidates ayant une influence sur la virulence d'*U. maydis* constitue la divergence principale qui pourrait expliquer la nature non pathogène de l'agent de lutte biologique. Ces résultats suggèrent que *P. flocculosa* pourrait avoir été un agent causal d'une maladie végétale, mais que sa virulence pourrait avoir été compromise à la suite de la perte de protéines-effecteurs spécifiques et nécessaires à la compatibilité avec l'hôte. De plus, il est intéressant d'observer que l'agent de lutte biologique semble posséder des gènes absents chez les Ustilaginales auxquelles il a été comparé et qui codent des protéines sécrétées, incluant des protéines contenant des motifs de la *necrosis-inducing-Phytophthora-protein* et de lysine. Ces effecteurs candidats pourraient avoir un impact direct sur son mode de vie. La disponibilité de la séquence génomique nouvellement annotée de *P. flocculosa* devrait contribuer à mettre en lumière de nouvelles avenues permettant de comprendre les subtiles différences qui mènent à un changement de mode de vie aussi drastique chez ce champignon.

Abstract

Pseudozyma flocculosa is related to the model plant pathogen *Ustilago maydis* yet is not a phytopathogen but rather a biocontrol agent of powdery mildews; this relationship makes it unique for the study of the evolution of plant pathogenicity factors. The *P. flocculosa* genome of ~23 Mb includes 6877 predicted protein coding genes. Genome features, including hallmarks of pathogenicity, are very similar in *P. flocculosa* and *U. maydis*, *Sporisorium reilianum*, and *Ustilago hordei*. Furthermore, *P. flocculosa*, a strict anamorph, revealed conserved and seemingly intact mating-type and meiosis loci typical of Ustilaginales. By contrast, we observed the loss of a specific subset of candidate secreted effector proteins reported to influence virulence in *U. maydis* as the singular divergence that could explain its nonpathogenic nature. These results suggest that *P. flocculosa* could have once been a virulent smut fungus that lost the specific effectors necessary for host compatibility. Interestingly, the biocontrol agent appears to have acquired genes encoding secreted proteins not found in the compared Ustilaginales, including necrosis-inducing-Phytophthora-protein- and Lysin-motif- containing proteins believed to have direct relevance to its lifestyle. The genome sequence should contribute to new insights into the subtle genetic differences that can lead to drastic changes in fungal pathogen lifestyles.

1. Introduction

The Ustilaginomycetes, commonly known as smut fungi, comprise important plant pathogens that have been extensively studied for their pathogenic determinants. The smut fungi studied share common features essential for pathogenicity, such as mating, and the different mating types intrinsic to smut fungi have been reviewed in detail Bakkeren *et al.*, 2008. Indeed, as part of their sexual life cycle, haploid cells of compatible mating type must fuse and generate dikaryotic hyphae in order to infect and penetrate the host. Once inside the plant, these smut fungi establish a biotrophic interaction with their host in which infected cells remain alive throughout the disease cycle.

Over the years, the *Ustilago maydis* interaction with maize (Zea mays) has become the model pathosystem to investigate factors essential for the establishment of the biotrophic interaction typical of Ustilaginomycetes. In 2006, the generation and analysis of the genome sequence of *U. maydis* has provided unprecedented insight into novel genes that play key roles during pathogenesis (Kämper *et al.*, 2006). Among these is a distinctive set of genes that code for secreted proteins referred to as effector proteins (or effectors), of which many have unknown function. However, some have since been reported to be essential for infection and several seem to interfere with plant defense responses, thus facilitating infection by the pathogen (Brefort *et al.*, 2009; Doehlemann *et al.*, 2011, 2009).

First described in prokaryotes, effectors have now been found in all important groups of plant pathogens, including powdery mildews, rusts, and oomycetes. In the latter group, many effectors share a common conserved Arg-X-Leu-Arg (RXLR) motif that plays an important role in translocation of the effectors into the plant cell (Whisson *et al.*, 2007). In the case of *U. maydis*, the secreted effectors lack such a motif, but many are arranged in clusters and are upregulated upon recognition of the host plant, invasion, and in tumor tissue. Through cluster deletion analysis, the importance in pathogenicity of these novel effectors was established, the first such demonstration for a eukaryotic pathogen (Kämper *et al.*, 2006; Schirawski *et al.*, 2010).

Comparison of the *U. maydis* genome to genomes of other sequenced basidiomycetes, including *Laccaria bicolor*, *Cryptococcus neoformans*, *Phanerochaete chrysosporium*, and *Malassezia globosa*, revealed great variability in terms of genome size, synteny, structure,

amount of repetitive elements, and number and homology of effector proteins. These observations reinforced the relevance of using phylogenetically related species to improve our understanding into the evolution of pathogenic factors through comparative genomic approaches. Additionally, the recent genome annotation of *Sporisorium reilianum* (Schirawski *et al.*, 2010) and *Ustilago hordei* (Laurie *et al.*, 2012) has provided an opportunity to perform comparative analyses that highlight common and divergent features among these closely related pathogens. Of particular interest was the finding that all three fungi share many conserved effectors but also possess sets of species-specific effectors. These results suggest that common effectors are essential for basic plant pathogenicity, while species-specific effectors may be responsible for divergent interaction strategies, targeting different host tissues and molecules and thereby dictating host specificity and pathogen development.

In earlier classification, all smuts were ecologically characterized by their ability to infect plants and shared a similar life cycle with a yeast-like haploid phase and a parasitic dikaryophase, culminating in the production of numerous powdery black teliospores, hence, their common name (Begerow *et al.*, 2006). However, a number of anamorphic fungi lacking sexual development, initially placed in deuteromycetous taxa, were found to be morphologically and phylogenetically related to the Ustilaginales. In order to integrate these anamorphs into the general phylogenetic system of Ustilaginomycetes, Begerow *et al.* (2000) analyzed and compared diagnostic ribosomal DNA sequences of teleomorphic and anamorphic species of Ustilaginomycetes. Their analyses confirmed that species of Pseudozyma and Ustilaginales parasitizing grasses form a monophyletic group. Pseudozyma species thus represent the sole known members of the Ustilaginales that cannot parasitize plants. On the basis of these findings, it appears that Pseudozyma spp are either anamorphic vestiges or descendants of plant pathogenic smut species that have lost their phytoparasitic and/or sexual capacity. It was recently proposed that comparative genomics, in the context of understanding the evolution of biotrophic pathogens, should target phylogenetically relevant organisms that are not pathogens (Kemen et Jones, 2012). As such, Pseudozyma spp constitute a rare and unique model to investigate the determinants that differentiate them from pathogenic Ustilaginales.

Pseudozyma spp have received little attention in the literature. One notable exception is *Pseudozyma flocculosa* (Traquair, Shaw, and Jarvis) Boekhout and Traquair, a species that has been studied for its ability to antagonize powdery mildews (Figure 1). *P. flocculosa* was discovered in 1987 and originally identified as *Sporothrix flocculosa*, an ascomycetous yeast (Traquair *et al.*, 1988). Because of its potential as a biofungicide, it was extensively characterized for its properties, mode of action, secondary metabolites, and environmental fate, before being registered by the Environmental Protection Agency and Pest Management Regulatory Agency in the US and Canada, respectively (Bélanger *et al.*, 2012). *P. flocculosa* has no known sexual stage and is considered a leaf epiphyte although the sole report of its isolation specified it was found in association with clover powdery mildew. This particularity may explain its rarity in nature and the scarcity of known isolates, with only six reported but most of them being isolated from the same strain following biocontrol trials (Avis *et al.*, 2001). However, its reclassification among Ustilaginomycetes in 2000 (Begerow *et al.*, 2000) launched unexpected opportunities to better characterize *P. flocculosa*. At first, the notion that a biocontrol agent was related to smut fungi, including the well-studied model species *U. maydis*, appeared biologically incongruous. Nevertheless, many observations have since confirmed its link to *U. maydis*. Cheng *et al.* (2001) were the first to bring forth biological evidence by observing that the fungus was only amenable to genetic transformation when using promoters from *U. maydis* to drive expression of a selectable marker. This was followed by the discovery of an unusual glycolipid produced by *P. flocculosa* that was nearly identical to ustilagic acid, a rare metabolite identified in culture filtrates of *U. maydis* in 1951(Lemieux *et al.*, 1951; Mimee *et al.*, 2005). Taking advantage of the annotated genome of *U. maydis* and based on specific chemical characteristics of the molecule, Bölker's group was able to find a cluster of 10 genes responsible for the synthesis of ustilagic acid in *U. maydis* (Teichmann *et al.*, 2007). As expected, a similar cluster, containing homologous genes and responsible for the synthesis of the glycolipid, was recently reported in *P. flocculosa* (Teichmann *et al.*, 2011a).

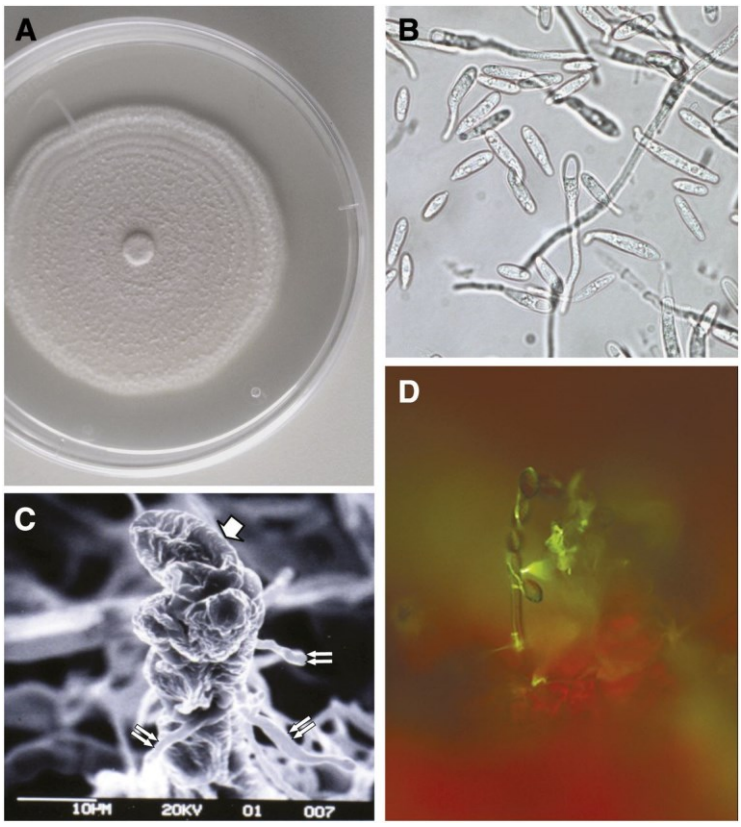


Figure 1 : Different life forms of the biocontrol agent *P. flocculosa*. **A,** *P. flocculosa* growing on potato dextrose agar medium. **B,** sporidia in 24-h-old culture of *P. flocculosa*. **C,** scanning electron micrograph of collapsed cucumber powdery mildew conidia (*Podosphaera xanthii*; wide arrow) antagonized by *P. flocculosa* hyphae (double arrow). **D,** green fluorescent protein–transformed *P. flocculosa* coiling around conidia of cucumber powdery mildew.

Here, we report the generation and analysis, including annotation, of the complete genome of *P. flocculosa*. Comparison to the genomes of phytopathogenic Ustilaginales (*U. maydis*, *S. reilianum*, and *U. hordei*) highlighted similarities and differences. Particular emphasis was placed on reported hallmarks of pathogenicity and biocontrol activity, such as mating factors, cell wall–degrading enzymes (CWDEs), secondary metabolites, and effector proteins in an effort to identify the conserved and/or lost elements that account for the divergent lifestyle of *P. flocculosa*.

2. Results

Genome Statistics

The genome of *P. flocculosa* was sequenced using the Roche 454 sequencing platform. Three sequencing runs yielded 522 Mb of whole-genome sequencing data and 240 Mb of paired-end data to generate an average genome coverage of 28 X (Table 1). The assembly yielded 1583 contigs from which 1187 were oriented and ordered into 37 scaffolds to which three contigs larger than 2 kb were added. The N50 value shows that 50% of the bases that compose the assembly are contained in scaffolds larger than 920 kb, the longest scaffold being 1.975 Mb long.

Table 1 : Comparative Genome Statistics of *P. flocculosa*, *U. maydis*, *U. hordei*, and *S. reilianum*

	P. flocculosa	*U. maydis*	*U. hordei*	*S. reilianum*
Assembly Statistics				
Total contig length (Mb)	23.2	19.7	20.7	18.2
Total scaffold length (Mb)	23.4	19.7	21.2	18.5
Average base coverage	28x	10x	20x	20x
N50 contig (kb)	38.6	127.4	48.7	50.3
N50 scaffold (kb)	919.9	885.1	307.7	772.4
GC content (%)	65.1	54.0	52.0	59.7
Assembled	65.2	54.0	52.2	59.7
Coding	66.3	56.3	54.3	62.6
Noncoding	63.7	50.5	49.0	54.3
Coding sequences				
Percentage coding (%)	54.3	61.1	57.3	64.7
Average gene size (bp)	2,097	1,836	1,782	1,858
Average gene density (genes/Mb)	291.5	344.5	335.4	360.7
Protein coding genes	6,877	6,787	7,111	6,673
Homology-supported gene calls[a]	6,239 (90.5%)	6,475 (95.4%)	6,443 (90.6%)	6,450 (96.7%)
RNA-seq–supported gene calls[b]	6,142 (89.3%)	n.a.[c]	n.a	n.a
RNA-seq and homology-supported predictions[ab]	5,673 (82.3%)	n.a.	n.a	n.a
Ab initio only predictions	183 (2.7%)	312 (4.6%)	668 (9.4%)	223 (3.3%)
Exons	19,318	9,778	10,995	9,777
Average exon size	658	1,230	1,103	1,222
Exons/gene	2.8	1.44	1.55	1.47
tRNA genes	176	111	110	96
Noncoding sequences				
Introns	12,427	2,991	3,884	3,104
Introns/gene	1.8	0.44	0.55	0.47
Intron-free genes	1,565 (22.7%)	4,910 (72.3%)	4,821 (67.8%)	4,732 (70.9%)
Average intron length (base)	141	146	141	146
Average intergenic length (bp)	1,273	1,061	1,098	919
Transposable elements and repeats (%)	3.4	2.0	10.0	0.8

[a.] Homology-supported gene calls are based on BLASTP similarity search to the nonredundant GenBank database (cutoffs: e-value 1e-05, bits 80).

[b.] RNA-sequencing–supported gene calls are based on tBLASTN similarity search to database of retrieved transcript sequences after manual curation (cutoffs: e-value 1e-30, alignment on >70% of protein).

[c.] n.a., not available.

Of the 40 genome scaffolds, seven are considered to represent complete chromosomes since telomeric motifs (TTAGGG) are present at both ends and 17 present such a motif at one end. The total assembly of the *P. flocculosa* genome is slightly bigger in size than that of the available phytopathogenic Ustilaginales, but the number of protein coding genes is remarkably similar (Table 1). Accordingly, a lower proportion of coding sequences is

found in the genome, correlating with lower gene density. Regarding structural annotation, more than 80% of the gene calls are supported by homology to known proteins and manually curated transcriptomic data, while 8.2 and 6.8% are supported by homology to known proteins or transcriptomic data alone. Less than 3% of the genes are based solely on ab initio predictions.

The most important differences between the genome structures of *P. flocculosa* and the three Ustilaginales were found in the proportion of guanine and cytosine (GC) residues and in the structure of genes. First, *P. flocculosa* has a substantially higher GC content than the three phytopathogenic Ustilaginales spp, and this proportion exceeded 66% in the coding regions (Table 1). All codon positions are richer in GC with the highest proportion reaching 82.3% at the third position. Second, genes identified in *P. flocculosa* contained an average of 4 times more introns than *U. maydis* (Table 1). Consequently, *P. flocculosa* contains roughly three times fewer genes without any intron, though the size of the introns is well conserved across all four species. Rapid analysis of intron position within genes showed that introns of all three phytopathogenic Ustilaginales are depleted toward the 3′ end of genes, while introns of *P. flocculosa* are more equally distributed across genes (see Supplemental Figures 1 and 2 online).

The number of transposable elements and simple repeats found in the *P. flocculosa* genome is similar to that reported in *U. maydis* (Table 1; see Supplemental Tables 1 and 2 online). It is 3 times lower than in *U. hordei*, in which species-specific transposable element families seem to have expanded over time (Laurie *et al.*, 2012). Repetitive elements do not cluster in specific areas of the genome but appear to be randomly scattered (data not shown). Some of them bear signatures specific to some of the elements in the other Ustilaginales, such as the long terminal repeat (LTR) retrotransposon and LINE-type elements. Repeat-induced point–like mutation signatures as described for *U. hordei* could not be found in *P. flocculosa*, a situation similar to the one reported in *U. maydis* and *S. reilianum* (see Supplemental Figure 3 online).

Genome Synteny

Genome synteny evaluated by genome-to-genome alignments is an efficient way to assess the level of conservation in gene order and content. It allows the visualization of

translocation, inversion, and gene loss events that occurred during evolution from a common ancestor. In the case of *P. flocculosa*, its level of genome synteny with *U. maydis* is much lower than that observed between *U. maydis* and *S. reilianum* (Figure 2; Schirawski *et al.*, 2010). However, many smaller *P. flocculosa* genome regions do share a common content and gene orientation with genome regions in *U. maydis*. Figure 2 is visually similar to mesosynteny described by Hane *et al.* (2011). These observations are further supported by OrthoCluster analysis (see Supplemental Table 3 online). The lowest levels of synteny conservation are obtained when comparing any of the three phytopathogenic species with *P. flocculosa*.

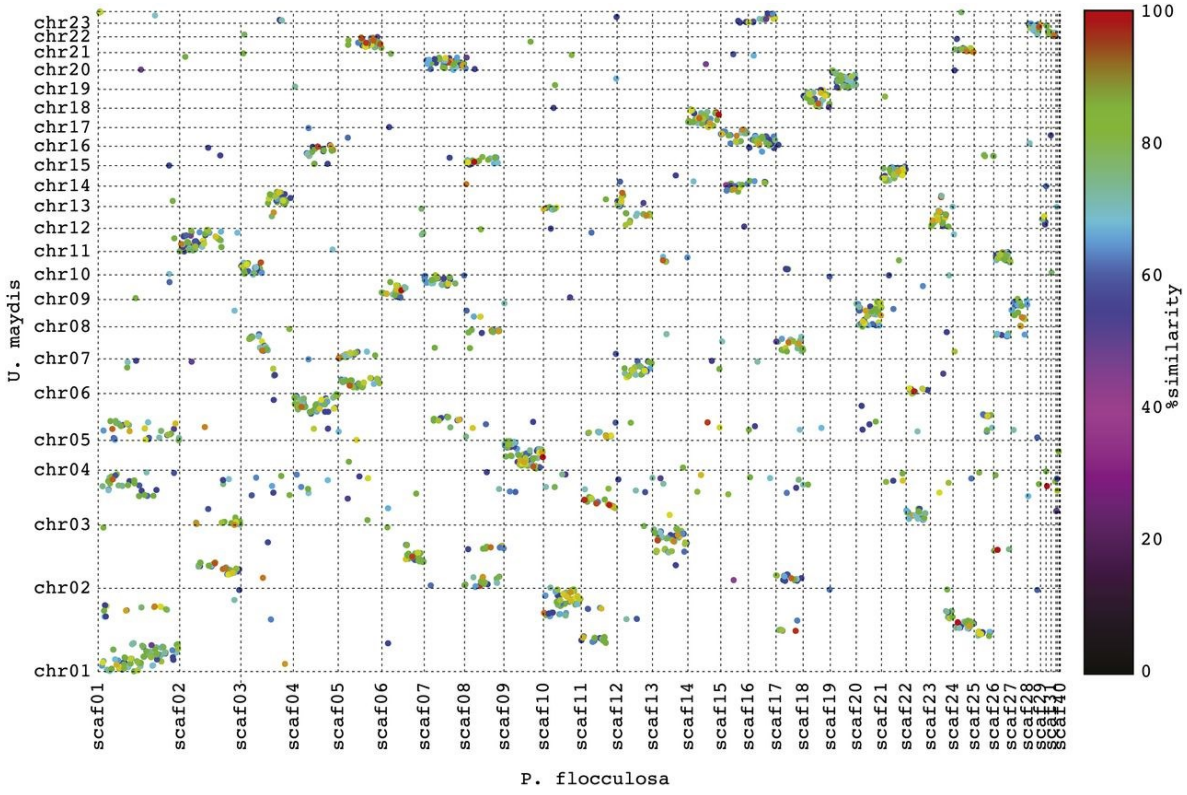


Figure 2 : Alignment of *P. flocculosa* and *U. maydis* genome assemblies. Regions of the genomes were compared on the basis of their amino acid sequences generated from six-frame translations. The graph was plotted using the promer program from MUMmer 3.0 with the following options: -mum (promer) and -color (mummerplot). Colors refer to the percent similarity of each alignment as shown by the color chart on the right.

Gene Families

In order to group genes with similar functions regardless of whether that function is documented or not, we used BLASTP (Atschul *et al.*, 1990) to perform an all-against-all

similarity search with the 27,448 genes from the four organisms and Tribe-MCL (Enright *et al.*, 2002) to generate clusters of genes based on their similarity. All genes were grouped into 5841 clusters, further referred to as gene families.

A total of 602 genes (2.5%) were grouped in species-specific families, while the bulk of genes (24,105, and 87.8%) were grouped in core families present in all four organisms (Figure 3A; see Supplemental Data Set 1 online). Nearly 91% (6238) of *P. flocculosa* genes are part of the core gene families, while only 4.0% (275) are specific to this biocontrol agent, the vast majority (>90%) lacking known functions. Based on gene families, *P. flocculosa* has the least diversified genome among the four species, with 5288 gene families (Figure 3A).

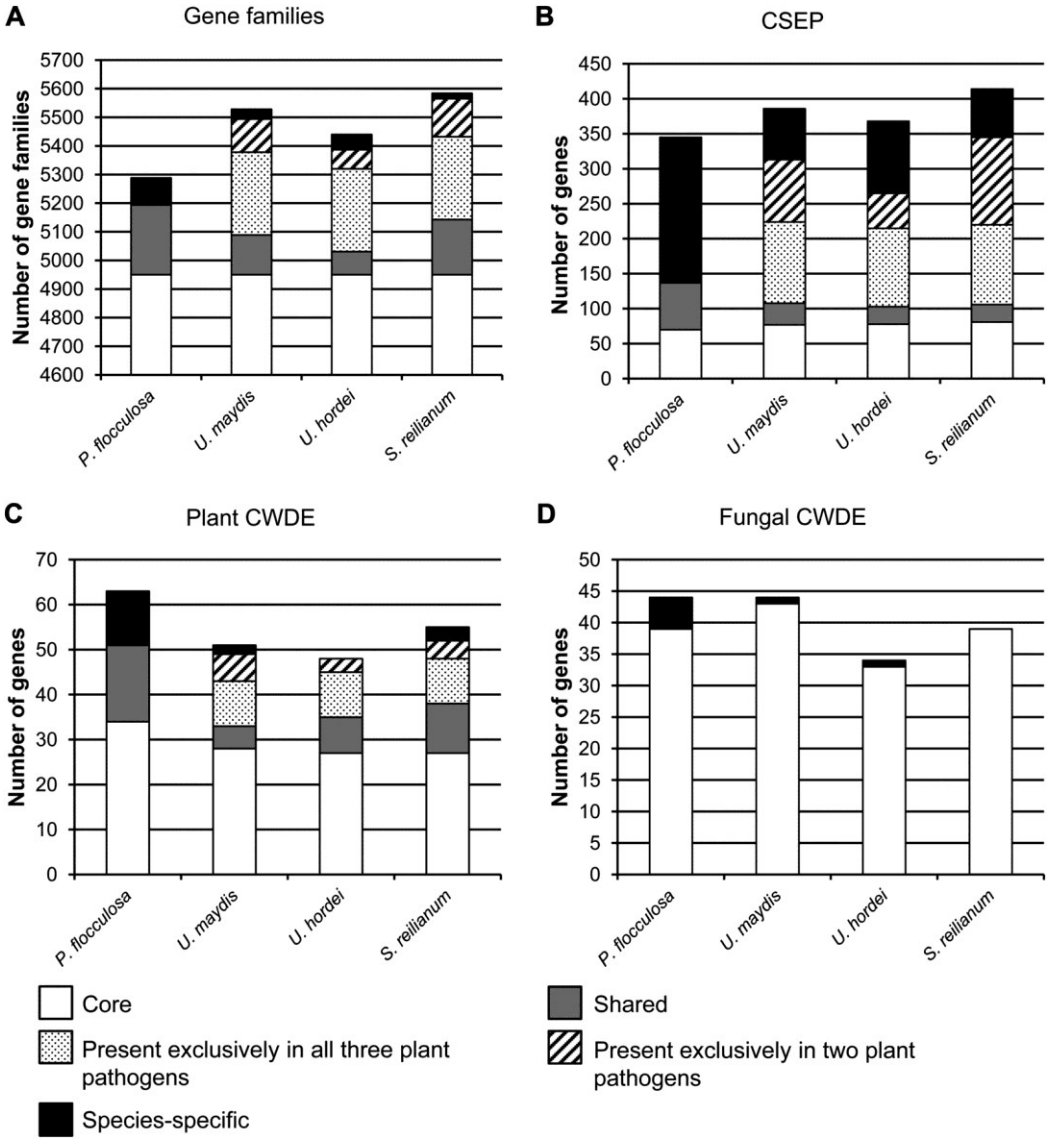


Figure 3 : Composition of four gene categories based on gene families in four ustilaginomycete fungi. Four types of gene families were defined: “Core” refers to gene families composed of genes present in all four species. “Shared” represents gene families composed of genes found in two or three species including *P. flocculosa*. The two other types of gene families refer to genes families “present exclusively in two plant pathogens” or “present exclusively in all three plant pathogens.” “Species-specific” refers to gene families composed of genes found in only one species. A, graph of gene families reports frequencies of all gene families present in species; B to D, otherwise, genes were classified according to the category (CSEP, plant CWDEs, and fungal CWDE), the species, and the type of gene family to which they belong. For specific gene product categories, types of gene families were defined exclusively on the basis of genes belonging to the targeted category and being present in the family.

The data generated in this exhaustive similarity search were used to derive with a high level of confidence and using 306 core gene families, the phylogenetic relationship between the

four species and M. globosa (Figure 4). Based on the consensus tree, *P. flocculosa* is indeed more closely related to Ustilaginales than Malassezia but does branch outside of the three pathogens.

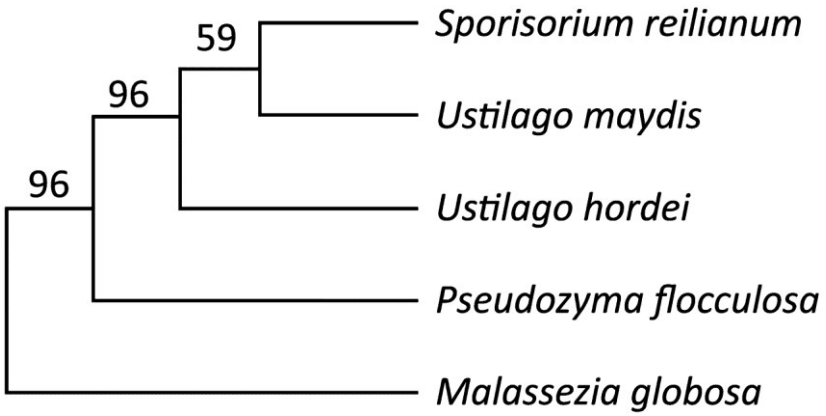


Figure 4 : Molecular phylogeny of five sequenced Ustilaginomycota. Consensus phylogenetic tree constructed from 306,000 trees (306 gene families × 1000 bootstrap replicates) estimated by the maximum likelihood method under the constraint of a molecular clock and the JTT matrix-based model (Jones *et al.*, 1992). Each tree was estimated from one bootstrap replicate generated from the alignment of amino acid sequences from a gene family containing only one member per species. Families had to group genes encoding an ortholog involved in Metabolism or Cellular Process as defined by KEGG (Kanehisa et Goto, 2000). Trees were rooted using the midpoint criterion. Branch lengths are unscaled.

Mating, Dimorphic Switch, and Meiosis

P. flocculosa is not known to be a plant pathogen, and no sexual stage has ever been discovered; the biocontrol activity toward mildews is achieved by the haploid form (Bélanger et Deacon, 1996). Therefore, a particular emphasis was placed on examining the presence or absence in *P. flocculosa* of genes typical of pathogenic activity in Ustilaginales. Since successful mating between haploid yeast-like cells is essential to trigger a switch to dikaryotic hyphal growth and pathogenicity in the Ustilaginales, genes involved in these processes are of great importance. Complete mating-type genes constituting the Ustilaginales a and b loci were localized in the *P. flocculosa* genome assembly, on scaffolds 4 and 10, respectively. The b locus was similarly organized in two divergently transcribed, homeodomain-containing heterodimeric transcription factor open reading frames, bE1 and bW1. The a locus housed a pheromone receptor gene, pra1, and two mating pheromone genes, mfa1.1 and mfa1.2, arranged in an orientation and spacing

similar to the *S. reilianum* a3 locus (Schirawski *et al.*, 2005) (see Supplemental Table 4 online). A phylogenetic analysis based on amino acid sequences of Pra (STE3 homolog) revealed a similar relationship between *P. flocculosa* and other Ustilaginales species (Figure 5) as derived from the previous comparison of the 306 core gene families (Figure 4). Despite a close phylogenetic relationship, there was no indication *P. flocculosa* was responsive to mating interactions with various *U. maydis* strains and mating types (see Supplemental Figure 4 online). Second, all genes coding for the proteins involved in the mitogen-activated protein kinase and cAMP signaling cascades downstream of the G protein–coupled pheromone receptor and leading to crucial morphological and physiological modifications in pathogenic Ustilaginales were present and mostly highly conserved in *P. flocculosa* (see Supplemental Data Set 2 online). In *U. maydis*, meiosis occurs after the mating of compatible haploid mating types and in planta development of diploid teliospores (Saville *et al.*, 2012). All genes reported to be involved in meiosis and sporulation were found in the genome of *P. flocculosa* (see Supplemental Data Set 3 online).

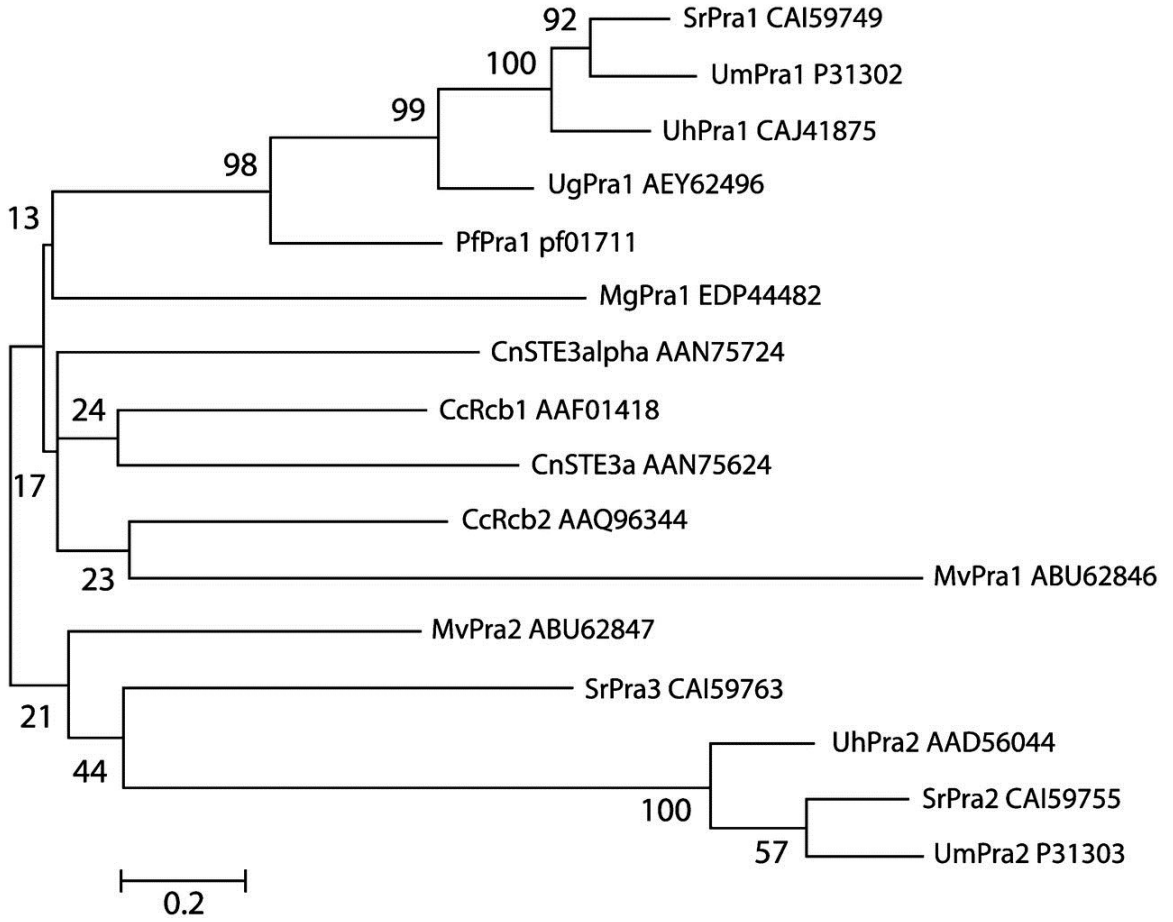


Figure 5 : Molecular phylogeny of basidiomycete pheromone receptor protein sequences. Maximum likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992) and evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011). Pheromone receptor sequences were C-terminally truncated to exclude the cytoplasmic tail and to optimize the alignment and were from the following species (protein name, accession number, number of amino acids): Coprinopsis cinerea (CcRcb1, GenBank AAF01418, 293) and (CcRcb2, GenBank AAQ96344, 298); Cryptococcus neoformans (CnSte3a, GenBank AAN75624, 295) and (CnSte3alpha, GenBank AAN75724, 295); Malassezia globosa (MgPra1, GenBank EDP44482, 298); Microbotryum violaceum (MvPra1, GenBank ABU62846, 300) and (MvPra2, GenBank ABU62847, 299); *Pseudozyma flocculosa* (PfPra1, pf01711, 279); *Sporisorium reilianum* (SrPra1, GenBank CAI59749, 296), (SrPra2, GenBank CAI59755, 299), and (SrPra3, GenBank CAI59763, 303); Ustanciosporium gigantosporum (UgPra1, GenBank AEY62496, 279); *Ustilago hordei* (UhPra1, GenBank CAJ41875, 296) and (UhPra2, GenBank AAD56044, 300); *U. maydis* (UmPra1, GenBank P31302, 296) and (UmPra2, GenBank P31303, 298). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed.

Components necessary for transcriptional gene silencing and chromatin remodeling were also found in *P. flocculosa* (see Supplemental Table 4 online). These genes are missing

from the *U. maydis* genome but are present in *U. hordei* and *S. reilianum* (Laurie *et al.*, 2012).

Cell Wall–Degrading and Carbohydrate-Active Enzymes

CWDEs play important roles in plant pathogenicity (Walton 1994) and biocontrol activity (Vinale *et al.*, 2008). Analysis of CWDEs in the *P. flocculosa*, *U. maydis*, *U. hordei*, and *S. reilianum* genomes revealed that all four organisms possess fairly similar sets of genes (Figures 3C and 3D).

However, when comparing individual carbohydrate-active enzyme (CAZy) families, *P. flocculosa* revealed interesting differences with respect to possible functions in fungal cell wall degradation, specifically dealing with chitosanase and chitinase activities (see Supplemental Data Set 4 online). For example, the glycoside hydrolase family 18 (GH18) associated with fungal cell wall degradation (chitinase) clearly differentiates *P. flocculosa* from the phytopathogenic Ustilaginales. First, *P. flocculosa* possesses seven genes coding for members of the GH18 family compared with only four in the phytopathogenic species. Second, in *P. flocculosa* only, two of those GH18 chitinase-encoding genes are each positioned right next to a gene coding for a protein bearing five (pf01239) or seven (pf06228) lysin motifs (LysM, CAZy CBM50), known for their affinity for chitin, and to a gene encoding a protein harboring a calcium binding EF-hand motif. Both of the latter genes also encode secreted proteins and neither have homologs in the three smuts.

Regarding plant CWDEs, no apparent differences could be found between gene sets. However, a motif that was associated with pectin degradation by Amselem *et al.* (2011) led to the identification of two *P. flocculosa*–specific putative lipases with a consensus Gly-Asp-Ser-Leu (GDSL) motif and a secretion signal (pf01013 and pf03087). Both have a carbohydrate esterase family 16 motif, but, surprisingly, they share a very low level of similarity with one another and are not homologous to any protein from any of the three plant pathogenic species.

Biosynthesis of Secondary Metabolites

Secreted secondary metabolites also play an important role in pathogenicity and biocontrol. Using the JCVI Secondary Metabolite Unique Regions Finder server (SMURF) (Khaldi *et*

al., 2010), we identified pivotal genes associated with the synthesis of secondary metabolites, also referred to as backbone genes, and predicted their presence in a broader cluster of genes involved in biosynthesis.

P. flocculosa, *U. hordei*, and *S. reilianum* species had a nearly identical number of SMURF clusters and backbone genes, while *U. maydis* had approximately 50% more in each category (Table 2). All backbone genes have at least one homolog in all other species except for two *U. maydis* dimethylallyltryptophan synthases (um05796 and um06508). Based on homology between their gene content, 13 SMURF-predicted clusters could be grouped into six groups, two of which contain a predicted cluster in *P. flocculosa* (see Supplemental Table 5 online). Not surprisingly, one group linked the flocculosin gene cluster with the ustilagic acid gene cluster.

Table 2 : Secondary metabolite backbone genes found and clusters predicted in four ustilaginomycete fungi

Species	PKS[a]	PKS-Like	NRPS[b]	NRPS-Like	DMAT[c]	Total	No. of Predicted Clusters
P. flocculosa	4	1	1	4	0	10	8
U. maydis	3	2	3	6	2	16	12
U. hordei	2	1	3	4	0	10	7
S. reilianum	2	1	2	4	0	11	9

[a.] Polyketide synthase.
[b.] Nonribosomal peptide synthetase.
[c.] Dimethylallyl Trp synthase.

Secretome

Based on the method described by Mueller *et al.* (2008), 547 secreted proteins were identified in *P. flocculosa*. Of this number, 345 could not be assigned an enzymatic function and were therefore considered as candidate secreted effector proteins (CSEPs). In comparison to *U. maydis*, the genome of *P. flocculosa* has nearly the same number of predicted secreted proteins overall but ~10% less identified CSEPs (554 secreted proteins and 386 CSEPs for *U. maydis* according to Mueller *et al.*, 2008) (see Supplemental Data Set 5 online). Clustering is also similar in *P. flocculosa* with 112 (20%) secreted proteins being clustered into 21 clusters versus 137 and 139 (28%) into 22 clusters for *U. maydis* and *S. reilianum*, respectively. In comparison, only seven clusters grouping 30 genes (7%) could be found in *U. hordei*.

When focusing on potential effectors, 60% (209 genes) of the identified *P. flocculosa* CSEPs are distributed over its 138 species-specific gene families (Figure 3B). This proportion is just 21% in the plant pathogens, on average. The closer relatedness among the CSEP sets in the plant pathogens is also reflected in the number of gene families they share. On average, 75% of gene families containing CSEPs are shared in the plant pathogenic species, whereas this number is only 43% between *P. flocculosa* and the other plant pathogens. Surprisingly, a NPP1 (for necrosis-inducing Phytophthora protein) domain was identified in two *P. flocculosa* CSEPs (pf04735 and pf04834), a feature not found in the smuts.

Orthology to *U. maydis* Genes Proven to Affect Virulence

From the Pathogen-Host Interaction database (PHI-base) (Winnenburg *et al.*, 2006), we extracted 97 genes that had been biologically characterized as playing a role in the ability of *U. maydis* to infect maize or affect the level of virulence after infection. Using this *U. maydis* set, we then verified the presence of orthologs in the other three organisms and grouped them into four categories based on the type of response they provoked in the *U. maydis*–maize interaction (loss of pathogenicity versus increased or reduced virulence) and on whether they are secreted (secreted versus nonsecreted) (Figure 6; see Supplemental Data Set 6 online). Roughly half of the genes analyzed in this manner are conserved and represent equal numbers in each of three categories. Conserved genes are, for the most part, genes coding for nonsecreted components of signal transduction pathways, chitin synthases, and proteins with nucleic acid binding motifs. For example, components of the cAMP and mitogen-activated protein kinase signaling cascades downstream of the pheromone receptor are all conserved. Of the four gene categories created, the one represented by secreted proteins influencing the level of virulence clearly showed the highest level of variation among the species (Figure 6). Most notably, *P. flocculosa* contained only two genes sharing homology with the 51 found in *U. maydis*, a striking difference considering the high conservation of genes in all other categories studied.

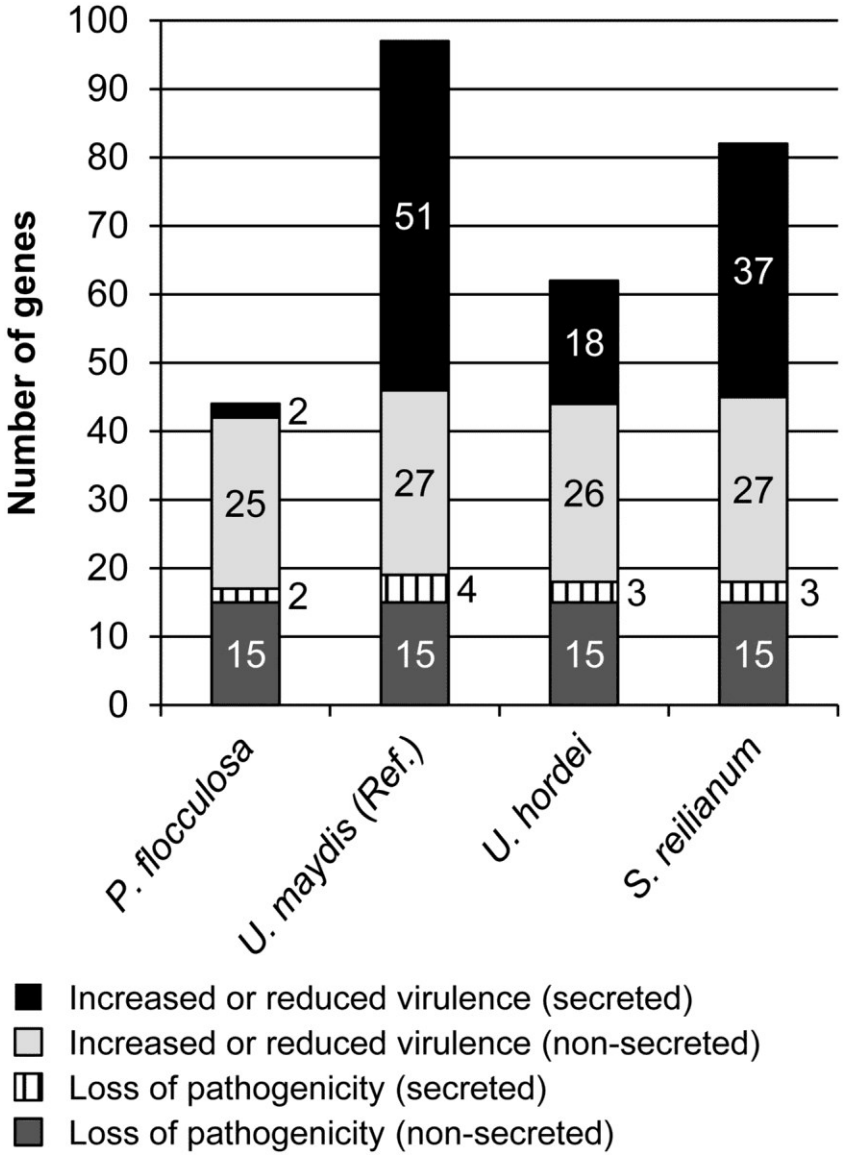


Figure 6 : Conservation of *U. maydis* genes proven to affect virulence. For each of the reported 97 *U. maydis* genes, presence of an ortholog in *P. flocculosa*, *U. hordei*, or *S. reilianum* was assessed. Genes were grouped according to the type of response observed and the secretion status of the coded protein.

3. Discussion

Comparative genomics has been extensively exploited to unravel specific virulence determinants and their evolution among the phytopathogenic Ustilaginomycetes *U. maydis*, *S. reilianum*, and *U. hordei* (Kämper *et al.*, 2006; Laurie *et al.*, 2012; Schirawski *et al.*, 2010). In this work, the genome release of the closely related fungus *P. flocculosa* constitutes a new step in the study of Ustilaginomycete fungi evolution and

phytopathogenicity-associated traits. In addition, because *P. flocculosa* cannot parasitize plants but is a powerful antagonist of powdery mildews (Clément-Mathieu *et al.*, 2008; Hajlaoui et Bélanger, 1993; Jarvis *et al.*, 1989), it represents a unique model to assess the presence and role of virulence factors inherent to pathogenic activity against both plants and fungal pathogens.

In order to generate high-confidence gene structures for a maximum of protein-coding genes, RNA sequencing was performed based on four previously defined growth phases of *P. flocculosa* (Hammami *et al.*, 2008). The resulting contigs were aligned against the *P. flocculosa* genome along with homologous predicted protein sequences from the closely related Ustilaginales species already sequenced. This determined, with accuracy, gene models for more than 90% of the predicted genes composing the *P. flocculosa* genome.

Although *P. flocculosa* has never been reported to infect plants, its genome sports several conserved features similar to those in the genomes of the three phytopathogenic Ustilaginales. The genome was comparable in terms of size and gene number, and many regions revealed mesosynteny to the *U. maydis* genome. Also, the repeat analysis showed a low proportion of TEs and repetitive sequences similar to the fairly streamlined genomes of *U. maydis* and *S. reilianum*. Of the four, *U. hordei* seems to stand out with over 10% of TEs and repeats. We did not find evidence of a repeat-induced point mutation mechanism as in *U. hordei*, but, as in *S. reilianum*, too few available homologous repeats made a comparative analysis difficult to assess mutation rates (Laurie *et al.*, 2012). In general, repeat content in genomes of fungi seems to vary widely and is much higher in the biotrophic rusts and mildews (Spanu 2012). Even though TEs are now seen as genome modifiers creating variation in gene structure and expression, streamlined genomes having few TEs and repeats may be a result of combinations of mating strategies, sex, and recombination potential (Laurie *et al.*, 2013). In contrast with *U. maydis*, *P. flocculosa* does not have a high rate of homologous recombination (Teichmann *et al.*, 2011b), a feature it shares with *U. hordei*.

A striking feature unique to *P. flocculosa* is the high number of introns per gene. It is unclear if this is the result of intron gain in the case of *P. flocculosa* or intron loss in phytopathogenic Ustilaginales. Interestingly, the three Ustilaginales species revealed a

biased intron distribution toward the 5′ end of genes, a characteristic common to intron-poor species, possibly associated with a mechanism of reverse transcriptase–mediated intron loss (RTMIL) (Mourier et Jeffares, 2003). Alternatively, the 5′ bias observed in intron-poor species could be linked to selection pressure favoring the conservation of the 5′-proximal intron having a functional role (Jeffares *et al.*, 2006; Nielsen *et al.*, 2004) even though different studies showed lineage-specific variability in the pattern of intron gain and loss and contradicted the thesis of a simple RTMIL mechanism (Sharpton *et al.*, 2008). However, considering that evolution of eukaryotes is characterized more by intron loss than gain, Roy and Gilbert (2006) suggested that if RTMIL occurred in some species, then the rate of paralogous recombination will directly influence the rate of intron loss. If this holds true for Ustilaginomycetes, since recombination events are favored during meiosis in fungi, it would explain why species with a sexual cycle possess fewer introns than others. The possible lack of a sexual cycle in *P. flocculosa* could also explain the state of mesosynteny observed between its genome and that of *U. maydis*. As reported for Magnaporthe grisea, a fungus that exclusively propagates asexually in nature, genome rearrangements are suspected to be maintained as long as vegetative fitness is conserved (Dean *et al.*, 2005).

Phylogenetic reconstruction based on more than 300 core gene families supports the relatively close proximity of *P. flocculosa* to phytopathogenic Ustilaginales. However, its exclusion from a common branch with the three smut fungi may indeed suggest that it has adopted a divergent lifestyle after the speciation event from a common ancestor. When classified by Begerow *et al.* (2006), *P. flocculosa* was described as an anamorph lacking a sexual stage. One hypothesis and possible differentiating factor with pathogenic Ustilaginales is that loss of mating-type genes accounts for the inability of *P. flocculosa* to infect plants. However, mating-type genes, constituting the pheromone-receptor module and the homeodomain-containing heterodimeric transcription factors, are present in its genome. Although we currently cannot confirm functionality based on compatibility assays with *U. maydis* (see Supplemental Figure 4 online), initial RNA sequencing analysis revealed that the receptor and both pheromone genes are transcribed (data not shown). The presence in the *P. flocculosa* genome of all other genes shown to be necessary for the initiation of mating, maintenance of the dikaryotic pathogenic state, and meiosis in *U. maydis* suggests that *P. flocculosa* is either capable of mating and sex under certain

circumstances or has recently lost this capability. One would expect that selection pressure would alter the presence or function of mating genes in strict anamorphs, but recent genome sequence analyses have shown that fungal species reproducing asexually in nature possessed all genes involved in sexual reproduction and were even capable of mating under *in vitro* condition (Böhm *et al.*, 2013; O'Gorman *et al.*, 2009; Xu *et al.*, 2007). Interestingly, *P. flocculosa* does not require mating to unleash its pathogenic potential onto mildew fungi, although this mechanism of epiphytic competition is likely different from the host infection requirement of the smuts.

The *P. flocculosa* a and b loci were localized in the current genome assembly to scaffolds 4 and 10, respectively. Scaffold 4 (1.1 Mb) reveals patchy synteny with interrupted stretches to 600-kb regions in *U. maydis* chromosome 5 and *S. reilianum* chromosome 20, and to a much shorter region in *U. hordei* scaffold 5.00019, all of which house the homologous pra and mfa genes (Laurie *et al.*, 2012). Incidentally, the *P. flocculosa* pra and two mfa genes are arranged similarly as at the *S. reilianum* a3 locus (Schirawski *et al.*, 2005). The *P. flocculosa* b locus localizes to scaffold 10, which, over a length of ~920 kb, matches homologs present in a 830-kb region of chromosome 1 in both *U. maydis* and *S. reilianum*; only a very short 9-kb stretch containing the bE and bW genes matches *U. hordei* scaffold 5.00019. Interestingly, genes in between the homologs on chromosome 1 locate to various other chromosomes but which are often the same in both *U. maydis* and *S. reilianum*. Overall, this indicates an arrangement closer to the tetrapolar mating-type organization of the *U. maydis*/*S. reilianum* lineage. The presence of two mating pheromone genes, also found at the *S. reilianum* a3 locus, could indicate that this arrangement with multiple pheromone genes is ancestral; the structure of the mating-type gene-containing chromosomes in *S. reilianum* also pointed to this fungus being closer to an ancestral lineage (Laurie *et al.*, 2012).

In addition to the presence of pivotal genes known to trigger the drastic morphological and physiological changes that enables the smut fungi to infect their host, *P. flocculosa* has also conserved all other typical plant pathogenicity traits found in smuts. This finding indicates that *P. flocculosa* likely has evolved from a plant pathogenic ancestor common to all four species.

In previous studies of fungal genomes, the presence and number of CAZy have been associated with the specific lifestyle of the fungus under study. For instance, the strict biotrophic pathogen *Blumeria graminis* f. sp *hordei* contains very few CAZy capable of plant cell wall degradation (Spanu *et al.*, 2010). Other biotrophs, such as Puccinia graminis f. sp tritici (Duplessis *et al.*, 2011) and *U. maydis* (Kämper *et al.*, 2006), are also characterized by reduced enzyme systems capable of degrading plant cell walls but do possess predicted cellulases and xylanases. Analysis of CAZy in *P. flocculosa* revealed a similar set of CAZy, including plant CWDEs as found in plant pathogenic Ustilaginales. It is unclear why it would have conserved these enzymes given that they are not deemed necessary for its epiphytic lifestyle on the leaf surface (Bélanger et Avis, 2002) nor for its parasitic activity on powdery mildews.

Ever since the establishment of the phylogenetic link between *P. flocculosa* and *U. maydis*, their shared property for production of secondary metabolites, namely, unusual glycolipids, has been an intriguing biological feature. It was established that both organisms contain an elaborate cluster of 10 genes regulating the synthesis of highly similar glycolipids termed flocculosin (*P. flocculosa*) and ustilagic acid (*U. maydis*) (Teichmann *et al.*, 2011a, 2011b, 2007). Owing to the antifungal property of flocculosin, it was originally hypothesized that it played a determinant role in the biocontrol activity of *P. flocculosa*. This hypothesis came under scrutiny when *U. maydis* was tested and found to be unable to antagonize powdery mildews under specific assay conditions in spite of producing antifungal ustilagic acids and when genes involved in flocculosin synthesis were observed to be repressed in spores that came into contact with powdery mildew (Hammami *et al.*, 2011). While these glycolipids are believed to contribute to the fitness of both organisms, possibly in fending off competitors, their true ecological role remains to be determined.

Following the release of the first fungal genomes and the development of reliable bioinformatic tools to predict secretion signals in protein sequences, emphasis has been placed on the study of effector proteins as determinants of pathogenicity (Torto *et al.*, 2003). In a variety of plant pathogens, including smut fungi, powdery mildews, rusts, and oomycetes, effectors have been found to affect virulence, suppress plant defense responses, dictate host specificity, and/or to maintain a biotrophic interaction. In this study, 345 CSEPs were identified in *P. flocculosa*, a number similar to those found in the plant

pathogenic Ustilaginales. However, the single most striking difference between *P. flocculosa* and the three smut fungi was found in terms of gene homology between CSEPs (Figure 5). Borrowing from the wealth of information available from sequenced Ustilaginales genomes and functional analyses in *U. maydis*, it appeared that orthologs for nearly all *U. maydis* genes (51 out of 55) encoding secreted proteins deemed to influence pathogenicity and virulence are absent in *P. flocculosa*. Considering that *P. flocculosa* has conserved all the necessary elements to mate and that it shows a high level of conservation for all other pathogenicity-related genes (e.g., CWDEs and biosynthesis of secondary metabolites), this result strongly suggests that loss of effectors represent the single most determinant factor that relates to the nonpathogenic lifestyle of *P. flocculosa*. At the same time, the presence of many orthologs in both *U. hordei* (21) and *S. reilianum* (40) strengthens the role that those proteins play in the establishment of a compatible pathogenic interaction in their respective host plants.

Allowing for this absence of homologs to a large number of virulence-promoting effectors in the phytopathogenic Ustilaginales, *P. flocculosa* seems to have evolved its own set of secreted proteins, the role of which is unknown at present. While not a phytopathogen, a feature shared with the other described Pseudozyma spp (Avis *et al.*, 2001), *P. flocculosa* possesses nonetheless the unique ability to antagonize powdery mildews. Based on our results, this interaction between a biocontrol agent and a fungal pathogen might also be dictated by effector proteins. For example, the secretome of *P. flocculosa* includes two NPP1-containing proteins that were not found in plant pathogenic Ustilaginales. In basidiomycetes, such proteins, labeled necrosis and ethylene-inducing peptide 1–like proteins, have only been identified in Moniliophthora perniciosa, the causal agent of witches'-broom disease of cacao. These proteins are absent from all other sequenced basidiomycetes, including Ustilaginomycetes (Kämper *et al.*, 2006; Laurie *et al.*, 2012; Schirawski *et al.*, 2010). Horizontal gene transfer has been suggested as one mechanism to account for the phylogenetic distribution of NPP1-containing proteins (Gijzen et Nürnberger, 2006). However, neither the GC content nor the codon bias supports this hypothesis. These proteins were recently determined to exhibit structural similarities to cytolytic toxins produced by marine organisms (actinoporins), which forms transmembrane pores (Ottmann *et al.*, 2009). Interestingly, the rapid collapse of powdery mildew colonies

by *P. flocculosa* has always been associated with an alteration of the plasma membrane and cytoplasmic leaking (Hajlaoui *et al.*, 1994; Hajlaoui et Bélanger, 1991; Mimee *et al.*, 2009). Thus, NPP1-containing proteins could be key elements explaining the antagonism of *P. flocculosa* toward powdery mildews. For now, the paucity of data detailing the interaction has prevented the validation of this hypothesis, a challenge that this new set of data will help to address.

Other genes from species-specific gene families revealed here might explain how *P. flocculosa* acquired the potential to antagonize powdery mildews. For instance, the discovery of two divergent GDSL lipases/esterases, containing a carbohydrate esterase family 16 motif that is exclusive to *P. flocculosa*, may be of relevance to its activity as an epiphytic competitor. GDSL lipases are multifunctional enzymes with a broad range of substrate specificities (Akoh *et al.*, 2004). Oh *et al.* (2005) reported that a GDSL produced by *Arabidopsis thaliana* had direct antifungal activity against *Alternaria brassicicola*. Of particular interest, the effect of the protein on fungal spores was reminiscent of the effect observed on powdery mildew spores when antagonized by *P. flocculosa* (Hajlaoui *et al.*, 1992; Hajlaoui et Bélanger, 1993; Mimee *et al.*, 2009). This distinctive feature may therefore be linked to the biocontrol activity of *P. flocculosa*.

Another important observation differentiating *P. flocculosa* from the plant pathogens was the identification of two loci similar to others found in mycoparasitic Trichoderma species (Kubicek *et al.*, 2011). These loci contain a gene encoding a subgroup C GH18 chitinase adjacent to another gene on the opposite strand encoding a LysM protein. Interestingly, LysM protein TAL6 was shown to inhibit spore germination of Trichoderma atroviride while having no effect on Aspergillus niger or Neurospora crassa (Seidl-Seiboth *et al.*, 2013). Since tal6 is mainly expressed in mature hyphae, the authors suggested that TAL6 could have a self-regulatory role in fungal growth and development. TAL6 could also act to protect the fungus against self-degradation by its own TAC6 during mycoparasitism. A similar protective function for LysM against wheat (*Triticum aestivum*) chitinases was described during infection by *Mycosphaerella graminicola* (Marshall *et al.*, 2011). A role for chitinases in the biocontrol activity of *P. flocculosa* has never been established (Bélanger *et al.*, 2012), but this finding suggests a feature shared by mycoparasites that warrants further investigation.

In conclusion, this work highlighted how *P. flocculosa* represents a unique model for comparative genomics because it provides insights into both the virulence determinants of fungal plant pathogens and the elusive properties of biocontrol agents. It can thus serve as a powerful tool to advance concerted efforts toward plant protection.

4. Methods

Sequencing, Assembly, and Alignment of Genomes

Genomic DNA was extracted from *Pseudozyma flocculosa* (DAOM 196992) according to the protocol of the Qiagen DNeasy plant mini kit (Venlo) for fungal cells. Two sequencing runs were performed by GenomeQuebec (McGill University, Montreal, Canada) and yielded 522 Mb data from a whole-genome shotgun library and 43 Mb data from a 2.5-kb paired-end library, respectively. A third run was performed at the Institut de Biologie Integrative et des Systèmes (Université Laval, Quebec City, Canada) and yielded 197 Mb data from 4.5-kb paired-end library. Sequences obtained represent an average coverage of 28 times the expected 25-Mb genome. All runs used the Roche 454 Life Sciences technology of pyrosequencing. Sequencing information was assembled using Roche proprietary software Newbler v2.6. MUMmer v.3.0 was used with promer program and –mum parameter for the alignment of genomes. Metrics for genome synteny showed in Supplemental Table 3 online were generated using OrthoCluster (version 2013/02/14; Zeng *et al.*, 2008). Genomes from plant pathogenic species were retrieved from the Munich information center for protein sequences (ftp://ftpmips.gsf.de).

Structural Annotation

Structural annotation of the *P. flocculosa* genome relied on several sources of information. First and foremost, the sequencing of the transcriptome was used as the most reliable source of information. The data were obtained from an experiment specifically designed to yield an exhaustive bank of transcripts. Briefly, *P. flocculosa* was cultured in yeast malt peptone dextrose broth and harvested at four different times (4, 8, 16, and 30 h) spanning the different growth phases of the fungus (Hammami *et al.*, 2008). Total RNA from the four samples was extracted using the RNeasy mini kit from Qiagen. Library construction was done by Genome British Columbia at the Michael Smith Genome Sciences Centre.

Only the poly(A)+ fraction was used, and cDNA were synthetized from random hexamers to cover entire transcripts. Both ends of fragments from the four libraries were sequenced on a single lane using the Illumina HiSeq sequencer and yielded 2 × 192 M 100-bp sequences.

Sequences from RNA sequencing were de novo assembled using CLC Genomics Workbench v5.0.1 (CLC bio, Aarhus, Denmark) and Trinity assembler v2012-01-25 (Grabherr *et al.*, 2011). All assembled contigs, mapped to the genome with GMAP v2007-09-28 (Wu et Watanabe, 2005), were selected based on reads mapped to the genome by TopHat v1.3.3 (Trapnell *et al.*, 2009), alignment of protein sequences from *Ustilago maydis* and *Sporisorium reilianum* by Exonerate v2.2.0 (Slater et Birney, 2005), and a provisory gene prediction by AUGUSTUS v.2.5.5 (Stanke *et al.*, 2004) based on training information from *U. maydis* genome annotation. All this information was visualized with GBrowse v2.0 (Stein *et al.*, 2002). Following manual curation, 6458 highly reliable coding sequences were retrieved. A subset of 400 handpicked annotations was then used to train ab initio gene predictors AUGUSTUS and SNAP v2006-07-28 (Korf 2004). The nontrainable gene predictor GeneMark-ES v2.3e (Lomsadze *et al.*, 2005) was also used as a supplementary source of evidence.

Annotation files from transcript sequences processed by PASA (Haas *et al.*, 2003), protein sequences of *U. maydis* and *S. reilianum* aligned with Exonerate, and predictions from ab initio gene predictors were fed to EvidenceModeler (Haas *et al.*, 2008) to produce an initial annotation that was further improved by manual curating. TRNA gene predictions were obtained using tRNAScan-SE v1.3.1 (Lowe et Eddy, 1997) with default settings for eukaryotes.

Repetitive Sequences

A custom library of repeat sequences was created from de novo calling of repeats by RepeatModeler (http://repeatmasker.org) on the target genome, repeats obtained via RepeatModeler run performed on *Ustilago hordei*, repeats identified from the *U. hordei* MAT-1 locus, and the entire RepeatMasker library of repeats. RepeatMasker (http://repeatmasker.org) was run with the custom library on the assembled 40 scaffolds and on the contigs that were not included in the assembly (contigs 1191 and higher).

Analysis of repeat induced point mutations was done using RIPCAL v1.0 (Hane *et al.*, 2008) on sequences retrieved from the analysis of repetitive sequences and coding sequences.

Functional Annotation

Gene naming was first established on the basis of a BLASTP v2.2.27 (Atschul *et al.*, 1990) similarity search against the Swiss-Prot database (Boeckmann *et al.*, 2003). At this point, only sequences with 70% identity and 70% coverage were assigned a name with the prefix "probable." Other names were obtained via the Blast2GO tool v2.6.2 (Conesa *et al.*, 2005) using default settings and were assigned the prefix "putative." Gene functions were also predicted using InterproScan v4.8 (database v38.0) (Quevillon *et al.*, 2005).

Annotation of CAZy was performed using the dbCAN Web server (Yin *et al.*, 2012). Original output was downloaded and parsed with custom python script to retrieve only matches the following parameters: e-value $< 10^{-3}$ if alignment length ≤ 80 amino acids, otherwise e-value $< 10^{-5}$; covered fraction > 0.3. Association of CAZy modules with substrates specificity followed the classification used by Amselem *et al.* (2011).

A search for genes involved in the biosynthesis of secondary metabolites was performed using JCVI Secondary Metabolite Unique Regions Finder Web server (SMURF) (Khaldi *et al.*, 2010). The server first identifies what is called backbones genes for which the coded protein harbors a domain related to polyketide synthases (PKS and PKS-Like), nonribosomal peptide synthetase (NRPS and NRPS-Like), or dimethylallyl Trp synthase (DMAT). From the located backbone gene, a region is identified as grouping the other genes usually associated to the biosynthesis and transport of secondary metabolites. Predictions of backbone genes are highly reliable. Otherwise, prediction of biosynthetic clusters is useful but results are approximate. Overall, both information outputs give a good portrait of global genetic resources putatively involved in the biosynthesis of secondary metabolites.

Annotation of secreted and CSEPs was accomplished according to the method described by Mueller *et al.* (2008). Secreted proteins were selected based upon SignalP v3.0 (Bendtsen *et al.*, 2004) D-value and Dmax cutoffs, TargetP v1.1 (Emanuelsson *et al.*, 2000) predicted location of protein, TMHMM v2.0 (Krogh *et al.*, 2001) predicted number of

transmembrane domains and position according to cleavage site, and finally, correlation to LocDB or PotLocDB ProtComp v9.0 (http://www.softberry.com) databases. Based on InterproScan-assigned domains, proteins lacking enzymatic functions were classified as candidate effectors (CSEP).

A list of all *P. flocculosa* genes and annotations is presented in Supplemental Data Set 7 online.

Clusters of Genes Encoding Secreted Proteins

Identification of physically clustered secreted proteins was based on a more permissive version of the definition given by Kämper *et al.* (2006). Here, clusters are defined as groups of three directly neighboring genes that are predicted to encode secreted proteins or groups of more than three genes with no more than three genes encoding nonsecreted gene products in between subgroups of at least two neighboring genes encoding secreted products.

Phylogeny Analysis

A first phylogenic tree was constructed using 306 gene families that were generated using Tribe-MCL (Enright *et al.*, 2002) and the protein sets from *P. flocculosa*, *Ustilago maydis*, *U. hordei*, *Sporisorium reilianum*, and *Malassezia globosa*. Gene families had to group only one gene from every species and to have a positive hit to a member of the Metabolism or Cellular Process categories of the KEGG database (Kanehisa et Goto, 2000) (see Supplemental Data Set 8 online). Alignments were performed using Clustal omega 1.1.0 with default parameters. Alignments were edited for optimization (see Supplemental Data Set 9 online). For each gene family, 1000 replicates were generated using the bootstrap method. For each replicate, a phylogenetic tree was estimated with the maximum likelihood method for protein amino acid sequences under the constraint of a molecular clock and the Jones-Taylor-Thornton model by the program Promlk from the Phylip package version 3.69 (http://evolution.genetics.washington.edu/phylip/) using default parameters. A consensus tree was constructed by a majority rule (extended) on the basis of the topologies of the 306,000 trees that were considered as rooted (midpoint) by the program Consense

(Phylip). Each branch of the consensus tree shows the percentage of the 306,000 trees that support it.

For the phylogenetic tree of basidiomycete pheromone receptor protein sequences, the maximum likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992) and evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011). Pheromone receptor sequences were C-terminally truncated to exclude the cytoplasmic tail and to optimize the alignment (see Supplemental Data Set 10 online). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in <50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Initial trees for the heuristic search were obtained automatically as follows: When the number of common sites was <100 or less than one-fourth of the total number of sites, the maximum parsimony method was used; otherwise, the BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 270 positions in the final data set.

Gene Families, Orthologs, and Characterized Genes for Virulence and Pathogenicity

All 27,462 sequences from *P. flocculosa*, *U. maydis*, *U. hordei*, and *S. reilianum* were searched for similarity, one against the other, using the BLASTP algorithm. Tribe-MCL (Enright *et al.*, 2002) was used to group sequences into gene families on the basis of the expectation value (e-value) of all hits. The inflation value was set to 2.1 as by Kämper *et al.* (2006). Each family was assigned to a category based on whether it contained genes from all four organisms (core), genes from two to three organisms including *P. flocculosa* (shared), genes from the three plant pathogens (present exclusively in all three plant pathogens), genes from two plant pathogens (present exclusively in two plant pathogens), and genes from only one species (species-specific). For the analysis of CSEP, non-CSEP genes were not considered in the assignation of categories.

Orthology was established on the basis of bidirectional best hits from the results of BLASTP all-against-all similarity search with a cutoff of $1e10^{-6}$. For homology, BLASTP local alignment had to cover at least 50% of query sequence and show 25% identity.

Information concerning effects of *U. maydis* gene deletion on pathogenicity and virulence was retrieved from PHI-base (Winnenburg *et al.*, 2006).

Mating Assays

The mating test was done as described by Müller *et al.* (1999) by mixing *U. maydis* strain FB1 (a1 and b1) or MB215 (a2 and b13) with *P. flocculosa* to observe the formation of fluffy white colonies indicative of dikaryotic growth following successful mating. Using overnight cultures of each strain with an equal OD600 of 0.5, 5 μL of the strains were mixed with each other on a potato dextrose agar plate supplemented with 1% (w/v) activated charcoal. The assays were conducted for 48 h at room temperature.

Accession Numbers

This Whole Genome Shotgun project has been deposited at GenBank/DDBJ/EMBL under accession number AOUS00000000. The version described in this article is the first version, AOUS01000000.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Number of Introns in Four Ustilaginomycotina Fungi According to Relative Position in Genes.

Supplemental Figure 2. Relative Frequency of Introns According to Relative Position in Genes.

Supplemental Figure 3. Analysis of Dinucleotide Frequency in *P. flocculosa*.

Supplemental Figure 4. Mating Assays between *P. flocculosa* and *U. maydis*.

Supplemental Table 1. Inventory of Repetitive Sequences in the *P. flocculosa* Genome Assembly.

Supplemental Table 2. Inventory of Repetitive Sequences in Nonassembled *P. flocculosa* Sequences.

Supplemental Table 3. OrthoCluster Analysis.

Supplemental Table 4. Subset of Gene Correspondence for Important Biological Processes.

Supplemental Table 5. Conserved SMURF Secondary Metabolite Gene Clusters.

Supplemental Data Set 1. Genes and Gene Families Distribution.

Supplemental Data Set 2. Inventory of Genes Related to Mating and Pathogenicity.

Supplemental Data Set 3. Inventory of Genes Related to Meiosis and Other Functions.

Supplemental Data Set 4. Gene Counts for CAZy Families.

Supplemental Data Set 5. CSEP IDs.

Supplemental Data Set 6. Characterized Genes for Pathogenicity and Virulence in *U. maydis*.

Supplemental Data Set 7. Gene Set.

Supplemental Data Set 8. Genes for Phylogeny.

Supplemental Data Set 9. Amino Acid Sequence Alignments of the 306 Gene Families Used for Phylogeny Reconstruction (Figure 4).

Supplemental Data Set 10. Amino Acid Sequence Alignment of Pheromone Receptor Proteins Used for Phylogeny Reconstruction (Figure 5).

5. Acknowledgments

We thank Michael Bölker for supplying *U. maydis* strains used in this study, Marie Vasse and Anaïs Pitarch for technical assistance, and Jérôme Laroche and Brian Boyle from the Institut de Biologie Integrative et des Systèmes at Laval University for help and support in bioinformatics. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, by grants from Fonds de Recherche du Québec–Nature et Technologies, and Natural Sciences and Engineering Research Council of Canada to F.L. and the Canada Research Chairs Program to R.R.B.

6. References

Akoh, C.C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H. and Shaw, J.-F. (2004). GDSL family of serine esterases/lipases. *Progress in Lipid Research*, *43*(6), 534-552.

Amselem, J., Cuomo, C.A., van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., et al. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea. PLoS Genetics*, *7*(8), e1002230.

Atschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.

Avis, T.J., Caron, S.J., Boekhout, T., Hamelin, R.C. and Bélanger, R.R. (2001). Molecular and physiological analysis of the powdery mildew antagonist *Pseudozyma flocculosa* and related fungi. *Phytopathology*, *91*(3), 249-254.

Bakkeren, G., Kämper, J. and Schirawski, J. (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. *Fungal Genetics and Biology*, *45*(Suppl. 1), S15-21.

Begerow, D., Bauer, R. and Boekhout, T. (2000). Phylogenetic placements of ustilaginomycetous anamorphs as deduced from nuclear LSU rDNA sequences. *Mycological Research*, *104*(1), 53-60.

Begerow, D., Stoll, M. and Bauer, R. (2006). A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data. *Mycologia*, *98*(6), 906-916.

Bendtsen, J.D., Nielsen, H., Heijne, G.V. and Brunak, S. (2004). Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology*, *340*(4), 783-795.

Bélanger, R.R., Labbé, C., Lefebvre, F. and Teichmann, B. (2012). Mode of action of biocontrol agents: all that glitters is not gold. *Canadian Journal of Plant Pathology*, *34*(4), 469-478.

Bélanger, R.R. and Avis, T.J. (2002). Ecological processes and interactions occuring in leaf surface fungi. In S. E. Lindow, E. I. Hecht-Poinar, and V. J. Elliott (Eds.), *Phyllosphere microbiology* (pp. 193-208). St. Paul, MN: The American Phytopathological Society.

Bélanger, R.R. and Deacon, J.W. (1996). Interaction specificity of the biocontrol agent *Sporothrix flocculosa*: a video microscopy study. *Phytopathology*, *86*(12), 1317-1323.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, *31*(1), 365-370.

Böhm, J., Hoff, B., O'Gorman, C.M., Wolfers, S., Klix, V., Binger, D., et al. (2013). Sexual reproduction and mating-type–mediated strain development in the penicillin-producing fungus *Penicillium chrysogenum. Proceedings of the National Academy of Sciences, 110(4), 1476-1481.*

Brefort, T., Doehlemann, G., Mendoza-Mendoza, A., Reissmann, S., Djamei, A. and Kahmann, R. (2009). *Ustilago maydis* as a pathogen. *Annual Review of Phytopathology*, *47*, 423-445.

Cheng, Y., Belzile, F., Tanguay, P., Bernier, L. and Bélanger, R.R. (2001). Establishment of a gene transfer system for *Pseudozyma flocculosa*, an antagonistic fungus of powdery mildew fungi. *Molecular Genetics and Genomics*, *266*(1), 96-102.

Clément-Mathieu, G., Chain, F., Marchand, G. and Bélanger, R.R. (2008). Leaf and powdery mildew colonization by glycolipid-producing *Pseudozyma* species. *Fungal Ecology*, *1*(2-3), 69-77.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674-3676.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea. Nature*, *434*(7036), 980-986.

Doehlemann, G., Reissmann, S., Aßmann, D., Fleckenstein, M. and Kahmann, R. (2011). Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. *Molecular Microbiology*, *81*(3), 751-766.

Doehlemann, G., van der Linde, K., Assmann, D., Schwammbach, D., Hof, A., Mohanty, A., et al. (2009). Pep1, a secreted effector protein of Ustilago maydis, is required for successful invasion of plant cells. *PLoS Pathogens*, *5*(2), e1000290.

Duplessis, S., Cuomo, C.A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9166-9171.

Emanuelsson, O., Nielsen, H., Brunak, S. and Heijne, G.V. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, *300*(4), 1005-1016.

Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575-1584.

Gijzen, M. and Nürnberger, T. (2006). Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry*, *67*(16), 1800-1807.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644-652.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Hannick, L.I., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*(19), 5654-5666.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, *9*(1), R7.

Hajlaoui, M.R., Benhamou, N. and Bélanger, R.R. (1992). Cytochemical study of the antagonistic activity of *Sporothrix flocculosa* on rose powdery mildew, *Sphaerotheca pannosa* var. *rosae. Cytology and Histology*, *82*(5), 583-589.

Hajlaoui, M.R., Traquair, J.A., Jarvis, W.R. and Bélanger, R.R. (1994). Antifungal activity of extracellular metabolites produced by *Sporothrix flocculosa. Biocontrol Science and Technology*, *4*(2), 229-237.

Hajlaoui, M.R. and Bélanger, R.R. (1991). Comparative effects of temperature and humidity on the activity of three potential antagonists of rose powdery mildew. *European Journal of Plant Pathology*, *97*(4), 203-208.

Hajlaoui, M.R. and Bélanger, R.R. (1993). Antagonism of the yeast-like phyllophane fungus *Sporothrix flocculosa* against *Erysiphe graminis* var *tritici. Biocontrol Science and Technology*, *3*(4), 427-434.

Hammami, W., Castro, C.Q., Remus-Borel, W., Labbe, C. and Bélanger, R.R. (2011). Ecological basis of the interaction between *Pseudozyma flocculosa* and powdery mildew fungi. *Applied and Environmental Microbiology*, *77*(3), 926-933.

Hammami, W., Labbé, C., Chain, F., Mimee, B. and Bélanger, R.R. (2008). Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa. Applied Microbiology and Biotechnology*, *80*(2), 307-315.

Hane, J., Oliver, R., Hane, J.K. and Oliver, R.P. (2008). RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics*, *9*, 478.

Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H., Goodwin, S.B. and Oliver, R.P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biology*, *12*(5), R45.

Jarvis, W.R., Shaw, L.A. and Traquair, J.A. (1989). Factors affecting antagonism of cucumber powdery mildew by *Stephanoascus flocculosus* and *S. rugulosus. Mycological Research*, *92*(2), 162-165.

Jeffares, D.C., Mourier, T. and Penny, D. (2006). The biology of intron gain and loss. *Trends in Genetics*, *22*(1), 16-22.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, *8*(3), 275-282.

Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27-30.

Kämper, J., Kahmann, R., Bölker, M., Ma, L.J., Brefort, T., Saville, B.J., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis. Nature*, *444*(7115), 97-101.

Kemen, E. and Jones, J.D. (2012). Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends in Plant Science*, *17*(8), 448-457.

Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H. and Fedorova, N.D. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology*, *47*(9), 736-741.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59.

Krogh, A., Larsson, B., Heijne, G.V. and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Fungal Genetics and Biology*, *305*(3), 567-580.

Kubicek, C.P., Herrera-Estrella, A., Seidl-Seiboth, V., Martinez, D.A., Druzhinina, I.S., Thon, M., et al. (2011). Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma. Genome Biology*, *12*(4), R40.

Laurie, J., Linning, R., Wong, P. and Bakkeren, G. (2013). Do TE activity and counteracting genome defenses, RNAi and methylation, shape the sex lives of smut fungi? *Plant Signaling & Behavior*, *8*(4), e23853.

Laurie, J.D., Ali, S., Linning, R., Mannhaupt, G., Wong, P., Güldener, U., et al. (2012). Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *The Plant Cell, 24(5), 1733-1745.*

Lemieux, R.U., Thorn, J.A., Brice, C. and Haskins, R.H. (1951). Biochemistry of the Ustilaginales: II. Isolation and partial characterization of ustilagic acid. *Canadian Journal of Chemistry*, *29*(5), 409-414.

Lowe, T.M. and Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955-964.

Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K.E., et al. (2011). Analysis of two *in planta* expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiology*, *156*(2), 756-769.

Mimee, B., Labbé, C., Pelletier, R. and Bélanger, R.R. (2005). Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa. Antimicrobial Agents and Chemotherapy*, *49*(4), 1597-1599.

Mimee, B., Pelletier, R. and Bélanger, R.R. (2009). *In vitro* antibacterial activity and antifungal mode of action of flocculosin, a membrane-active cellobiose lipid. *Journal of Applied Microbiology*, *107*(3), 989-996.

Mourier, T. and Jeffares, D.C. (2003). Eukaryotic intron loss. *Science*, *300*(5624), 1393-1393.

Mueller, O., Kahmann, R., Aguilar, G., Trejo-Aguilar, B., Wu, A. and de Vries, R.P. (2008). The secretome of the maize pathogen Ustilago maydis. *Fungal Genetics and Biology*, *45*(Suppl. 1), S63-S70.

Müller, P., Aichinger, C., Feldbrügge, M. and Kahmann, R. (1999). The MAP kinase Kpp2 regulates mating and pathogenic development in *Ustilago maydis*. *Molecular Microbiology*, *34*(5), 1007-1017.

Nielsen, C.B., Friedman, B., Birren, B., Burge, C.B. and Galagan, J.E. (2004). Patterns of intron gain and loss in fungi. *PLoS Biology*, *2*(12), e422.

Oh, I.S., Park, A.R., Bae, M.S., Kwon, S.J., Kim, Y.S., Lee, J.E., et al. (2005). Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternaria brassicicola*. *The Plant Cell*, *17*(10), 2832-2847.

Ottmann, C., Luberacki, B., Küfner, I., Koch, W., Brunner, F., Weyand, M., et al. (2009). A common toxin fold mediates microbial attack and plant defense. *Proceedings of the National Academy of Sciences*, *106*(25), 10359-10364.

O'Gorman, C.M., Fuller, H.T. and Dyer, P.S. (2009). Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Nature*, *457*(7228), 471-474.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research*, *33*(suppl 2), W116-W120.

Saville, B.J., Donaldson, M.E. and Doyle, C.E. (2012). Investigating host induced meiosis in a fungal plant pathogen. In A. Swan (Ed.), *Meiosis - molecular mechanisms and cytogenetic diversity* (pp. 411-460). Rijeka, Croatia: InTech.

Schirawski, J., Heinze, B., Wagenknecht, M. and Kahmann, R. (2005). Mating type loci of *Sporisorium reilianum*: Novel pattern with three a and multiple b specificities. *Eukaryotic Cell*, *4*(8), 1317-1327.

Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*, *330*(6010), 1546-1548.

Seidl-Seiboth, V., Zach, S., Frischmann, A., Spadiut, O., Dietzsch, C., Herwig, C., et al. (2013). Spore germination of *Trichoderma atroviride* is inhibited by its LysM protein TAL6. *The FEBS Journal*, *280*(5), 1226-1236.

Sharpton, T.J., Neafsey, D.E., Galagan, J.E. and Taylor, J.W. (2008). Mechanisms of intron gain and loss in *Cryptococcus*. *Genome Biology*, *9*(1), R24.

Slater, G. and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31.

Spanu, P.D. (2012). The genomics of obligate (and nonobligate) biotrophs. *Annual Review of Phytopathology*, *50*, 91-109.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, *330*(6010), 1543-1546.

Stanke, M., Steinkamp, R., Waack, S. and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*, *32*(suppl 2), W309-W312.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., et al. (2002). The Generic Genome Browser: A building block for a model organism system database. *Genome Research*, *12*(10), 1599-1610.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731-2739.

Teichmann, B., Labbé, C., Lefebvre, F., Bölker, M., Linne, U. and Bélanger, R.R. (2011a). Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa. Molecular Microbiology*, *79*(6), 1483-1495.

Teichmann, B., Lefebvre, F., Labbé, C., Bölker, M., Linne, U. and Bélanger, R.R. (2011b). Beta hydroxylation of glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-dependent β-hydroxylase. *Applied and Environmental Microbiology*, *77*(21), 7823-7829.

Teichmann, B., Linne, U., Hewald, S., Marahiel, M.A. and Bölker, M. (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis. Molecular Microbiology*, *66*(2), 525-533.

Torto, T.A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N.A., et al. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora. Genome Research*, *13*(7), 1675-1685.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105-1111.

Traquair, J.A., Shaw, L.A. and Jarvis, W.R. (1988). New species of *Stephanoascus* with *Sporothrix* anamorphs. *Canadian Journal of Botany*, *66*(5), 926-933.

Vinale, F., Sivasithamparam, K., Ghisalberti, E.L., Marra, R., Woo, S.L. and Lorito, M. (2008). *Trichoderma*–plant–pathogen interactions. *Soil Biology & Biochemistry*, *40*(1), 1-10.

Walton, J.D. (1994). Deconstructing the cell wall. *Plant Physiology*, *104*(4), 1113-1118.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M., et al. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*, *450*(7166), 115-118.

William Roy, S. and Gilbert, W. (2006). The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nature Reviews. Genetics*, *7*(3), 211-221.

Winnenburg, R., Baldwin, T.K., Urban, M., Rawlings, C., Köhler, J. and Hammond-Kosack, K.E. (2006). PHI-base: A new database for pathogen host interactions. *Nucleic Acids Research*, *34*(suppl 1), D459-D464.

Wu, T.D. and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*(9), 1859-1875.

Xu, J., Saunders, C.W., Hu, P., Grant, R.A., Boekhout, T., Kuramae, E.E., et al. (2007). Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proceedings of the National Academy of Sciences*, *104*(47), 18730-18735.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *40*(W1), W445-W451.

Zeng, X., Nesbitt, M.J., Pei, J., Wang, K., Vergara, I.A. and Chen, N. (2008). OrthoCluster: a new tool for mining synteny blocks and applications in comparative genomics. In *EDBT '08: Proceedings of the 11th international conference on extending database technology* (pp. 656-667). Nantes, France: ACM.

Chapitre 6 : Discussion générale et conclusion

1. Contexte général

Les chemins menant à la découverte de nouveaux outils dans la lutte aux parasites des cultures sont parsemés de découvertes enthousiasmantes et de défis de taille. Depuis de nombreuses années, un effort considérable est déployé à l'échelle mondiale afin de développer des alternatives plus efficaces aux pesticides conventionnels (Enserink *et al.*, 2013). Évidemment, les retombées de telles découvertes ne sont pas garanties, mais certaines d'entre elles pourraient néanmoins grandement réduire l'impact négatif lié à l'utilisation de certains produits nocifs.

Depuis ses débuts, le travail de recherche effectué au sujet de *Pseudozyma flocculosa* s'inscrit dans cette mouvance. Bien que son succès commercial ait été modeste, ses propriétés d'agent de lutte biologique (ALB) et son positionnement taxonomique en font un organisme modèle fascinant. C'est pourquoi depuis sa découverte, un effort continu a été déployé afin d'approfondir nos connaissances de ce champignon pour en tirer un nouvel outil de lutte contre le blanc, mais aussi afin de mieux comprendre les bases génétiques liées au caractère phytopathogène des Ustilaginales.

Au cours des années passées à étudier cet ALB, notre conception de son mode d'action n'a cessé d'évoluer. Dès les débuts, l'idée d'un mycoparasitisme semblable à celui observé chez *Trichoderma* ssp. a été écartée au profit de l'hypothèse favorisant l'antibiose comme mode d'action principal (Hajlaoui et Bélanger, 1993). Par la suite, la vraisemblance de cette hypothèse a été renforcée par la découverte d'une molécule aux propriétés antifongiques produite en grandes quantités par l'ALB et qu'on nomma flocculosine (Cheng *et al.*, 2003; Mimee *et al.*, 2005).

C'est à cette époque que la découverte inattendue d'une parenté entre *P. flocculosa* et *Ustilago maydis* a été clairement établie. En effet, non seulement un lien phylogénétique a-t-il pu être démontré de manière convaincante, mais leur capacité à synthétiser des versions hautement similaires d'un même glycolipide n'a fait qu'ajouter à la conviction que tous deux découlaient d'un récent événement de spéciation (Begerow *et al.*, 2006; Marchand *et al.*, 2009).

Toutefois, la découverte de ce lien de parenté a aussi eu pour conséquence de remettre en question plusieurs idées avancées quant au mode d'action privilégié par l'ALB pour

s'attaquer au blanc. En effet, pourquoi ces deux organismes ayant adopté des modes de vie si différents auraient-ils conservé la capacité de synthétiser une molécule considérée être associée à une activité observée seulement chez *P. flocculosa*? D'autres hypothèses concernant le rôle de l'acide ustilagique durant le cycle sexué d'*U. maydis* ont aussi été ébranlées, cette forme de vie n'ayant jamais été observée chez *P. flocculosa*.

La question du rôle écologique de la flocculosine et de l'acide ustilagique demeure donc entière. Jusqu'à aujourd'hui, la production de flocculosine n'a pu être associée de manière évidente à l'activité antagoniste de *P. flocculosa*. De récents travaux semblent même démontrer le contraire (Hammami *et al.*, 2011). Comment alors confirmer ou infirmer le rôle de la molécule dans le processus de lutte biologique, et s'il s'avère négatif, comment identifier le mode d'action qui mène à l'effondrement du blanc? Ces questions demeurent centrales en ce qui concerne la mise au point d'approches innovantes pour combattre le blanc basées sur la connaissance de la biologie de l'ALB *P. flocculosa*.

Jusqu'à maintenant, la résolution de ces questions s'est butée à un obstacle de taille. En effet, *P. flocculosa* demeure réfractaire à la transformation génétique, particulièrement par recombinaison homologue (Bélanger *et al.*, 2012). Cette approche serait certainement la plus élégante pour prouver l'absence de lien entre la flocculosine et le processus de biocontrôle. L'issue d'un bio-essai où un mutant incapable de synthétiser une molécule active de flocculosine serait mis en contact avec le blanc nous indiquerait alors si la molécule possède un rôle de premier plan dans la destruction du blanc.

2. Approche génomique

Pour contourner ce problème, l'approche la plus porteuse envisagée jusqu'à maintenant a été celle qui fait appel aux outils de la génomique. En effet, les technologies de séquençage nous ont permis de générer une séquence génomique annotée de l'ALB. L'analyse de cette dernière nous a fait découvrir toute la richesse du génome de l'organisme. De manière systématique et en nous appuyant sur les annotations génomiques d'organismes apparentés déjà publiées, nous avons été en mesure de faire une analyse systématique de différentes catégories de gènes. L'analyse comparative a porté sur les gènes dont le produit présente une activité enzymatique envers les polymères de glucose (CAZy), les gènes associés à la

synthèse de métabolites secondaires, les gènes associés aux loci de compatibilité sexuelle (*mating*) et les gènes dont les produits sont sécrétés et pour lesquels aucune fonction n'a pu être déterminée par homologie. Il est apparu que ce dernier groupe, maintenant reconnu comme un des plus importants en ce qui a trait aux interactions entre organismes, démontrait un niveau de diversité inégalé par rapport à celui des protéines sécrétées des agents phytopathogènes apparentés. Cette découverte indique que le groupe de gènes codant des protéines sécrétées est, comme plusieurs le soupçonnent déjà, central en ce qui concerne les interactions entre un agent pathogène et son hôte. Dans le cas de *P. flocculosa*, il semble que l'évolution de l'organisme à partir d'un ancêtre commun aux quatre espèces comparées ait principalement touché le groupe des gènes codant des protéines sécrétées excluant celles qui sont associées à la synthèse d'enzymes de dégradation.

3. Protéines-effecteurs

Depuis la publication du génome des premiers agents pathogènes, il est apparu que plusieurs espèces possèdent un grand nombre de protéines de faible taille dont la fonction demeure inconnue à ce jour (Kamoun, 2009). À la suite de cette découverte, il est devenu de plus en plus clair que ce groupe de gènes est composé en grande partie de membres dont les produits interviennent de manière considérable dans le processus d'infection et de colonisation des tissus végétaux. En effet, ces protéines, souvent riches en cystéine, jouent un rôle central en ce qui a trait à la compatibilité entre l'agent pathogène et son hôte et, conséquemment, à l'issue de leur interaction. Ce groupe de protéines s'est vu donner le nom d'effecteurs.

L'analyse des génomes d'agents pathogènes et d'organismes fongiques symbiotiques a démontré que ce groupe de protéines sécrétées est très diversifié. En effet, il semble qu'ils soient, pour la plupart, sujets à une forte pression de diversification qui se vérifie à la fois par la diversité des sites d'action déjà répertoriés, qu'ils soient cellulaires et extracellulaires, ou par la grande spécificité qui unit souvent les effecteurs à leur cible (Stergiopoulos et de Wit, 2009). Certains d'entre eux interviennent pour masquer la présence de l'envahisseur, en séquestrant les particules issues de la dégradation de ses parois, pour prendre le contrôle de processus cellulaires chez l'hôte afin d'en modifier le

comportement à son profit ou pour inactiver les réponses de défense de la plante de manière à compléter de son cycle de vie (de Jonge et Thomma, 2009).

De plus, il est maintenant reconnu que les champignons parasites et mutualistes utilisent des mécanismes semblables pour entrer en relation avec leur hôte (Martin *et al.*, 2008). De tels mécanismes pourraient-ils intervenir, eux aussi, dans le cas d'interactions entre champignons? C'est une question qui mérite d'être considérée sachant que *P. flocculosa* possède une parenté avec des agents pathogènes dont l'arsenal de protéines-effecteurs joue un rôle central dans l'envahissement de l'hôte.

La recherche sur les protéines-effecteurs n'en est qu'à ses débuts. Beaucoup de découvertes restent à faire, mais déjà la quantité de connaissances sur le sujet nous permet d'apprécier la diversité des moyens utilisés par les champignons et les plantes pour parvenir à se protéger tout en profitant des meilleures conditions pour se disséminer. C'est peut-être en s'appuyant sur la diversité fonctionnelle des protéines-effecteurs que de nouvelles stratégies de protection des cultures pourront être mises de l'avant dans l'avenir.

4. Approche transcriptomique

Bien que les protéines sécrétées pourraient jouer un rôle déterminant en ce qui a trait à la lutte biologique à l'aide de *P. flocculosa*, plusieurs autres candidats intéressants ont été identifiés par l'analyse du génome. Il demeure toutefois nécessaire d'associer ces gènes à une activité transcriptionnelle spécifique à la lutte biologique, manière la plus efficace de réduire le nombre de candidats et de faire avancer notre investigation.

En effet, pour déterminer avec certitude si un gène possède un rôle particulier dans un processus biologique, il est nécessaire, même si cela est insuffisant, de démontrer que son expression est corrélée à l'activité observée. Une méthode des plus élégantes consiste à éteindre l'expression d'un gène de manière à démontrer son rôle incontournable dans le processus étudié. Cette approche est toute désignée pour les espèces qui peuvent être transformées génétiquement et pour les processus biologiques qui font appel à un sentier métabolique linéaire où il est impossible que d'autres gènes puissent compenser l'extinction génique induite, par l'emprunt d'un sentier alternatif ou par simple redondance fonctionnelle.

Dans le cas de *P. flocculosa*, puisqu'il s'agit d'un organisme difficile à transformer, il est nécessaire d'opter pour une approche alternative à l'extinction génique. De plus, le bassin de gènes candidats pouvant être impliqués dans la lutte biologique étant très large, nous croyons que l'approche transcriptomique par séquençage des transcrits (RNA-Seq) pourrait faire progresser la recherche d'éléments génétiques reliés à l'interaction antagoniste contre le blanc en réduisant le nombre de gènes candidats considérés pour une étude plus ciblée.

Toutefois, cette approche demandera plusieurs ajustements étant donné la complexité du système étudié. En effet, puisque le blanc est une maladie causée par un agent pathogène biotrophe, il est impossible de cultiver l'organisme en milieu artificiel. Le processus de lutte biologique doit donc se dérouler à la surface d'une feuille lors de l'échantillonnage. Bien évidemment, les conditions de l'expérience doivent être reproductibles, c'est-à-dire que le recouvrement de la feuille par l'agent pathogène doit permettre une colonisation de l'ALB semblable d'une feuille à l'autre et les conditions environnementales doivent être bien contrôlées dans tous les cas. De plus, les temps d'inoculation et de prélèvement tout comme la technique d'échantillonnage devront permettre de récupérer une quantité suffisante de *P. flocculosa* pour les besoins de l'analyse. En effet, au moment où s'établissent la reconnaissance et l'induction de la réponse de l'ALB à la présence du blanc, il est à prévoir que la quantité de cellules provenant de ce dernier surpasse de beaucoup celle du premier. Une fois ces questions résolues par la conduite d'essais prélimiaires, l'expérience pourra être menée de manière concluante et générer un jeu de données qui permettra de mieux circonscrire les éléments à la base du mode d'action de *P. flocculosa*.

5. Conclusion

La publication du génome de *P. flocculosa* constitue une étape importante en ce qui a trait à la connaissance de la biologie de l'organisme. Depuis ses débuts, la recherche s'intéressant à cet ALB a su tirer profit des outils disponibles et à chaque fois, leur adoption a mené à des découvertes marquantes qui ont fait progresser rapidement nos connaissances.

Aujourd'hui, la puissance des outils de séquençage et des systèmes informatiques est mise à contribution afin de poursuivre les efforts visant à découvrir le mode d'action de l'ALB. Dans le futur, il est à prévoir que de nombreuses découvertes découleront du projet

d'analyse transcriptomique de l'interaction entre *P. flocculosa* et le blanc. Parmi celles-ci, plusieurs pourraient mener au développement de solutions novatrices dans la lutte à la maladie du blanc.

Bibliographie

Akoh, C.C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H. et Shaw, J.-F. (2004). GDSL family of serine esterases/lipases. *Progress in Lipid Research*, *43*(6), 534-552.

Amselem, J., Cuomo, C.A., van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., et al. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea. PLoS Genetics*, *7*(8), e1002230.

Atschul, S.F., Gish, W., Miller, W., Myers, E.W. et Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.

Avis, T.J., Caron, S.J., Boekhout, T., Hamelin, R.C. et Bélanger, R.R. (2001). Molecular and physiological analysis of the powdery mildew antagonist *Pseudozyma flocculosa* and related fungi. *Phytopathology*, *91*(3), 249-254.

Avis, T.J. et Bélanger, R.R. (2001). Specificity and mode of action of the antifungal fatty acid *cis*-9-heptadecenoic acid produced by *Pseudozyma flocculosa. Applied and Environmental Microbiology*, *67*(2), 956-960.

Avis, T.J. et Bélanger, R.R. (2002). Mechanisms and means of detection of biocontrol activity of *Pseudozyma* yeasts against plant-pathogenic fungi. *FEMS Yeast Research*, *2*(1), 5-8.

Bakkeren, G., Kämper, J. et Schirawski, J. (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. *Fungal Genetics and Biology*, *45 Suppl 1*, S15-S21.

Banuett, F. et Herskowitz, I. (1989). Different a alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(15), 5878-5882.

Begerow, D., Bauer, R. et Boekhout, T. (2000). Phylogenetic placements of ustilaginomycetous anamorphs as deduced from nuclear LSU rDNA sequences. *Mycological Research*, *104*(1), 53-60.

Begerow, D., Stoll, M. et Bauer, R. (2006). A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data. *Mycologia*, *98*(6), 906-916.

Bendtsen, J.D., Nielsen, H., Heijne, G.V. et Brunak, S. (2004). Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology*, *340*(4), 783-795.

Benyagoub, M., Willemot, C. et Bélanger, R.R. (1996). Influence of a subinhibitory dose of antifungal fatty acids from *Sporothrix flocculosa* on cellular lipid composition in fungi. *Lipids*, *31*(10), 1077-1082.

Bélanger, R.R., Labbé, C., Lefebvre, F. et Teichmann, B. (2012). Mode of action of biocontrol agents: all that glitters is not gold. *Canadian Journal of Plant Pathology*, *34*(4), 469-478.

Bélanger, R.R. et Avis, T.J. (2002). Ecological processes and interactions occuring in leaf surface fungi. In S. E. Lindow, E. I. Hecht-Poinar, et V. J. Elliott (Eds.), *Phyllosphere microbiology* (pp. 193-208). St. Paul, MN: The American Phytopathological Society.

Bélanger, R.R. et Deacon, J.W. (1996). Interaction specificity of the biocontrol agent *Sporothrix flocculosa*: a video microscopy study. *Phytopathology*, *86*(12), 1317-1323.

Bitto, N.J., Graichen, F.H. et Monahan, B.J. (2009). Functionality at the end of a fatty acid chain chemical and biological routes to ω-hydroxylated fatty acids. *Lipid Technology*, *21*(10), 216-219.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, *31*(1), 365-370.

Boothroyd, B., Thorn, J.A. et Haskins, R.H. (1956). Biochemistry of the Ustilaginales. XII. Characterization of extracellular glycolipids produced by *Ustilago* sp. *Canadian Journal of Biochemistry and Physiology*, *34*(1), 10-14.

Böhm, J., Hoff, B., O'Gorman, C.M., Wolfers, S., Klix, V., Binger, D., et al. (2013). Sexual reproduction and mating-type–mediated strain development in the penicillin-producing fungus *Penicillium chrysogenum. Proceedings of the National Academy of Sciences*, *110*(4), 1476-1481.

Brachmann, A., König, J., Julius, C. et Feldbrügge, M. (2004). A reverse genetic approach for generating gene replacement mutants in *Ustilago maydis. Molecular Genetics and Genomics*, *272*(2), 216-226.

Bray, J.E., Marsden, B.D. et Oppermann, U. (2009). The human short-chain dehydrogenase/reductase (SDR) superfamily: a bioinformatics summary. *Chemico-biological Interactions*, *178*(1), 99-109.

Brefort, T., Doehlemann, G., Mendoza-Mendoza, A., Reissmann, S., Djamei, A. et Kahmann, R. (2009). *Ustilago maydis* as a pathogen. *Annual Review of Phytopathology*, *47*, 423-445.

Cameotra, S.S. et Makkar, R.S. (2004). Recent applications of biosurfactants as biological and immunological molecules. *Current Opinion in Microbiology*, *7*(3), 262-266.

Castoria, R., De Curtis, F., Lima, G. et De Cicco, V. (1997). β-1,3-glucanase activity of two saprophytic yeasts and possible mode of action as biocontrol agents against postharvest diseases. *Postharvest Biology and Technology*, *12*(3), 293-300.

Cheng, Y., Belzile, F., Tanguay, P., Bernier, L. et Bélanger, R.R. (2001). Establishment of a gene transfer system for *Pseudozyma flocculosa*, an antagonistic fungus of powdery mildew fungi. *Molecular Genetics and Genomics*, *266*(1), 96-102.

Cheng, Y., McNally, D.J., Labbe, C., Voyer, N., Belzile, F. et Bélanger, R.R. (2003). Insertional mutagenesis of a fungal biocontrol agent led to discovery of a rare cellobiose lipid with antifungal activity. *Applied and Environmental Microbiology*, *69*(5), 2595-2602.

Choudhury, S.R., Traquair, J.A. et Jarvis, W.R. (1994). 4-Methyl-7,11-heptadecadienal and 4-methyl-7,11-heptadecadienoic acid: new antibiotics from *Sporothrix flocculosa* and *Sporothrix rugulosa. Journal of Natural Products, 57(6), 700-704.*

Choudhury, S.R., Traquair, J.A. et Jarvis, W.R. (1995). New extracellular fatty acids in culture filtrates of *Sporothrix flocculosa* and *S. rugulosa. Canadian Journal of Chemistry*, *73*(1), 84-87.

Clément-Mathieu, G., Chain, F., Marchand, G. et Bélanger, R.R. (2008). Leaf and powdery mildew colonization by glycolipid-producing *Pseudozyma* species. *Fungal Ecology*, *1*(2-3), 69-77.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. et Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674-3676.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea. Nature*, *434*(7036), 980-986.

Déziel, E., Lépine, F., Milot, S. et Villemur, R. (2003). *rhlA* is required for the production of a novel biosurfactant promoting swarming motility in *Pseudomonas aeruginosa*: 3-(3-hydroxyalkanoyloxy)alkanoic acids (HAAs), the precursors of rhamnolipids. *Microbiology*, *149*(8), 2005-2013.

Doehlemann, G., Reissmann, S., Aßmann, D., Fleckenstein, M. et Kahmann, R. (2011). Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. *Molecular Microbiology*, *81*(3), 751-766.

Doehlemann, G., van der Linde, K., Assmann, D., Schwammbach, D., Hof, A., Mohanty, A., et al. (2009). Pep1, a secreted effector protein of Ustilago maydis, is required for successful invasion of plant cells. *PLoS Pathogens*, *5*(2), e1000290.

Duplessis, S., Cuomo, C.A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9166-9171.

Elad, Y., Chet, I. et Henis, Y. (1982). Degradation of plant pathogenic fungi by *Trichoderma harzianum. Canadian Journal of Microbiology*, *28*(7), 719-725.

Elad, Y., Kirshner, B., Yehuda, N. et Sztejnberg, A. (1998). Management of powdery mildew and gray mold of cucumber by *Trichoderma harzianum* T39 and *Ampelomyces quisqualis* AQ10. *BioControl*, *43*(2), 241-251.

Emanuelsson, O., Nielsen, H., Brunak, S. et Heijne, G.V. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, *300*(4), 1005-1016.

Enright, A.J., Van Dongen, S. et Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575-1584.

Enserink, M., Hines, P.J., Vignieri, S.N., Wigginton, N.S. et Yeston, J.S. (2013). The Pesticide Paradox. *Science*, *341*(6147), 728-729.

Fravel, D.R. (2005). Commercialization and implementation of biocontrol. *Annual Review of Phytopathology*, *43*, 337-359.

Gijzen, M. et Nürnberger, T. (2006). Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry*, *67*(16), 1800-1807.

Golubev, V.I., Kulakovskaia, T.V., Shashkov, A.S., Kulakovskaia, E.V. et Golubev, N.V. (2008). [Antifungal cellobiose lipid secreted by the epiphytic yeast *Pseudozyma graminicola*]. *Mikrobiologiia*, *77*(2), 201-206.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644-652.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Hannick, L.I., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*(19), 5654-5666.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, *9*(1), R7.

Hajlaoui, M.R., Benhamou, N. et Bélanger, R.R. (1992). Cytochemical study of the antagonistic activity of *Sporothrix flocculosa* on rose powdery mildew, *Sphaerotheca pannosa* var. *rosae. Cytology and Histology*, *82*(5), 583-589.

Hajlaoui, M.R., Traquair, J.A., Jarvis, W.R. et Bélanger, R.R. (1994). Antifungal activity of extracellular metabolites produced by *Sporothrix flocculosa. Biocontrol Science and Technology*, *4*(2), 229-237.

Hajlaoui, M.R. et Bélanger, R.R. (1991). Comparative effects of temperature and humidity on the activity of three potential antagonists of rose powdery mildew. *European Journal of Plant Pathology*, *97*(4), 203-208.

Hajlaoui, M.R. et Bélanger, R.R. (1993). Antagonism of the yeast-like phyllophane fungus *Sporothrix flocculosa* against *Erysiphe graminis* var *tritici. Biocontrol Science and Technology*, *3*(4), 427-434.

Hammami, W., Castro, C.Q., Remus-Borel, W., Labbe, C. et Bélanger, R.R. (2011). Ecological basis of the interaction between *Pseudozyma flocculosa* and powdery mildew fungi. *Applied and Environmental Microbiology*, *77*(3), 926-933.

Hammami, W., Labbé, C., Chain, F., Mimee, B. et Bélanger, R.R. (2008). Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa. Applied Microbiology and Biotechnology*, *80*(2), 307-315.

Hane, J., Oliver, R., Hane, J.K. et Oliver, R.P. (2008). RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics*, *9*, 478.

Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H., Goodwin, S.B. et Oliver, R.P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biology*, *12*(5), R45.

Haskins, R.H. et Thorn, J.A. (1951). Biochemistry of the Ustilaginales VII. Antibiotic activity of ustilagic acid. *Canadian Journal of Botany*, *29*(6), 585-592.

Hewald, S., Josephs, K. et Bolker, M. (2005). Genetic analysis of biosurfactant production in *Ustilago maydis. Applied and Environmental Microbiology*, *71*(6), 3033-3040.

Hoffmeister, D. et Keller, N.P. (2007). Natural products of filamentous fungi: enzymes, genes, and their regulation. *Natural Product Reports*, *24*(2), 393-416.

Holton, T.A. et Cornish, E.C. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *The Plant Cell*, *7*(7), 1071-1083.

Jarvis, W.R., Shaw, L.A. et Traquair, J.A. (1989). Factors affecting antagonism of cucumber powdery mildew by *Stephanoascus flocculosus* and *S. rugulosus*. *Mycological Research*, *92*(2), 162-165.

Jeffares, D.C., Mourier, T. et Penny, D. (2006). The biology of intron gain and loss. *Trends in Genetics*, *22*(1), 16-22.

Jones, D.T., Taylor, W.R. et Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, *8*(3), 275-282.

de Jonge, R. et Thomma, B.P. (2009). Fungal LysM effectors: extinguishers of host immunity? *Trends in Microbiology*, *17*(4), 151-157.

Kallberg, Y., Oppermann, U., Jörnvall, H. et Persson, B. (2002a). Short-chain dehydrogenase/reductase (SDR) relationships: A large family with eight clusters common to human, animal, and plant genomes. *Protein Science*, *11*(3), 636-641.

Kallberg, Y., Oppermann, U., Jörnvall, H. et Persson, B. (2002b). Short-chain dehydrogenases/reductases (SDRs). *European Journal of Biochemistry*, *269*(18), 4409-4417.

Kamoun, S. (2009). The Secretome of Plant-Associated Fungi and Oomycetes. In H. B. Deising (Ed.), *The Mycota: Vol. 5. Plant relationships* (pp. 173-180). New York, USA: Springer Berlin Heidelberg

Kanehisa, M. et Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27-30.

Kämper, J. (2004). A PCR-based system for highly efficient generation of gene replacement mutants in *Ustilago maydis*. *Molecular Genetics and Genomics*, *271*(1), 103-110.

Kämper, J., Kahmann, R., Bölker, M., Ma, L.J., Brefort, T., Saville, B.J., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*, *444*(7115), 97-101.

Kemen, E. et Jones, J.D. (2012). Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends in Plant Science*, *17*(8), 448-457.

Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H. et Fedorova, N.D. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology*, *47*(9), 736-741.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59.

Krogh, A., Larsson, B., Heijne, G.V. et Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Fungal Genetics and Biology*, *305*(3), 567-580.

Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C. et Turgeon, B.G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15670-15675.

Kubicek, C.P., Herrera-Estrella, A., Seidl-Seiboth, V., Martinez, D.A., Druzhinina, I.S., Thon, M., et al. (2011). Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma. Genome Biology*, *12*(4), R40.

Kubicek, C.P., Mach, R.L., Peterbauer, C.K. et Lorito, M. (2001). *Trichoderma*: from genes to biocontrol. *Journal of Plant Pathology*, *83*(2), 11-23.

Kulakovskaya, T.V., Shashkov, A.S., Kulakovskaya, E.V. et Golubev, W.I. (2005). Ustilagic acid secretion by *Pseudozyma fusiformata* strains. *FEMS Yeast Research*, *5*(10), 919-923.

Lacombe, E., Hawkins, S., Doorsselaere, J., Piquemal, J., Goffner, D., Poeydomenge, O., et al. (1997). Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *The Plant Journal*, *11*(3), 429-441.

Laurie, J., Linning, R., Wong, P. et Bakkeren, G. (2013). Do TE activity and counteracting genome defenses, RNAi and methylation, shape the sex lives of smut fungi? *Plant Signaling & Behavior*, *8*(4), e23853.

Laurie, J.D., Ali, S., Linning, R., Mannhaupt, G., Wong, P., Güldener, U., et al. (2012). Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *The Plant Cell*, *24*(5), 1733-1745.

Lemieux, R.U., Thorn, J.A., Brice, C. et Haskins, R.H. (1951). Biochemistry of the Ustilaginales: II. Isolation and partial characterization of ustilagic acid. *Canadian Journal of Chemistry*, *29*(5), 409-414.

Lorito, M., Scala, F., Zoina, A. et Woo, S. (2001). Enhancing biocontrol of fungal pests by exploiting the *Trichoderma* genome. In M. Vurro et al. (Eds.), *NATO Science Series I: Life and Behavioural Sciences: Enhancing biocontrol agents and handling risks* (Vol. 339, pp. 248-259). Amsterdam, Netherlands: IOS Press.

Lowe, T.M. et Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955-964.

Marchand, G., Clément-Mathieu, G., Neveu, B. et Bélanger, R.R. (2007). Enhancing biological control efficacy of yeasts to control fungal diseases through biotechnology. In Z. K. Punja, S. H. De Boer, et H. Sanfaçon (Eds.), *Biotechnology and plant disease management* (pp. 518-531). Wallingford, UK: CABI.

Marchand, G., Rémus-Borel, W., Chain, F., Hammami, W., Belzile, F. et Bélanger, R.R. (2009). Identification of genes potentially involved in the biocontrol activity of *Pseudozyma flocculosa. Phytopathology*, *99*(10), 1142-1149.

Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K.E., et al. (2011). Analysis of two *in planta* expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiology*, *156*(2), 756-769.

Martens, S., Teeri, T. et Forkmann, G. (2002). Heterologous expression of dihydroflavonol 4-reductases from various plants. *FEBS Letters*, *531*(3), 453-458.

Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E.G., Duchaussoy, F., et al. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, *452*(7183), 88-92.

Martinez, C., Jauneau, A., Roux, C., Savy, C. et Dargent, R. (2000). Early infection of maize roots by *Sporisorium reilianum* f. sp. *zeae. Protoplasma*, *213*(1-2), 83-92.

Martinez, C., Roux, C. et Dargent, R. (1999). Biotrophic development of *Sporisorium reilianum* f. sp. *zeae* in vegetative shoot apex of maize. *Phytopathology*, *89*(3), 247-253.

Matheussen, A.M., Morgan, P.W. et Frederiksen, R.A. (1991). Implication of gibberellins in head smut (*Sporisorium reilianum*) of *Sorghum bicolor. Plant Physiology*, *96*(2), 537-544.

Matyac, C.A. et Kommedahl, T. (1985). Factors affecting the development of head smut caused by *Sphacelotheca reiliana* on corn. *Phytopathology*, *75*(5), 577-581.

Mimee, B., Labbé, C., Pelletier, R. et Bélanger, R.R. (2005). Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa. Antimicrobial Agents and Chemotherapy*, *49*(4), 1597-1599.

Mimee, B., Pelletier, R. et Bélanger, R.R. (2009). *In vitro* antibacterial activity and antifungal mode of action of flocculosin, a membrane-active cellobiose lipid. *Journal of Applied Microbiology*, *107*(3), 989-996.

Mourier, T. et Jeffares, D.C. (2003). Eukaryotic intron loss. *Science*, *300*(5624), 1393-1393.

Mueller, O., Kahmann, R., Aguilar, G., Trejo-Aguilar, B., Wu, A. et de Vries, R.P. (2008). The secretome of the maize pathogen Ustilago maydis. *Fungal Genetics and Biology*, *45*(Suppl. 1), S63-S70.

Müller, P., Aichinger, C., Feldbrügge, M. et Kahmann, R. (1999). The MAP kinase Kpp2 regulates mating and pathogenic development in *Ustilago maydis. Molecular Microbiology*, *34*(5), 1007-1017.

Nielsen, C.B., Friedman, B., Birren, B., Burge, C.B. et Galagan, J.E. (2004). Patterns of intron gain and loss in fungi. *PLoS Biology*, *2*(12), e422.

Oh, I.S., Park, A.R., Bae, M.S., Kwon, S.J., Kim, Y.S., Lee, J.E., et al. (2005). Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternaria brassicicola. The Plant Cell*, *17*(10), 2832-2847.

Ottmann, C., Luberacki, B., Küfner, I., Koch, W., Brunner, F., Weyand, M., et al. (2009). A common toxin fold mediates microbial attack and plant defense. *Proceedings of the National Academy of Sciences*, *106*(25), 10359-10364.

O'Gorman, C.M., Fuller, H.T. et Dyer, P.S. (2009). Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Nature*, *457*(7228), 471-474.

Paulitz, T.C. et Bélanger, R.R. (2001). Biological control in greenhouse systems. *Annual Review of Phytopathology*, *39*(1), 103-133.

Potter, A.A. (1914). Head smut of sorghum and maize. *Journal of Agricultural Research*, *2*(5), 339-372.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. et Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research*, *33*(suppl 2), W116-W120.

Rehm, B.H.A., Krüger, N. et Steinbüchel, A. (1998). A new metabolic link between fatty acid *de novo* synthesis and polyhydroxyalkanoic acid synthesis. *The Journal of Biological Chemistry*, *273*(37), 24044-24051.

Ron, E.Z. et Rosenberg, E. (2001). Natural roles of biosurfactants. *Environmental Microbiology*, *3*(4), 229-236.

Rossmann, M.G., Moras, D. et Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature*, *250*(5463), 194-199.

Sambrook, J., Fritsch, E.F. et Maniatis, T. (1989). *Molecular cloning* (Vol. 2). New York, NY, USA: Cold Spring Harbor Laboratory Press.

Saville, B.J., Donaldson, M.E. et Doyle, C.E. (2012). Investigating host induced meiosis in a fungal plant pathogen. In A. Swan (Ed.), *Meiosis - molecular mechanisms and cytogenetic diversity* (pp. 411-460). Rijeka, Croatia: InTech.

Schirawski, J., Heinze, B., Wagenknecht, M. et Kahmann, R. (2005). Mating type loci of *Sporisorium reilianum*: Novel pattern with three a and multiple b specificities. *Eukaryotic Cell*, *4*(8), 1317-1327.

Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*, *330*(6010), 1546-1548.

Seidl-Seiboth, V., Zach, S., Frischmann, A., Spadiut, O., Dietzsch, C., Herwig, C., et al. (2013). Spore germination of *Trichoderma atroviride* is inhibited by its LysM protein TAL6. *The FEBS Journal*, *280*(5), 1226-1236.

Sharpton, T.J., Neafsey, D.E., Galagan, J.E. et Taylor, J.W. (2008). Mechanisms of intron gain and loss in *Cryptococcus*. *Genome Biology*, *9*(1), R24.

Shoda, M. (2000). Bacterial control of plant diseases. *Journal of Bioscience and Bioengineering*, *89*(6), 515-521.

Slater, G. et Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31.

Spanu, P.D. (2012). The genomics of obligate (and nonobligate) biotrophs. *Annual Review of Phytopathology*, *50*, 91-109.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, *330*(6010), 1543-1546.

Sparvoli, F., Martin, C., Scienza, A., Gavazzi, G. et Tonelli, C. (1994). Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). *Plant Molecular Biology*, *24*(5), 743-755.

Spellig, T., Bottin, A. et Kahmann, R. (1996). Green fluorescent protein (GFP) as a new vital marker in the phytopathogenic fungus *Ustilago maydis*. *Molecular & General Genetics*, *252*(5), 503-509.

Spoeckner, S., Wray, V., Nimtz, M. et Lang, S. (1999). Glycolipids of the smut fungus *Ustilago maydis* from cultivation on renewable resources. *Applied Microbiology and Biotechnology*, *51*(1), 33-39.

Stanke, M., Steinkamp, R., Waack, S. et Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*, *32*(suppl 2), W309-W312.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., et al. (2002). The Generic Genome Browser: A building block for a model organism system database. *Genome Research*, *12*(10), 1599-1610.

Stergiopoulos, I. et de Wit, P.J. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, *47*, 233-263.

St-Pierre, B., Laflamme, P., Alarco, A.-M., D, V. et Luca, E. (1998). The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. *The Plant Journal*, *14*(6), 703-713.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. et Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731-2739.

Teichmann, B., Labbé, C., Lefebvre, F., Bölker, M., Linne, U. et Bélanger, R.R. (2011a). Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa*. *Molecular Microbiology*, *79*(6), 1483-1495.

Teichmann, B., Lefebvre, F., Labbé, C., Bölker, M., Linne, U. et Bélanger, R.R. (2011b). Beta hydroxylation of glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-dependent β-hydroxylase. *Applied and Environmental Microbiology*, *77*(21), 7823-7829.

Teichmann, B., Linne, U., Hewald, S., Marahiel, M.A. et Bölker, M. (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Molecular Microbiology*, *66*(2), 525-533.

Teichmann, B., Liu, L., Schink, K.O. et Bölker, M. (2010). Activation of the ustilagic acid biosynthesis gene cluster in *Ustilago maydis* by the C2H2 zinc finger transcription factor Rua1. *Applied and Environmental Microbiology*, *76*(8), 2633-2640.

Torto, T.A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N.A., et al. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora. Genome Research*, *13*(7), 1675-1685.

Trapnell, C., Pachter, L. et Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105-1111.

Traquair, J.A., Shaw, L.A. et Jarvis, W.R. (1988). New species of *Stephanoascus* with *Sporothrix* anamorphs. *Canadian Journal of Botany*, *66*(5), 926-933.

Vinale, F., Sivasithamparam, K., Ghisalberti, E.L., Marra, R., Woo, S.L. et Lorito, M. (2008). *Trichoderma*–plant–pathogen interactions. *Soil Biology & Biochemistry*, *40*(1), 1-10.

Walton, J.D. (1994). Deconstructing the cell wall. *Plant Physiology*, *104*(4), 1113-1118.

Walton, J.D. (2000). Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: An hypothesis. *Fungal Genetics and Biology*, *30*(3), 167-171.

Whipps, J.M. (2001). Microbial interactions and biocontrol in the rhizosphere. *Journal of Experimental Botany*, *52*(suppl 1), 487-511.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M., et al. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*, *450*(7166), 115-118.

William Roy, S. et Gilbert, W. (2006). The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nature Reviews. Genetics*, *7*(3), 211-221.

Winnenburg, R., Baldwin, T.K., Urban, M., Rawlings, C., Köhler, J. et Hammond-Kosack, K.E. (2006). PHI-base: A new database for pathogen host interactions. *Nucleic Acids Research*, *34*(suppl 1), D459-D464.

Woo, S.L., Scala, F., Ruocco, M. et Lorito, M. (2006). The molecular biology of the interactions between *Trichoderma* spp., phytopathogenic fungi, and plants. *Phytopathology*, *96*(2), 181-185.

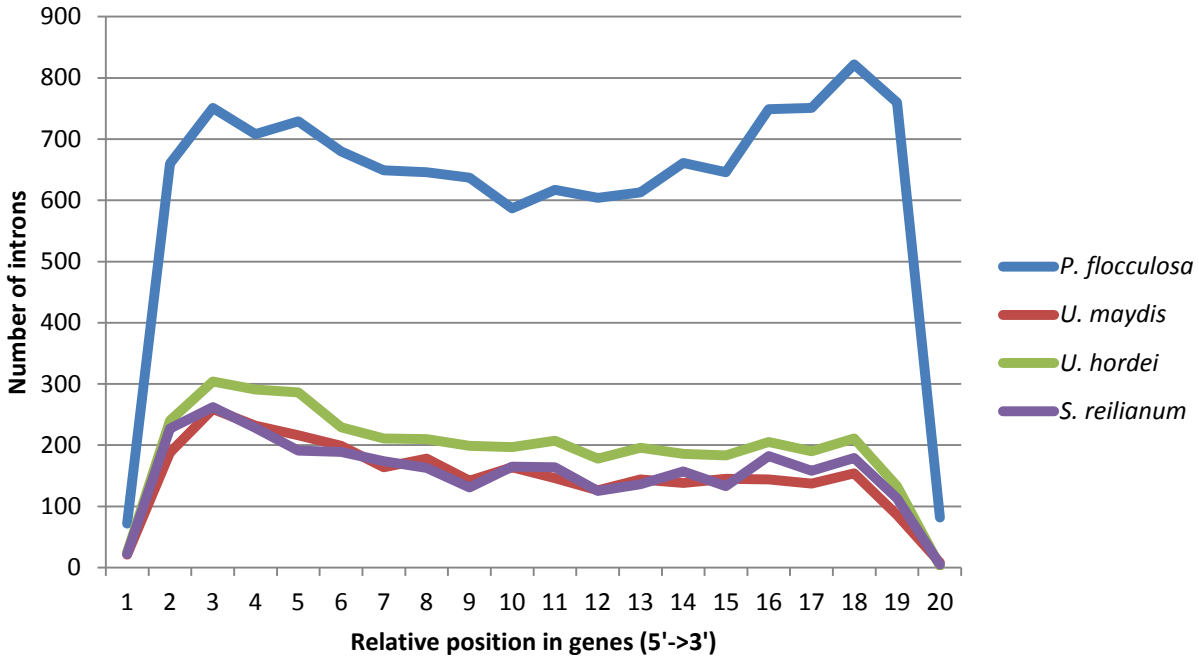
Wu, T.D. et Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*(9), 1859-1875.

Xu, J., Saunders, C.W., Hu, P., Grant, R.A., Boekhout, T., Kuramae, E.E., et al. (2007). Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(47), 18730-18735.

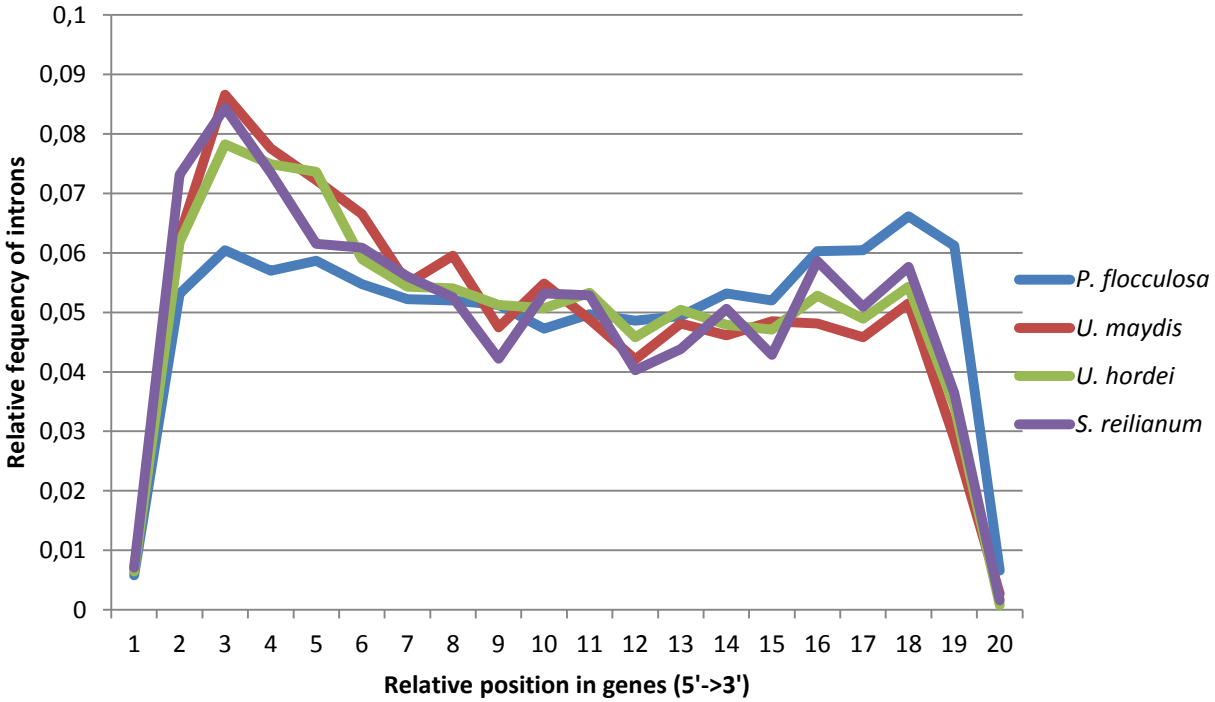
Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. et Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *40*(W1), W445-W451.

Zeng, X., Nesbitt, M.J., Pei, J., Wang, K., Vergara, I.A. et Chen, N. (2008). OrthoCluster: a new tool for mining synteny blocks and applications in comparative genomics. In *EDBT '08: Proceedings of the 11th international conference on extending database technology* (pp. 656-667). Nantes, France: ACM.

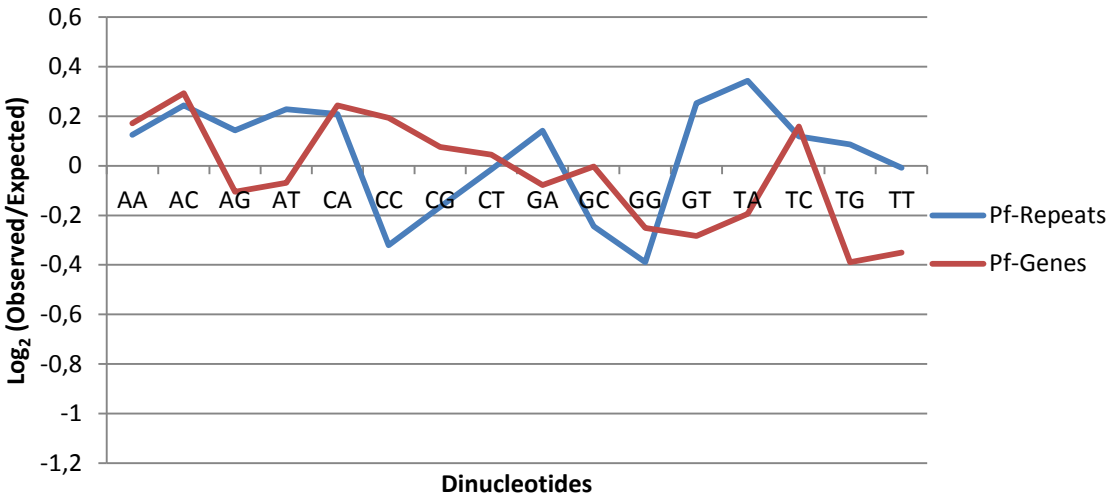
Annexe I : Chapter 5 – Supplemental tables and figures



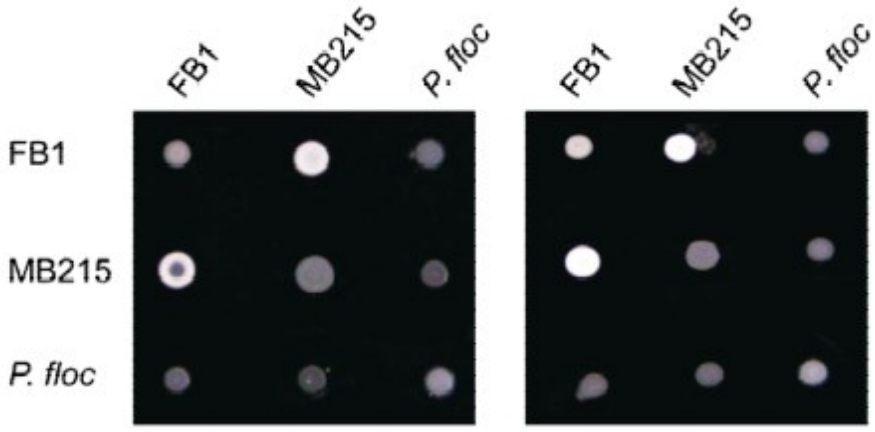
Supplemental figure 1 : Number of introns in four Ustilaginomycotina fungi according to relative position in genes.



Supplemental figure 2 : Relative frequency of introns according to relative position in genes.



Supplemental figure 3 : Analysis of dinucleotide frequency in *P. flocculosa*. Dinucleotide frequencies for repeat sequences and gene sequences show no specific depletion of certain pairs; GC-depletion was shown to be a sign of inactivation through repeat induced point mutation (RIP) (Laurie et al., 2012).



Supplemental figure 4 : Mating assays between *P. flocculosa* and *U. maydis*. *P. flocculosa* and *U. maydis* strains were tested for their ability to mate. For this assay, haploid cultures were placed as drops on a charcoal-containing medium. Successful mating produces dikaryotic aerial hyphae which appear as a white fluffy colony phenotype. While strong dikaryotic filamentous growth could be observed when *U. maydis* strains MB215 and FB1 were mated, no aerial hyphae were ever observed when *P. flocculosa* was mixed with either *U. maydis* strain or when the strains were selfed. This indicates that *P. flocculosa* is not able to mate with *U. maydis*.

Supplemental table 1 : Inventory of repetitive sequences in the *P. flocculosa* genome assembly

		Number of elements[a]	Length occupied (bp)	Percentage of sequence (%)
SINEs		1	68	0.00
	ALUs	0	0	0.00
	MIRs	0	0	0.00
LINEs		226	22271	0.10
	LINE1	13	818	0.00
	LINE2	2	174	0.00
	L3/CR1	4	421	0.00
LTR elements		413	46452	0.20
	ERVL	0	0	0.00
	ERVL-MaLRs	0	0	0.00
	ERV class I	4	494	0.00
	ERV class II	7	359	0.00
DNA elements		266	19267	0.08
	hAT-Charlie	0	0	0.00
	TcMar-Tigger	0	0	0.00
Unclassified		269	75783	0.32
Total interspersed repeats			163841	0.70
Small RNA		23	5090	0.02
Satellites		18	1583	0.01
Simple repeats		11748	521283	2.23
Low complexity		1913	97309	0.42

[a.] most repeats fragmented by insertions or deletions have been counted as one element

Supplemental table 2 : Inventory of repetitive sequences in non-assembled *P. flocculosa* sequences

		Number of elements[a]	Length occupied (bp)	Percentage of sequence (%)
SINEs		0	0	0.00
	ALUs	0	0	0.00
	MIRs	0	0	0.00
LINEs		7	782	0.47
	LINE1	1	80	0.05
	LINE2	0	0	0.00
	L3/CR1	0	0	0.00
LTR elements		40	9226	5.55
	ERVL			
	ERVL-MaLRs	0	0	0.00
	ERV class I	1	24	0.01
	ERV class II	0	0	0.00
DNA elements				
	hAT-Charlie	0	0	0.00
	TcMar-Tigger	0	0	0.00
Unclassified		67	17012	10.24
Total interspersed repeats			27020	16.27
Small RNA		13	3567	2.15
Satellites		0	0	0.00
Simple repeats		121	11098	6.68
Low complexity		13	658	0.40

[a.] most repeats fragmented by insertions or deletions have been counted as one element

Supplemental table 3 : OthoCluster analysis

Synteny conservation			
	P. flocculosa	*U. maydis*	*U. hordei*
U. maydis	0.409		
U. hordei	0.484	0.778	
S. reilianum	0.308	0.768	0.511

Number of blocks			
	P. flocculosa	*U. maydis*	*U. hordei*
U. maydis	672		
U. hordei	683	759	
S. reilianum	645	456	612

Largest block size (Genes*)			
	P. flocculosa	*U. maydis*	*U. hordei*
U. maydis	25		
U. hordei	34	59	
S. reilianum	43	103	87

Average block size (Genes[a])			
	P. flocculosa	*U. maydis*	*U. hordei*
U. maydis	4		
U. hordei	4	8	
S. reilianum	5	14	10

Insertions/deletions			
	P. flocculosa	*U. maydis*	*U. hordei*
U. maydis	12		
U. hordei	9	93	
S. reilianum	13	125	155

Transpositions			
	P. flocculosa	*U. maydis*	*U. hordei*
U. maydis	67		
U. hordei	59	65	
S. reilianum	77	38	75

[a.] Only genes from gene families present in both species (homologs) are counted.

Supplemental table 4 : Subset of gene correspondence for important biological processes

Gene	Reference gene	Gene ID in *P. flocculosa*	BlastP e-value
Mating related genes			
Pheromone receptor (*pra*)	um02383	pf01711	2.9e-85
Mating pheromone (*mfa*)	S. reilianum16900	pf06826	4.1e-06
Mating pheromone (*mfa*)	UHOR_03899	pf06827	1.7
rba1	um02384	pf01707	2.1e-49
bW	um00578	pf03265	5.9e-14
bE	um12052	pf03264	3.7e-24
Pheromone response factor (*prf*)	um02713	pf03871	5.8e-51
Ste-20 like protein kinase (*smu*)	um12272	pf06042	0
lga2 (mitochondrial inheritance)	AAA99769.1[a]	-	-
rga2 (mitochondrial inheritance)	AAA99770.1[a]	-	-
Transcriptional gene silencing and chromatin remodeling			
ARGONAUTE (*ago*)	UHOR_06256	pf06699	0
DICER (*dcl*)	UHOR_08937	pf04802	0
RNA dependent RNA polymerase – (*RdRP*)	UHOR_08874	pf06447	0
RNA dependent RNA polymerase – (*RdRP*)	UHOR_01631	-	-
RNA dependent RNA polymerase – (*RdRP*)	UHOR_15740	pf06710	0
Chromodomain-coding HP1-like (*Chp*)	UHOR_05116	-	-
Chromodomain-coding HP1-like (*Chp*)	UHOR_07750	pf04343	8.1e-40
Chromodomain-coding (*Chp*)	UHOR_16420	pf06213	1.5e-73
C5-cytosine methyltransferase (*DNAme*)	UHOR_08509	pf06913	1.1e-100

[a.] Genbank accession number

Supplemental table 5 : Conserved SMURF secondary metabolite gene clusters

	Type of backbone gene	ID of backbone gene associated to SMURF[a] cluster	Lowest fraction of shared gene families for all possible combinationsb	Function
1	NRPS	pf03176, um10189, UHOR_03116, sr13070	3/5	siderophore biosynthesis (Sid2)
2	NRPS-Like	um01830, UHOR_02719, sr12912	10/16	unknown
3	PKS	um10339, sr14030	8/11	unknown
4	PKS	um06418, UHOR_08970, sr16861	7/14	unknown
5	NRPS	um01434, sr12503	7/13	siderophore biosynthesis (Fer3)
6	PKS	pf00007, um06460	5/7	flocculosin/ustilagic acid biosynthesis

[a.] Predictions of SMURF (Secondary Metabolite Unique Regions Finder, JCVI) clusters are approximative due to the inclusion of genes not involved in the biosynthesis but physically present in the cluster, or due to the exclusion of genes involved in the biosynthesis but having a function that is not considered by SMURF as a hallmark of secondary metabolite biosynthesis.

[b.] For all combinations of species sharing a SMURF cluster, the lowest proportion of gene families being shared between the clusters of two species is reported.

Annexe II : Liste des publications et communications

Publications scientifiques

Lefebvre, F., Joly, D.L., Labbé, C., Teichmann, B., Linning, R., Belzile, F., et al. (2013). The transition from a phytopathogenic smut ancestor to an anamorphic biocontrol agent deciphered by comparative whole-genome analysis. *The Plant Cell*, *25*(6), 1946-1959.

Teichmann, B., Labbé, C., Lefebvre, F., Bölker, M., Linne, U. et Bélanger, R.R. (2011). Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa. Molecular Microbiology*, *79*(6), 1483-95.

Teichmann, B., Lefebvre, F., Labbé, C., Bölker, M., Linne, U. et Bélanger, R.R. (2011). Beta hydroxylation of glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-dependent β-hydroxylase. *Applied and Environmental Microbiology*, *77*(21), 7823-7829

Revue de littérature

Bélanger, R.R., Labbé, C., Lefebvre, F. et Teichmann, B. (2012). Mode of action of biocontrol agents: all that glitters is not gold. *Canadian Journal of Plant Pathology*, *34*(4), 469–478.

Communications orales

Lefebvre, F., Joly, D.L., Teichmann, B., Bakkeren, G. et Bélanger, R.R. (2014). Genome of *Pseudozyma flocculosa* reveals singular path towards biocontrol activity. [Résumé] *Phytopathology*, *104*(3 Suppl.), S1.2.

(Réalisée dans le cadre du 2013 Northeastern Division Meeting of the American Phytopathological Society, Southbury, CT, É-UA)

Lefebvre, F., Teichmann, B., Labbé, C. et Bélanger, R.R. (2011). *Ustilago maydis* as a model system for the study of a glycolipid gene cluster in the biocontrol agent *Pseudozyma flocculosa.* [Résumé] *Phytopathology*. *101*(6 suppl.), S260.

(Réalisée dans le cadre du 2010 Northeastern Division Meeting of the American Phytopathological Society, Northampton, MA, É-UA)

Affiches scientifiques

Lefebvre, F., Joly, D.L., Bakkeren, G., Belzile, F. et Bélanger, R.R. (2013). A biocontrol agent among pathogens : How *Pseudozyma flocculosa* genome relates to singular lifestyle. [Résumé] *Fungal Genetics Reports. 60*(suppl.), 201-202.

(Présentée dans le cadre du 27th Fungal Genetics Conference, Asilomar, CA, É-UA)

Lefebvre, F., Teichmann, B., Labbé, C. et Bélanger, R.R. (2011). Chimeric production of flocculosin by *Ustilago maydis*. [Résumé] *Fungal Genetics Reports. 58*(suppl.), 199.

(Présentée dans le cadre du 26th Fungal Genetics Conference, Asilomar, CA, É-UA)