

*Ecole Doctorale Agriculture, Alimentation, Biologie, Environnement,
Santé*

INRA-Micalis & Genomic Vision

résumé de la thèse de doctorat en Sciences de la Vie

soutenue le 20/11/2013
par

Kevin Cheeseman

**Aspects of *Penicillium* Genomics: Molecular Combing Genome
Assembly, Genetic exchange in food and potential for secondary
metabolite production.**

**Aspects de la Génomique du genre *Penicillium*: Assemblage de
génomme par Peignage Moléculaire, Echange génétique dans
l'environnement alimentaire et potentiel de production en
métabolites secondaires.**

Directeur de thèse :
Co-encadrant :

Pierre Renault
Maurizio Ceppi

Directeur de recherches, INRA, Jouy-en-Josas
Chef de Projet, Genomic Vision, Bagneux

Composition du jury :

Président du jury :

Armel Guyonvarch

Professeur université Paris Sud XI, Orsay

Rapporteurs :

Sylvie Dequin
Ronald De Vries

Directrice de recherches, INRA, Montpellier
Directeur de recherches, CBS-KNAWL, Utrecht

Examineurs :

Cecile Neuvéglise
Eric Record

Directrice de recherches, INRA, Grignon
Directeur de recherches, INRA, Marseille

Sommaire

Objectif du Résumé :	- 3 -
Contexte et objectifs:	- 3 -
Etat de l'art et introduction bibliographique :	- 6 -
Séquençage et assemblage de génomes complet :	- 6 -
Peignage moléculaire :	- 7 -
Les champignons:	- 9 -
Les transferts horizontaux de gènes ou l'acquisition rapide de caractère innovant par l'Evolution... 11 -	-
Résultats	- 13 -
Amélioration de l'assemblage de <i>Penicillium roqueforti</i> : Développement d'une méthodologie innovante pour améliorer les assemblages génomiques.....	- 13 -
Transfert latéral de matériel génétique dans la chaîne alimentaire : Flux de gènes dans le fromage... 14 -	-
Evaluation du potentiel génomique pour la production de métabolites secondaires par les espèces de <i>Penicillium</i> et leur impact sur la sécurité alimentaire : résultats préliminaires.....	- 16 -
Discussion	- 18 -

Objectif du Résumé :

Ce résumé en français a pour vocation d'indiquer quelles informations sont présentes dans les différentes parties du manuscrit de thèse, lui-même rédigé en anglais. Il permettra ainsi au lecteur éventuel de décider quelles parties du manuscrit sont susceptibles de l'intéresser et d'y accéder directement. Le résumé suit donc une structure similaire à celle du manuscrit. Cette dernière est constituée d'une courte préface relatant les découvertes et aspects des sciences naturelles ayant conduit au développement de la biologie moderne et de la génomique, suivie d'une introduction bibliographique en quatre parties distinctes : une introduction au séquençage et à l'assemblage de génome, une revue sur la technologie du Peignage moléculaire, une revue sur la biologie générale des champignons et leur importance pour l'homme, et enfin une présentation exhaustive de l'état des recherches sur les transferts horizontaux de matériel génétique. La partie suivante détaille les résultats obtenus durant la thèse, sous trois aspects distincts : l'amélioration de l'assemblage du génome de *Penicillium roqueforti* par le développement d'une nouvelle méthodologie d'assemblage sur molécule d'ADN peignée, la mise en évidence d'un transfert horizontal de plus de 250 gènes dans les *Penicillia* de l'environnement fromager, et une étude préliminaire du potentiel en métabolites secondaires des *Penicillia*. Enfin, une discussion sur l'ensemble du travail est présentée en dernière partie.

Contexte et objectifs:

Le travail rapporté ici a été financé par une convention CIFRE. Les conventions CIFRE, acronyme de Convention Industrielles de Formation par la Recherche, supportée par l'ANRT sont destinées à créer et favoriser les liens industrie – académie en portant conjointement un projet de recherche et en permettant l'accueil d'un doctorant dans les deux environnements. En conséquence, le travail présenté ici a été réalisé dans deux laboratoires partenaires : un laboratoire industriel, Genomic Vision, et un laboratoire académique, l'équipe de recherche INRA Bactéries Alimentaires et Commensales (Bac). La conséquence directe de ce type de convention et collaboration réside dans l'attente de résultats de différentes natures : Le projet doit générer des connaissances fondamentales et de nouvelles applications technologiques, de la propriété intellectuelle ou du savoir-faire technologique d'intérêt industriel.

Genomic Vision est une société de biotechnologie fondée en 2004 par Aaron Bensimon et Daniel Nerson en tant que spin-off de l'Institut Pasteur. Sa mission première et son activité principale concerne le développement et l'application d'une technologie propriétaire, le peignage moléculaire, au diagnostic clinique, à la pharmaco-génomique et à la recherche en sciences de la vie. Genomic

Vision concentre ses activités sur le développement de tests diagnostiques répondant à des besoins spécifiques et actuels dans le cadre des maladies génétiques et génomiques, ainsi que sur l'étude de la dynamique et cinétique de la réplication de l'ADN au travers du « Replication combing Assay ». Le peignage moléculaire est une technologie de biologie moléculaire permettant d'étirer de manière constante des molécules d'ADN sur des lamelles de microscopie. Des expériences d'hybridation *in-situ* fluorescentes sur l'ADN ainsi peigné permettent l'étude du paysage génomique à des loci d'intérêt sur des molécules uniques à l'échelle génome complet, tout en permettant d'étudier de nombreux exemplaire du génome sur une même lamelle. Cette technologie propose ainsi une fenêtre unique d'observation au sein des méthodes d'étude de l'ADN. Le travail réalisé durant cette thèse réside l'exploration et le développement d'une nouvelle application du Peignage Moléculaire pour répondre aux problématiques spécifique du domaine de l'assemblage de génome, fournissant ainsi un moyen d'améliorer la qualité des assemblages bioinformatique de génomes complets.

L'équipe INRA Bactéries Alimentaires et Commensales concentre ses activités de recherche sur l'étude de la génétique microbienne dans un contexte de santé humaine et de sécurité alimentaire. Son domaine d'expertise comprend les bactéries commensales, les bactéries acides lactiques et la metagénomique des environnements alimentaires et des environnements humains associés. L'équipe est impliquée dans plusieurs projets de recherche à fort potentiel en santé humaine et sécurité alimentaire. Le projet dans lequel cette thèse a pris place est le projet Food Microbiomes, coordonnée par le Dr. Pierre Renault, chef de l'équipe Bac et directeur de cette thèse. Le projet Food Microbiomes vise à générer un large corpus de connaissance par l'application de méthodologies microbiologique conventionnelles et innovantes de type metagénomique à l'écosystème fromager. Ce projet aidera ainsi à caractériser plus finement l'écosystème fromager ainsi qu'à évaluer l'innocuité des microorganismes utilisés dans les procédés de fabrication du fromage. Pour nombre de microorganismes survenant dans l'environnement fromager, et notamment pour les emblématiques et traditionnels *Penicillia*, nous ne disposons pas à l'heure actuelle de connaissances suffisantes pour les associer à des normes de sécurité alimentaires et sanitaires. L'un des objectifs du projet Food Microbiomes est donc d'améliorer ce corpus de connaissances autour de l'environnement fromager et de ses microorganismes associés aux chaînes alimentaires de nos sociétés. L'un des objectifs du travail de thèse rapporté ici est donc d'apporter une meilleure connaissance des génomes d'espèces de *Penicillium*, impactant ainsi sur la taxonomie peu connue d'un groupe de champignons filamenteux, et aidant à établir l'innocuité de leur utilisation dans la chaîne alimentaire.

Ainsi, le manuscrit de thèse rapporte et commente plusieurs aspects de la génomique des *Penicillia*, par l'amélioration de la séquence génomique d'une espèce de *Penicillium*, *Penicillium roqueforti*, et par l'observation de flux horizontaux de gènes entre différentes espèces de *Penicillium* dans l'environnement fromager. Ce manuscrit rapporte aussi une caractérisation préliminaire du potentiel de production de métabolites secondaires par les *Penicillia* de l'environnement fromager.

Etat de l'art et introduction bibliographique :

Séquençage et assemblage de génomes complet :

Le séquençage, et son but ultime, la reconstruction de génomes complets était jusqu'à il y a peu le territoire de grands centres de séquençages. L'arrivée de la seconde génération de technologies de séquençage a profondément bouleversé cet ordre préétabli en fournissant, vers le milieu des années 2000, un accès au séquençage à de nombreux laboratoires pour un coût réduits et un meilleur débit. Ainsi, l'assemblage de génome *de novo* s'est démocratisé. Ceci a conduit d'une part à une augmentation du nombre de séquences génomiques, et d'autre part, à une diminution globale de la qualité de ces séquences, principalement due au manque d'expertises, d'infrastructure et de standards de qualité appropriés pour la publication des génomes. Le premier chapitre de cette introduction bibliographique vise à donner au lecteur une appréciation de ce qu'est le séquençage et l'assemblage de génomes de nos jours, ainsi qu'à expliquer pourquoi nous assistons à une telle diminution de la qualité des génomes publiés. Après un bref historique du séquençage, nous introduisons les principales stratégies utilisées pour séquencer un génome, et l'assembler. Enfin nous aborderons les aspects pratiques affectant la qualité finale de l'assemblage du génome d'intérêt. Ce chapitre n'a pas pour vocation de passer en revue tous les aspects techniques des diverses technologies existantes, mais plutôt de mettre en lumière le paysage complexe du séquençage de nos jours, et de discuter comment les concepts et techniques actuelles mènent à une augmentation spectaculaire du nombre de génomes publiés, tout en amenant une réduction globale de la qualité de ces génomes.

La diminution de la longueur des "reads" ou lectures d'ADN produit par l'ADN a entraîné la génération de diverses approches, tant pour le séquençage en lui-même, que pour la reconstruction de la séquence complète d'ADN à partir de ces fragments lus par les séquenceurs. Le chapitre décrit ces différentes approches, en commençant par les approches de séquençage de génome complet dites "Whole Genome Shotgun" et "Hierarchical shotgun". Sont ensuite abordés les différentes approches bioinformatiques de reconstruction de la séquence complète ou "assemblage". Ici seront brièvement introduites les approches algorithmiques générales ("greedy", "Overlap-Layout-Consensus" et "De Bruijn graph") avec leurs avantages respectifs. Enfin les problèmes causés par les séquences répétées sont ensuite exposés. Pour finir, une discussion sur la complexité actuelle des choix qui gouvernent les approches choisies par les chercheurs, et leur répercussion sur la qualité actuelle et future des génomes publiés est proposée.

Peignage moléculaire :

Le peignage moléculaire représente un important aspect de cette thèse. Dans ce chapitre introductif, nous positionnons le peignage moléculaire parmi les autres technologies d'étude de l'ADN à l'échelle de la molécule unique. Ces dernières années, les technologies simple molécules sont devenues de plus en plus utilisées pour l'étude des acides nucléiques. Ainsi, aujourd'hui, de nombreuses technologies sont à disposition du chercheur. Ces technologies dépendent toutes de principes particuliers et, par conséquent, ont donc chacune leurs avantages et leurs limitations, les rendant ainsi plus ou moins adaptées à certaines applications. Dans cette revue, nous présenterons brièvement certaines des technologies simple molécules applicable à l'étude de l'ADN, puis nous décrirons de manière plus précise le peignage moléculaire et ses applications possibles. Nous décrirons aussi comment certaines de ces technologies ont été appliquées à la problématique de l'assemblage de génomes, décrite dans le chapitre précédent. Ce chapitre a pour vocation d'introduire le rationnel pour utiliser le peignage moléculaire à l'assemblage de génomes.

Le peignage moléculaire est une nanotechnologie qui permet d'étudier l'organisation des molécules d'ADN, simultanément à l'échelle du génome entier et de la molécule unique, sur plusieurs copies du génome. Le procédé de peignage moléculaire permet l'obtention de longues molécules d'ADN étirées de manière uniforme et homogène, sur une lamelle de microscopie. La première étape de ce procédé inclut l'obtention d'ADN de haut poids moléculaire en solution par la lyse de cellules incluse dans un bloc d'agarose. Une lamelle de microscopie silanisée est ensuite immergée et incubée dans la solution d'ADN avec un pH légèrement acide préalablement obtenue. Les molécules d'ADN dans un environnement acide s'organisent en « pelote » avec les domaines hydrophobiques des bases exposées aux extrémités de la molécule, à cause de la dénaturation induite par le pH. Ceci permet une interaction forte entre la surface silanisée qui est hydrophobique et les extrémités de l'ADN. La lamelle est ensuite retirée de la solution à vitesse constante ($300\mu\text{m}\cdot\text{sec}^{-1}$). Les molécules sont ainsi étirées par la force appliquée par le ménisque. La triple interface air-verre-solution étant la même tout le long de la lamelle de microscopie, la force résultante de la tension de surface appliquée par le ménisque est la même, et les molécules sont donc étirées de manière homogène et constante sur toute la lamelle de microscopie. Ce procédé permet ainsi d'obtenir de nombreuses longues molécules d'ADN étirées de la même manière, quelle que soit leur séquence. Les molécules font en moyenne entre 500 et 700 kilobases. Selon la taille du génome de l'organisme et la concentration cellulaire utilisée initialement, il en résulte donc quelques centaines à plusieurs milliers de génomes étirés sur une simple lamelle de microscopie. Des expériences d'hybridation in situ fluorescentes peuvent ensuite être appliquées à ces molécules étirées. Ceci permet la visualisation directe par

microscopie à fluorescence de l'organisation génomique de nombreux loci, sur des molécules uniques.

Dans ce chapitre, nous repositionnons le peignage moléculaire parmi les technologies simple molécule appliquées à l'étude de l'ADN. Nous décrivons donc brièvement les technologies de séquençage de molécules uniques, les pinces optiques et magnétiques, les technologies de FISH et les technologies basées sur l'étirement de molécules d'ADN par l'utilisation de technologies de micro-fluidique. Sont ensuite décrites les différentes applications existantes du Peignage moléculaire pour l'étude de l'ADN. Plus particulièrement, nous illustrons les cas suivants :

- La détection de grands réarrangements dans des gènes ou autres régions génomiques d'intérêt, avec une emphase sur leur application en diagnostic.
- L'étude la réplication de l'ADN, tant du point de vue de sa cartographie, que de sa cinétique, au travers du « replication combing assay ».
- l'Etude de l'organisation structurale de régions complexes, type « copy number variation ».
- La détection d'ADN viral dans les génomes, et l'étude de leur intégration.

Enfin, le rationnel pour l'application du peignage moléculaire à la résolution des problèmes d'assemblage de génomes qui constitue un des aspects technologique de cette thèse est exposé.

Les champignons:

Ce chapitre propose une vision générale du royaume des champignons, ou *fungi*. Après avoir décrit certains importants aspects de leur physiologie et de leur mode de vie, ce chapitre aborde l'impact que ce royaume a sur les écosystèmes, la recherche, et nos sociétés. Le but de ce chapitre est de proposer et lier suffisamment d'informations sur les *fungi* pour mettre en lumière l'importance du royaume *fungi* sur les écosystèmes, tant naturels que ceux créés par l'homme. La première partie du chapitre fournit un aperçu de la biologie générale des champignons et des écosystèmes dont ils font partie. Dans un second temps, leur taxonomie est présentée. La troisième section présente un focus sur leur génomique et les découvertes importantes que ce domaine de recherche a permis. Enfin, la dernière section met en perspective l'importance des champignons pour l'Homme et discute de manière plus approfondie les champignons étudiés dans cette thèse et leur environnement anthropomorphe, le fromage et la chaîne alimentaire.

Le royaume des *fungi* comprend diverses formes de vie établies dans des niches écologiques très variées. Ce royaume comprend ainsi des organismes aux modes de vie très différents. En effet, si la plupart des espèces appartenant aux *fungi* sont multicellulaires, il existe aussi de nombreuses espèces unicellulaires. Les *fungi*, ou champignons au sens large du terme, jouent un rôle clé dans la plupart des écosystèmes, en agissant comme les principaux décomposeurs de biomasse. Les *fungi* sont retrouvés dans de nombreux environnements, tel le sol, les eaux marines et douces, dans le système digestif des animaux, sur les plantes... Ils sont, selon l'espèce et l'environnement, saprophytes, parasites, symbiotes... Plus de 99 000 espèces ont été décrites à l'heure d'aujourd'hui, mais l'on estime que la grande majorité des *fungi* n'a pas encore été décrite ou même rencontrée par l'homme. En effet, les estimations sur la diversité des champignons varient entre 1.5 et 5.1 millions. Ce succès écologique peut s'expliquer par le mode de nutrition des *fungi*. En effet, tous les champignons sont hétérotrophes, c'est-à-dire qu'ils se nourrissent par absorption de nutriments. Des exo-enzymes hydrolytiques sont secrétées dans l'environnement extracellulaire ou elles décomposent les macromolécules organiques en composés plus petits. Ces derniers sont ensuite absorbés par la cellule, fournissant les nutriments et l'énergie nécessaires à la survie de l'organisme et à sa prolifération. Une fois la source d'énergie du milieu épuisée, les hyphes croissent jusqu'à atteindre un environnement riche en macromolécules organiques.

Cette association entre immobilité et hétérotrophie est particulièrement adaptée aux nombreux modes de vie des diverses espèces fongiques. Ce chapitre décrit les différents environnements dans

lesquels on retrouve les *fungi*, ses principales associations avec les autres organismes vivant, puis aborde certains aspects de sa taxonomie. Enfin, pour finir, l'importance des champignons pour l'homme, tant en termes de maladies émergentes qu'en termes d'intérêts biotechnologiques et agroalimentaires est abordé. La dernière partie du chapitre concerne le genre *Penicillium*, l'objet d'étude de cette thèse.

Les transferts horizontaux de gènes ou l'acquisition rapide de caractère innovant par l'Evolution.

Les transferts horizontaux de gènes, aussi connus sous le nom de transfert latéraux, sont définis comme étant des flux d'informations génétiques de manière asexuelle entre deux organismes. Appelé horizontaux ou latéraux, par opposition au flux vertical d'héritage génétique de parents vers la descendance, ces phénomènes ont été découverts au milieu des années 90. L'existence d'un tel phénomène, fréquent chez les organismes procaryotes, et dont l'importance et la fréquence sont toujours débattues chez les eucaryotes, remet en question de nombreuses idées sur l'arbre de la vie, la synthèse néo-darwinienne de l'évolution, ainsi que les pratiques médicales (diffusion de la résistance aux antibiotiques) et agro-alimentaires. D'années en années, l'appréciation et la compréhension de l'étendue et de l'impact de ce phénomène a conduit à un important débat sur la possibilité de reconstruire un arbre phylogénétique des procaryotes ainsi que son utilité. Ce débat est toujours d'actualité. Plus important encore, ce débat a permis une révision des pratiques médicales et agricole pour éviter la diffusion de résistance aux antibiotiques ou le contact d'organismes écologiquement isolés qui pourrait devenir de nouveaux pathogènes ou des pathogènes plus résistants.

La transmission du support et de l'information génétique des géniteurs vers leur descendance est un concept bien connu et bien ancré en biologie. La découverte de l'ADN en tant que support de l'information génétique dans les années 50, puis le développement de la phylogénétique dans les années 70 ont permis la création d'arbre du vivant par la comparaison de marqueurs universels. Durant ces années, la proximité des espèces entre elles a permis de nombreux progrès dans l'étude de nombreux organismes ainsi qu'une meilleure compréhension des mécanismes évolutifs. L'observation de l'incongruence entre certains arbres phylogénétiques a rapidement donné lieu à un débat sur la pertinence de certains marqueurs utilisés pour la reconstruction de ces arbres. L'observation de ces incongruences est en désaccord avec le principe de l'hérédité génétique verticale. Cette observation, combinée à la démonstration expérimentale que certaines formes de vie (les bactéries) peuvent échanger du matériel génétique entre elles ont permis le développement de l'idée de transfert horizontal de gène. Ainsi, dans les années 1980, le concept de transfert horizontal de gène s'est développé et établi comme étant un phénomène anecdotique pouvant expliquer l'incongruence des arbres phylogénétiques. Peu, en ces temps-là, réalisèrent l'importance qu'un tel phénomène pourrait avoir si il ne s'avérait pas limité et anecdotique. Avec l'ère de la génomique et la démocratisation du séquençage à haut débit, de nombreux autres exemples de transfert horizontaux furent découverts. La découverte d'une forte fréquence de ces transferts

horizontaux a donné lieu a de nombreuses controverses touchant a divers dogmes de la biologie, notamment l'idée d'un arbre du vivant chez les procaryotes.

Ce chapitre présente et discute l'importance de ce phénomène de transfert horizontal de matériel génétique. Il commence par une introduction historique sur la découverte de ce phénomène chez les procaryotes, puis présente les différentes étapes mécanistiques qui permettent a un organisme d'acquérir du matériel génétique libre dans l'environnement et n'appartenant pas a ses parents. Le chapitre présente ensuite l'importance de ce phénomène pour l'homme, en prenant notamment comme exemple la diffusion de la résistance aux antibiotiques. Nous abordons ensuite la découverte de transferts horizontaux chez les eucaryotes, et présentons divers cas de transferts. Enfin, le chapitre se termine en faisant un état des lieux de ce phénomène dans le royaume des *Fungi*.

Résultats

Amélioration de l'assemblage de *Penicillium roqueforti* : Développement d'une méthodologie innovante pour améliorer les assemblages génomiques.

Ce chapitre rapporte la conception et le développement d'une approche d'amélioration des assemblages génomiques par peignage moléculaire et son application au génome de *Penicillium roqueforti* FM164. L'approche repose sur l'utilisation de la génomique comparative entre les différents assemblages d'un même génome ou de génomes d'espèces proches pour émettre un maximum d'hypothèse concernant la localisation possible des scaffolds issus de l'assemblage bioinformatique initial, suivi d'un premier test rapide par PCR long range pour la validation de ces hypothèses, et enfin d'une étape de validation par peignage moléculaire pour les scaffolds non positionnés à l'issue de la première validation.

Ce chapitre présente d'une part le développement d'un protocole d'extraction robuste permettant de peigner l'ADN des champignons filamenteux du genre *Penicillium*, la conception d'une méthodologie innovante d'amélioration des assemblages en combinant génomique comparative, PCR long-range et peignage moléculaire, ainsi que quatre exemples de solution que le peignage moléculaire a apporté aux problématiques d'assemblage de génomes. Ces quatre exemples sont :

- la résolution de trous (ou « gaps ») de séquençages dans les séquences de références, ici en s'appuyant sur le génome humain et la région de BRCA1.

- Le positionnement de deux scaffolds séparés par un très long gap dans le séquençage (ou « long distance scaffolding »). Ce type de situation peut typiquement intervenir au niveau des centromères chez les mammifères par exemple.

- la mise en évidence et la résolution d'erreurs dans les assemblages bioinformatiques.

- Le positionnement de plusieurs scaffolds en une seule expérience de peignage moléculaire, nécessaire pour l'utilisation de la méthode sur de grands génomes de plusieurs mégabases.

Enfin, le résultat de l'application de la méthode développée durant cette thèse est discuté par l'obtention d'un assemblage de haute qualité pour le génome de *Penicillium roqueforti*, regroupant plus de 92% du génome en 6 super-scaffolds de tailles chromosomales.

Transfert latéral de matériel génétique dans la chaîne alimentaire : Flux de gènes dans le fromage.

Ce chapitre rapporte l'occurrence multiple et récurrente de récents transferts horizontaux d'un large ilot génomique entre plusieurs espèces de *Penicillium* impliquées dans la fabrication du fromage. Ce chapitre, présent ici sous la forme d'une version soumise au journal *Nature communications* a été accepté et publié depuis. Le transfert horizontal de gène peut être défini comme étant la transmission non sexuelle d'information génétique entre deux organismes. Ce phénomène de transfert horizontal est largement décrit et accepté chez les procaryotes et est reconnu comme étant fréquent et ayant un impact très important sur l'écologie et l'évolution de ces organismes. Si des exemples de transferts horizontaux ont aussi été décrit chez les eucaryotes, et ce avec divers impacts fonctionnels ou écologiques, leur fréquence, étendue et impact est encore ouvertement débattue. Ils sont en effet fréquemment rapporté comme étant des événements anciens, sporadiques et assez rare, et les revues les plus récentes rapportent que "ce flux de gène représente une forme de variation génétique donc les implications sont encore mal perçues" (Syvanen, 2012).

Les résultats présentés dans ce chapitre indiquent qu'un vaste ilot génomique a été horizontalement transféré plusieurs fois, entre plusieurs espèces de *Penicillium* appartenant à l'environnement fromager. Les espèces de *Penicillium* sont des champignons filamenteux ubiquitaires important pour la production de nombreux fromages.

Comparé aux autres cas décrits de transferts horizontaux chez les eucaryotes, nos résultats démontrent :

- le transfert d'un très grand ilot génomique (0.5 mégabase, 250 gènes).
- son insertion dans le chromosome de l'espèce réceptrice (par opposition aux transferts de chromosomes entiers déjà décrits chez certains champignons).
- Son occurrence répétitive dans une fenêtre de temps réduite et récente (il s'agit de séquences identiques retrouvées dans plusieurs espèces différentes.)
- Son probable déclenchement imputable à l'activité humaine (les transferts ont eu lieu dans l'environnement fromager).

Ceci a des implications évidentes pour l'évaluation de la sécurité alimentaire dans l'environnement fromager, mais aussi dans d'autres environnements dus à l'activité humaines, comme les champs céréaliers. Ce cas de transfert supporte aussi l'idée que la fréquence et l'impact des transferts

horizontaux eucaryotes est sous-évaluée, et cette étude pourrait aider à accélérer une refonte de la perception actuelle de l'importance du transfert horizontal de gène chez les eucaryotes, tant dans des environnements anthropomorphiques que dans la nature.

En effet, alors que l'importance du transfert latéral de matériel génétique a été discuté dans le chapitre quatre, la découverte de transferts multiple et récent dans l'environnement fromager rajoute encore à la vision émergente d'un phénomène sous étudié et sous-évalué chez les champignons. Ce travail souligne l'importance de rechercher ce type d'événement dans les procédés alimentaires, et plus encore lorsque combiné aux autres rapports de transferts horizontaux dans la chaîne alimentaire, comme par exemple, le transfert de gènes entre levures de l'environnement viticole. Les transferts horizontaux sont des événements perçus comme étant intrinsèquement liés au processus d'adaptation à de nouveaux environnements ou "niche shift". Ce phénomène a donc de nombreuses implications pour l'industrie alimentaire, tant en terme d'évaluation de l'innocuité des produits et microorganismes, qu'en terme d'innovation. L'utilisation des champignons pour produire de nouvelles saveurs ou pour modifier des textures et des apparences pourrait ainsi mettre en contact des espèces précédemment isolées écologiquement, et donner lieu à des échanges génétiques non anticipés entre différentes souches.

De plus, dans une ère de mondialisation et de commerce global, comme indiqué dans le chapitre trois, de nouvelles souches d'espèces similaires ou différentes sont de plus en plus susceptibles de se retrouver en contact, et pourraient ainsi éventuellement échanger du matériel génétique, à la manière des transferts décrits dans ce chapitre...

Evaluation du potentiel génomique pour la production de métabolites secondaires par les espèces de *Penicillium* et leur impact sur la sécurité alimentaire : résultats préliminaires.

Ce chapitre présente un travail préliminaire conduit à la fin de la thèse. Il décrit les premiers résultats sur la capacité des champignons du genre *Penicillium* à produire des métabolites secondaires. Cette étude utilise les génomes générés durant le projet Food Microbiomes et a donc une emphase particulière sur les espèces présentes dans les environnements alimentaires. Cette étude inclut aussi d'autres génomes précédemment publiés, notamment *Penicillium chrysogenum*, utilisé par l'industrie pharmaceutique pour produire de la pénicilline et d'autres composés, et deux souches de *Penicillium digitatum*, un important pathogène du citronnier et des orangers.

Les métabolites secondaires sont de petites molécules produites par de nombreux organismes, comme certaines bactéries, plantes et champignons. Les *fungi* produisent beaucoup de ces composés non essentiels pour leur croissance en laboratoire. Certains de ces composés ont été retrouvés dans l'air, le sol, les champs céréaliers, mais aussi dans la chaîne alimentaire, comme par exemple dans le fromage. Ces métabolites secondaires jouent dans la nature un rôle important pour les interactions entre les champignons et leurs écosystèmes en prenant le rôle de molécule signal, d'antibiotiques ou d'antifongiques. Ces molécules revêtent un aspect particulier pour l'homme, de deux manières : certains sont toxiques, tandis que d'autres peuvent être utilisés pour des traitements pharmaceutiques, en tant qu'antibiotique, antifongique, ou anti tumeur. La plupart des métabolites secondaires ne sont pas connus ou pas caractérisés, et représentent ainsi un vaste réservoir inexploité d'innovation biotechnologique. De surcroît, les recherches sur ces composés sont relativement récentes, datant de la fin des années 1940, avec la découverte de l'impact de la pénicilline sur la santé humaine. Depuis, l'Homme les utilise en tant qu'antibiotique, antifongique, suppresseur de tumeurs, pesticides, anti cholestérol... Les métabolites secondaires toxiques produits par les champignons sont communément appelés mycotoxines, et de nombreux exemples sont connus, tels l'aflatoxine, la fumonisine, la zearalone, etc. Certains ont des effets neurotoxiques, carcinogéniques ou néphrotoxiques. Malgré ces effets détritimaux pour la santé humaine, les connaissances concernant les mycotoxines dans l'environnement alimentaire restent faibles, et sont essentiellement due à des expériences biochimiques pour évaluer la production de mycotoxine par une espèce. L'apport de la génomique permet de mieux caractériser le potentiel de production des métabolites secondaires de chaque organisme en détectant les clusters de gènes responsable de leur production. A l'heure actuelle, seuls les *Aspergillus* et les *Fusarium* ont fait l'objet de telles études.

Les *Penicillia* sont connus pour la production de certains de ces métabolites. Ainsi *P. carneum* produit de la patuline, *P. roqueforti* la toxine PR, et *P. paneum* et *carneum* produisent l'andrastine A et la roquefortine C. Si ces métabolites ne sont pas forcément produits dans les conditions de cultures expérimentales, il se peut qu'ils soient produits lors de l'utilisation de ces champignons dans les procédés de production alimentaire, ou lors de contamination. Ceci est en partie responsable de la non obtention du statut de présomption d'innocuité (QPS – Qualified Presumption of Safety) de l'EFSA.

Les résultats présentés dans ce chapitre sont issus de l'identification bioinformatique des clusters de gènes susceptibles de produire des métabolites secondaires. Ainsi, selon les espèces, entre 24 et 58 clusters ont été détectés par génome. Une recherche d'orthologie a aussi montré clairement que chaque espèce possède de nombreux clusters de gènes codant pour des métabolites secondaires uniques. Le potentiel de production en métabolites secondaire de ces espèces est donc très élevé, et représente donc potentiellement un obstacle à leur utilisation dans l'alimentation, mais aussi potentiellement de nouvelles molécules aux propriétés intéressantes d'un point de vue agroalimentaire ou biotechnologique. Une identification des métabolites effectivement produit, tant en recherchant des similitudes avec les métabolites déjà décrit dans la littérature et les bases de données, que par des expériences de biologie moléculaire et de biochimie reste à effectuer pour caractériser finement le répertoire en métabolique secondaire de chaque organisme. L'identification des gènes codant pour ces métabolites facilitera ces expériences futures.

Discussion

Ce chapitre résume les principaux résultats de cette thèse et présente des perspectives de développement éventuelles. Nous discutons ici aussi des retombés potentielles du travail effectué pendant cette thèse. Nous nous concentrons ici sur les deux aspects principaux de la thèse, présentés dans les chapitres 5 et 6. Le chapitre commence par une discussion sur le développement d'une nouvelle méthodologie pour améliorer les assemblages en utilisant le peignage moléculaire, et discute ensuite les résultats obtenus lors de son application au génome de *Penicillium roqueforti*. Des perspectives possibles (ainsi que leur implémentation) concernant son amélioration et son développement sont proposées, notamment sur l'intégration de logiciels de génomique comparative et d'assemblage déjà existant au procédé développé ici. Dans un second temps, nous revenons brièvement sur les résultats du chapitre 6, la découverte de transferts récurrents, fréquents et récents d'un très grand îlot génomique dans un environnement alimentaire. Nous discutons ensuite de l'importance d'une meilleure évaluation de la fréquence de ce type d'événement, non seulement dans les environnements créés par l'homme, mais aussi dans la nature. L'impact potentiel de ce type de transferts sur les écosystèmes est aussi présenté et discuté, en prenant en compte les interactions qu'ont les fungi avec les autres organismes présents dans leurs écosystèmes.