



Année 2013

N°

UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE
ECOLE DOCTORALE SCIENCES TECHNOLOGIE SANTE
THESE

Présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE

Discipline : Biologie-Biophysique

Soutenue publiquement le 02/12/2013

Par

Aurélie LECELLIER

Née le 18 mars 1983 à Issy les Moulineaux

Titre :

**Caractérisation et identification des champignons
filamenteux par spectroscopie vibrationnelle**

JURY

Président :	Dr. Jérôme MOUNIER (Brest, France)
Rapporteurs :	Pr. Boualem SENDID (Lille, France) Pr. Olivier SIRE (Vannes, France)
Examineurs :	Dr. Wilfried ABLAIN (Rennes, France) Dr. Caroline AMIEL (Caen, France) Pr. Michel MANFAIT (Reims, France)
Directeurs de thèse :	Pr. Ganesh SOCKALINGUM (Reims, France) Dr. Dominique TOUBAS (Reims, France)

« Le rôle de l'infiniment petit est infiniment grand. »

Louis Pasteur

Remerciements

A Messieurs le Professeur Michel Manfait et le Professeur Olivier Piot,

Je vous remercie sincèrement pour m'avoir accueillie et pour m'avoir permis de réaliser ce travail au sein de votre unité que vous avez dirigée successivement lors de ces trois années de thèse. Je vous suis très reconnaissante pour m'avoir donné l'occasion de présenter mon travail dans des congrès nationaux et internationaux.

A Monsieur le Professeur Boualem Sendid,

Je vous suis très reconnaissante d'avoir accepté d'être rapporteur de cette thèse, je vous remercie pour votre participation au Jury de soutenance et pour l'intérêt que vous avez porté à mon travail.

A Monsieur le Professeur Olivier Sire,

Je vous suis très reconnaissante d'avoir accepté d'être rapporteur de cette thèse, je vous remercie pour votre participation au Jury de soutenance et pour l'intérêt que vous avez porté à mon travail.

A Monsieur le Docteur Wilfried Ablain,

Je vous remercie sincèrement d'avoir accepté de faire partie du Jury de soutenance. Je vous remercie également pour votre contribution à l'élaboration de ce travail.

A Madame le Docteur Caroline Amiel,

Je vous exprime toute ma reconnaissance de m'avoir initié à la spectroscopie vibrationnelle lors de mon passage au sein de votre équipe pour y effectuer mon stage de Master 2 ainsi que pour avoir guidé mes premiers pas dans le monde de la recherche. Je vous remercie pour votre soutien et vos encouragements que vous m'avez apporté. Je n'oublie pas que c'est grâce à vous si j'ai pu réaliser cette thèse. Cela me fait extrêmement plaisir que vous ayez accepté d'examiner mon travail et de me faire l'honneur de participer au Jury de soutenance.

A Monsieur le Professeur Michel Manfait,

Je vous adresse mes remerciements les plus respectueux pour avoir accepté d'être membre du Jury de soutenance. Je vous remercie pour l'intérêt que vous avez porté à ce travail, ainsi que pour votre sympathie.

A Monsieur le Docteur Jérôme Mounier,

Je tiens à t'exprimer toute ma gratitude pour l'aide précieuse que tu m'as apporté tout au long de cette thèse et ainsi que pour ta disponibilité malgré la distance. Je te remercie pour ton implication dans ce projet, pour tes conseils avisés et ta sympathie à mon égard. Enfin, je te remercie énormément pour m'avoir fait l'honneur de participer au Jury de soutenance.

A Monsieur le Professeur Ganesh Sockalingum,

Je vous suis sincèrement reconnaissante d'avoir encadré ce travail de thèse. Je vous remercie pour la confiance que vous m'avez accordée pendant ces trois années et pour avoir cru en mes capacités tout au long de ce travail. Je vous remercie pour tous vos conseils stimulants, votre patience, ainsi que pour la gentillesse que vous avez manifestée à mon égard. J'ai beaucoup apprécié travailler à vos côtés tant sur le plan scientifique que sur le plan humain. Je garde toujours beaucoup de plaisir à discuter avec vous et à bénéficier de votre expérience.

A Madame le Docteur Dominique Toubas,

Je vous adresse mes plus sincères remerciements pour votre encadrement durant cette thèse. Je suis ravie d'avoir travaillé en votre compagnie tout au long de ces années durant lesquelles j'ai pu apprécier vos qualités tant pédagogiques et scientifiques qu'humaines. Je vous remercie également pour votre disponibilité, ainsi que pour votre aide et vos précieux conseils que vous m'avez apportés lors de l'élaboration de cette thèse. Pour tout ce que vous m'avez appris, je vous remercie très sincèrement.

A Monsieur le Docteur Vincent Gaydou,

Je te remercie sincèrement pour toute l'aide que tu m'as apportée durant ma dernière année de thèse. L'aboutissement de ce travail est le fruit de la complémentarité de nos compétences respectives. Je te remercie pour ton soutien et tes encouragements ainsi que pour ta patience.

A Monsieur le Docteur Jayakrupakar Nalalla,

Nuvvu naku chesina prathi daniki am forever gratefull to you. Nenu neku eppatiki runapadiuntanu. Nevichina encouragement, advices ki chala thanks. Nrnnu opikaga vinnaduku neku ela thanks cheppalo ardam avvatledu. Nuvvu naa work environment ni fun environment ga marchavu with your humor. Mana madhyalo oka viluvaina sambandham erpadindi. Mana projects aipoiundochu kani mana friendship will last forever.

Je tiens à remercier également :

Stéphane Huet et Nadia Leden pour leur collaboration au sein de ce projet. Les discussions que nous avons pu avoir lors des réunions de projets m'ont beaucoup apporté.

L'équipe Médian et plus particulièrement : Céline, Christophe, Cyril, David, Denise Pisani, Goutam, Imane, Irène, Julien, Joan, Laurence, Mathieu, Mickaël, Mireille Cousina, Mohammed, Nathalie, Qué et Valérie.

Georges Barbier, Loïc Castrec, Marie-Anne Le Bras, Valérie Vasseur, Amélie Weill et Olivia Le Bourhis de l'équipe du Lubem pour leur contribution à l'élaboration de ce travail.

Un grand merci à tous mes amis : Caroline, Georges, Hadrien, Jean-Claude, Lucie, Mai-Ahn, Marie, Mathilde, Sunny, Teddy et The Thuong, pour tous les bons moments que l'on a passé ensemble. Ces trois années passées à Reims n'auraient pas été les mêmes sans votre présence, votre sympathie et votre bonne humeur.

Je remercie respectueusement mes parents, Catherine et Didier, pour leur soutien sans faille et leurs encouragements, non seulement au cours de ces trois années mais aussi tout au long de mon parcours, et sans lesquels je n'en serais pas là aujourd'hui. Vous m'avez donné toutes les chances pour réussir et c'est grâce à vous que je suis ce que je suis maintenant. Je remercie sincèrement ma sœur Aurore pour son aide précieuse et pour avoir toujours été à mes cotés durant ces années. Enfin, je remercie Arnaud, mon mari, pour son soutien indéfectible, et pour avoir fait preuve de compréhension et d'écoute malgré l'éloignement. Merci à Cannelle pour sa tendresse et son réconfort. Je ne pourrais vous remercier assez...

Ces remerciements ne peuvent s'achever sans une pensée pour Alexandre Mazine.

Je dédie cette thèse à Claude Grimoult.

Sommaire

Liste des abréviations	7
Liste des figures et des tableaux	9
Avant-propos	11
Chapitre I : Introduction	14
I.1- Les champignons filamenteux	15
I.1.a- Définition	15
I.1.b- Caractéristiques morphologiques des champignons filamenteux	15
I.1.c- Développement des champignons filamenteux	17
I.1.d- Phylogénie des champignons filamenteux	19
I.1.e- Habitat et conditions de développement	21
I.1.f- Problèmes liés aux champignons filamenteux et à leurs métabolites	22
- secteur agroalimentaire	22
- secteur pharmaceutique et médical	24
I.1.g- Méthodes d'identification des champignons filamenteux	26
- Analyse macroscopique	26
- Analyse microscopique	26
- Analyse moléculaire	27
I.2- La spectroscopie infrarouge à transformée de Fourier (IRTF)	29
I.2.a- Principe	29
- Rayonnement infrarouge	29
- Interaction rayonnement-matière	31
- Absorption du rayonnement IR par l'échantillon	33
- Modes de vibration	34
- Instrumentation	35
- Résolution spectrale	38

- Avantages de la spectroscopie IRTF.....	38
I.2.b- Application de la spectroscopie IRTF en microbiologie.....	39
Chapitre II : Méthodologie.....	44
II.1- Préparation des échantillons de champignons filamenteux.....	45
II.1.a- Souches de champignons filamenteux utilisées.....	45
II.1.b- Conservation des souches de champignons filamenteux.....	45
II.1.c- Mise en culture des champignons filamenteux à partir de cryobilles et d’implants	46
II.1.d- Extraction d’ADN, amplification, séquençage et assignement taxonomique des souches de champignons filamenteux.....	48
II.1.e- Préparation des cultures liquides des champignons filamenteux.....	49
II.1.f- Vérification de la pureté des cultures de mycélium.....	49
II.1.g- Préparation des suspensions de mycélium.....	49
II.2- Analyse spectrale des échantillons de champignons filamenteux.....	51
II.2.a- Acquisition spectrale.....	51
II.2.b- Test qualité et prétraitements des spectres infrarouge.....	52
- Test qualité.....	52
- Prétraitements des spectres infrarouge.....	53
II.3- Analyses statistiques des données spectrales.....	57
II.3.a- Analyse discriminante des moindres carrés partiels : PLS-DA.....	57
II.3.b- Validation croisée.....	60
II.4- Etablissement d’un score et d’un seuil de validation.....	62
II.5- Etablissement d’une fonction de standardisation.....	63
Chapitre III : Résultats et discussion.....	65
III.1- Préambule.....	66
III.2- Article 1.....	67

- Préambule à l'article 1	68
Differentiation and identification of filamentous fungi by high-throughput FTIR spectroscopic analysis of mycelia.....	70
III.3- Article 2.....	98
- Préambule à l'article 2	99
Implementation of an FTIR spectral library of 486 filamentous fungi strains for rapid identification of molds.....	102
Chapitre IV : Travaux supplémentaires	129
IV.1.Article 3.....	130
- Préambule à l'article 3	131
Chapitre V : Conclusion générale.....	170
Références bibliographiques	174
Publications et communications.....	182
Annexes.....	185

Liste des abréviations

°C : Degrés Celsius

ADN : Acide désoxyribonucléique

AFLP : Amplified Fragment Length Polymorphism

ANN : Artificial Neural Network

ARN : Acide ribonucléique

a_w : activité de l'eau

BLAST : Basic Local Alignment Search Tool

CBS : Centraalbureau voor Schimmelcultures

CO₂ : Dioxyde de carbone

COV : Composés organiques volatils

D : Distance Euclidienne

DNA : Deoxyribonucleic acid

DO : Densité Optique

FAO : Food and agriculture organization

FDA : Factorial Discriminant Analysis

FTIR : Fourier transform infrared

h : homogénéité

H₂O : Monoxyde de dihydrogène

HCA : Hierarchical Cluster Analysis

HCl : Chlorure d'hydrogène

HeNe : Hélium-néon

HR : Humidité Relative

HTS-Xt : High-Throughput Screening eXTension

IN : Iteration number

Inst1 : Instrument 1

Inst2 : Instrument 2

IR : Infrarouge

IRTF : Infrarouge à transformée de Fourier

ITS : Internal Transcribed Spacer

KNN : k-Nearest Neighbor

LDA : Linear Discriminant Analysis

MAFFT : Multiple Alignment using Fast Fourier Transform

MALDI-TOF MS : Matrix-assisted laser desorption ionization time of flight mass spectroscopy

MCM : Mini Chromosome Maintenance

McN : MacNemar
MEGA : Molecular evolutionary genetics analysis
MIC : Microbiologically Influenced Corrosion
M-tube : Molecular tube
NaCl : Chlorure de sodium
NCBI : National Center for Biotechnology Information
O₂ : Dioxygène
PC : Principal Component
PCA : Principal Component Analysis
PCR : Polymerase Chain Reaction
PDA : Potato dextrose agar
PDB : Potato Dextrose Broth
PGP : Percentage of Good Prediction
pH : Potentiel Hydrogène
PLS : Partial Least Square
PLS-DA : Partial Least Square Discriminant Analysis
PNN : Probabilistic Neural Network
QDA : Quadratic Discriminant Analysis
QT : Quality test
RAPD : Random Amplified Polymorphic DNA
RBF : Radial Basis Function
RFLP : Restriction Fragment Length Polymorphisms
S : Score
SF : Fonction de standardisation
SiC : Carbure de silicium
SIMCA : Soft Independent Modeling of Class Analogy
SRB : Sulfate-reducing bacteria
SVM : Support Vector Machine
TEF : Translation elongation factor
TRB : Thiosulfate-reducing bacteria
TSR : Twenty SrRNA accumulation
UBOCC : Université de Bretagne Occidentale
YM : Yeast Malt

Liste des figures et des tableaux

Figures :

<i>Figure 1 : Structure de l'hyphe. (A) hyphe coenocytique, (B) hyphe cloisonné.</i>	16
<i>Figure 2 : Schématisation de la structure de la paroi fongique.</i>	16
<i>Figure 3 : Les différents modes de sporulation et les différents types de spores associées.....</i>	18
<i>Figure 4 : Schématisation de la reproduction asexuée et sexuée d'une moisissure.</i>	19
<i>Figure 5 : Phylogénie et classification des Fungi.....</i>	20
<i>Figure 6 : Quelques exemples d'aliments contaminés par des moisissures.</i>	23
<i>Figure 7 : Organes de fructifications du genre Aspergillus (A) et du genre Rhizopus (B).</i>	27
<i>Figure 8 : Représentation d'une onde électromagnétique.....</i>	29
<i>Figure 9 : Spectre électromagnétique.....</i>	30
<i>Figure 10 : Barycentres (G) de la molécule H₂O et de la molécule CO₂ (: charge).</i>	31
<i>Figure 11 : Moment dipolaire de la molécule HCl et de la molécule O₂.</i>	32
<i>Figure 12 : Schéma du phénomène d'absorption et transition entre deux niveaux vibrationnels.....</i>	32
<i>Figure 13 : Différents modes fondamentaux de vibration d'une molécule triatomique non-linéaire.</i>	35
<i>Figure 14 : Schéma d'un interféromètre de Michelson.</i>	36
<i>Figure 15 : (A) Représentation de la différence du trajet par deux ondes de même longueur d'onde. (B) Représentation d'une interférence constructives où $\Delta L = k \lambda$. (C) Représentation d'une interférence destructive où $\Delta L = (k + 1/2) \lambda$.</i>	37
<i>Figure 16 : Interférogramme (: position du miroir).</i>	37
<i>Figure 17 : Spectres d'absorption IR caractéristiques d'un champignon filamenteux (Aspergillus flavus) et sa dérivée seconde inverse.....</i>	40
<i>Figure 18 : Schématisation du protocole de conservation des champignons filamenteux par cryoconservation à l'aide de cryobilles.</i>	46
<i>Figure 19 : Schématisation du protocole de conservation des champignons filamenteux par cryoconservation à partir d'implants de gélose.....</i>	46
<i>Figure 20 : Colonies de moisissures obtenues sur tubes de gélose Sabouraud inclinée à partir d'une cryobille (à gauche) ou d'un implant de gélose (à droite).....</i>	47
<i>Figure 21 : Schématisation du protocole de préparation des suspensions de mycélium.</i>	50

Figure 22 : Schématisation du protocole de l'analyse spectrale des souches de champignons filamenteux. 51

Figure 23 : Récapitulatif des gammes spectrales et des valeurs utilisées pour les tests qualité. 53

Figure 24 : Illustration des prétraitements appliqués aux spectres infrarouge. 56

Figure 25 : Schématisation du principe de la PLS-DA : partial least square discriminant analysis (A : calibration, B : validation). 59

Figure 26 : Schématisation de la régression PLS (A : calibration). 60

Figure 27 : Choix du nombre de dimensions utilisé par la méthode de validation croisée interne ou externe (B : validation). 61

Tableaux :

Tableau 1 : Fréquences d'absorption observées dans les spectres infrarouges des microorganismes et leur attribution biomoléculaire. 39

Avant-propos

Ce travail de thèse s'inscrit dans le cadre du projet MOLDID pour « **MOLDS** **ID**entification » et est intitulé « Caractérisation et identification des champignons filamenteux par spectroscopie vibrationnelle ». Ce projet, labellisé par le « Pôle de Compétitivité VALORIAL : l'Aliment de Demain », est porté par la région Bretagne et la région Champagne Ardenne. Cette labellisation a pour objectifs de développer coopération, alliance et synergie entre entreprises et centres de recherche. Il s'agit d'un partenariat, permettant de regrouper des compétences complémentaires, entre un établissement industriel spécialisé dans l'analyse microbiologique (AES Chemunex/Biomérieux, Région Bretagne), un laboratoire spécialisé dans la recherche en biophotonique et biophysique (MÉDIAN, Région Champagne-Ardenne) et un laboratoire spécialisé dans la recherche en microbiologie agroalimentaire et environnementale (Lubem, Région Bretagne).

Les champignons filamenteux sont des microorganismes ubiquitaires dont le rôle principal est le recyclage de la matière organique. Leur diversité est très importante (environ 1,5 millions d'espèces dans le monde entier). Certains d'entre eux peuvent être utiles à l'homme alors que d'autres s'avèrent très dangereux. En effet, ils peuvent être utilisés pour la production d'aliments et la synthèse de médicaments. Cependant, la contamination par les moisissures représente un réel problème dans le secteur agroalimentaire, dans les industries pharmaceutiques et cosmétologiques ainsi que dans le domaine de la santé publique. Un des problèmes majeurs dans le secteur agro-alimentaire est la contamination, souvent constatée à la surface, des produits destinés à l'alimentation, et en particulier des denrées stockées, entraînant des pertes économiques importantes. La production de mycotoxines par les champignons filamenteux est la principale préoccupation. Au sein du secteur médical, les infections fongiques ont augmenté de façon considérable au cours de ces vingt dernières années, notamment chez les patients immunodéprimés, et sont responsables d'un taux de mortalité élevé. Cette émergence s'explique en partie par le développement de nouvelles thérapeutiques entraînant une immunodépression profonde et prolongée chez les patients atteints de pathologies sévères (hémopathies,...) les exposant au risque d'infection fongique invasive.

Actuellement, les méthodes d'identification des champignons filamenteux utilisées en routine sont essentiellement basées sur l'analyse des caractéristiques morphologiques macroscopiques et microscopiques. Ces méthodes sont souvent longues et laborieuses, manquent de précision et d'objectivité, et requièrent une bonne connaissance du domaine de la mycologie en raison de la grande diversité des souches de champignons filamenteux. Depuis plusieurs années, les techniques de biologie moléculaire se sont considérablement développées en tant qu'outil complémentaire d'identification des champignons filamenteux. Ces nouvelles méthodes basées sur le séquençage moléculaire sont coûteuses et complexes à mettre en œuvre. Plus récemment, une méthode basée sur la spectrométrie de masse (MALDI-TOF) est apparue comme une technique alternative d'identification des champignons filamenteux. Cette méthode présente certains avantages comme la rapidité, la fiabilité et son faible coût en ce qui concerne les consommables. Cependant, l'investissement en termes d'équipement est élevé. De plus, cette technique, reposant sur l'identification à partir de banques de données spectrales et ayant déjà fait ses preuves pour l'identification des bactéries et des levures, est en cours d'optimisation en ce qui concerne les moisissures. Du fait de l'absence d'outils simples et rapide à mettre en œuvre pour l'identification des champignons filamenteux et adaptés aux besoins des industriels, il s'en suit une mauvaise appréciation des contaminations liées aux moisissures générant un risque non maîtrisé.

Dans ce contexte, il existe un intérêt majeur à développer d'autres méthodes pour l'identification des champignons filamenteux. Dans ce travail, une nouvelle méthode biophysique, basée sur la spectroscopie vibrationnelle infrarouge à transformée de Fourier (IRTF), est proposée. Cette approche biophotonique, basée sur l'interaction onde-matière, permet la caractérisation des variations spectrales liées aux modifications des différents constituants moléculaires. Cette méthode consiste en l'absorption d'un rayonnement infrarouge par un échantillon et en l'analyse des différents modes de vibration fondamentaux des liaisons moléculaires présentes au sein de l'échantillon. Le résultat obtenu est un spectre qui indique l'absorbance en fonction de la longueur d'onde ou du nombre d'onde. Cette technique est rapide, reproductible, sensible et présente une haute résolution spectrale.

Le premier objectif de cette étude est de développer un protocole simple, rapide et standardisé pour l'analyse des champignons filamenteux par spectroscopie IRTF à haut débit. Les souches de notre collection retenues pour cette étude font partie en majorité de la collection du laboratoire du Lubem de l'Université "Bretagne Occidentale", cette collection

ayant été complétée par l'achat de souches CBS. Le deuxième objectif est de construire une bibliothèque de données spectrales en utilisant une méthode d'analyse statistique supervisée pour le traitement des données IR. Le troisième objectif est d'évaluer le potentiel de la spectroscopie IRTF couplée à une méthode d'analyse statistique comme technique d'analyse pour l'identification des champignons filamenteux. Pour ce faire, la validation de la banque de données spectrales est envisagée en établissant un score et seuil permettant de valider les résultats de prédiction et en réalisant une étude de transférabilité de la méthode sur d'autres appareils IRTF.

Ce manuscrit est présenté en cinq parties.

Le chapitre 1 comporte des données générales concernant les champignons filamenteux, le principe de la spectroscopie infrarouge à transformée de Fourier ainsi que les différentes applications de cette méthode en microbiologie.

Le chapitre 2 présente la méthodologie développée pour cette étude. Il décrit les différentes étapes de la préparation des échantillons de champignons filamenteux et de l'analyse spectrale des différents échantillons obtenus. Ce chapitre décrit également la méthodologie développée pour l'analyse statistique des données spectrales, pour l'établissement d'un score et d'un seuil de validation des résultats de prédiction, ainsi que pour l'application d'une fonction de standardisation.

Le chapitre 3 présente les résultats obtenus au cours de ce travail sous forme de deux articles scientifiques soumis dans des journaux internationaux.

Le chapitre 4 porte sur une étude complémentaire visant à comparer et à évaluer le potentiel de différentes méthodes chimiométriques supervisées linéaires et non linéaires, couplées à la spectroscopie IRTF, dans la discrimination et l'identification des champignons filamenteux.

Le chapitre 5 décrit les conclusions les plus importantes, ainsi que les différentes perspectives qu'apporte ce travail tant au niveau de la recherche qu'au niveau des applications industrielles et médicales.

Chapitre I : Introduction

I.1- Les champignons filamenteux

I.1.a- Définition

Les micromycètes sont des champignons microscopiques regroupant les levures et les champignons filamenteux. Ce sont des microorganismes eucaryotes caractérisés par la présence d'une membrane nucléaire et de mitochondries.

Ils sont ubiquitaires et très répandus dans la nature, notamment au niveau des végétaux en décomposition.

Les champignons filamenteux sont hétérotrophes, et plus particulièrement absorbotrophes puisqu'ils absorbent les éléments, digérés de manière extracellulaire, au travers de leur appareil végétatif présentant une perméabilité pariétale. Les champignons filamenteux ne peuvent synthétiser de matière organique à partir du gaz carbonique atmosphérique. En effet, ils sont incapables d'assurer la photosynthèse. Une source de carbone organique est donc nécessaire à leur développement. Ils synthétisent leurs propres nutriments à partir de l'eau et des éléments nutritifs et minéraux qu'ils puisent dans leur environnement. Il joue un rôle important dans le recyclage des matières organiques en puisant leur énergie à partir de ces sources carbonées externes.

I.1.b- Caractéristiques morphologiques des champignons filamenteux

Les champignons filamenteux sont composé d'un appareil végétatif appelé thalle. Il est composé de filaments ou hyphes enchevêtrés les uns par rapport aux autres, et l'ensemble des hyphes constituent un réseau appelé mycélium. Les hyphes sont diffus, tubulaires et fins avec un diamètre compris entre 2 et 15 μm et sont plus ou moins ramifiés. Chez certaines moisissures, comme par exemple *Mucor*, les cellules ne sont pas séparées par une cloison transversale, le thalle est alors dit coenocytique ou « siphonné » alors que chez d'autres, comme par exemple *Aspergillus*, le thalle est cloisonné ou « septé » (Figure 1). Les cloisons, appelées septa possèdent des perforations assurant la communication entre les cellules. Les caractéristiques morphologiques de ces microorganismes sont liées à leur substrat nutritif. La colonisation du substrat est réalisée par extension et ramification des hyphes.

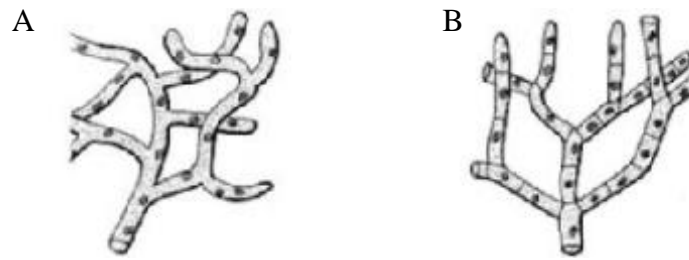


Figure 1 : Structure de l'hyphe. (A) hyphe coenocytique, (B) hyphe cloisonné.

Les champignons filamenteux possèdent une paroi constituée essentiellement de polysaccharides, de glycoprotéines et de mannoprotéines (Figure 2). Les polysaccharides sont majoritairement la chitine, polymère de molécules de N-acétylglucosamine liées entre elles par une liaison de type β -1,4, et les glucanes, polymères de molécules de D-glucose liées entre elles par des liaisons α -1,4. Ces deux polysaccharides assurent la protection des moisissures vis-à-vis des agressions du milieu extérieur. La chitine joue un rôle dans la rigidité de la paroi cellulaire, les glycoprotéines jouent un rôle dans l'adhérence et les mannoprotéines forment une matrice autour de la paroi.

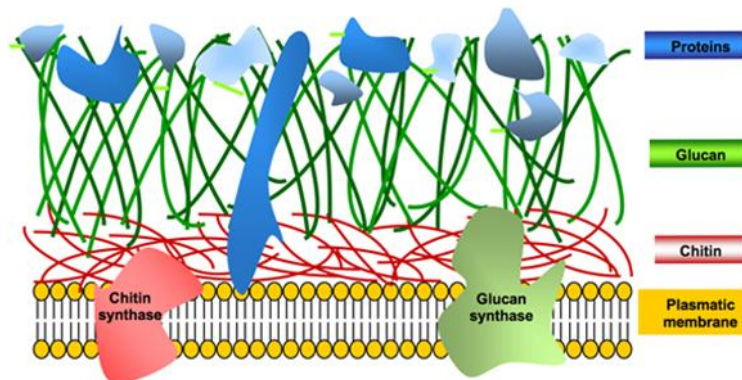


Figure 2 : Schématisation de la structure de la paroi fongique.

I.1.c- Développement des champignons filamenteux

Le développement des moisissures comprend deux phases : une phase végétative et une phase reproductive.

Pendant la phase végétative, qui correspond à la phase de croissance, l'appareil végétatif colonise le substrat par extension et ramification des hyphes. Il existe deux types de ramifications, la ramification par dichotomie (apex) ou par bourgeonnement (latéral). Cette phase correspond également à la phase de nutrition, les hyphes absorbant à travers leur paroi, l'eau ainsi que les éléments nutritifs contenus au sein du substrat tout en dégradant le substrat par émission d'enzymes et d'acides. La forme mycélienne en expansion, qui constitue une phase active de développement, est responsable de la dégradation et de l'altération du substrat.

La phase reproductive comprend deux types de reproduction : la reproduction asexuée, correspondant à la forme anamorphe, et la reproduction sexuée, correspondant à la forme téléomorphe.

La reproduction asexuée se fait sans fusion de gamètes. Elle correspond majoritairement à la dispersion de spores asexuées, permettant la propagation des moisissures afin de coloniser d'autres substrats. Cette forme de reproduction asexuée est appelée la sporulation. Au cours de la sporulation, ces spores, petites cellules déshydratées au métabolisme réduit et entourées d'une paroi protectrice les isolant du milieu environnant, sont produites en grande quantité par des structures spécialisées développées à partir du mycélium. Leur diamètre varie de 2 à 250 μm . Il existe différentes formes de reproduction asexuée et différents types de spores (Figure 3). Les spores peuvent être le résultat de la fragmentation. Dans ce cas, un nouvel organisme se développe à partir d'un fragment parent de mycélium (arthrospores). Les spores peuvent aussi être produites de manière endogène à l'intérieur du sporocyste (sporocystiospores), ou de manière exogène en continu à l'extrémité des structures spécialisées appelées phialides (conidiospores). Ensuite, les spores se détachent du mycélium sous l'effet d'un petit choc mécanique, d'un frôlement ou d'un courant d'air. Il existe différents modes de propagation des spores. Les spores appelées gloeiospores sont véhiculées sur un nouveau substrat soit par contact, soit par des insectes ou soit par l'eau. Ces spores présentent une paroi épaisse et humide leur permettant de rester collées entre elles par un mucus et de former ainsi des amas difficilement transportables par l'air. Les spores appelées xérospores sont dissociables et légères leur permettant d'être facilement dispersées par l'air.

Après propagation et lorsque les spores se sont déposées sur un nouveau substrat, celles-ci peuvent rester inertes tant que l'environnement n'est pas favorable à leur développement. Ce n'est que lorsque les conditions environnementales deviennent favorables, qu'elles « germent » comme des graines et émettent du mycélium.

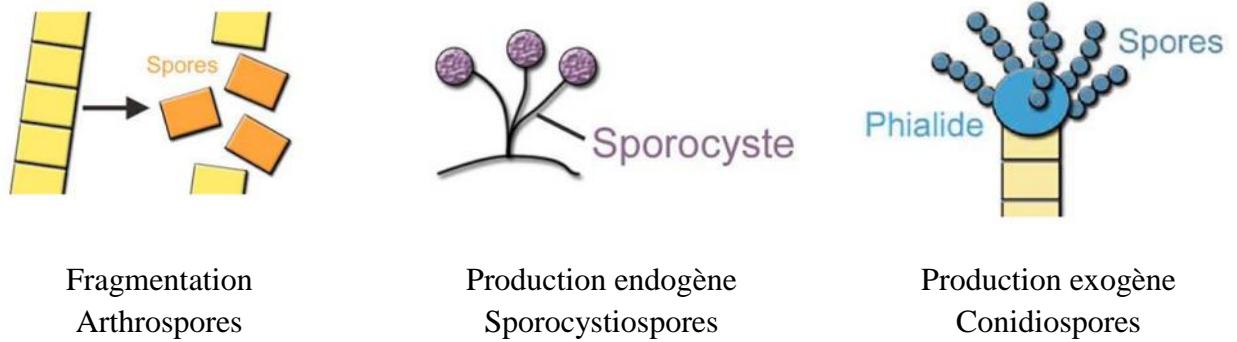


Figure 3 : Les différents modes de sporulation et les différents types de spores associées.

La reproduction sexuée se base sur la fusion de deux gamètes haploïdes (n) donnant un zygote diploïde ($2n$). Une structure (+) à n chromosomes rencontre un autre structure (-) et la fusion des cytoplasmes donne naissance à un nouveau mycélium à $2n$ chromosomes (Figure 4). Il existe différents modes de fécondation en fonction des champignons : planogamie (fusion de gamètes complémentaires et flagellés), oogamie-siphonogamie (fécondation des gamètes femelles dans le sporocyste femelle par des gamètes flagellés provenant des gamétocystes mâles par l'intermédiaire d'un siphon), siphonogamie (accolement du gamétocyste mâle non flagellé au gamétocyste femelle puis émission d'un siphon, il n'y a pas de production de gamète flagellé mâle), trichogamie (fusion des parois du gamète mâle non flagellé appelé spermatie et de l'organe femelle appelé ascogone puis injection du noyau mâle), cystogamie (émission d'un diverticule par deux mycélium compatibles, accolement des deux diverticules et production d'une cloison appelée gamétocyste permettant le mélange des cytoplasmes) et la somatogamie (fusion de deux mycélium compatibles).

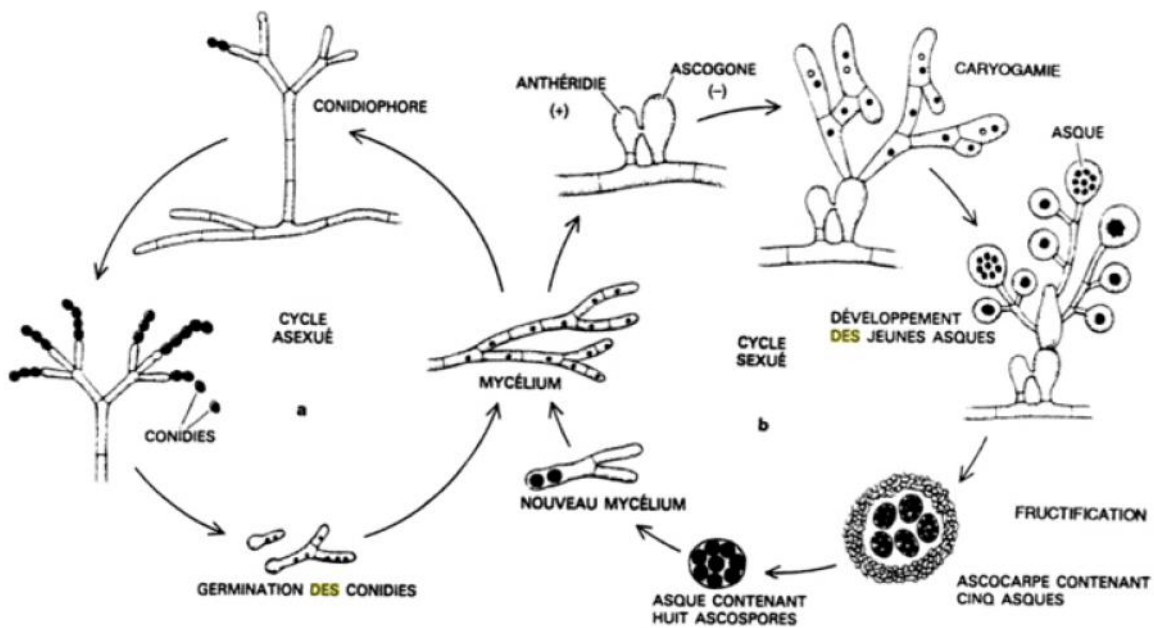


Figure 4 : Schématisation de la reproduction asexuée et sexuée d'une moisissure.

I.1.d- Phylogénie des champignons filamenteux

Dans le monde du vivant, les champignons (microscopiques et macroscopiques) représentent un règne à part au sein des eucaryotes, bien distinct du règne des protistes, des plantes ou de celui des animaux. La classification traditionnelle des champignons filamenteux basée sur les critères phénotypiques a été supplantée par le développement des méthodes de biologie moléculaire. La classification phylogénétique basée sur des critères génotypiques a permis une révision et une modification de la classification des champignons filamenteux.

Actuellement, les champignons filamenteux sont classés parmi les Opisthokontes constituant un groupe particulier d'eucaryotes (2). Différents rangs taxonomiques sont utilisés pour la classification des êtres vivants. Ces rangs hiérarchiques sont : le règne, l'embranchement ou division, la classe, l'ordre, la famille, le genre et l'espèce. Il existe des rangs intermédiaires comme par exemple le sous-embranchement ou la sous-division, la sous-famille ou la sous-espèce qui peut elle-même se diviser en variétés. La nomenclature utilisée pour déterminer le nom scientifique des espèces est binomiale. Elle fait référence au genre puis à l'espèce. Cette nomenclature suit les règles énoncées par le naturaliste Carl Von Linné en 1753. La nomenclature des différents taxons présente une terminologie codifiée, le suffixe

permettant de définir chaque rang taxonomique dans la classification hiérarchique. On distingue ainsi : *-mycota* pour la division, *-mycotina* pour la sous-division, *-mycètes* pour la classe, *-ales* pour l'ordre, et *-aceae* pour la famille.

Le règne des « champignons vrais » ou *Fungi*, appelé *Eumycota*, comprend actuellement 1 sous-règne, 7 divisions et 10 sous-divisions (3). Le sous règne des *Dikarya* se divise en deux divisions, *Ascomycota* et *Basidiomycota*. Les autres divisions sont les *Glomeromycota*, les *Chytridiomycota*, les *Neocallimastigomycota*, les *Blastocladiomycota* et les *Microsporidiomycota*. La division des *Ascomycota* se scinde en trois sous-divisions *Pezizomycotina*, *Saccharomycotina* et *Taphrinomycotina*, tandis que la division des *Basidiomycota* se décompose en trois sous-divisions *Pucciniomycotina*, *Ustilaginomycotina* et *Agaricomycotina*. Les autres sous-divisions sont les *Mucoromycotina*, les *Entomophthoromycotina*, les *Kickxellomycotina* et les *Zoopagomycotina* (Figure 5).

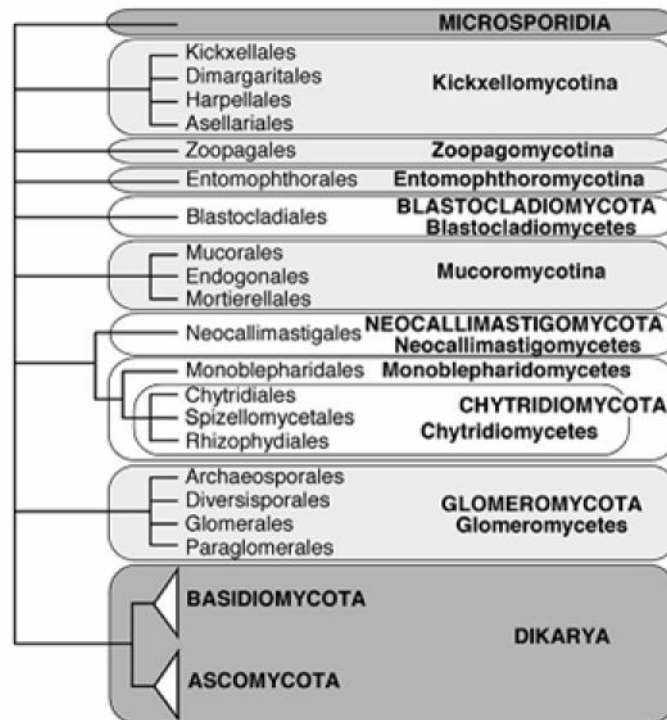


Figure 5 : Phylogénie et classification des *Fungi*.

La classification des champignons est basée sur le mode de reproduction sexuée. Les champignons appartenant au règne des *Eumycota* sont des champignons pour lesquels le mode de reproduction sexuée est connu. On parle alors de champignons téléomorphes. Pour certains champignons appelés anamorphes, le mode de reproduction sexuée est inconnu et seule une multiplication asexuée ou végétative est observée. L'ensemble de ces champignons

sont regroupés au sein de la division des *Deuteromycota*, que l'on appelle aussi « champignons imparfaits » ou *Fungi imperfecti*. Le développement des méthodes moléculaires a permis de classer certains d'entre eux dans le règne des *Eumycota* et plus particulièrement dans les *Ascomycota* en les rattachant à une forme sexuée connue. Ce groupe ne représentant pas un véritable groupe de champignons, il est constitué de nombreuses espèces, il est très hétérogène et il ne constitue pas un ensemble phylogénétique. Les *Deuteromycota* sont divisés en trois classes : les *Blastomycètes*, regroupant les levures, les *Hyphomycètes* et les *Coelomycètes*. Dans cette étude, seuls les champignons filamenteux microscopiques sont étudiés.

I.1.e- Habitat et conditions de développement

Le développement des moisissures est dépendant de facteurs nutritifs et environnementaux. Les champignons filamenteux étant cosmopolites et hétérotrophes, ils présentent différents types d'habitat au sein desquels ils vont établir des interactions différentes avec leur environnement et il existe donc différents modes de nutrition des champignons filamenteux. Les éléments nutritifs les plus importants sont le carbone et l'azote comme composés organiques, les ions minéraux comme le potassium, le phosphore, le magnésium, le fer ou le soufre. Les acides aminés peuvent pénétrer dans la cellule sans transformation, alors que des molécules complexes comme l'amidon, la cellulose ou les protéines nécessitent une digestion enzymatique préalable. Cette digestion s'effectue par production d'enzymes ou d'acides par les moisissures, permettant ainsi une altération du substrat.

Le premier mode de nutrition est le saprophytisme. Dans ce cas, les champignons dégradent la matière organique morte ou en décomposition afin de prélever les éléments minéraux essentiels. Ils jouent un rôle très important dans le recyclage des matières mortes comme les débris végétaux et animaux. Le deuxième mode de nutrition est la symbiose. Les champignons filamenteux obtiennent leurs nutriments grâce à un autre organisme en échange de certains bénéfices, tels une protection, de l'eau ou des sels minéraux. Les deux organismes sont alors qualifiés de symbiotes. Le troisième mode de nutrition est le commensalisme. Les champignons filamenteux dits commensaux tirent un bénéfice de leur hôte sans leur nuire et sans leur apporter un quelconque avantage. Les bénéfices observés peuvent être la récupération des débris nutritifs de l'hôte, le transport ou le refuge. Enfin, le quatrième mode de nutrition est le parasitisme. Dans ce cas, les champignons filamenteux extraient leurs

nutriments de la matière vivante. Ils tirent profit de leurs hôtes en vivant à leurs dépens entraînant parfois leur mort.

Le développement des moisissures est également dépendant de l'environnement. Le facteur environnemental le plus important est l'humidité relative (HR) qui est évaluée par la formule $HR = a_w * 100$, a_w représentant l'activité de l'eau, c'est-à-dire la disponibilité en eau d'un substrat. La majorité des moisissures se développent pour une activité de l'eau comprise entre 0,85 et 0,99. Le deuxième facteur est la température, et les exigences thermiques pour le développement des moisissures diffèrent d'une moisissure à l'autre. La majorité des moisissures sont mésophiles, c'est-à-dire qu'elles se développent aux alentours de 20-25°C, mais il existe certaines moisissures dites thermophiles, thermotolérantes ou psychrophiles. Les moisissures étant des microorganismes aérobies, le troisième facteur est l'oxygène. Enfin, le pH peut avoir une influence sur la croissance et le développement des moisissures. Elles peuvent se développer avec un pH compris entre 4,5 et 8, bien que le pH optimum soit compris entre 5,5 et 7,5.

I.1.f- Problèmes liés aux champignons filamenteux et à leurs métabolites

Les champignons filamenteux présentent un intérêt au sein de l'environnement humain, de manière bénéfique ou néfaste, avec des conséquences économiques. Ils sont impliqués dans différents domaines tels que l'industrie agroalimentaire, pharmaceutique et cosmétique, ainsi que dans le secteur médical.

- secteur agroalimentaire

Au sein de l'industrie agroalimentaire, certaines moisissures sont utilisées pour la production de fromage comme le roquefort (*Penicillium roqueforti*) ou le camembert (*Penicillium camemberti*) (4). Elles peuvent également servir à la synthèse d'acides organiques comme l'acide citrique ou l'acide gluconique (*Aspergillus niger*). Ces deux types d'acides sont utilisés comme additifs alimentaires (5). Certaines moisissures sont utilisées pour la synthèse d'enzymes telles la maltase et la dextrinase servant à transformer le maltose et l'amidon en alcool (*Rhizopus oryzae*). Ce processus de fermentation alcoolique est rencontré dans la fabrication de l'alcool de riz en Asie (6). Les champignons filamenteux présentent donc un intérêt industriel. Néanmoins, ils représentent un risque dans le domaine

de l'industrie agroalimentaire sous forme de contamination des denrées alimentaires. En effet, elles peuvent être à l'origine d'importantes dégradations des propriétés physicochimiques entraînant une altération de la qualité des denrées alimentaires. Le premier type d'altération de la qualité des aliments concerne la qualité dite « marchande ». La prolifération des moisissures, qu'elles soient pathogènes ou non, entraîne des modifications défavorables des caractéristiques diététiques et organoleptiques, tels l'aspect, la texture, l'odeur et la saveur des aliments, avec des conséquences économiques importantes dans l'industrie agroalimentaire (Figure 6).



Figure 6 : Quelques exemples d'aliments contaminés par des moisissures.

Le deuxième type d'altération de la qualité des aliments concerne la qualité dite « sanitaire ». La prolifération des moisissures pathogènes entraîne une diminution de l'innocuité des aliments et représentent un risque pour la santé du consommateur. La production de métabolites secondaires comme les mycotoxines est responsable d'un taux élevé de toxicité et la présence de ces métabolites représente donc un risque sanitaire majeur pour la santé humaine et animale. Une espèce donnée de champignon microscopique peut générer plusieurs types de mycotoxines, et une même mycotoxine peut être produite par plusieurs espèces de moisissures. L'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) estime qu'environ 25% des céréales produites dans le monde sont contaminées par des mycotoxines (7). Les mycotoxines sont soit d'origine endogène et dans ce cas on peut les retrouver dans les spores ou le thalle, soit d'origine exogène, on les retrouve alors au sein du substrat de développement de la moisissure. Les mycotoxines les plus répandues sont les

aflatoxines produites par des espèces d'*Aspergillus*, les ochratoxines produites par certaines espèces d'*Aspergillus* et de *Penicillium*, les fumonisines, les trichotécènes et les zéaralénones produites par des espèces de *Fusarium*, et la patuline produite par un certain nombre d'espèces d'*Aspergillus*, de *Penicillium* et de *Paecilomyces* (8). L'ochratoxine A, considérée comme l'une des mycotoxines les plus nocives pour la santé humaine, est présente essentiellement dans les produits tels que les céréales, le cacao et le café, et c'est la mycotoxine principalement retrouvée dans le raisin, le vin et le jus de raisin (9). La patuline, très toxique pour la consommation humaine, est présente au niveau des fruits et notamment dans les pommes et les poires (10).

- secteur pharmaceutique et médical

Au sein de l'industrie pharmaceutique, certaines moisissures sont utilisées pour la synthèse de médicaments, notamment d'antibiotiques telles la pénicilline (*Penicillium chrysogenum*) ou les céphalosporines (*Cephalosporium acremonium*). Bien que ces composés présentent un intérêt médical majeur, certains champignons filamenteux représentent une source de contamination avec un risque élevé pour la santé humaine. On distingue donc deux types de contamination chez l'homme : alimentaire et infectieuse.

La consommation d'aliments contenant des mycotoxines est responsable d'une intoxication alimentaire appelée mycotoxicose. Il existe deux types de mycotoxicose : la mycotoxicose aiguë (une seule ingestion d'une forte dose) et la mycotoxicose chronique (ingestions répétées de faibles doses). Les symptômes observés sont dépendant du type de mycotoxine, de la dose et de la durée d'exposition, ainsi que des caractéristiques de l'individu. Dans le cas par exemple des aflatoxines, principalement produites par les espèces *Aspergillus flavus* et *Aspergillus parasiticus*, l'intoxication aiguë se traduit par différents symptômes (dépression, anorexie, diarrhée, ictère ou anémie) et peut être fatale. Lors d'une intoxication chronique, ces mycotoxines sont hautement hépatotoxiques, tératogènes, mutagènes et cancérigènes (11). Chez l'homme, l'aflatoxine B1, composé d'origine naturel, représente l'agent hépato-cancérigène le plus puissant connu actuellement.

Les infections fongiques chez l'homme dues aux champignons filamenteux sont en augmentation depuis ces dernières décennies. La contamination se fait principalement par inhalation de spores en suspension dans l'air, ou plus rarement par ingestion ou encore par inoculation cutanée post traumatique. Elles peuvent être d'origine nosocomiale. Les infections

fongiques invasives sont principalement retrouvées chez les patients immunodéprimés et sont associées à un taux de mortalité élevé.

Chez les patients immunodéprimés, il existe différents types d'infections fongiques invasives selon les champignons filamenteux. La plus fréquente est l'aspergillose invasive. Cette pathologie est due à la contamination par *Aspergillus*, 80 à 90 % des aspergilloses humaines étant dues à l'espèce *Aspergillus fumigatus* (12). La porte d'entrée est principalement respiratoire et les aspergilloses invasives se rencontrent principalement chez les patients présentant une neutropénie profonde et sous corticothérapie (patients atteints de leucémie ou ayant reçu une allogreffe de moëlle osseuse) (13-15). D'autres champignons filamenteux sont responsables d'infections fongiques invasives. Les champignons de l'ordre des *Mucorales* sont responsables de mucormycoses, notamment chez les patients immunodéprimés, comme les patients atteints d'hémopathies ou encore les diabétiques en acidocétose (16). Ils peuvent entraîner des atteintes rhinocérébrales, cutanées ou viscérales. La contamination se fait essentiellement par voie respiratoire, mais aussi par voie cutanée et digestive. D'autres genres tels *Fusarium* ou *Scedosporium* sont responsables d'infections invasives sévères chez les patients immunodéprimés (17, 18).

Chez les patients immunocompétents, les contaminations fongiques peuvent se faire par voie cutanée en particulier après souillure tellurique de plaies traumatiques sévères. En effet, de nombreux champignons filamenteux sont présents dans le sol comme par exemple *Aspergillus*, *Fusarium*, *Rhizopus*, *Absidia* ou *Rhizomucor*. Ces contaminations fongiques peuvent également se faire par le biais d'une plaie chirurgicale, d'une brûlure ou d'une injection intraveineuse ou sous cutanée (19). Les contaminations par voie aérienne sont également observées chez les patients immunocompétents. Une aspergillose peut survenir suite à l'inhalation de spores d'*Aspergillus*. L'inhalation de spores fongiques peut être associée à des symptômes allergiques et irritatifs se manifestant cliniquement par des irritations des muqueuses, des rhinites, de l'asthme, voire des pneumopathies d'hypersensibilité. La colonisation des voies respiratoires chez les patients atteints d'affection respiratoire chronique comme la mucoviscidose peut être observée et être responsable d'aggravations des signes pulmonaires (20). Enfin, certains champignons filamenteux peuvent également être responsables d'infections superficielles. Outre les dermatophytes qui sont des champignons kératinophiles responsables de dermatophytoses (21), les moisissures environnementales peuvent être responsables de kératites (*Fusarium*, *Aspergillus*), d'onychomycoses (*Fusarium*, *Aspergillus*, *Alternaria*, Dermatophytes) ou d'otomycoses (*Aspergillus*).

I.1.g- Méthodes d'identification des champignons filamenteux

L'identification des champignons filamenteux en routine repose essentiellement sur l'analyse des caractères morphologiques macroscopiques et microscopiques. Ces méthodes d'identification peuvent être complétées par une analyse moléculaire.

- Analyse macroscopique

Lors de l'analyse macroscopique des colonies obtenues après culture des champignons filamenteux, plusieurs aspects de l'appareil végétatif sont observés :

- l'aspect : duveteux, laineux, cotonneux, velouté, poudreux, granuleux ou glabre.
- le relief : plat, plissé ou cérébriforme.
- la taille : petite, étendue ou envahissante.
- la couleur : blanche, crème ou colorée (verte, brune, orangée, violette, grises...).

La présence d'un pigment diffusant dans la gélose ainsi que certains paramètres telle la vitesse de la pousse des colonies ou la température de développement peuvent être de bons indicateurs pour l'identification d'une moisissure.

- Analyse microscopique

Lors de l'analyse microscopique des colonies, plusieurs structures des champignons filamenteux sont observées comme l'appareil végétatif, les organes de fructification et les spores :

- le thalle végétatif : septé (diamètre étroit et régulier de 2 à 5 μm) ou siphonné (filaments peu ou pas ramifiés, diamètre large et irrégulier de 5 à 15 μm), paroi pigmentée (mélanisée) ou non (hyaline).
- les organes de fructifications (Figure 7) : présence ou non d'organes protecteurs des conidies, modes de formation des conidies (issues directement du thalle, solitaires (aleuriospores) ou en chaînes (arthrospores), ou produites par bourgeonnement et regroupées soit en grappes, en masse, en têtes ou en chaînes basipètes ou acropètes), modes d'implantation des cellules conidiogènes [indifférenciée ou peu indifférenciée, différenciées (sur le filament végétatif, porté sur les conidiophores dispersés ou groupés)].
- les spores : endogènes (endospores) ou exogènes (conidiospores ou conidies), l'aspect des spores [améropores (unicellulaires et de petite taille), didymospores (bicellulaires),

phragmospores (pluricellulaires à cloisons transversales), dictyospores (pluricellulaires à cloisons transversales et longitudinales), scolécospores (étroites et effilées)], présence ou non de chlamydospores.

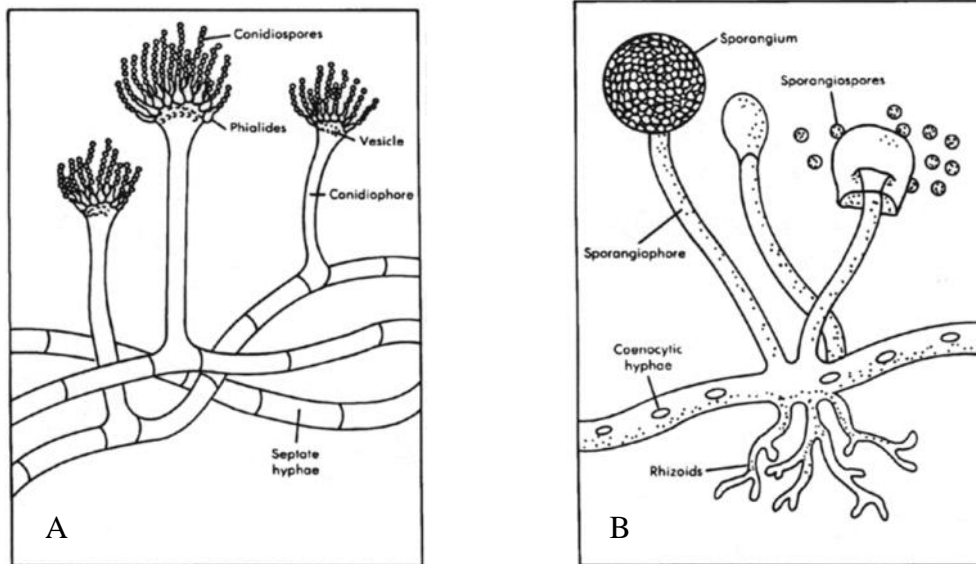


Figure 7 : Organes de fructifications du genre *Aspergillus* (A) et du genre *Rhizopus* (B).

- Analyse moléculaire

Les méthodes d'identification des champignons filamenteux par biologie moléculaire reposent sur l'analyse des séquences portant l'information génétique (22). Les techniques de biologie moléculaires s'intègrent progressivement aux côtés des méthodes mycologiques classiques, et tendent à se généraliser dans les laboratoires spécialisés (23). L'émergence de la PCR (Polymerase Chain Reaction) a permis d'importants progrès des techniques moléculaires. Les différentes méthodes proposées permettent d'étudier le polymorphisme génétique des différents champignons filamenteux et de les discriminer à différents niveaux taxonomiques par l'étude de l'ensemble du génome, d'un ou plusieurs gènes ou d'un fragment d'ADN bien définis. Plusieurs techniques sont appliquées : la RFLPs (Restriction Fragment Length Polymorphisms) est basée sur le polymorphisme de taille des fragments de restriction et a été utilisée pour la discrimination d'espèce d'*Aspergillus* (24), la RAPD (Random Amplified Polymorphic DNA), basée sur le polymorphisme de l'ADN amplifié au hasard, a permis de mettre en évidence une différenciation des souches de *Penicillium roqueforti* (25), l'AFLP (Amplified Fragment Length Polymorphism), qui est une

combinaison de la PCR et de la RFLP, a permis de discriminer différentes espèces d'*Aspergillus* (26). Ces méthodes sont généralement assez coûteuses et longues à mettre en œuvre. De plus, certaines de ces méthodes présentent des limites dues au manque de sensibilité et de reproductibilité et à la nécessité d'une standardisation des protocoles, notamment lors de l'extraction de l'ADN.

Récemment, une nouvelle méthode d'analyse phénotypique a émergé comme outil d'identification des champignons filamenteux en routine. Cette technique qui repose sur la spectrométrie de masse (MALDI-TOF MS), permet d'obtenir sous forme de spectre le profil protéique des champignons filamenteux (27). La moisissure est identifiée en comparant son spectre avec ceux d'une librairie de spectres de référence. Cette méthode permet de discriminer des champignons filamenteux au niveau de l'espèce, en donnant des résultats comparables à ceux obtenus par les méthodes d'identification moléculaire (28, 29). Cette technique simple, rapide et pouvant être utilisée à haut débit fait l'objet d'une standardisation tant au niveau du protocole (conditions de culture, méthodes d'extraction) qu'au niveau de la construction des banques de données (30). Le coût de l'appareillage nécessite un investissement assez lourd.

I.2- La spectroscopie infrarouge à transformée de Fourier (IRTF)

I.2.a- Principe

- Rayonnement infrarouge

La spectroscopie infrarouge à transformée de Fourier, une méthode analytique physicochimique utilisée depuis longtemps par les chimistes, trouve plus en plus d'applications dans les domaines de la biologie, biochimie et biomédical. Cette méthode est basée sur l'absorption d'un rayonnement infrarouge par l'échantillon analysé. Le rayonnement infrarouge (IR) a été découvert en 1800 par l'astronome Frédéric Wilhelm Hershel. C'est une onde électromagnétique (Figure 8) de fréquence inférieure à celle de la lumière visible : le rouge, d'où son nom tiré du latin « infra » signifiant « plus bas ». Les radiations du rayonnement infrarouge, dont le domaine s'étend de $0,8 \mu\text{m}$ à $1000 \mu\text{m}$, sont situées entre la région du spectre visible et les ondes hertziennes. Ce domaine est divisé en 3 régions : le proche, le moyen et le lointain infrarouge. Le proche infrarouge se situe entre $0,8$ et $2,5 \mu\text{m}$ ($12500\text{-}4000 \text{ cm}^{-1}$), le moyen infrarouge entre $2,5$ et $25 \mu\text{m}$ ($4000\text{-}400 \text{ cm}^{-1}$) et le lointain infrarouge entre 25 et $1000 \mu\text{m}$ ($400\text{-}10 \text{ cm}^{-1}$) (Figure 9).

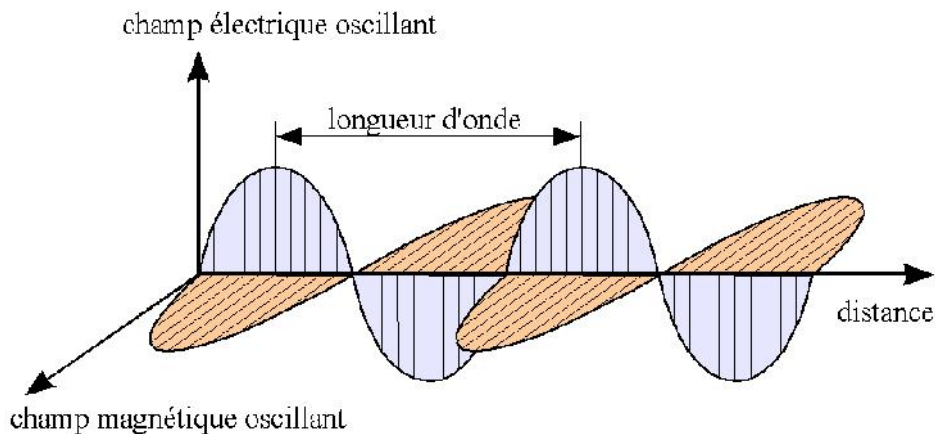


Figure 8 : Représentation d'une onde électromagnétique.

La relation entre l'énergie, la longueur d'onde et le nombre d'onde est exprimée par la formule ci-dessous :

$$E = h \cdot \nu = (h \cdot c) / \lambda = h \cdot c \cdot \tilde{\nu}$$

Avec : E : énergie (J)

h : constante de Planck : $6.624 \cdot 10^{-34}$ J/s

ν : fréquence de l'onde (Hz)

c : vitesse de la lumière dans le vide : $2.998 \cdot 10^8$ m/s

λ : longueur d'onde (m)

$\tilde{\nu}$: nombre d'onde (m^{-1})

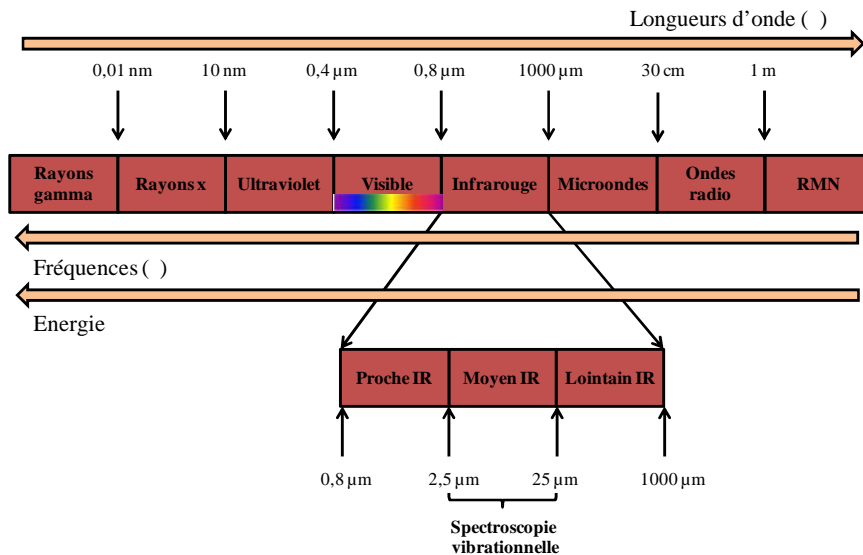


Figure 9 : Spectre électromagnétique.

L'énergie du rayonnement infrarouge moyen coïncide avec l'énergie des mouvements internes des molécules. Le rayonnement infrarouge peut donc être utilisé pour étudier les vibrations fondamentales et caractéristiques ainsi que la structure vibrationnelle associée des liaisons chimiques. Il permet l'analyse des fonctions chimiques présentes au sein d'un échantillon. Il existe une relation entre l'absorption du rayonnement infrarouge moyen par une molécule et la structure moléculaire de cette molécule. Le rayonnement du proche infrarouge, qui est plus énergétique, permet d'étudier les vibrations harmoniques, alors que le rayonnement de l'infrarouge lointain, qui est moins énergétique, permet d'étudier les

vibrations rotationnelles. L'utilisation de la spectroscopie dans le moyen infrarouge est donc la plus adaptée pour l'étude de la composition moléculaire d'un échantillon.

- Interaction rayonnement-matière

En ce qui concerne la spectroscopie IR, l'interaction entre le rayonnement électromagnétique et les molécules ne s'effectue que dans le cas où la molécule possède au moins une liaison présentant un moment dipolaire. Une molécule est une entité neutre constituée d'atomes, assimilés à des charges positives ou négatives et reliés entre eux par des liaisons covalentes. Chaque atome est caractérisé par son électronégativité c'est-à-dire sa capacité à attirer plus ou moins les électrons des liaisons qu'il forme avec un autre atome au sein de la molécule.

Une molécule est dite à caractère dipolaire lorsque le barycentre de ses « charges » positives ne coïncide pas avec celui de ses « charges » négatives. Un exemple de molécule dipolaire est la molécule d'eau. Au sein de cette molécule, les deux atomes d'hydrogène sont « chargés » $+ \delta/2$ et leur barycentre se trouve sur l'axe de symétrie de la molécule H_2O . L'atome d'oxygène est chargé $- \delta$ et son centre est confondu avec le barycentre de la « charge » négative. Comme les deux barycentres ne coïncident pas, la molécule d'eau est donc un dipôle électrique et est caractérisée par son moment dipolaire. Un exemple de molécule non dipolaire est la molécule de dioxyde de carbone (CO_2). Les atomes d'oxygène sont « chargés » $- \delta/2$ et leur barycentre est confondu avec le centre de l'atome de carbone chargé $+ \delta$, étant donné que la molécule est linéaire et symétrique par rapport à l'axe du carbone (Figure 10).

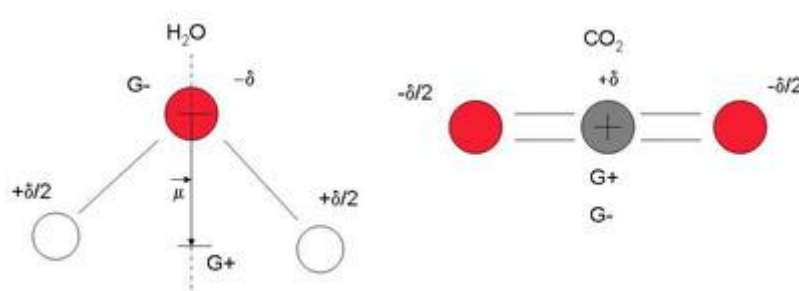


Figure 10 : Barycentres (G) de la molécule H_2O et de la molécule CO_2 (: charge).

Bien que la molécule de dioxyde de carbone soit une molécule non polaire, elle possède deux liaisons $C=O$ qui elles, possèdent un moment dipolaire. La polarité d'une

liaison est due à la différence d'électronégativité entre les deux éléments chimiques qui composent cette liaison.

Pour les molécules diatomiques homonucléaires, comme par exemple la molécule du dioxygène, le moment dipolaire est nul car il n'y a pas de différence d'électronégativité entre les deux atomes d'oxygène (Figure 11).

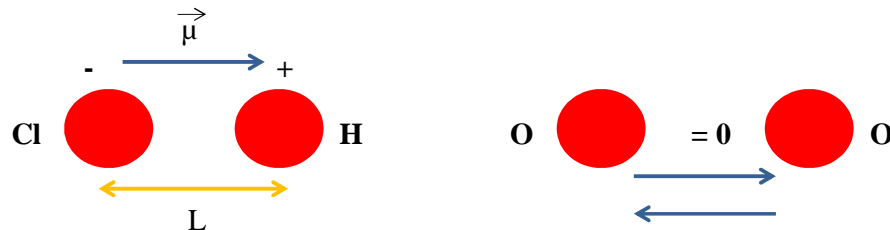


Figure 11 : Moment dipolaire de la molécule HCl et de la molécule O₂.

La spectroscopie infrarouge repose sur le fait qu'une liaison moléculaire possède une fréquence spécifique à laquelle elle absorbe de l'énergie provoquant la vibration de cette liaison. La fréquence de vibration dépend de la force de la liaison et de la masse des atomes composant la liaison. La vibration se fait autour de la position d'équilibre et dans le cas de la spectroscopie infrarouge, implique une variation du moment dipolaire autour de cette position. Les niveaux d'énergie associés aux fréquences de vibration correspondent à l'énergie des ondes moyen-infrarouge du spectre électromagnétique.

Lorsqu'une onde électromagnétique ayant une fréquence donnée interagit avec une molécule, cette dernière peut absorber une partie de l'énergie, entraînant la vibration de la liaison moléculaire. L'onde transmise aura une intensité diminuée par rapport à l'onde incidente. L'absorption d'énergie permet à la molécule de passer du niveau fondamental à un niveau vibrationnel excité (Figure 12).

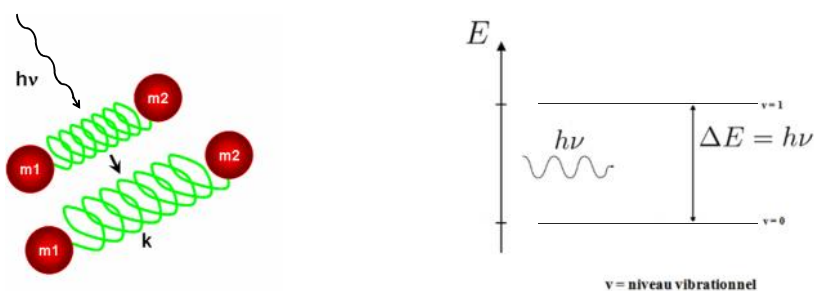


Figure 12 : Schéma du phénomène d'absorption et transition entre deux niveaux vibrationnels.

Pour qu'un mode de vibration soit actif en infrarouge, il faut que la variation de son moment dipolaire ($d\mu/dQ$) soit non nulle. Une liaison ayant une variation du moment dipolaire nulle n'absorbe pas et le mode vibrationnel sera inactif en infrarouge.

- Absorption du rayonnement IR par l'échantillon

Lorsque l'énergie apportée par le rayonnement IR incident est voisine de l'énergie de vibration d'une liaison, cette dernière va absorber le rayonnement et on observera alors une diminution de l'énergie transmise.

$$T = I / I_0$$

$$A = \text{Log } 1 / T = \text{Log } (I_0 / I)$$

Avec : T : facteur de transmission ou transmittance (en %)

A : absorbance

I_0 : intensité du rayon incident

I : intensité du rayon transmis

Ces grandeurs sont reliées à la composition de l'échantillon par la loi de Beer-Lambert :

$$A = \epsilon l c$$

Avec : ϵ : coefficient d'absorption molaire (en $\text{L mol}^{-1} \text{cm}^{-1}$)

l : épaisseur de l'échantillon traversée par l'onde (en cm)

c : concentration de l'échantillon (mol L^{-1})

Le spectre infrarouge d'un échantillon est représenté par son absorbance (ou transmittance) en fonction du nombre d'onde. Ceci est mesuré par un spectromètre. Pour des raisons pratiques, en spectroscopie vibrationnelle, le nombre d'onde noté $\tilde{\nu}$ est souvent utilisé à la place de ν ou λ . C'est une grandeur proportionnelle au nombre d'oscillations qu'effectue une onde par unité de longueur : c'est le nombre de longueurs d'onde présentes sur une distance donnée. Ce nombre d'onde est ainsi une grandeur inversement proportionnelle à la longueur d'onde.

$$= 1/$$

Avec : ν : nombre d'onde (m^{-1})

λ : longueur d'onde (m)

Dans le domaine moyen-infrarouge, les longueurs d'onde varient de 2,5 à 25 μm , correspondant à une gamme spectrale de 4000 à 400 cm^{-1} .

Il existe donc un lien entre les pics observés dans un spectre infrarouge et les liaisons chimiques. Pour un composé simple et pur, chaque pic peut être associé à un mode de vibration de liaison spécifique, et le spectre représente l'empreinte de tous les constituants moléculaires de l'échantillon. Par conséquent, à un échantillon de composition chimique et de structure donnée va correspondre un ensemble de bandes d'absorption caractéristiques permettant de déterminer sa structure et sa composition moléculaire.

- Modes de vibration

Dans le moyen infrarouge, les bandes d'absorption observées correspondent à différents modes de vibration fondamentaux (Figure 13). Les modes les plus rencontrés sont :

- les vibrations de valence ou d'élongation (stretching) notées ν correspondant aux déplacements de deux atomes dans l'axe de leur liaison. Elles peuvent être symétriques ou asymétriques. Les énergies (fréquences) impliquées sont différentes.

- les vibrations de déformation (bending) notée ν_2 correspondant aux oscillations de deux atomes produisant une variation de l'angle entre les deux liaisons. Ces vibrations comprennent celles dans le plan (cisaillement ou rotation plane) ou hors du plan (balancement ou torsion).

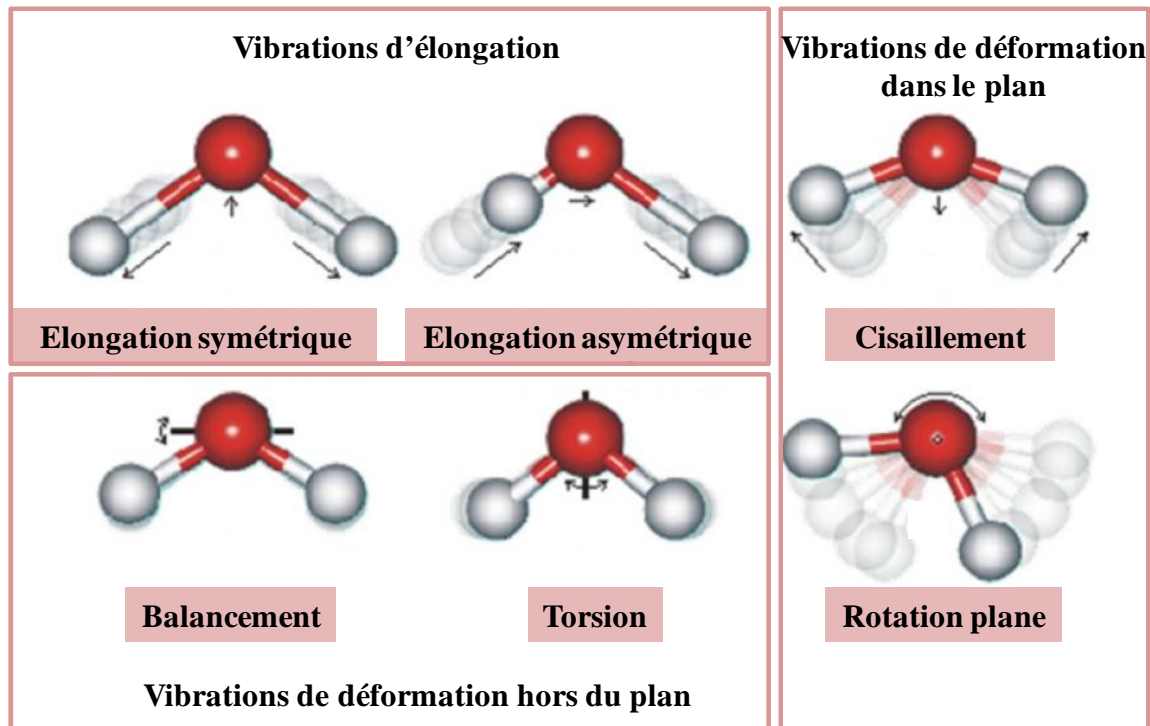


Figure 13 : Différents modes fondamentaux de vibration d'une molécule triatomique non-linéaire.

- Instrumentation

L'analyse s'effectue à l'aide d'un spectromètre infrarouge à transformée de Fourier (IRTF). Une source IR polychromatique éclaire l'échantillon avec des longueurs d'ondes comprises entre 2,5 et 25 μm .

Le faisceau infrarouge provient d'une source thermique. Cette source émet de l'énergie sous la forme d'une radiation lumineuse qui est le résultat de l'échauffement d'un filament métallique par un courant électrique. Les filaments les plus rencontrés dans le moyen infrarouge sont les filaments de Global constitués de carbure de silicium (SiC) et les filaments de nichrome constitués de nickel et de chrome enroulé en spirale.

Le faisceau infrarouge est ensuite dirigé vers l'interféromètre, qui est un dispositif optique permettant de moduler chaque longueur d'onde du faisceau infrarouge à une fréquence différente de manière simultanée. Ainsi, il permet de mesurer les longueurs d'onde par production d'interférences (Figure 14).

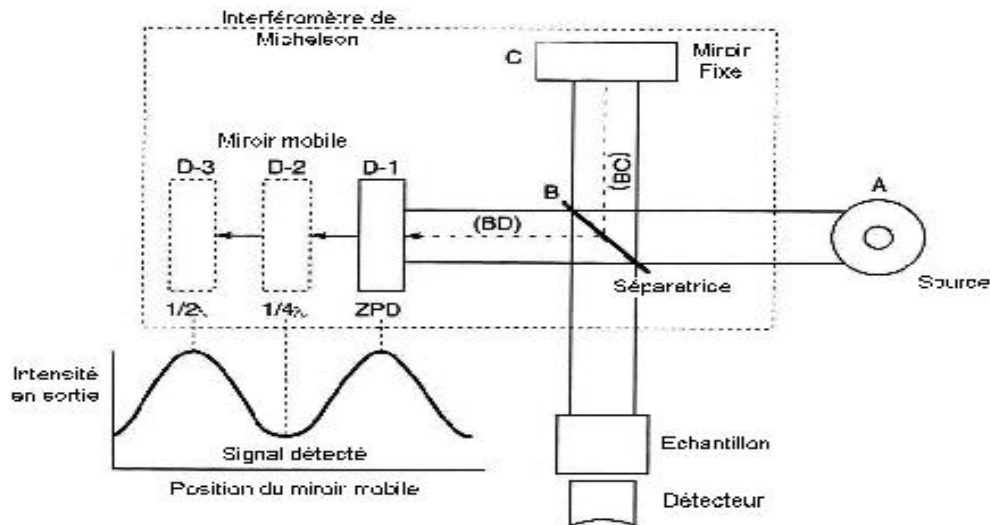


Figure 14 : Schéma d'un interféromètre de Michelson.

Ce dispositif est constitué d'une lame séparatrice semi-transparente (fluorure de calcium pour le moyen IR) sur laquelle arrive le faisceau. Le faisceau est alors divisé en deux : une moitié est réfléchi sur un miroir fixe, et l'autre moitié passe à travers la séparatrice puis est dirigé sur un miroir mobile. Les deux miroirs sont perpendiculaires l'un par rapport à l'autre. Le miroir mobile se déplace à vitesse constante le long de son axe. Le premier faisceau parcourt donc un trajet optique fixe (BC) et le deuxième, un trajet optique de longueur variable (BD) selon la position du miroir. La différence de longueur des bras de l'interféromètre (distance miroir-séparatrice) induit une différence de trajet entre les deux rayons égale à $2 \times$, étant donné que chacun des deux faisceaux effectue un aller-retour entre la séparatrice et les miroirs.

Les deux faisceaux se recombinaient ensuite sur la séparatrice créant ainsi des interférences. On parle d'interférence lorsque deux ondes de même type se rencontrent et interagissent l'une avec l'autre. Il existe des interférences constructives (positives) ou destructives (négatives) en fonction de la position du miroir mobile. Lorsque deux ondes de même nature, c'est-à-dire de même longueur d'onde, atteignent le même point de l'espace mais en ayant parcouru un chemin différent, les deux ondes arrivent en ce point de manière déphasée. Lorsque les deux ondes parcourent le même trajet, l'intensité obtenue est maximale puisque les deux ondes sont en phase et celles-ci s'ajoutent. C'est également le cas lorsque la différence du trajet entre les deux ondes est égale à un nombre entier de la longueur d'onde. On parle alors dans ce cas d'interférences constructives. À l'inverse, lorsque la différence de trajet est égale à une demi-longueur d'onde ou à un nombre impair de la demi-longueur

d'onde, les ondes arrivent en opposition de phase et l'intensité est alors nulle. On parle dans ce cas d'interférences destructives (Figure 15). L'ensemble des interférences constructives et destructives constituent un signal appelé interférogramme (Figure 16). Chaque point de ce signal est fonction des longueurs d'onde du faisceau infrarouge.

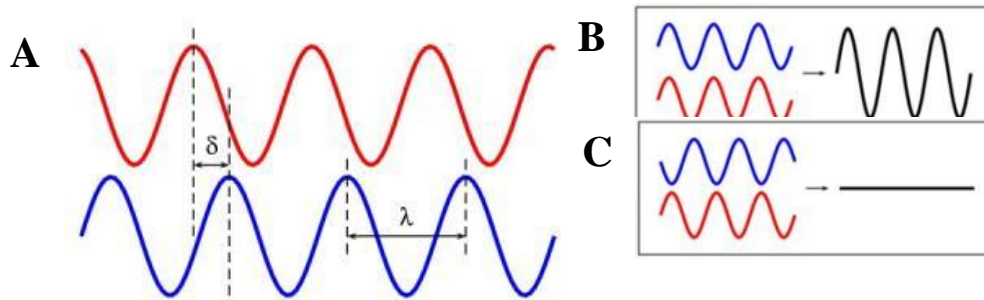


Figure 15 : (A) Représentation de la différence du trajet par deux ondes de même longueur d'onde. (B) Représentation d'une interférence constructives où $\delta = k \cdot \lambda$. (C) Représentation d'une interférence destructive où $\delta = (k+1/2) \cdot \lambda$.

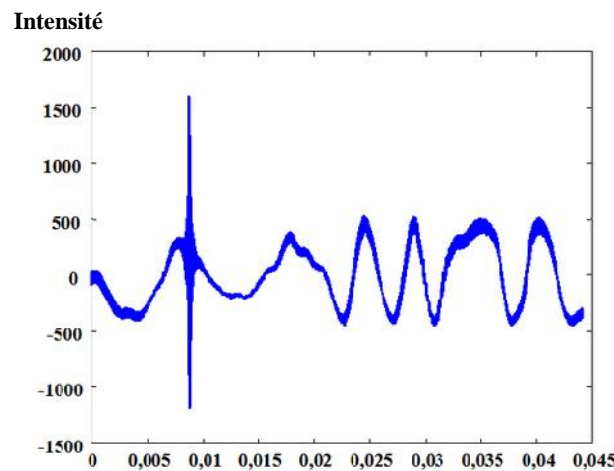


Figure 16 : Interférogramme (x : position du miroir).

Le faisceau modulé est ensuite réfléchi vers l'échantillon où des absorptions interviennent. Le faisceau arrive ensuite sur le détecteur pour être transformé en signal électrique. Le signal du détecteur apparaît comme un interférogramme, c'est-à-dire une signature de l'intensité en fonction de la position du miroir ($I = f(x)$). L'interférogramme est la somme de toutes les fréquences du faisceau. Afin de pouvoir être interprété, cet

interférogramme est ensuite converti en un spectre infrarouge par une opération mathématique appelée transformée de Fourier, d'où l'appellation IRFT. Cet algorithme est couramment utilisé en traitement numérique du signal et permet de convertir les unités de temps (ou de déplacement comme la vitesse du miroir mobile est connue) en unités de fréquence. On obtient un spectre exprimant l'intensité en fonction de la fréquence ou nombre d'onde ($I = f(\nu)$).

- Résolution spectrale

Pour obtenir une séparation correcte de \mathbf{d} entre deux bandes sur le spectre dans le domaine des nombres d'onde, il faut au minimum mesurer l'interférogramme sur une distance $= \mathbf{1/d}$. Par exemple, une résolution spectrale de 4 cm^{-1} demandera au minimum un retard de 0,25 cm. En conclusion, plus on augmente le parcours du miroir mobile, plus on augmente la résolution spectrale.

- Avantages de la spectroscopie IRTF

- rapidité (avantage de Fellgett ou avantage multiplexe permettant de mesurer toutes les fréquences de manière simultanée par rapport un à système dispersif)
- reproductibilité (avantage de Connes permettant un échantillonnage précis sur l'échelle des fréquences grâce au laser He:Ne, utilisé comme une horloge pour déclencher le « décodage » de l'interférogramme à des intervalles réguliers de λ lors de la conversion analogique-digitale)
- sensibilité (avantage de Jacquinot permettant l'amélioration du signal sur bruit en respectant la géométrie cylindrique du faisceau par rapport un à système dispersif)
- haute résolution spectrale (fonction du déplacement maximum du miroir mobile)
- simplicité mécanique (seul le miroir mobile se déplace)
- autocalibration

I.2.b- Application de la spectroscopie IRTF en microbiologie

Les récents développements de la spectroscopie IRTF ont permis de nombreuses applications dans le domaine de la microbiologie, et notamment en tant qu'outil complémentaire de discrimination et d'identification adapté à un large spectre de microorganismes comme les bactéries, les levures et les champignons filamenteux. Les différences entre les taxons, exprimés sous forme de différences quantitatives et qualitatives de la composition cellulaire de chaque groupe peut se traduire en différences moléculaires et donc en différences spectrales caractéristiques. Ces différences spectrales peuvent être utilisées comme des caractères (marqueurs) spectraux discriminants pour l'identification (31). Pour les microorganismes, il est possible de faire correspondre certaines bandes caractéristiques de spectre infrarouge aux différentes fonctions chimiques des molécules présentes au sein de l'échantillon (32) (Tableau 1). Le couplage de la spectroscopie IRTF avec les méthodes d'analyse statistique multivariée, donne lieu à un outil performant pour l'étude des microorganismes (33, 34).

Tableau 1 : Fréquences d'absorption observées dans les spectres infrarouges des microorganismes et leur attribution biomoléculaire.

Fréquences (cm ⁻¹)	Liaison moléculaire	Mode de vibration	Attribution biomoléculaire
3200-2800	CH ₂ , CH ₃ N-H	élongation symétrique et asymétrique élongation symétrique	Lipides, Protéines
1780-1700	C=O	élongation symétrique	Acides gras
1695-1625	C=O, C-N N-H	élongation symétrique déformation	Protéines (amide I)
1560-1525	C-N N-H	élongation symétrique déformation	Protéines (amide II)
1480-1400	CH ₃ , CH ₂ C=O	déformation élongation asymétrique	Lipides
1300-1200	P=O,	élongation asymétrique	Acides nucléiques
1200-900	C-O-C, C-O, P=O, C-C/C-O	élongation symétrique	Ribose, Glycogène, Chitine, Glucane, Mananne, Acides nucléiques
900-700	C-H	déformation	Groupes aromatiques

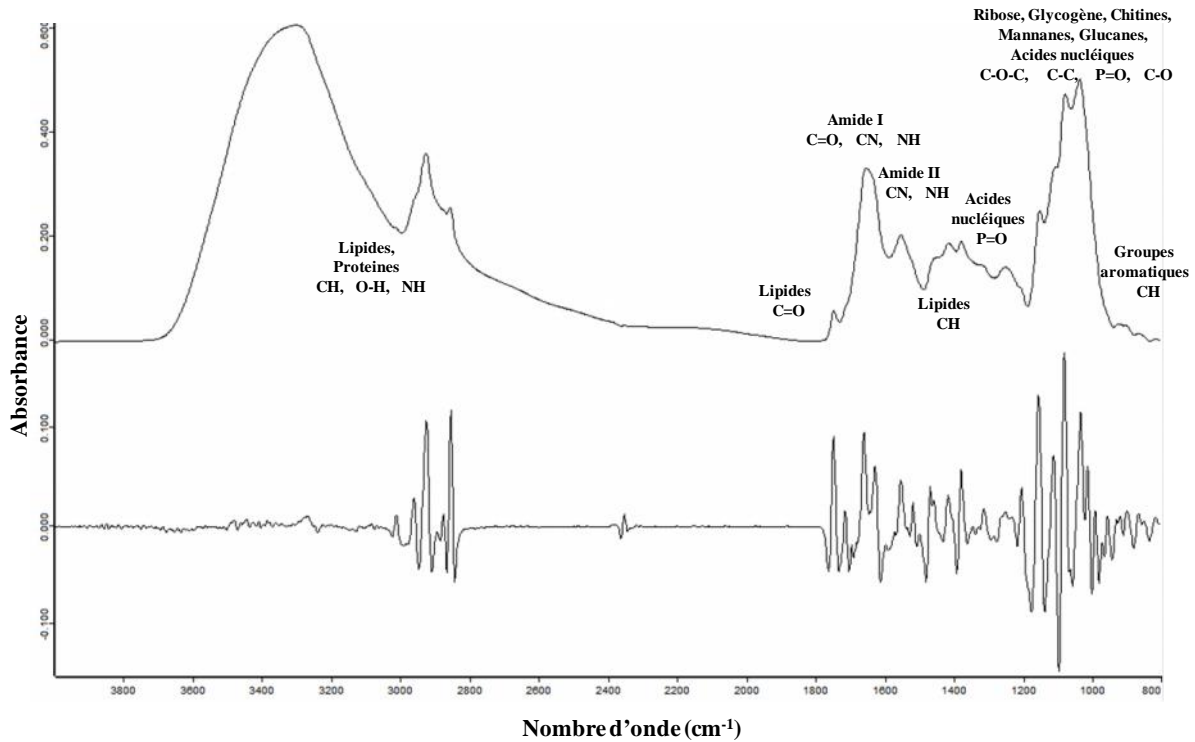


Figure 17 : Spectres d'absorption IR caractéristiques d'un champignon filamenteux (*Aspergillus flavus*) et sa dérivée seconde inverse.

En ce qui concerne les bactéries et les levures, le potentiel de la spectroscopie IRTF pour la discrimination et l'identification a été démontré dans de nombreuses études. Cette méthode a notamment permis la discrimination et l'identification de souches de bactéries lactiques d'origine laitière au niveau genre et espèce (35, 36). Elle a permis également de différencier et d'identifier des espèces de bactéries pathogènes retrouvées dans le domaine agroalimentaire comme *Listeria* (37) et *Campylobacter* (38). L'identification de bactéries d'intérêt clinique au niveau de l'espèce comme *Staphylococcus*, *Enterococcus*, *Escherichia*, *Enterobacter*, *Klebsiella*, *Pseudomonas*, *Proteus* ou *Citrobacter* a pu être réalisée par cette méthode (39-41). Des bactéries d'intérêt environnemental comme les bactéries marines sulfatoreductrices (SRB) et thiosulfatoreductrices (TRB) impliquées dans la corrosion marine ont également fait l'objet d'analyse par spectroscopie IRTF. Cette technique a permis la caractérisation et l'identification de souches impliquées dans les phénomènes de corrosion induite par les microorganismes (MIC) et la mise en évidence d'une corrélation entre la biodiversité et l'importance de la corrosion (42, 43). Cette méthode d'analyse a aussi été utilisée pour l'analyse intraspécifique, c'est-à-dire pour la discrimination de sérotypes au sein d'une espèce donnée. Elle a permis la caractérisation et l'identification de souches

d'actinomycètes (44) avec des résultats comparables à ceux obtenus par des méthodes classiques d'analyses taxonomiques. La spectroscopie IRTF permet aussi de discriminer les microorganismes eucaryotes telles les levures (45, 46). En microbiologie médicale, la spectroscopie infrarouge offre une identification rapide et une caractérisation de levures du genre *Candida* d'intérêt clinique impliquées dans les infections humaines (47). Elle a aussi été utilisée pour l'analyse intraspécifique, c'est-à-dire pour la discrimination de souches au sein même d'une espèce donnée (48-50).

Cette technique peut être aussi utilisée pour l'étude des prions et notamment pour l'étude des protéines de la scrapie du hamster (51), pour l'étude des microalgues marines (Giordano et al, 2001), et en virologie pour la détection et l'identification des cellules infectées par les virus de l'herpès (52).

En ce qui concerne les champignons filamenteux, la spectroscopie IRTF a déjà été appliquée pour la discrimination de 3 espèces de dermatophytes, *Trichophyton rubrum*, *Trichophyton mentagrophytes* et *Microsporum canis* (53). Une étude récente a évalué la capacité de la spectroscopie IRTF pour la différenciation et la classification d'espèces de *Trichophyton* (54). Cette méthode a également permis la différenciation des 3 espèces *Aspergillus fumigatus*, *Aspergillus flavus* et *Aspergillus parasiticus* et la discrimination entre les souches toxigènes et non-toxigènes de l'environnement agricole (55). La capacité de la spectroscopie IRTF à différencier 3 espèces d'*Aspergillus* morphologiquement semblables, *Aspergillus niger*, *Aspergillus ochraceus* et *Aspergillus westerdijkiae* a également été démontrée (56). La spectroscopie IRTF a été utilisée avec succès pour la différenciation de 16 isolats appartenant à cinq espèces de *Fusarium* (57).

Toutes les études précédemment citées et concernant l'analyse des champignons filamenteux par spectroscopie IRTF portent sur la différenciation et l'identification au niveau de l'espèce pour un seul genre et ne portent que sur un nombre limité d'espèces par étude. Peu d'études ont trait à la capacité de la spectroscopie IRTF pour la discrimination fongique et l'identification de plusieurs genres. Ainsi, cette méthode a été appliquée pour l'identification de dix espèces fongiques appartenant aux genres *Aspergillus*, *Emericella* et *Penicillium*, d'origine clinique et contaminants de l'air (58). La caractérisation et l'identification de champignons filamenteux rencontrés dans le domaine alimentaire appartenant à 11 espèces de 5 genres (*Alternaria*, *Aspergillus*, *Mucor*, *Paecilomyces* et *Phoma*) ont pu être obtenues par l'analyse spectrale de ces souches (59). Dans une étude récente, la capacité de la spectroscopie IRTF pour la caractérisation et l'identification des champignons filamenteux au

niveau genre et espèce a été démontrée à partir de 59 souches fongiques relatives à 19 espèces et 10 genres couramment impliqués dans l'altération des aliments (60).

Ces études ont montré que la spectroscopie IR constitue une véritable alternative à d'autres méthodes d'identification et de discrimination des champignons filamenteux. En utilisant différentes méthodes de traitement statistique de classification des données spectrales développées récemment, il est possible d'optimiser des modèles de discrimination et d'identification. Par sa simplicité de mise en œuvre, cette méthode représente un véritable progrès alliant gain de temps, fiabilité, spécificité et sensibilité. De plus, cette technique est non destructive, peu coûteuse et nécessite aucun réactif ou consommable pour l'analyse elle-même.

Bien que la spectroscopie IRTF soit une technique rapide et simple, cette méthode présente quelques limites. La première limite est la standardisation des protocoles et la qualité des banques de données spectrales (61). Il est indispensable de normaliser les conditions de culture, la procédure de préparation des échantillons et de normaliser les conditions d'enregistrement des spectres (62). Les conditions de culture comme la durée et la température d'incubation, le milieu de culture, le pH ainsi que les techniques d'échantillonnage doivent être standardisées et optimisées. Des études montrent l'impact des conditions environnementales et du stress chez les moisissures *Aspergillus nidulans* ASK 30, *Rhizopus ssp* et *Neurospora ssp*. Ces études sont basées sur l'analyse, par microspectroscopie IRTF, des changements subcellulaires en fonction du pH et de la température (63, 64). Les résultats sont exprimés en fonction des conditions de culture et du stade de développement. La durée d'incubation varie d'une étude à l'autre en fonction du matériel biologique utilisé pour l'analyse. Celle-ci est d'environ 14 jours lorsque que les spores sont utilisées comme matériel biologique. Lorsque le mycélium est utilisé comme matériel biologique, l'analyse peut être réalisée après 5 jours de culture en milieu liquide (60). Dans ce contexte, la mise au point d'un protocole standardisé et optimisé pour l'analyse des champignons filamenteux par spectroscopie IRTF pour un usage en routine dans un environnement industriel reste une étape essentielle. La préparation des échantillons doit être aussi simple que possible avec un minimum de temps de culture possible de souches fongiques répondant au mieux aux attentes des industriels.

La deuxième limite est la non-exhaustivité des banques de données spectrales (44). De nombreuses applications nécessitent des bases de données larges et complètes, dans les domaines agroalimentaire, clinique, pharmaceutique ou environnemental. Le nombre d'espèces est gigantesque et les bases de données de référence doivent faire face à cette

diversité. La nécessité de couvrir un nombre élevé de différents genres et d'espèces pour la création de bibliothèques est impératif (34). Des bases de données IR de référence pour l'identification en routine sont déjà disponibles (Bruker Optics). Ces bibliothèques contiennent des milliers de spectres de différentes espèces et souches de bactéries comme par exemple *Staphylococcus*, *Enterococcus*, *Pseudomonas*, *Bacillus*, *Clostridium* et de levures (40). En revanche, il reste indispensable de constituer des librairies spectrales spécifiques de référence à l'aide de la spectroscopie IR pour l'identification des champignons filamenteux les plus fréquemment rencontrés dans le secteur agroalimentaire, médical, pharmaceutique et cosmétique. De plus, lorsqu'un échantillon est confronté à une banque de données spectrales au sein de laquelle il n'y a pas d'homologue de cet échantillon, celui-ci sera attribué à l'espèce la plus proche conduisant à un mauvais résultat. Par conséquent, il est important d'établir un score et un seuil de prédiction afin de valider ou d'invalider le résultat. Cette technique a déjà été développée pour d'autres méthodes d'identification basées sur des banques de données spectrales telles que MALDI-TOF MS (matrix-assisted laser desorption ionization time of flight mass spectroscopy) (65, 66).

Enfin, la troisième limite de la spectroscopie IR concerne le transfert de la méthode, et plus particulièrement le transfert de la banque de données d'un appareil à un autre appareil de même type (67). En effet, les caractéristiques de différents appareils de même type peuvent donner des résultats différents selon les conditions dans lesquelles les mesures sont réalisées indépendamment de l'autocontrôle et de l'auto-étalonnage effectués par certains instruments (68). En spectroscopie IR, la stabilité de l'énergie, la reproductibilité des nombres d'onde, les variations de pression et de température, et le taux d'humidité et de dioxyde de carbone dans l'atmosphère, peuvent influencer la qualité spectrale (69). Cela peut entraîner un changement dans la réponse spectrale et l'utilisation de la calibration initialement mesurée sur un instrument donné peut nécessiter une correction. Une banque de données spectrales réalisée sur un instrument et les différents modèles de prédiction associés à cette banque de données peuvent donner de bons résultats de prédiction et une identification correcte pour un échantillon inconnu lorsque celui-ci est analysé l'instrument utilisé pour la réalisation de la banque de données. Si cet échantillon inconnu est analysé sur un instrument différent de celui utilisé pour réaliser la banque de données, les résultats de l'identification peuvent être de qualité inférieure. Le transfert d'une fonction de standardisation d'un instrument de laboratoire de recherche à un instrument d'un autre laboratoire de recherche ou un instrument utilisé dans un contexte industriel est donc nécessaire afin de mieux optimiser les résultats de prédiction d'un échantillon inconnu (70, 71).

Chapitre II : Méthodologie

II.1- Préparation des échantillons de champignons filamenteux

II.1.a- Souches de champignons filamenteux utilisées

Les souches de champignons filamenteux sélectionnées pour l'étude font partie de la collection de l'Université de Bretagne Occidentale (UBOCC, Plouzané, France) et de la collection du Centraalbureau voor Schimmelcultures (CBS, Utrecht, Pays-Bas). Au total, 498 souches, appartenant à 45 genres et 140 espèces, ont été analysées (Annexe 1).

L'ensemble des manipulations de ces souches, qui sont réalisées sous hotte à flux laminaire, le sont à proximité d'un bec bunsen afin de limiter au maximum les risques de contamination.

II.1.b- Conservation des souches de champignons filamenteux

Les souches de champignons filamenteux sont conservées à -80°C soit sous forme de cryobilles soit sous forme d'implants.

Pour conserver les souches de champignons filamenteux par cryoconservation en utilisant des cryobilles, 5 ml de glycérol à 10 % sont déposés à la surface d'un tube de gélose en pente contenant une culture fongique. Le tube est agité au vortex afin de mettre les spores en suspension. Ensuite, 2 ml de suspension sont récupérés au sein d'un tube Eppendorf et une centrifugation de 12 minutes à 3500 g est réalisée. Le surnageant est éliminé jusqu'à ce qu'il ne reste plus qu'environ 0,5 ml de celui-ci, le culot est réhomogénéisé et 0,5 ml sont ensuite prélevés à l'aide d'une pipette de transfert et déposés au sein d'un cryotube contenant 25 cryobilles (AES Chemunex/Biomérieux). Le cryotube est homogénéisé par retournement et après un temps de d'attente de 30 secondes, le surnageant est éliminé à l'aide d'une pipette de transfert. Le cryotube est ensuite stocké à -80°C (Figure 18).

Pour conserver les souches de champignons filamenteux par cryoconservation en utilisant des implants, des petits carrés de gélose de 0,5 sur 0,5 cm sont découpés à partir d'une gélose sur boîte de Pétri contenant à sa surface une culture fongique. Les petits carrés de gélose sont ensuite transférés au sein d'un cryotube et 1 ml de glycérol à 10 % est ajouté. Le cryotube est ensuite stocké à -80°C (Figure 19). Cette méthode de cryoconservation est essentiellement utilisée pour les souches ne sporulant que très peu et/ou très lentement.

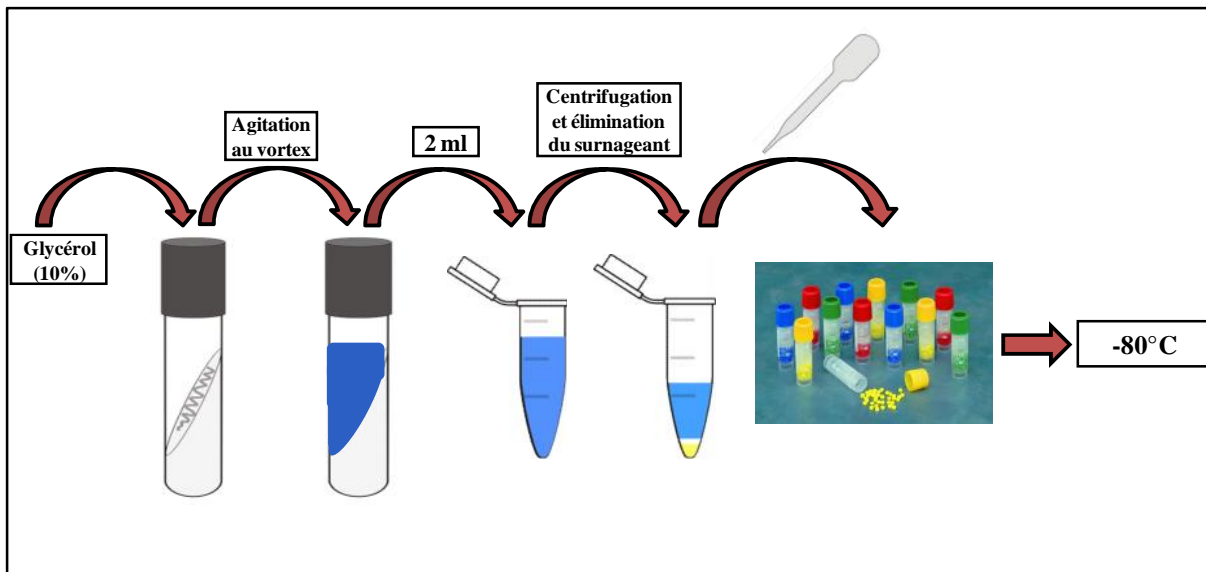


Figure 18 : Schématisation du protocole de conservation des champignons filamenteux par cryoconservation à l'aide de cryobilles.

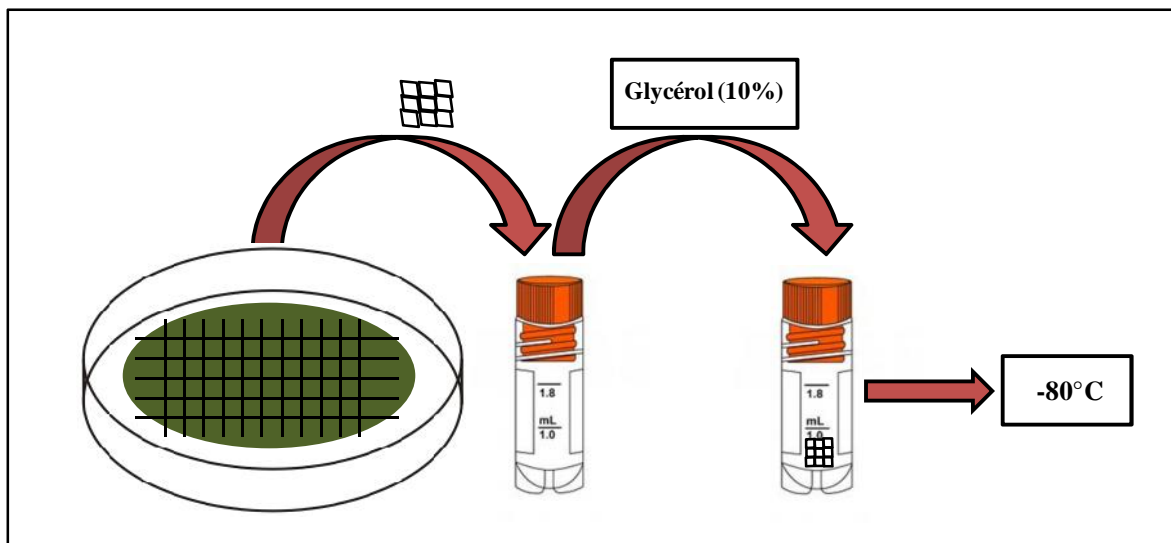


Figure 19 : Schématisation du protocole de conservation des champignons filamenteux par cryoconservation à partir d'implants de gélose.

II.1.c- Mise en culture des champignons filamenteux à partir de cryobilles et d'implants

La mise en culture des souches à partir des cryobilles se fait en prélevant une cryobille au sein du cryotube à l'aide d'une oëse stérile. La cryobille est ensuite déposée dans un tube contenant un milieu gélosé de Sabouraud (Becton Dickinson) et est roulée sur la surface inclinée du milieu gélosé au moyen de l'oëse afin que les spores se déposent sur la totalité de

la surface, et elle est laissée dans le tube. Les différents tubes sont ensuite incubés à 25°C dans une étuve pendant 4 à 7 jours en fonction des souches. Chaque jour, les tubes sont vérifiés pour le suivi du développement des champignons (Figure 20). Les souches sont repiquées sur gélose Sabouraud (Becton Dickinson) inclinée à partir de colonies présentes sur gélose Sabouraud (Becton Dickinson) inclinée à l'aide d'une oëse stérile. Les différents tubes sont ensuite incubés à 25°C dans une étuve pendant 4 à 7 jours, selon les souches. Chaque jour, les tubes sont examinés pour le suivi du développement des champignons.

La mise en culture des souches à partir des implants se fait en déposant, en conditions stériles, les implants d'une même souche sur une feuille de papier filtre stérile. Les implants sont ensuite séchés sur le papier filtre, en les déplaçant sur celui-ci, de manière à enlever l'excédent de glycérol. Les implants de gélose sont ensuite déposés à la surface d'un milieu gélosé incliné Sabouraud (Becton Dickinson) contenu au sein d'un tube au moyen d'une oëse stérile et les implants sont laissés au sein du tube. Les différents tubes sont ensuite incubés à 25°C dans une étuve pendant 4 à 7 jours selon les souches. Chaque jour, les tubes sont examinés pour le suivi du développement des champignons (Figure 20). Les souches sont repiquées sur gélose Sabouraud (Becton Dickinson) inclinée à partir de colonies présentes sur gélose Sabouraud (Becton Dickinson) inclinée à l'aide d'une oëse stérile. Les différents tubes sont ensuite incubés à 25°C dans une étuve pendant 4 à 7 jours, dépendant des souches. Chaque jour, les tubes sont observés pour le suivi du développement des champignons.

Après incubation et si la souche est pure, les tubes sont conservés à 4°C

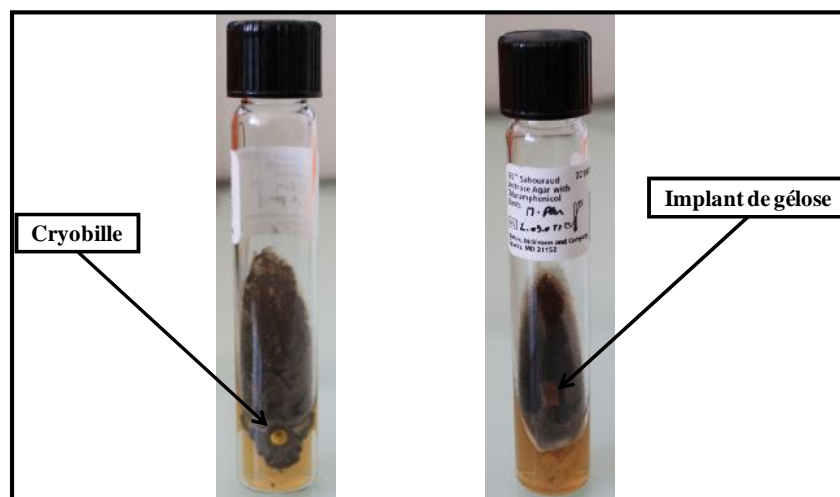


Figure 20 : Colonies de moisissures obtenues sur tubes de gélose Sabouraud inclinée à partir d'une cryobille (à gauche) ou d'un implant de gélose (à droite).

II.1.d- Extraction d'ADN, amplification, séquençage et assignement taxonomique des souches de champignons filamenteux

L'ADN génomique total des souches de champignons filamenteux a été extrait en utilisant le « FastDNA Kit SPIN » (MPBio, Illkirch) et selon les instructions du fabricant. L'ADN a été extrait à partir de mycélium obtenu en réalisant des cultures au sein d'un bouillon PDB (potato dextrose broth) pendant 2 à 4 jours à 25 °C et sous agitation à 120 rpm. Cinq régions différentes ont été amplifiées en fonction des genres fongiques analysés.

Pour toutes les souches, excepté celles appartenant au genre *Fusarium*, l'ADN ribosomal de la région ITS (Internal Transcribed Spacer), incluant le gène ARNr 5,8S, a été amplifié. La région ITS est considérée comme la région de référence par les taxonomistes pour l'identification des espèces fongiques (72). En effet, une partie de cette région est très conservée chez la majorité des espèces fongiques et une partie de cette région présente une variabilité pouvant être utilisée pour la phylogénie des champignons filamenteux.

La région correspondant au gène de la β -tubuline a été amplifiée chez les souches appartenant aux genres *Penicillium* et *Aspergillus*. Les séquences de cette région sont riches en introns où le taux de variabilité semble approprié pour une bonne discrimination au sein de genres conservés tels que *Penicillium* et *Aspergillus* (73).

Pour les souches appartenant au genre *Fusarium*, une amplification de la région correspondant au gène du facteur d'élongation de la traduction 1 (TEF-1) a été réalisée (74).

Puis pour les souches appartenant au genre *Mucor*, la région correspondant au gène *mcm7* codant pour la protéine MCM7 (Mini Chromosome Maintenance) nécessaire pour la réplication de l'ADN et la prolifération cellulaire (75) a été amplifiée ainsi que la région correspondant au gène *tsr1* codant pour la protéine TSR1 (Twenty SrRNA accumulation) nécessaire pour l'accumulation des ARNr pendant la synthèse des ribosomes (76).

Les amorces utilisées pour l'amplification des différentes régions ITS, β -tubuline, TEF-1, *mcm7* et *tsr1* sont : ITS4 et ITS5 (77), BT2A et BT2B (78), EF1F et EF1R (79), MCM7-709for et MCM7-1348rev (80) et TSR1-f1 et TSR1-r2 (81). Le séquençage des amplicons obtenus a été réalisé au sein de la plate-forme Biogenouest de la « Station Biologique de Roscoff » et en utilisant les mêmes paires d'amorces. L'assemblage des séquences a été réalisé à l'aide du logiciel DNA Baser (Heracle Software, Allemagne). Ensuite, les séquences ont été comparées à la base de données GenBank en utilisant la méthode de recherche BLAST (basic local alignment search tool) afin de trouver des régions

ayant des zones de similitude entre deux ou plusieurs séquences. Des alignements des séquences obtenues et des séquences de la base de données NCBI ont été effectués en utilisant le serveur MAFFT (version 7) qui est un programme d'alignement multiple de séquences et en utilisant la méthode de raffinement itératif E-INS-i. Les arbres phylogénétiques ont ensuite été construits dans MEGA5 (82) en utilisant la méthode Neighbor-Joining avec 1000 jeux de données bootstrap.

II.1.e- Préparation des cultures liquides des champignons filamenteux

A l'aide d'une oëse stérile, la surface des colonies présentes au sein des tubes gélosés est grattée, puis les spores et/ou le mycélium sont resuspendus dans 20 ml de chemboost YM (AES Chemunex/Biomérieux) contenu dans le flacon d'origine. Les cultures sont incubées 48h exactement, à 25°C dans une étuve rotative à 150 rpm prévue à cet effet. Chaque souche fait l'objet de 3 cultures (réplicats biologiques) effectuées sur 3 jours différents (J1, J2 et J3) afin de vérifier la reproductibilité de l'étude. Chaque jour, un flacon de chemboost YM (AES Chemunex/Biomérieux) non ensemencé est incubé dans les mêmes conditions comme témoin négatif afin de s'affranchir d'éventuelles contaminations lors de la période d'incubation ou une éventuelle contamination des milieux de culture.

II.1.f- Vérification de la pureté des cultures de mycélium

A l'issue de la période d'incubation des cultures en milieu liquide de 48 h, la pureté des cultures liquides est vérifiée, pour deux à trois souches, en transférant du mycélium de ces cultures sur une boîte de Pétri contenant de la gélose PDA (AES Chemunex/Biomérieux) à l'aide d'une oëse stérile. Les boîtes de Pétri sont incubées pendant 7 jours à 25°C et la pureté est vérifiée à l'issue de cette période d'incubation.

II.1.g- Préparation des suspensions de mycélium

La première étape de la préparation des suspensions de mycélium consiste en un broyage du mycélium obtenu après 48 h de culture. Cette étape est effectuée dans le but d'obtenir une suspension homogène nécessaire pour la réalisation des suspensions sur la plaque de silicium, support utilisé pour l'analyse spectrale. Les cultures sont transférées dans un tube de broyage (M-tube) adapté au broyeur Dispomix « gentleMACS Octo Dissociator »

(Miltenyi Biotec). Un cycle de broyage de 100 secondes à 4000 rpm par échantillon est ensuite effectué. Un second cycle de 40 secondes à 4000 rpm est parfois réalisé dans le cas où le broyat n'est pas homogène, selon les souches de champignons filamenteux analysés. La seconde étape de la préparation des échantillons consiste en un lavage des suspensions de mycélia broyés. Pour ce faire, 2 ml de suspension de mycélia broyés sont transférés au sein d'un tube Eppendorf. Une première centrifugation de 30 secondes à 430 g est réalisée afin d'éliminer le milieu de culture en veillant à ne pas perdre de mycélium. Les culots de mycélium sont resuspendus dans 1 ml d'eau physiologique à 0,9 %. Chaque échantillon est ensuite passé au vortex de manière à resuspendre les mycélia. Puis une deuxième centrifugation de 30 secondes à 430 g est effectuée dans le but de laver les mycélia broyés, le surnageant est éliminé en veillant à ne pas perdre de mycélium, et les culots de mycélia sont repris dans environ 300 µl d'eau physiologique à 0,9 %. Les échantillons sont passés au vortex afin d'obtenir des suspensions homogènes avant l'analyse spectrale (Figure 21).

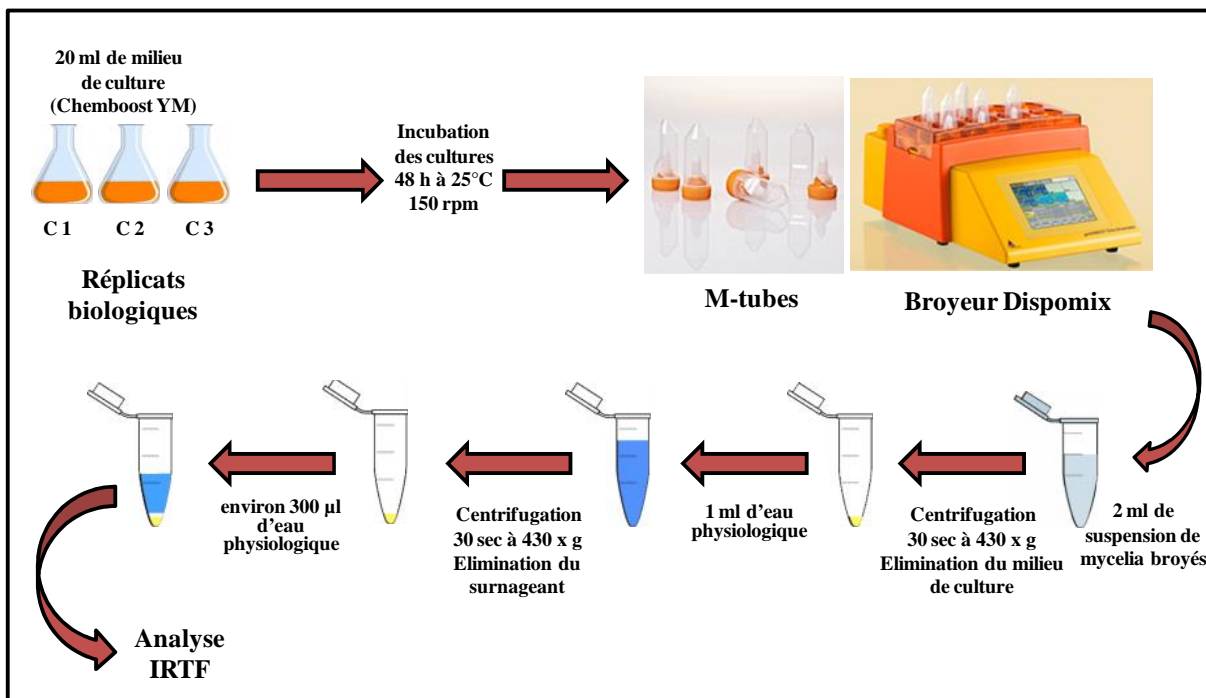


Figure 21 : Schématisation du protocole de préparation des suspensions de mycélium.

II.2- Analyse spectrale des échantillons de champignons filamenteux

II.2.a- Acquisition spectrale

Cinq microlitres de chaque échantillon sont déposés sur une plaque de silicium de 384 puits en sur 8 puits dans le but de vérifier la répétabilité de l'étude. Ainsi, huit réplicats instrumentaux sont réalisés par échantillon. Les différents échantillons sont passés au vortex entre chaque dépôt afin d'homogénéiser les suspensions. La plaque est ensuite placée au sein d'un dessiccateur (Schott) pendant 1 heure afin d'éliminer l'excédent d'eau au sein des dépôts, correspondant à l'eau extracellulaire. Un spectromètre IRTF Tensor 27 couplé au module à haut débit HTS-XT (Bruker Optics) est utilisé pour l'acquisition spectrale (Figure 22). L'enregistrement des spectres s'effectue via le logiciel OPUS 6.5 (Bruker) en définissant les différents paramètres d'acquisition suivants : un nombre d'accumulations de 64 par dépôts, une résolution spectrale de 4 cm^{-1} , une gamme spectrale d'acquisition utilisée comprise entre 4000 et 400 cm^{-1} et un facteur de zéro filling de 2. Un spectre du support correspondant au background de la plaque de silicium est enregistré dans les mêmes conditions avant chaque enregistrement des spectres des différents échantillons dans le but de réduire l'influence des signaux parasites dus à l'humidité ambiante et au CO_2 atmosphérique.

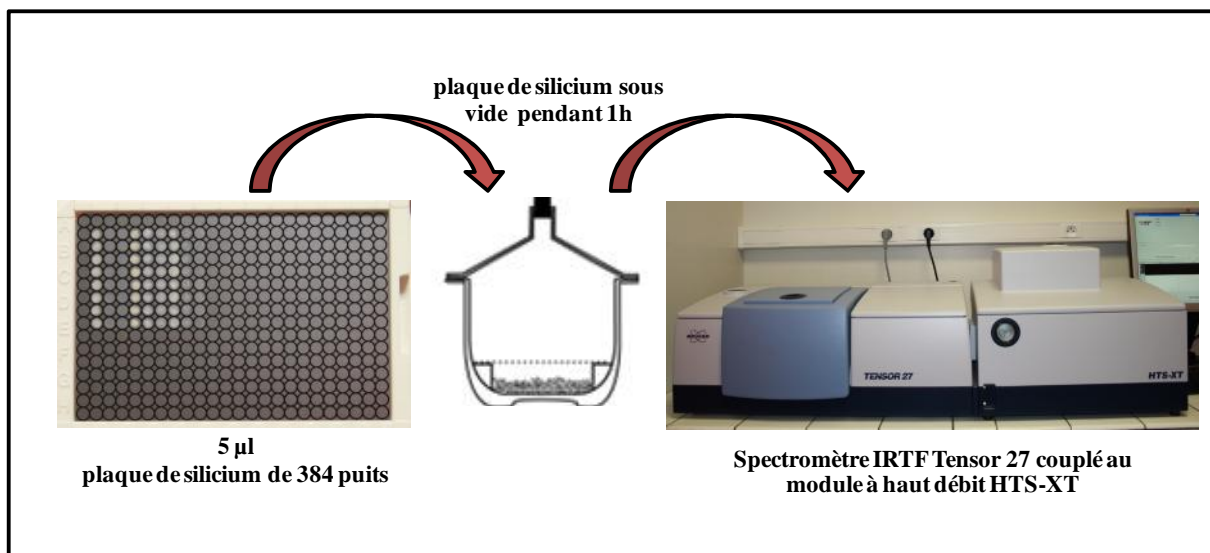


Figure 22 : Schématisation du protocole de l'analyse spectrale des souches de champignons filamenteux.

II.2.b- Test qualité et prétraitements des spectres infrarouge

- Test qualité

Avant d'analyser les spectres IRTF, il est important d'effectuer une série de tests afin de vérifier leur qualité et de permettre de valider ou non les spectres. Les spectres bruts subissent donc ces tests qui sont effectués à l'aide du logiciel OPUS (version 5.5) et qui sont basés sur les tests qualité développés par le groupe de Helm (83).

Dans la région $1600-2100\text{ cm}^{-1}$, l'absorbance doit être comprise entre 0,17 et 1 (unités arbitraires). Le rapport signal sur bruit (S/N) doit être suffisant. Il est calculé sur la dérivée première du spectre. Deux régions sont utilisées pour la mesure du signal, celle avec la plus forte absorbance comprise entre $1600\text{ et }1700\text{ cm}^{-1}$ (valeur S1), et celle comprise entre $960\text{ et }1260\text{ cm}^{-1}$ (valeur S2). L'intensité du bruit (valeur N) est définie sur la région $2000-2100\text{ cm}^{-1}$, région ne présentant pas de pic d'absorption. Le signal du bruit dans cette région doit être inférieur à 0,00016. Un spectre de bonne qualité est déterminé pour une valeur de S1/N supérieure à 50 et une valeur de S2/N supérieure à 10. L'eau résiduelle doit être minimisée. Elle est quantifiée par les rapports du signal S1 et S2 sur la vapeur d'eau (valeur W). La vapeur d'eau est mesurée dans la zone spectrale comprise entre $1837\text{ et }1847\text{ cm}^{-1}$ où se trouvent les bandes d'absorption de la vapeur d'eau. L'intensité d'absorption dans cette région doit être inférieure à 0,0003. Un spectre de bonne qualité est déterminé pour une valeur de S1/W supérieure à 20 et de S2/W supérieure à 4 (Figure 23).

Les spectres bruts ayant une ou plusieurs valeurs en dehors des seuils du test de qualité sont automatiquement retirés de l'ensemble de données. A la fin de cette étape, il est possible d'avoir un maximum de 24 spectres par souche (3 cultures par souches et 8 dépôts par culture). Pour certaines souches, ce nombre est réduit en raison de l'élimination de certains spectres après le test de qualité. Pour chaque culture, l'analyse est validée si au moins 5 spectres sur 8 ont passé le test de qualité. Par conséquent, pour chaque souche le nombre minimum de spectres est de 15.

Quality-Test

Select Files | Load / Save Method | Parameters

	X-Range 1	X-Range 2	X-Range 3	X-Range 4	X-Range 5	X-Range 6	0 Test
x-Start:	2100	1700	1200	2100	1847	2300	
x-End:	1600	1600	960	2000	1837	2000	
Delta y min:	0.17						
Delta y max:	1						
Noise max:				0.00016			
Signal/Noise:		50	10				
Water max:					0.0003		
Signal/Water:		20	4				
Fringes:						0.001	

Quality-Test Cancel Help

Figure 23 : Récapitulatif des gammes spectrales et des valeurs utilisées pour les tests qualité.

- Prétraitements des spectres infrarouge

Les prétraitements ont pour objectifs l'amélioration du signal et l'homogénéisation des données. Ces prétraitements sont nécessaires pour une meilleure analyse des différentes données spectrales. En effet, le signal correspondant à l'échantillon analysé peut être parasité par du signal correspondant à du bruit aléatoire, à des variations de la ligne de base entraînant des déformations du spectre, à des variations d'échelle incontrôlées de l'intensité générale entre les différents spectres et à la redondance de l'information qui doit être réduite. Les prétraitements sont réalisés en utilisant des fonctions mathématiques du logiciel OPUS 5,5 (Bruker) et sont appliqués aux spectres bruts de la manière suivante (Figure 24).

Tout d'abord, les spectres IR ont été tronqués dans la région $4000-800 \text{ cm}^{-1}$, région qui comprend la majorité des informations biochimiques de l'échantillon.

Ensuite, une correction de la ligne de base est appliquée de façon indépendante sur chaque spectre. La correction de la ligne de base permet d'éliminer les distorsions et les

dérives dues aux effets physiques. Lors de l'analyse d'un échantillon par spectroscopie infrarouge en mode transmission, un phénomène d'absorption de la lumière par l'échantillon est observé. D'autres phénomènes optiques peuvent également être observés comme le phénomène de diffusion et de diffraction de lumière, ainsi que des aberrations chromatiques. Une partie de la lumière transmise est alors déviée et celle-ci n'est donc pas détectée. Ces phénomènes, entraînant une variation de la ligne de base, dépendent de la longueur du trajet optique (longueur d'onde) à travers l'échantillon et des propriétés physiques de l'échantillon comme la taille et l'épaisseur des particules et leur distribution au sein de l'échantillon. Ces phénomènes entraînent une déformation de la ligne de base des spectres. Cette dérive peut être rectifiée en procédant à la correction de la ligne de base. La correction de ligne de base consiste à modéliser, sous la forme d'équations, les différentes variations souvent retrouvées dans les régions où il n'y a aucune bande d'absorption. La modélisation de ces variations spectrales est réalisée à partir de quelques points du spectre. Le nombre de points est défini par l'opérateur et la ligne de base passant par ces points est assimilée à une fonction polynômiale. L'allure de cette ligne de base est dépendante du degré du polynôme choisi (84). Les variations spectrales ainsi modélisées sont ensuite soustraites point par point du signal observé. Dans cette étude, une correction de la ligne de base élastique a été réalisée sur l'ensemble des spectres en choisissant un nombre de point de 64 et cette correction a été effectuée sur la région spectrale 4000-800 cm^{-1} .

La dérivée seconde de chaque spectre a ensuite été calculée. La dérivation des spectres permet d'accroître les différences spectrales, c'est-à-dire, rendre certaines informations du spectre plus distinctes en augmentant la résolution spectrale. Ainsi, les pics peu résolus sur les spectres bruts peuvent être dévoilés sur la dérivée seconde des spectres (85). La dérivation des spectres permet également de réduire la dérive de la ligne de base observée au sein des spectres (86). Si la dérivée seconde est positive sur un intervalle, cela veut dire que la pente augmente, que la courbure est vers le haut, et la fonction est alors dite « convexe » sur cet intervalle. A l'inverse, si la dérivée seconde est négative sur un intervalle, cela veut dire que la pente diminue, que la courbure est vers le bas, et la fonction est alors dite « concave » sur cet intervalle. Les maxima d'absorption des spectres sont mieux résolus en calculant la dérivée seconde mais ceux-ci possèdent une intensité négative. La dérivée seconde permet donc de mesurer la concavité des spectres. La dérivation des spectres entraîne une perte du rapport entre la concentration de l'échantillon et l'intensité ou absorbance des spectres. Dans cette étude, la dérivée seconde des spectres est calculée à l'aide de l'algorithme de Savitsky-Golay (87) comprenant un filtre de 9 points de lissage sur la gamme spectrale 4000-800 cm^{-1} .

Grâce à la fonction de lissage inhérente à cette fonction, le bruit, préalablement amplifié par le calcul de la dérivée seconde, est considérablement réduit.

Enfin, une normalisation vectorielle des dérivées secondes est réalisée. La normalisation vectorielle est une opération mathématique qui consiste à ramener les spectres à la même intensité (ou même échelle) afin de mieux pouvoir les comparer aussi bien sur le plan quantitatif que sur le plan qualitatif. La méthode de normalisation vectorielle calcule d'abord la moyenne des valeurs d'absorbance (y) du spectre dans la gamme sélectionnée. Cette moyenne est ensuite soustraite à chaque valeur du spectre, de manière à positionner le milieu du spectre à $y = 0$. Ensuite, le logiciel calcule la somme des carrés des ordonnées (y), puis divise le spectre par la racine de cette somme (88). Après la normalisation des spectres, seules les intensités relatives peuvent être comparées. La normalisation peut être faite sur toute la gamme spectrale ou sur une fenêtre spectrale choisie par l'opérateur. Dans cette étude, la normalisation vectorielle a été réalisée sur la fenêtre spectrale $4000\text{-}800\text{ cm}^{-1}$.

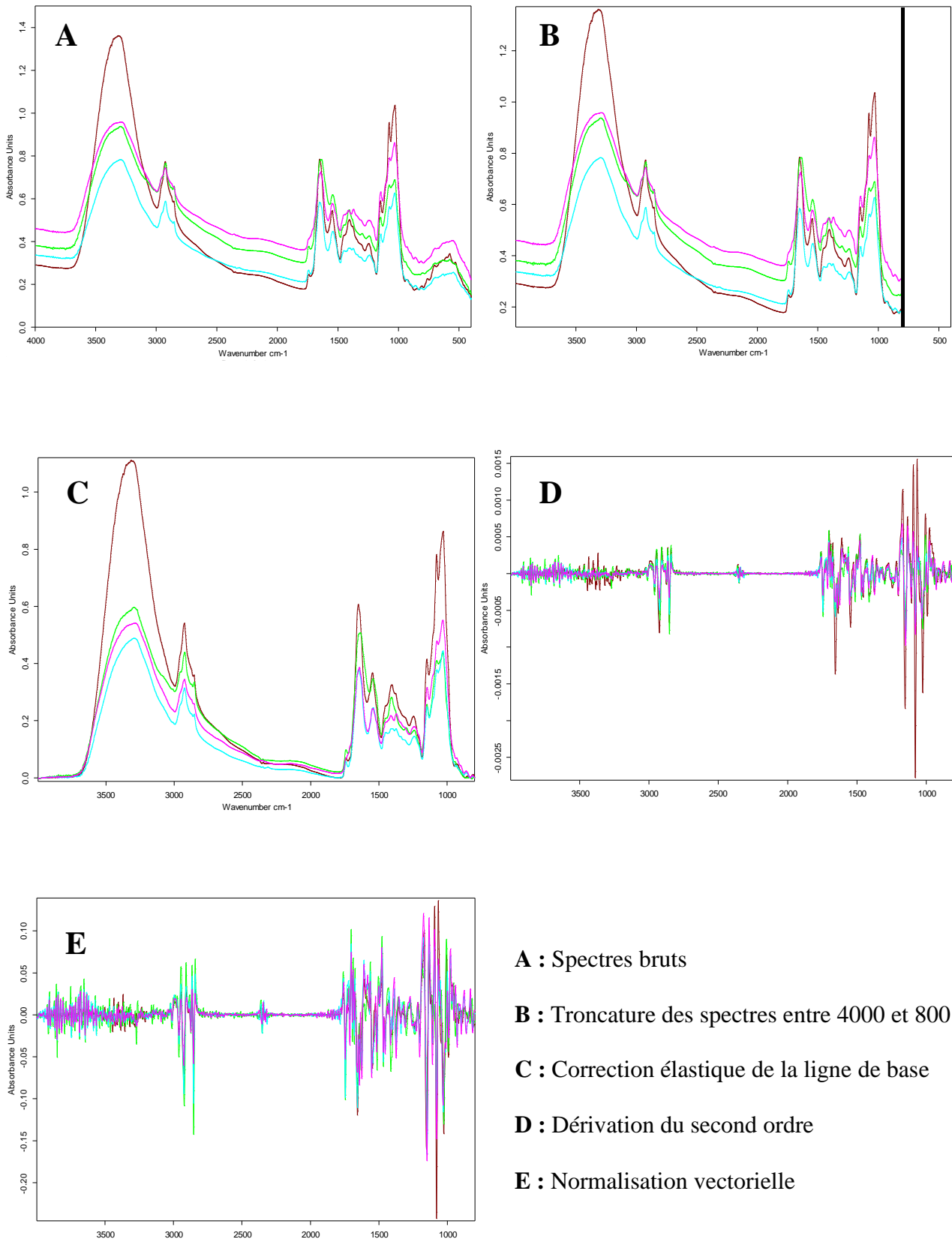


Figure 24 : Illustration des prétraitements appliqués aux spectres infrarouge.

II.3- Analyses statistiques des données spectrales

II.3.a- Analyse discriminante des moindres carrés partiels : PLS-DA

L'analyse chimiométrique utilisée dans notre étude est l'analyse discriminante des moindres carrés partiels (PLS-DA : partial least square discriminant analysis). Cette analyse est effectuée à l'aide du logiciel Matlab (version 7.2, Mathwork, USA). Cette méthode de classification est une méthode d'analyse linéaire supervisée basée sur l'algorithme de régression PLS. La régression PLS, méthode de modélisation linéaire, a été développée à l'origine pour l'analyse quantitative mais elle peut également être utilisée pour l'identification d'un échantillon, c'est-à-dire pour la prédiction d'appartenance d'un échantillon inconnu à un groupe de référence (89). L'adaptation de la régression PLS comme méthode de classification est réalisée en utilisant un code binaire. Chaque groupe est codé par une combinaison de 0 et de 1, sur la base de si oui ou non l'échantillon appartient à ce groupe.

La PLS-DA permet de mettre en relation les données des échantillons appelées « variables explicatives » correspondant à des mesures expérimentales et les différents groupes des échantillons appelés « variables à expliquer ». Dans notre étude, les variables explicatives correspondent aux valeurs d'absorbances spectroscopiques IR et les variables à expliquer correspondent aux différents rangs taxonomiques des échantillons. L'algorithme PLS-DA permet donc de corrélérer de manière mathématique les données explicatives avec une matrice de propriétés des différents groupes des échantillons à l'aide d'une régression multivariée (90). Cet algorithme est utilisé pour la construction de modèles de régression, appelé aussi modèles de prédiction, permettant de prédire au mieux les variables à expliquer à partir des variables explicatives d'échantillons inconnus.

La PLS-DA a été appliquée sur les fenêtres spectrales 3200-2800 et 1800-800 cm^{-1} . Ces deux larges gammes spectrales sont associées aux différents niveaux d'énergie IR des différentes fonctions biochimiques présentes au sein des échantillons. Les spectres ont donc été classés en prenant en compte l'ensemble de l'information moléculaire présente au sein de chaque échantillon.

L'élaboration des modèles de régression pour la prédiction d'échantillons inconnus se fait selon les étapes suivantes (Figure 25) :

Etape 1 :

La première étape consiste dans un premier temps à sélectionner des spectres afin de constituer un jeu d'échantillons de référence ou de calibration pour la construction des différents modèles de calibration. Ensuite, les valeurs à expliquer ou les groupes sont déterminés pour le jeu d'échantillon de référence. Puis, une matrice des indicateurs est créée selon un code binaire avec un bit par groupe.

Etape 2 :

La deuxième étape consiste en la construction des modèles de régression appelés matrice des vecteurs de régression à partir du jeu d'échantillons de référence (matrice de données) et à l'aide de la matrice des indicateurs. La matrice des vecteurs de régression représente la projection des spectres dans un nouvel espace permettant de maximiser la covariance entre la matrice de données et la matrice des indicateurs. Un vecteur de régression comporte autant de variables qu'il y a de variables au sein de la matrice de données, c'est-à-dire de nombres d'onde, pour une dimension donnée. Une matrice des scores A est créée, représentant la projection de la matrice de données à l'aide du vecteur de régression construit pour une dimension donnée.

Plusieurs itérations de la construction des modèles de régression et de la matrice des scores en fonction du nombre de dimension sont alors réalisées. En effet, la régression PLS est itérative et permet d'arrêter la construction des modèles de régression lorsque la covariance entre la matrice de données et la matrice des indicateurs est maximum. Lorsque le nombre d'itérations tend vers l'infini, la covariance entre la matrice de données et la matrice des indicateurs est maximisée. Le nombre de dimensions maximum est obtenu lorsque la matrice des scores tend à la matrice des indicateurs (Figure 26). L'erreur résiduelle à chaque itération correspond à la variance non prise en compte par les itérations précédentes. D'une itération à l'autre, les vecteurs de régression sont orthogonaux au précédent. Seule la variance qui n'a pas pu être prise en compte est projetée sur le nouveau vecteur.

Etape 3 :

La troisième étape consiste tout d'abord à la sélection de spectres constituant un jeu d'échantillons inconnus appelé jeu de validation externe. Ce jeu est constitué de spectres dont l'identification est à prédire. Une matrice de scores B, représentant la projection de la matrice de données du jeu d'échantillons inconnus à l'aide de la matrice des vecteurs de régression construits pour l'ensemble des dimensions, est créée. La matrice des scores B nous donne un résultat de prédiction pour l'ensemble des spectres du jeu d'échantillons inconnus.

Etape 4 :

La quatrième étape correspond à la validation externe des modèles de régression en confrontant la matrice des scores B du jeu d'échantillons inconnus aux valeurs à expliquer ou les groupes des échantillons de ce jeu.

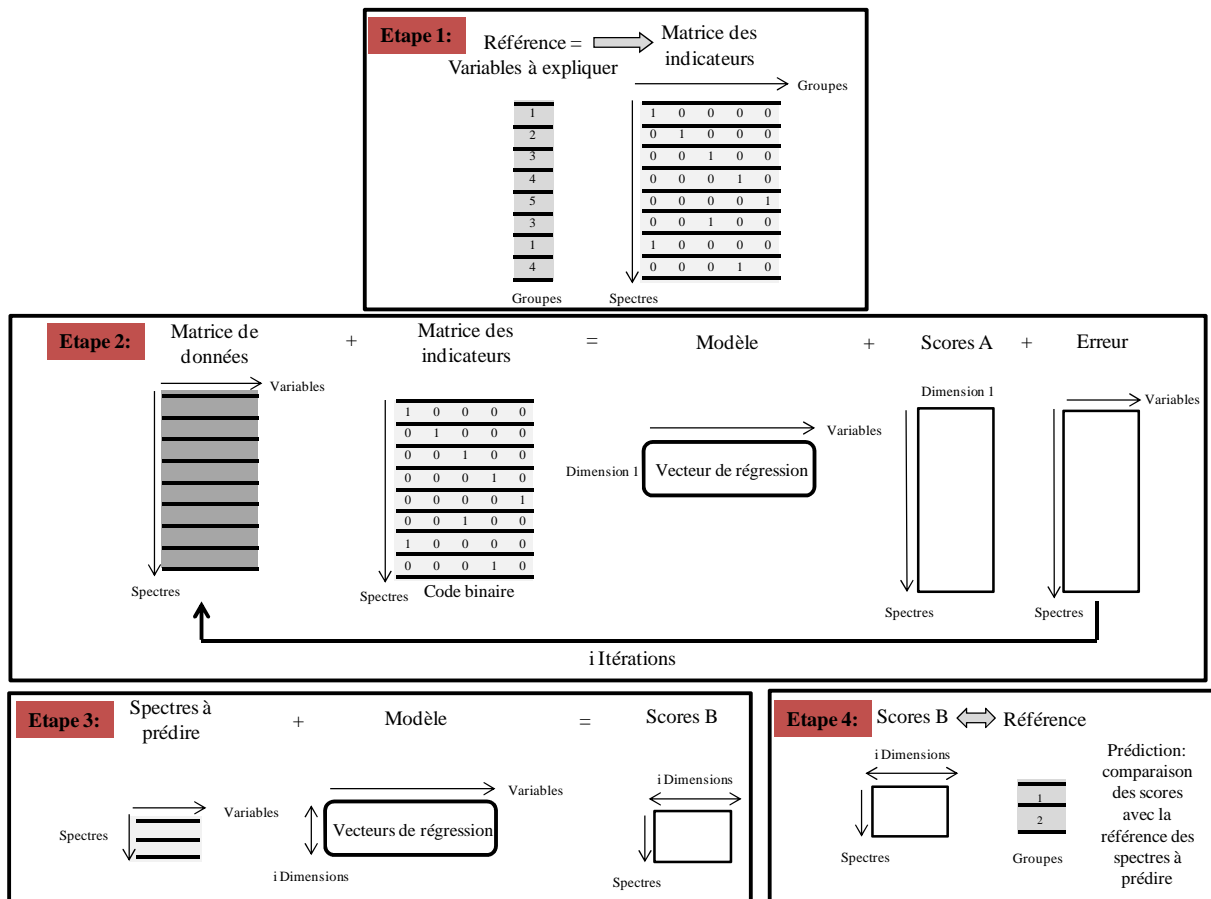


Figure 25 : Schématisation du principe de la PLS-DA : partial least square discriminant analysis (A : calibration, B : validation).

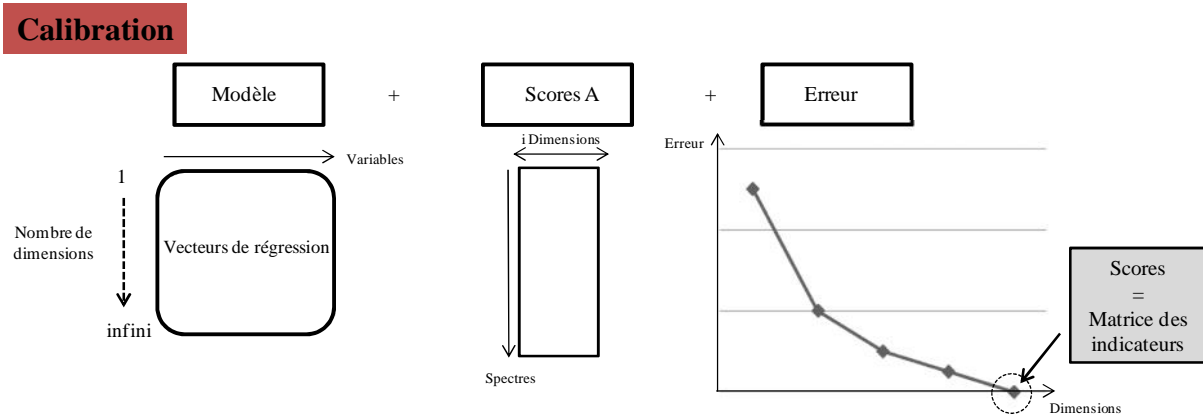


Figure 26 : Schématisation de la régression PLS (A : calibration).

Dans notre étude, les souches fongiques ont été attribuées à un groupe prédéfini sur la base des résultats de l'identification moléculaire et en fonction de la taxonomie actuelle (Annexe 2). Ainsi, des modèles de régression ont été construits à partir des spectres du jeu d'échantillons de calibration et ceci pour chacun des rangs taxonomiques suivants, auxquels appartiennent les souches fongiques: division / sous-division, classe, ordre, famille, genre, sous-genre, section, série et espèce.

II.3.b- Validation croisée

La méthode de validation croisée « leave-one-out » (91) a été utilisée pour évaluer la qualité des modèles d'étalonnage et fournir des informations sur les paramètres des modèles. Le principe de cette méthode est le suivant. Tous les spectres du jeu d'échantillons de calibration servent à la fois à la calibration et à la validation des modèles de régression. Dans notre étude, une validation croisée partielle a été réalisée, c'est-à-dire que chaque groupe de spectres correspondant à une même culture est écarté tour à tour, puis les différents modèles de régression sont construits avec les spectres restants. Les modèles ainsi créés sont ensuite testés par chaque spectre de la culture qui a été écarté. Ces différents spectres, représentant alors un jeu de validation interne, permettent d'estimer les caractéristiques des modèles de régression obtenus.

Cette méthode nous a également permis de déterminer le nombre optimal de dimensions à utiliser. En effet, le choix du nombre de dimension est un facteur essentiel. La dimensionnalité optimale correspond au nombre de dimensions donnant un minimum d'écart

entre les variables explicatives et les variables expliquées. Lorsque le nombre de dimensions choisi est trop faible, toute l'information de la matrice de données de départ nécessaire pour prédire les variables à expliquer d'un jeu de validation interne ou externe n'est pas prise en compte. Dans ce cas, on parle de phénomène d'underfitting ou de sous-information. A l'inverse, lorsque le nombre de dimensions sélectionné est trop élevé, de l'information non-explicative et contenant du bruit risque d'être prise en compte dans le calcul des valeurs à expliquer. On parle alors de phénomène d'overfitting ou de surinformation (Figure 27). Dans notre étude, la validation croisée partielle a été testé de façon cumulative de 1 à 35 itérations. Selon les différents modèles de régression, les nombre de dimensions qui ont fourni le meilleur pourcentage de prédiction des échantillons inconnus du jeu de validation interne ont été choisis pour la construction des modèles. Ces différents modèles ont ensuite servi à la validation externe. Cette validation externe a été réalisée à l'aide des spectres correspondant aux souches de champignons constituant le jeu de validation externe.

Validation (interne ou externe)

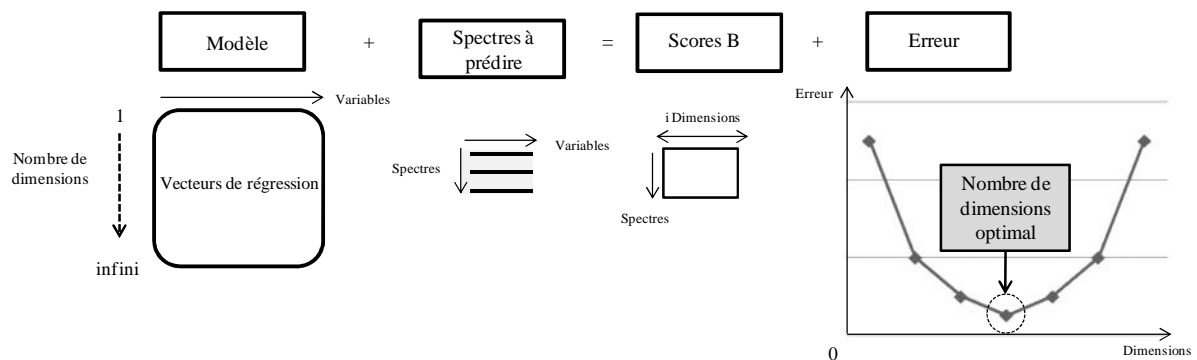


Figure 27 : Choix du nombre de dimensions utilisé par la méthode de validation croisée interne ou externe (B : validation).

II.4- Etablissement d'un score et d'un seuil de validation

Dans le but de valider les résultats de prédiction obtenus pour des échantillons inconnus, il est nécessaire d'établir un score et un seuil de validation afin de confirmer ou d'infirmer les résultats. Le calcul du score (S) pour chaque résultat de prédiction a été calculé avec Matlab (Version 7.2, Mathwork, USA). Dans notre étude, pour chaque souche dont l'identification est à prédire, 3 réplicats biologiques sont réalisés et pour chaque réplicat biologique, 8 réplicats instrumentaux sont effectués. Un total de 24 spectres est donc obtenu pour chaque souche inconnue. Pour chaque spectre, un résultat de prédiction est donné, et le résultat de prédiction de la souche correspond à celui obtenu majoritairement pour l'ensemble des spectres.

Le calcul du score est basé sur le calcul des distances Euclidiennes entre chaque spectre d'un échantillon inconnu à prédire et chaque spectre du cluster dans lequel l'échantillon a été prédit majoritairement. La moyenne des scores obtenus pour chaque spectre d'une même souche à prédire a été calculée. Ensuite, les scores ont été multipliés par la variable h, correspondant au pourcentage de spectres prédits majoritairement sur l'ensemble des spectres réalisés pour une souche, cette variable est définie comme l'homogénéité. Cette multiplication permet de pondérer les scores obtenus en fonction de l'homogénéité des résultats de prédiction. Les scores ainsi calculés sont compris entre 0 et 100.

Le calcul des scores a été réalisé pour deux jeux de spectres différents. Le premier jeu de spectres correspond à 105 souches, appartenant à 18 genres et 54 espèces, ayant un homologue dans la base de données spectrales. Le deuxième jeu de spectres correspond à 72 souches ne possédant pas d'homologue au sein de la banque de données spectrales, soit au niveau du genre pour 27 d'entre elles (17 genres et 27 espèces) soit au niveau de l'espèce pour 45 d'entre elles (17 genres et 45 espèces). En fonction du résultat d'identification par spectroscopie IRTF, des scores obtenus, et en fonction de l'identification de référence obtenue par séquençage moléculaire, un seuil de validation des résultats d'identification a été fixé.

$$S = (1-D) \times h$$

Avec:

S = Score

D = distance Euclidiennes

h = homogénéité

II.5- Etablissement d'une fonction de standardisation

Afin de confirmer la robustesse et la précision de la méthode d'identification de champignons filamenteux développée dans cette étude, la transférabilité de la méthode et donc de la banque de données d'un appareil IRTF à un autre a été vérifiée. Pour ce faire, 14 souches (3 genres et 7 espèces) ont été analysées sur deux spectromètres IRTF à haut débit. L'instrument 1 correspond à celui qui a été utilisé pour l'analyse des souches ayant servi à la construction des différents modèles de calibration et constituant la librairie de spectres, celui-ci se trouvant au laboratoire MÉDIAN de l'Université de Reims Champagne Ardenne. L'instrument 2 se trouve au sein du laboratoire du Lubem à l'Université de Brest, et les analyses effectuées sur cet appareil ont été réalisées par des opérateurs différents de ceux du site de Reims. La préparation des 14 souches ainsi que l'analyse spectrale ont été réalisées dans les mêmes conditions et en suivant le protocole développé et standardisé dans cette étude. Une fonction de standardisation (SF) a été calculée à partir de l'ensemble des spectres de ces 14 souches.

Dans un premier temps, pour chaque souche et pour les deux instruments, la médiane des spectres dérivés, dont le résultat de prédiction correspond à celui obtenu majoritairement pour une souche donnée, a été calculée. Ensuite, les 14 spectres médians obtenus pour les souches analysées sur l'instrument 1 ont été soustraits aux 14 spectres médians obtenus pour les souches analysées sur l'instrument 2. Puis la médiane de l'ensemble des médianes a été calculée. La qualité de la calibration de la fonction de standardisation a été vérifiée et testée par la méthode du leave-one-out cross validation. Tous les spectres d'une même souche et pour les deux instruments ont été écartés du jeu de calibration de la fonction de standardisation et la fonction a été calculée avec les spectres des 13 souches restantes. Ceci a été réalisé pour les 14 souches et au total 14 fonctions de standardisation ont été calculées. La précision des différentes fonctions de standardisation a été vérifiée à l'aide des spectres de chaque souche écartée. La fonction de standardisation, calculée à l'aide des spectres des 14 souches, a ensuite été appliquée sur un jeu de spectres correspondant à 7 nouvelles souches (2

genres et 5 espèces) analysées uniquement sur l'instrument 2 et n'ayant pas servi à l'implémentation de la fonction, dans le but de valider cette fonction.

$$\mathbf{SF} = \mathbf{median}(\mathbf{Inst}_1 - \mathbf{Inst}_2)$$

Avec :

Inst1 = matrice incluant les spectres dérivés médian de chaque souche analysée sur l'instrument 1

Inst2 = matrice incluant les spectres dérivés médian de chaque souche analysée sur l'instrument 2

Chapitre III : Résultats et discussion

III.1- Préambule

Ce chapitre comprend les résultats obtenus au cours de l'étude portant sur l'utilisation de la spectroscopie infrarouge à transformée de Fourier couplée à une méthode d'analyse chimiométrique pour la discrimination et l'identification des champignons filamenteux. Il se présente sous la forme de deux articles scientifiques.

L'article 1 porte d'une part sur la mise au point d'un protocole de préparation des souches de champignons filamenteux et d'un protocole d'analyse des différents échantillons par spectroscopie infrarouge à haut débit et d'autre part sur le développement d'une base de données spectrales à l'aide d'une méthode chimiométrique supervisée permettant la discrimination et l'identification des champignons filamenteux.

L'article 2 présente les résultats de l'analyse d'un plus grand nombre de souches et la construction d'une banque de données spectrales plus large grâce au protocole développé dans l'article 1, l'établissement d'un score et d'un seuil de validation des résultats obtenus et l'étude de la transférabilité de la méthode d'identification à un autre appareil IRTF.

III.2- Article 1

Differentiation and identification of filamentous fungi by high-throughput FTIR spectroscopic analysis of mycelia

International Journal of Food Microbiology, 168-169 (2014), pp. 32-41

- Préambule à l'article 1

Contexte de l'étude

Les moisissures sont responsables de contamination dans le secteur agro-alimentaire, dans l'industrie pharmaceutique et cosmétologique et représentent également en médecine, un risque infectieux sévère chez les patients immunodéprimés. L'identification des champignons filamenteux repose actuellement soit sur des méthodes phénotypiques qui nécessitent une expertise mycologique et peuvent manquer de sensibilité soit sur des techniques moléculaires qui sont coûteuses et lourdes. Les récents développements concernant la spectroscopie infrarouge à transformée de Fourier (IRTF) associée aux traitements chimiométriques ont permis de mettre en place des techniques d'identification alternatives adaptées à une grande diversité d'échantillons.

Objectifs

Les objectifs de cette étude sont de mettre au point un protocole d'analyse par spectroscopie infrarouge et de développer un modèle chimiométrique de discrimination adaptés à l'analyse des moisissures dans un contexte industriel.

Matériels et méthodes

Cent trente et une souches (14 genres et 32 espèces), dont l'identification a été validée par séquençage moléculaire, ont été analysées à l'aide d'un spectromètre IRTF à haut débit. Les filaments ont été obtenus en cultivant les souches en milieu liquide (Chemboost YM, AES Chemunex/Biomérieux) pendant 48h. Chaque souche a fait l'objet de trois cultures effectuées sur 3 jours différents afin de vérifier la reproductibilité de la méthode. Des tests de qualité ont permis d'éliminer les spectres aberrants parmi les spectres enregistrés et des prétraitements mathématiques (correction de la ligne de base, dérivée seconde et normalisation vectorielle) ont été appliqués afin d'optimiser la matrice de données. L'analyse discriminante par méthode des moindres carrés (PLS-DA), méthode d'analyse statistique supervisée mettant en jeu des régressions PLS dites multivariées, a été utilisée comme méthode d'analyse chimiométrique dans les gammes spectrales 3200-2800 et 1800-800 cm^{-1} .

A l'aide de 106 souches, différents modèles de calibration ont été construits en cascade et en suivant la taxonomie actuelle.

Résultats

La validation croisée des échantillons de calibration a permis d'optimiser les paramètres des modèles de calibration. L'identification de 25 souches de moisissures au niveau du genre et de l'espèce à respectivement 98,97% et à 98,77% ont permis la validation externe des différents modèles. Cette étude démontre d'une part les potentiels de la spectroscopie IR, en raison de sa rapidité et de son faible coût, et d'autre part les possibilités chimiométriques de la PLS-DA, comme méthode alternative intéressante pour l'identification rapide des champignons filamenteux.

Conclusion

L'obtention en 48 heures d'une quantité de biomasse mycélienne suffisante rend cette technique particulièrement attractive dans le contexte industriel. Ces résultats prometteurs nous engagent à poursuivre cette étude afin d'élargir notre base de données et d'obtenir une méthode d'identification d'intérêts agroalimentaire et médical.

Differentiation and identification of filamentous fungi by high-throughput FTIR spectroscopic analysis of mycelia

A. Lecellier^a, J. Mounier^b, V. Gaydou^a, L. Castrec^b, G. Barbier^b, W. Ablain^c, M. Manfait^a, D. Toubas^{a,d}, G.D. Sockalingum^a

^a*MéDIAN-Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, FRE CNRS 3481MEDyC, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 REIMS cedex, France*

^b*Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (EA3882), SFRI48 ScInBioS, Université Européenne de Bretagne, Université de Brest, ESIAB, Technopôle de Brest Iroise, 29280 Plouzané, France*

^c*AES CHEMUNEX/BIOMERIEUX, Rue Maryse Bastié, CS17219 Ker Lann, 35172 Bruz cedex, France*

^d*Laboratoire de Parasitologie-Mycologie, CHU de Reims, Hôpital Maison Blanche, 45 rue Cognacq Jay, 51092 Reims cedex, France*

Corresponding author:

Ganesh D Sockalingum

Université de Reims Champagne-Ardenne

Equipe MéDIAN, Biophotonique et Technologies pour la Santé

Unité MEDyC, CNRS FRE3481

UFR Pharmacie, SFR CAP-Santé FED4231

51 rue Cognacq-Jay, Reims, France.

Tel: +33 (0)3 26 91 35 53

Fax: +33 (0)3 26 91 35 50

Email: ganesh.sockalingum@univ-reims.fr

Abstract

Routine identification of fungi based on phenotypic and genotypic methods can be fastidious and time-consuming. In this context, there is a constant need for new approaches allowing the rapid identification of molds. Fourier-transform infrared (FTIR) spectroscopy appears as such an indicated method. The objective of this work was to evaluate the potential of FTIR spectroscopy for an early differentiation and identification of filamentous fungi. One hundred and thirty-one strains identified using DNA sequencing, were analyzed using FTIR spectroscopy of the mycelia obtained after a reduced culture time of 48 h compared to current conventional methods. Partial least square discriminant analysis was used as a chemometric method to analyze the spectral data and for identification of the fungal strains from the phylum to the species level. Calibration models were constructed using 106 strains pertaining to 14 different genera and 32 species and were used to identify 25 fungal strains in a blind manner. Identification levels of 98.97% and 98.77% achieved were correctly assigned to the genus and species levels respectively. FTIR spectroscopy with its high discriminating power and rapidity therefore shows strong promise for routine fungal identification. Upgrading of our database is ongoing to test the technique's robustness.

Keywords: Identification, Fungi, High-throughput FTIR spectroscopy, PLS-DA

1. Introduction

Filamentous fungi are important ubiquitous microorganisms in nature. Besides their pathogenic importance in the agricultural, veterinary and medical fields (Chaiwun, Vanittanakom, Jiviriyawat, Rojanasthien, Thorner, 2011; Inderbitzin et al., 2011; Nucci, Anaissie, 2007; Pitt, 1994; Skiada et al., 2011), they also pose serious problems in other areas such as the food, pharmaceutical, and cosmetic industries. In agriculture and the food industry, they are responsible for the spoilage of raw materials and processed food and may pose serious health risks due to their ability to produce mycotoxins (Pitt, Hocking, 2009).

The identification of fungi by traditional phenotypic methods is based mainly on their macroscopic and microscopic features. However, these methods are quite time-consuming, laborious and sometimes not sufficiently accurate and require a thorough knowledge and expertise in the morphological analysis of fungi (Verscheure, Lognay, Marlier, 2002). Molecular methods such as sequencing of the nuclear ribosomal internal transcribed spacer (ITS) region that was recently chosen as the universal DNA barcode marker for fungi (Schoch et al., 2012) or other genes of interest (translation elongation factor, tubulin) are considered the gold standard for fungal identification. Nevertheless, in routine applications, these methods require special laboratory skills and are still fastidious and expensive despite a decrease in the cost-per-analysis (Nilsson, Ryberg, Abarenkov, Sjokvist, Kristiansson, 2009; Rozynek, Gilges, Bruning, Wilhelm, 2004). In the last decade, a new emerging technique based on Matrix-assisted laser desorption ionization time-of-flight mass spectroscopy (MALDI-TOF MS) has been put forward for characterizing bacteria (Lay, 2001). Further, it has been shown to be capable of discriminating clinically relevant filamentous fungi at the species level, giving results that compare favorably with molecular identification methods (Cassagne et al., 2011; De Carolis et al., 2012). Reference MALDI-TOF spectrometric databases for filamentous fungi, particularly from the food industry, are still under development (Santos, Paterson, Venancio, Lima, 2010b).

The industry, driven by the productivity constraints of the laboratory and the increasing pressure from regulatory authorities regarding for example the presence of mycotoxins in food and feedstuffs, favors rapid and alternative reference methods. As a consequence, there is a need for simple, rapid and more straightforward techniques for fungal identification directly applicable in an industrial context.

In this regard, Fourier-transform infrared (FTIR) spectroscopy appears as a promising candidate method. FTIR spectroscopy is a physico-chemical analytical method based on the light-matter interaction that permits characterization of the energy levels of the atomic bond vibrations present in a sample. In addition, the spectral information obtained allows attribution, in a qualitative and quantitative manner, of the signal to the main macromolecular constituents (lipids, polysaccharides, nucleic acids, proteins, etc.). Thus, the spectrum obtained represents a global “molecular fingerprint” which can be used for characterization, differentiation and identification of microorganisms.

This method has already been applied for the discrimination of 3 fungal dermatophytes species (Bastert, Korting, Traenkle, Schmalreck, 1999) and for the identification and typing of a limited number of airborne fungal species (10 species) from clinical origin belonging to the *Aspergillus* and *Penicillium* genera (Fischer, Braun, Thissen, Dott, 2006). In another study, FTIR spectroscopy was successfully used for the differentiation of 16 isolates belonging to five *Fusarium* species (Nie et al., 2007). FTIR spectroscopy has also allowed the differentiation of 3 *Aspergillus* species and the discrimination between aflatoxin-producing and non-producing strains from the agricultural environment (Garon, El Kaddoumi, Carayon, Amiel, 2010). The most promising studies by Shapaval et al. on the applicability of FTIR spectroscopy for fungal identification pointed out that this technique could be coupled to a high throughput microcultivation protocol (Shapaval et al., 2010) and could well resolve at the species and genera levels 59 strains pertaining to 19 species and 10 genera commonly involved in food spoilage (Shapaval et al., 2013).

All these studies have demonstrated the potential of FTIR spectroscopy for identification of filamentous fungi. However, in order to utilize this method in an industrial set-up, the sample preparation protocol should be as simple as possible with the shortest feasible cultivation time of fungal strains. In the majority of the reports mentioned above, the culturing time was around 14 days because spores were used as biological material, except in the study of Nie et al. (2007), where the culturing time was not standardized (between 3 and 10 days). Again, in the study of Shapaval et al. (2013), a step forward was achieved and fungal mycelia could be analyzed after 5 days culture in a liquid medium. Although very good predictions were obtained in the latter study at the genus and species levels, the cultivation time before analysis appears still too long for industrial use. In addition, few fungal genera, species and strains per species have been investigated in most of the above studies. FTIR spectroscopy is a rapid and simple technique requiring few consumables but could be limited by the protocol standardization and by the spectral databank quality (Santos, Fraga, Kozakiewicz, Lima,

2010a). Therefore, a new protocol requires standardization to extend its applicability to a large range of filamentous fungi.

The objectives of this study were 1) to develop a simplified protocol for fungal mycelia preparation, standardized in accordance with industrial requirements, and allowing FTIR analysis of filamentous fungi in 2 days and 2) to evaluate the potential of FTIR spectroscopy as a high-throughput method for rapid differentiation and identification of molds using multivariate statistical method.

2. Material and Methods

In order to develop a robust microorganism identification method, it is important to optimize and standardize both the culture conditions, the sample preparation method for spectral acquisition (Bertrand, Dufour, 2006). It is well known that culture conditions can influence spectral information and in this experimental section we describe in detail the procedures used for mold sample preparation and acquisition of high quality FTIR spectral data. The main steps of the infrared-based methodology for the identification of filamentous fungi are schematically represented in the Fig. 1.

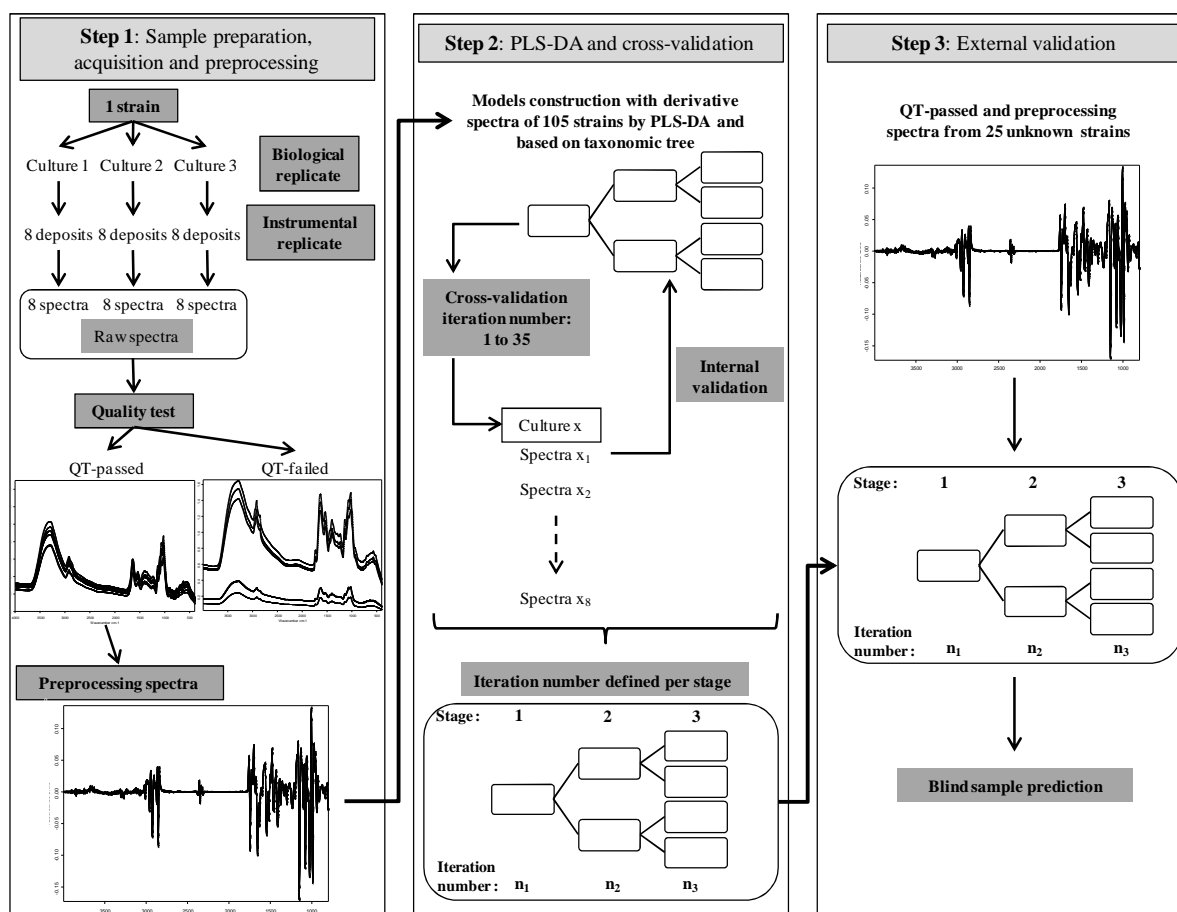


Figure 1: Schematic representation of infrared-based methodology for the identification of filamentous fungi. The first step consists of mycelia preparation, sample analysis by FTIR spectroscopy, and spectral preprocessing. The second step consists of building the calibration models for each stage of the taxonomic tree using second derivative spectra and PLS-DA method, and internal validation of the calibration models using the cross-validation. The third step consists of the external validation of calibration models using blind samples.

2.1. Fungal strains

One hundred and thirty-one filamentous fungal strains (14 genera and 32 species) were used in this study and were obtained from the Culture Collection of Université de Bretagne Occidentale (UBOCC, Plouzané, France) and the Culture Collection of Centraalbureau voor Schimmelcultures (CBS, Utrecht, The Netherlands) (Table 1). Mycelium plugs of these strains were stored in 10% glycerol at -80°C .

Table 1: List of fungal strains used to build the calibration models and for external validation of the models.

Species	Strain number†	Species	Strain number†
<i>Alternaria alternata</i>	CBS 116329	<i>Mucor spinosus</i>	UBOCC-A-103032
<i>Alternaria alternata</i>	UBOCC-A-111005	<i>Mucor spinosus</i>	UBOCC-A-109094
<i>Alternaria alternata*</i>	CBS 117143	<i>Mucor racemosus</i>	UBOCC-A-109083
<i>Alternaria alternata</i>	CBS 916.96	<i>Mucor racemosus</i>	UBOCC-A-109051
<i>Aspergillus flavus</i>	UBOCC-A-101063	<i>Mucor racemosus</i>	UBOCC-A-109063
<i>Aspergillus flavus</i>	UBOCC-A-106026	<i>Mucor racemosus*</i>	UBOCC-A-109068
<i>Aspergillus flavus*</i>	UBOCC-A-101060	<i>Mucor racemosus</i>	UBOCC-A-109056
<i>Aspergillus flavus</i>	UBOCC-A-106028	<i>Mucor racemosus</i>	UBOCC-A-101366
<i>Aspergillus niger</i>	UBOCC-A-101076	<i>Penicillium brevicompactum</i>	UBOCC-A-110065
<i>Aspergillus niger*</i>	UBOCC-A-101072	<i>Penicillium brevicompactum</i>	UBOCC-A-108094
<i>Aspergillus niger</i>	UBOCC-A-101089	<i>Penicillium brevicompactum</i>	CBS 257.29
<i>Aspergillus versicolor</i>	UBOCC-A-102012	<i>Penicillium brevicompactum*</i>	UBOCC-A-108093
<i>Aspergillus versicolor</i>	UBOCC-A-101087	<i>Penicillium brevicompactum</i>	UBOCC-A-108095
<i>Aspergillus versicolor*</i>	CBS 109274	<i>Penicillium chrysogenum</i>	UBOCC-A-110067
<i>Aspergillus versicolor</i>	UBOCC-A-101088	<i>Penicillium chrysogenum</i>	UBOCC-A-101399
<i>Aureobasidium pullulans</i>	UBOCC-A-101092	<i>Penicillium chrysogenum</i>	UBOCC-A-102022
<i>Aureobasidium pullulans</i>	UBOCC-A-108057	<i>Penicillium chrysogenum*</i>	UBOCC-A-106023
<i>Aureobasidium pullulans*</i>	UBOCC-A-101091	<i>Penicillium chrysogenum</i>	UBOCC-A-101400
<i>Aureobasidium pullulans</i>	UBOCC-A-108047	<i>Penicillium chrysogenum</i>	UBOCC-A-101393
<i>Aureobasidium pullulans</i>	UBOCC-A-108056	<i>Penicillium corylophilum</i>	UBOCC-A-109219
<i>Cunninghamella binariae</i>	UBOCC-A-101343	<i>Penicillium corylophilum</i>	UBOCC-A-109222
<i>Cunninghamella blakesleeana</i>	UBOCC-A-101341	<i>Penicillium corylophilum</i>	UBOCC-A-101405
<i>Cunninghamella elegans</i>	UBOCC-A-102008	<i>Penicillium corylophilum*</i>	UBOCC-A-109224
<i>Cunninghamella elegans</i>	UBOCC-A-101342	<i>Penicillium expansum</i>	UBOCC-A-108102
<i>Emericella nidulans</i>	CBS 589.65	<i>Penicillium expansum</i>	UBOCC-A-110070
<i>Emericella nidulans</i>	CBS 492.65	<i>Penicillium expansum*</i>	UBOCC-A-110021
<i>Emericella nidulans*</i>	CBS 121.35	<i>Penicillium expansum</i>	UBOCC-A-110024
<i>Emericella nidulans</i>	CBS 465.65	<i>Penicillium expansum</i>	UBOCC-A-110023
<i>Emericella nidulans</i>	CBS 119.55	<i>Penicillium expansum</i>	UBOCC-A-110032
<i>Eurotium amstelodami</i>	CBS 119376	<i>Penicillium glabrum</i>	UBOCC-A-108105
<i>Eurotium amstelodami</i>	CBS 117323	<i>Penicillium glabrum</i>	UBOCC-A-109098
<i>Eurotium chevalieri</i>	CBS 522.65	<i>Penicillium glabrum</i>	UBOCC-A-108114
<i>Eurotium chevalieri</i>	CBS 121704	<i>Penicillium glabrum*</i>	UBOCC-A-108107
<i>Fusarium equiseti</i>	CBS 414.86	<i>Penicillium glabrum</i>	UBOCC-A-109089
<i>Fusarium equiseti</i>	CBS 123566	<i>Penicillium glabrum</i>	UBOCC-A-108106
<i>Fusarium equiseti*</i>	UBOCC-A-109085	<i>Penicillium nalgiovense*</i>	UBOCC-A-108109
<i>Fusarium equiseti</i>	CBS 163.57	<i>Penicillium nalgiovense</i>	UBOCC-A-101430
<i>Fusarium equiseti</i>	CBS 791.70	<i>Penicillium nalgiovense</i>	UBOCC-A-101431
<i>Fusarium oxysporum</i>	UBOCC-A-108128	<i>Penicillium oxalicum</i>	UBOCC-A-101437
<i>Fusarium oxysporum</i>	UBOCC-A-101157	<i>Penicillium oxalicum</i>	UBOCC-A-101435
<i>Fusarium oxysporum*</i>	UBOCC-A-101154	<i>Penicillium oxalicum</i>	UBOCC-A-101436
<i>Fusarium oxysporum</i>	UBOCC-A-101135	<i>Penicillium oxalicum*</i>	UBOCC-A-102021
<i>Fusarium oxysporum</i>	UBOCC-A-109102	<i>Penicillium oxalicum</i>	CBS 301.97
<i>Fusarium verticillioides</i>	CBS 119825	<i>Penicillium oxalicum</i>	UBOCC-A-101438
<i>Fusarium verticillioides</i>	CBS 218.76	<i>Penicillium roqueforti</i>	UBOCC-A-109090
<i>Fusarium verticillioides*</i>	CBS 447.95	<i>Penicillium roqueforti</i>	UBOCC-A-108111
<i>Fusarium verticillioides</i>	UBOCC-A-101150	<i>Penicillium roqueforti</i>	UBOCC-A-108112
<i>Geotrichum candidum</i>	UBOCC-A-103039	<i>Penicillium roqueforti*</i>	CBS 221.30
<i>Geotrichum candidum</i>	UBOCC-A-101170	<i>Penicillium roqueforti</i>	UBOCC-A-108110
<i>Geotrichum candidum*</i>	UBOCC-A-108082	<i>Penicillium roqueforti</i>	UBOCC-A-101445
<i>Geotrichum candidum</i>	UBOCC-A-108081	<i>Penicillium verrucosum</i>	UBOCC-A-105004
<i>Litchtheimia corymbifera</i>	UBOCC-A-101328	<i>Penicillium verrucosum</i>	CBS 115508
<i>Litchtheimia corymbifera*</i>	UBOCC-A-103031	<i>Penicillium verrucosum</i>	UBOCC-A-105007
<i>Litchtheimia corymbifera</i>	UBOCC-A-102023	<i>Penicillium verrucosum*</i>	UBOCC-A-105014
<i>Mucor circinelloides</i>	UBOCC-A-109084	<i>Penicillium verrucosum</i>	UBOCC-A-109221
<i>Mucor circinelloides</i>	UBOCC-A-101354	<i>Rhizopus oryzae</i>	CBS 278.38
<i>Mucor circinelloides*</i>	UBOCC-A-108126	<i>Rhizopus oryzae</i>	UBOCC-A-101371
<i>Mucor circinelloides</i>	UBOCC-A-102003	<i>Rhizopus oryzae*</i>	UBOCC-A-101369
<i>Mucor circinelloides</i>	UBOCC-A-110124	<i>Rhizopus oryzae</i>	UBOCC-A-101372
<i>Mucor circinelloides</i>	UBOCC-A-110127	<i>Rhizopus oryzae</i>	CBS 146.90
<i>Mucor spinosus</i>	UBOCC-A-109053	<i>Syncephalastrum monosporum</i>	UBOCC-A-101373
<i>Mucor spinosus</i>	UBOCC-A-109052	<i>Syncephalastrum racemosum</i>	UBOCC-A-101374
<i>Mucor spinosus*</i>	UBOCC-A-101364	<i>Verticillium lecanii</i>	UBOCC-A-108023
<i>Mucor spinosus</i>	UBOCC-A-110133	<i>Verticillium lecanii*</i>	UBOCC-A-101320
<i>Mucor spinosus</i>	UBOCC-A-101363	<i>Verticillium lecanii</i>	UBOCC-A-108019
<i>Mucor spinosus</i>	UBOCC-A-102004		

*, Strains used for external validation of the calibration models,

†, UBOCC, Université de Bretagne Occidentale Culture Collection; CBS, Centraalbureau voor Schimmelcultures Culture Collection.

2.2. DNA extraction, amplification, sequencing and taxonomic assignment of fungal isolates

The total genomic DNA was extracted according to the manufacturer's instructions using the 'FastDNA SPIN Kit' (MPBio, Illkirch, France), from mycelia grown in potato dextrose broth for 2 to 4 days at 25°C on a rotary shaker at 120 rpm.

Five different regions were amplified depending on fungal genera : the rDNA internal transcribed spacer (ITS) region including the 5.8S rRNA gene (all genera except *Fusarium spp.*), partial β -tubulin gene (*Penicillium* and *Aspergillus spp.*), partial translation elongation factor-1 alpha (TEF-1) gene (*Fusarium spp.*) and, partial mcm7 and partial tsr1 genes (*Mucor spp.*). Amplification of the ITS region, β -tubulin, TEF-1, mcm7 and tsr1 genes was performed as described previously using primers ITS4 and ITS5 (White, Bruns, Lee, Taylor, 1990), Bt2a and Bt2b (Glass, Donaldson, 1995), EF1F and EF1R (O'Donnell, Kistler, Cigelnik, Ploetz, 1998), Mcm7-709for and Mcm7-1348rev (Schmitt et al., 2009) and Tsr1-f1 and Tsr1-r2 (Hermet, Meheust, Mounier, Barbier, Jany, 2012). After sequencing of the amplicons using the same primer pairs at the Biogenouest sequencing platform in the "Station Biologique de Roscoff" (<http://www.sb-roscoff.fr/SG/>) and contig assembly using DNA Baser (Heracle Software, Germany), sequences were compared to the GenBank database using the Basic Local Alignment Search Tool (BLAST) (<http://www.ncbi.nlm.nih.gov/BLAST>) to determine their closest known relatives. Alignments of the obtained sequences and sequences from the NCBI database were performed using the MAFFT online server (MAFFT version 7, <http://mafft.cbrc.jp/alignment/server/>) and the E-INS-i iterative refinement method. Phylogenetic trees were built in MEGA5 (Tamura et al., 2011) using the Neighbor-Joining method with 1000 bootstrapped data sets. The sequences have been deposited in the EMBL/GenBank database (Accession nos KF225013-KF225099; KF465750-KF465781; KF499564-KF499583).

2.3. FTIR analysis

2.3.1. Culture preparation

Cryopreserved strains were first sub-cultured on Sabouraud agar slants (Becton Dickinson, Le Pont de Claix, France) and incubated for 4 to 7 days at 25°C depending on the strain. Then, spores from the agar slant were inoculated in 20 ml of Chemboost YM broth (AES Chemunex, Bruz, France) and incubated on a rotary shaker at 150 rpm for 48 h at 25°C. For

each strain, 3 independent cultures prepared on 3 different days (biological replicates), were made.

2.3.2. Sample preparation

The cultures were transferred to an “M” tube adapted to a gentleMACS Octo Dissociator (Miltenyi Biotec, Paris, France). Each sample was dissociated at a speed set at 4000 rpm for 100 s. This step was necessary in order to obtain homogeneous mycelial suspensions for FTIR analysis. The homogeneous suspension consists of unaggregated pipettable mycelia, obtained after breakage of the latter into smaller sized particles, which can form easily reproducible deposits (in terms of thickness, mycelia size and distribution) for FTIR spectroscopy. Two millilitres of this suspension were then transferred into an Eppendorf tube. The culture supernatant was then eliminated by centrifugation at 430 xg for 30 s and the mycelia were resuspended in 1 ml of 0.9% NaCl. After another centrifugation at 430 xg for 30 s, the supernatant was discarded and the mycelial pellets were resuspended in approximately 300 μ l of 0.9% NaCl before FTIR spectroscopy analysis.

2.3.3. Spectral acquisition

Prior to spectral acquisition, each suspension was briefly vortexed. Five microliters of the suspension were deposited on a 384-well silicon plate in eight replicates. The plate was then dried under mild vacuum for 1 h to remove excess water (free water). A FTIR high-throughput system comprised of a spectrometer (Tensor 27, Bruker Optics, Champs sur Marne, France) coupled to a high-throughput module (HTS-XT, Bruker Optics) was used for spectral acquisition. The latter was performed using OPUS 6.5 software (Bruker Optics) with the following acquisition parameters: 64 accumulations per well, spectral resolution of 4 cm^{-1} , spectral range of 4000-400 cm^{-1} . The background spectrum of the silicon plate was recorded before each well was measured.

2.3.4. Analysis of FTIR spectra

Pre-processing of infrared spectra was composed of several steps including a quality test, baseline correction, derivation and vector normalization.

In the first step, the raw spectra were quality-tested using a slightly modified procedure initially described by Helm et al (Helm, Labischinski, Naumann, 1991). Each culture was validated if at least 5 out of 8 spectra passed the quality test. At the end of this step, between 15 and 24 spectra were available per strain.

FTIR spectra contain both biochemical information and information coming from physical effects due to light-matter interaction. The latter can introduce artifacts and large variabilities due to light scattering effects that can influence the classification model. Data mining is therefore an important step to extract biochemical information, whereby the necessity to mathematically pre-process the raw spectra and to perform data homogenization and reduction (Bertrand et al., 2006). All spectra were compiled in a table called data matrix for pre-processing using mathematical functions of the OPUS 5.5 software (Bruker Optics).

Mathematical pre-processing was performed according to the following procedures.

First, the FTIR spectra were truncated in the region 4000-800 cm^{-1} , which included most of the biochemical information. Then, a baseline correction was performed independently on each spectrum. This consists of modelling, in the form of equations, these different variations often found in the regions with no absorption bands. The modelling of these spectral variations is performed from few spectral data points and consists of fitting the baseline with a polynomial function passing through these points. The modelled spectral variations are then subtracted from the observed signal (Wu, Guo, Jouan-Rimbaud, Massart, 1999). Second derivative spectra were then computed with Savitzky-Golay 9 point smoothing (Savitzky, Golay, 1964). Finally, all the spectra were vector normalized. Briefly, the method consists of calculating the average absorbance values (y) of a spectrum in the selected spectral range. This average value is subtracted from each absorbance value of the spectrum and then divided by the root of the sum of the squares of all the absorbances (y) taken in the given spectral range (Bylesjö M, Cloarec O, M., 2009).

2.4. Chemometric analysis

Chemometric analysis performed with Matlab version 7.2 (Mathwork, USA) was used to classify the samples using their explained variables. The classification method used in our study is the partial least squares discriminant analysis (PLS-DA) in the spectral ranges 3200-2800 and 1800-800 cm^{-1} . These large spectral ranges can be associated to the IR energy levels of different biochemical functions present in the sample (Table 2). In our study, the strains were classified taking into account the ensemble of molecular information present in each

sample. This classification method is a supervised linear analysis method based on the algorithm of PLS regression. The PLS regression has been developed for quantitative analysis, but it can also be employed for class identification (Tenenhaus, 1998). The adaptation of the PLS regression as a method of classification is carried out using a binary code. Each class is coded by a combination of 0 and 1, based on whether or not the sample belongs to a given class. The PLS-DA is based on the relation between explained variables of the samples and the different classes of the samples. The PLS-DA algorithm permits to correlate the explained data with a matrix of class properties assisted by multivariate regression (Liang, Kvalheim, 1996). In our study, the fungal strains were assigned to a pre-established class based on the genetic results and according to the current taxonomy. Models were built for each of the taxonomic ranks from phylum/subphylum to species to which the fungal strains belonged and were constructed with the strains used for the calibration step. The strains of *Aspergillus* and of its teleomorphs *Eurotium* and *Emericella* were grouped under the generic name *Aspergillus* (Peterson, 2008).

Table 2: Characteristic infrared absorption frequencies typical of microorganisms and their biomolecular attribution.

Frequencies (cm ⁻¹)	Molecular bond	Vibrational mode	Biomolecular attribution
3200-2800	CH ₂ , CH ₃	symmetric and asymmetric stretching	Lipids
	N-H	symmetric stretching	Proteins
1780-1700	C=O	symmetric stretching	Fatty acids
1695-1625	C=O, C-N	symmetric stretching	Proteins (amide I)
	N-H	bending	
1560-1525	C-N	symmetric stretching	Proteins (amide II)
	N-H	bending	
1480-1400	CH ₃ , CH ₂	bending	Lipids
	C=O	asymmetric stretching	
1300-1200	P=O,	asymmetric stretching	Nucleic acids
1200-900	C-O-C, C-O, P=O, C-C/C-O	symmetric stretching	Ribose, Glycogen, Nucleic acids
900-700	C-H	bending	Aromatic groups

The partial leave-one-out cross-validation method (Stone, 1974) was used to evaluate the quality of the calibration models and to provide information on the parameters of the models. All replicate spectra from the same culture are removed in turn, and calibration models are developed with the remaining spectra. Models thus created are then tested for each spectrum of the removed culture (leave-one-out) representing the sample for internal validation. In this manner, all the spectra of the calibration models serve both for calibration and internal validation. The partial cross-validation was tested cumulatively from 1 to 35 iterations, also

known as the dimension number and corresponds to the computing of a latent variable associated to the PLS-DA model. Depending on different models of calibration, the number of iterations which provided the best percentage of sample prediction in internal validation was selected for the construction of models that served in the external validation. This external validation was realised with the remaining 25 fungal strains.

3. Results

3.1. Cross-validation and calibration models

Direct classification of the 106 strains at the genus and species level by PLS-DA was unsuccessful showing that this method cannot be applied on a high number of clusters. Therefore, in order to achieve discrimination at the genus and species levels, calibration models were constructed in cascade based on the current classification of fungi (Table 3). Thus, 16 calibration models were constructed using the second derivative spectra of 106 strains. The discriminant wavenumbers were selected for each taxonomic rank and this approach allowed a step-wise discrimination of the different strains from the highly general to the highly specific. Consequently, the number of spectra per cluster decreased from higher (phylum/subphylum) to lower taxonomic rank (species) since the number of clusters increased.

The cross-validation method used on the spectra of 106 calibration samples allowed control of the performance of the method, the robustness of the different models constructed along with the accuracy of prediction. In addition, this method enabled the optimization of the number of iterations providing the best percentage of correct prediction at each taxonomic rank in an internal validation and the construction of the best calibration models possible.

For example, a model was constructed in order to differentiate, at the family level, the strains that belonged to the families *Mucoraceae*, *Lichtheimiaceae*, *Syncephalastraceae* and *Cunninghamellaceae* of the *Mucorales* order. For this model, every culture of each strain was taken out in turn and each spectrum of the same culture was projected into the model one after the other. Fig. 2 shows the average percentages of correct prediction for the calibration and internal validation calculated as a function of the iteration number, which varies from 1 to 35. For this family model, it was observed that, using 10 iterations, an average of 100% and 99.45% of the spectra were correctly predicted in the calibration and the internal validation step, respectively. This optimal iteration number of 10 was therefore chosen for the

construction of calibration models which served in the external validation for this family model.

Table 3: Classification as a function of the current taxonomy of fungal species used for calibration models and external validation. The last two columns indicate the number of strains per species used for the calibration set and the external validation set respectively.

Phylum/ Subphylum	Class	Order	Family	Genus	Subgenus	Section	Species	Number of strains per species for calibration	Number of strains per species for validation	
Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	Aspergillus	-	-	<i>A. flavus</i>	3	1	
							<i>A. niger</i>	2	1	
							<i>E. nidulans</i>	4	1	
							<i>A. versicolor</i>	3	1	
							<i>E. amstelodami</i>	2	-	
							<i>E. chevalieri</i>	2	-	
				Penicillium	-	-	<i>Fasciculata</i>	<i>P. verrucosum</i>	4	1
							<i>Penicillium</i>	<i>P. expansum</i>	5	1
							<i>Roquefortorum</i>	<i>P. roqueforti</i>	5	1
							<i>Chrysogena</i>	<i>P. chrysogenum</i>	5	1
							<i>P. nalgiovense</i>	2	1	
							<i>Brevicompacta</i>	<i>P. brevicompactum</i>	4	1
	Aspergilloides	-	-	<i>P. glabrum</i>	5	1				
				<i>P. corylophilum</i>	3	1				
								<i>P. oxalicum</i>	5	1
	Saccharomycetes	Saccharomycetales	Endomycetaceae = Dipodascaceae	Geotrichum	-	-	-	<i>G. candidum</i>	3	1
	Sordariomycetes	Hypocreales	Nectriaceae	Fusarium	-	-	-	<i>F. oxysporum</i>	4	1
								<i>F. verticillioides</i>	3	1
Incertae sedis		Plectosphaerellaceae	Verticillium	-	-	-	<i>F. equiseti</i>	4	1	
							<i>V. lecanii</i>	2	1	
Dothideomycetes	Pleosporales	Pleosporaceae	Alternaria	-	-	-	<i>A. alternata</i>	3	1	
	Dothideales	Dothioraceae	Aureobasidium	-	-	-	<i>A. pullulans</i>	4	1	
Mucoromycotina	-	Mucorales	Mucoraceae	Mucor	-	-	<i>M. circinelloides</i>	5	1	
							<i>M. spinosus</i>	7	1	
							<i>M. racemosus</i>	5	1	
						<i>R. oryzae</i>	4	1		
			Lichtheimiaceae	Lichtheimia	-	-	-	<i>L. corymbifera</i>	2	1
			Syncephalastraceae	Syncephalastrum	-	-	-	<i>S. monosporum</i>	1	-
								<i>S. racemosum</i>	1	-
			Cunninghamellaceae	Cunninghamella	-	-	-	<i>C. elegans</i>	2	-
								<i>C. binariae</i>	1	-
								<i>C. blakesleeana</i>	1	-

References used to establish the taxonomic classification: Geiser et al., 2006; Hermet et al., 2012; Hibbett et al., 2007; Houbraken & Samson, 2011; Inderbitzin et al., 2011; J I Pitt & Hocking, 2009; Skiada et al., 2011; Suh, Blackwell, Kurtzman, & Lachance, 2006; Vitale et al., 2011; Walker, Castlebury, Rossman, & White, 2012; Wang, Geng, Ma, Wang, & Zhang, 2012; N. Zhang et al., 2006; Y. Zhang, Crous, Schoch, & Hyde, 2012.

Based on the same procedure, an iteration number was determined for each taxonomic rank (Table 4). These iteration numbers were utilized for the construction of the different calibration models. In the case where there were different models per taxonomic rank, i.e., there were 6 distinct models at the species level; an identical iteration number was selected for all the models. This optimum iteration number was selected as the one giving the best results for all models at a given taxonomic rank.

Discrimination planes were established in order to observe the regression vectors which permitted the best discrimination in a model. As shown in Fig. 3 for the 4 families of the

Mucorales order, the projection of the different spectra in a three-dimensional coordinate system according to the regression vectors 1, 3 and 4 resulted in four compact and distinct clusters corresponding to the 4 families of this model. The confusion matrix represented in Fig. 4 shows that a 100% correct classification of spectra could be achieved for the 4 families of this model.

Overall, the great majority of the calibration models yielded 100% of correct classification of the spectra, except those for *Saccharomyces* and *Eurotiomycetes* classes and *Aspergillus* and *Penicillium* genera, for which a correct classification rate of 99.93% and 99.71% are obtained respectively (Fig. 5).

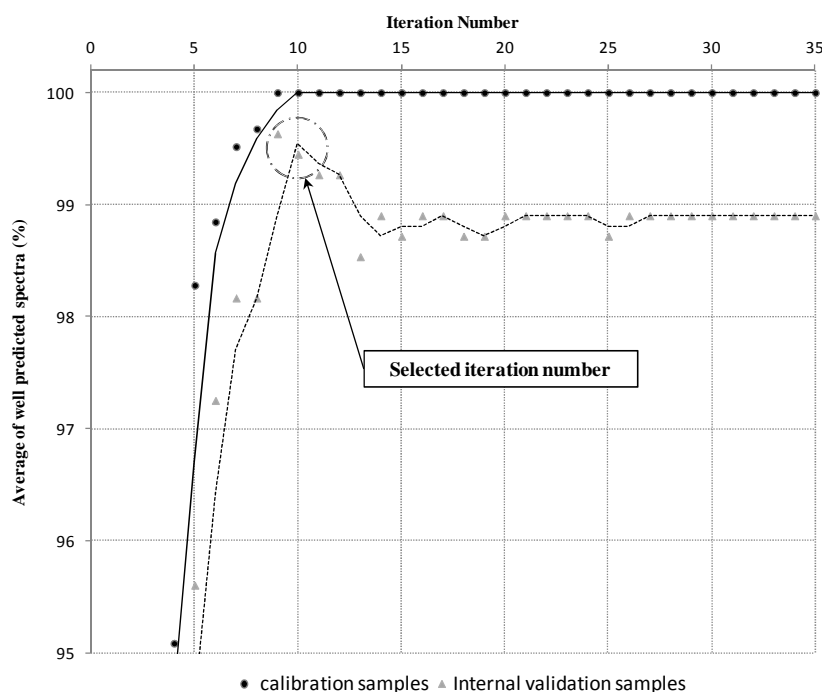


Figure 2: Cross-validation results of the calibration model allowing the discrimination of 4 families of the *Mucorales* order. The average percentage of well-predicted spectra as a function of the iteration number for the calibration strains and the internal validation strains are shown as well as the iteration number chosen for this model.

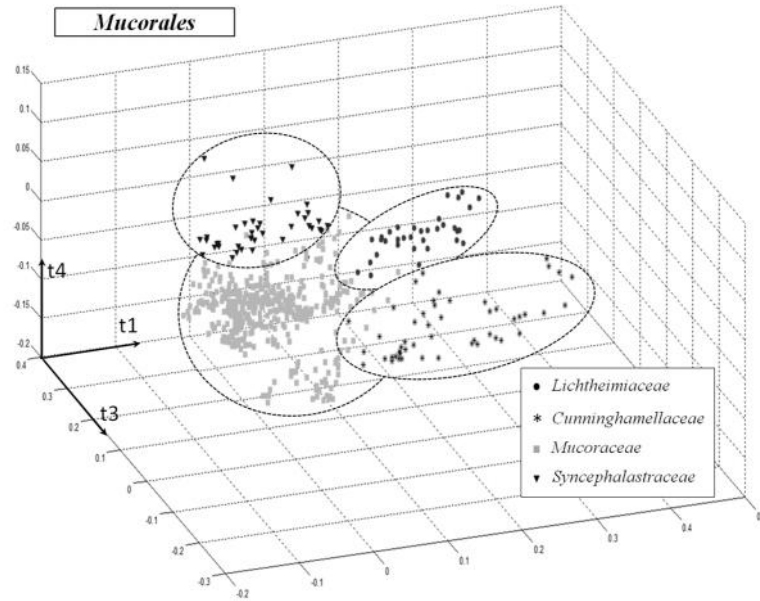


Figure 3: Projection of spectra in a three-dimensional coordinate system based on the PLS-DA regression vectors 1, 3 and 4, of the calibration model constructed to discriminate 4 families of the *Mucorales* order. This calibration model contains four groups of well discriminated spectra corresponding to each family, e.g., *Mucoraceae*, *Lichtheimiaceae*, *Syncephalastraceae* and *Cunninghamellaceae* of the *Mucorales* order.

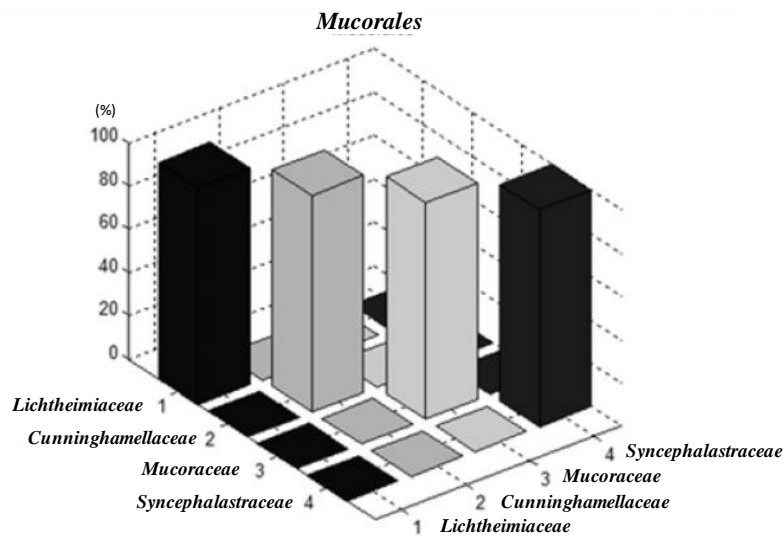


Figure 4: Confusion matrix obtained for the calibration model constructed to discriminate the families of the *Mucorales* order.

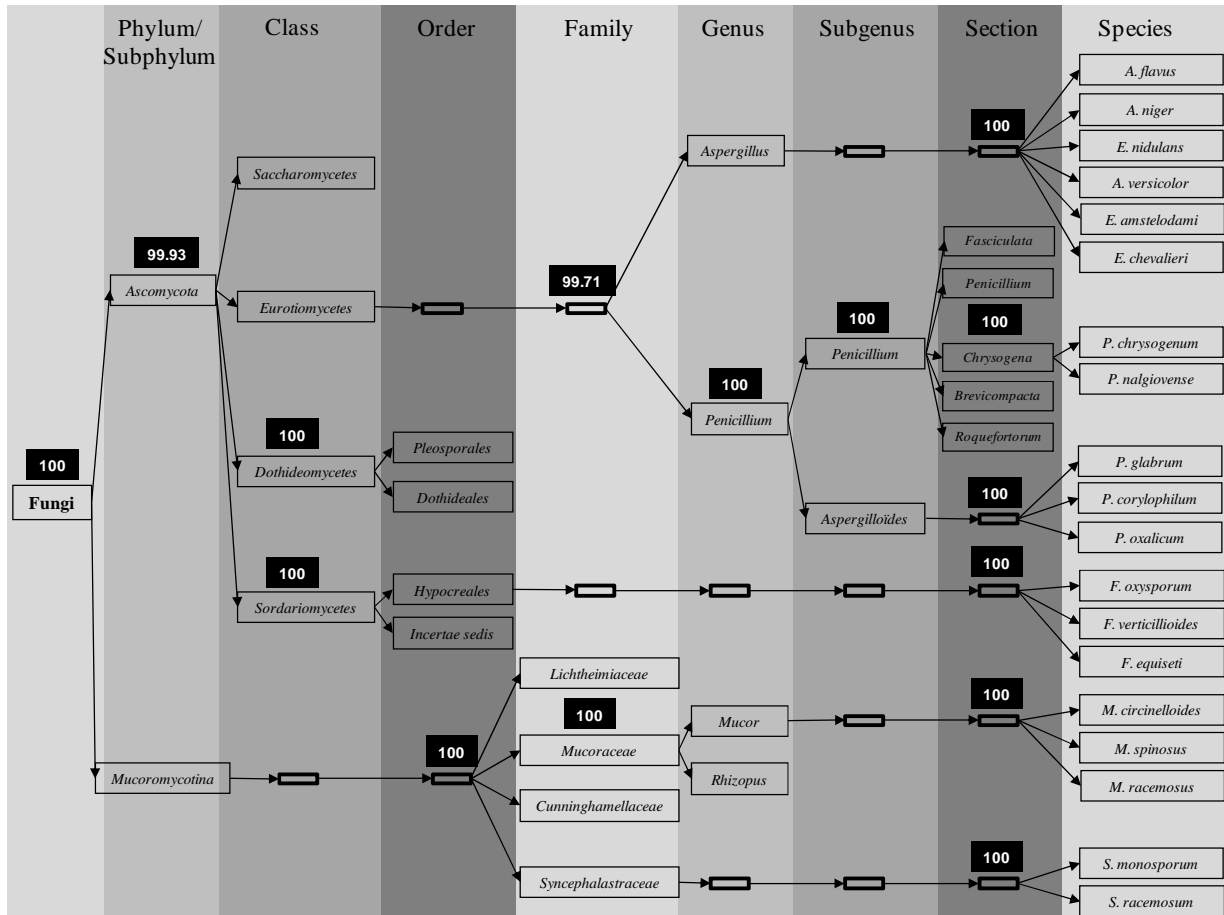


Figure 5: Percentage attribution of spectra for the sixteen calibration models constructed at the different taxonomic ranks.

3.2. External validation

The external validation of several models was carried out using twenty-five fungal strains from twenty-five species and twelve different genera. The percentages of correctly predicted spectra per taxonomic rank was 100% for phylum/subphylum, class and order, 99.79% for family, 98.97% for genera and subgenera, and 98.77% for section and species (Table 4).

Table 4: Optimal iteration number and average of well predicted spectra in percentage for the cross validation, the calibration, and the external validation for each stage of the taxonomic tree.

Stage	Optimal iteration number	Average of well predicted spectra (%)		
		Cross validation	Calibration	External Validation
Phylum/ Subphylum	10	100	100	100
Class	15	99,93	99,93	100
Order	15	100	100	100
Family	10	99,45	100	99,79
Genus	15	98,30	99,85	98,97
Subgenus	15	99,74	100	98,97
Section	30	97,01	100	98,77
Species	15	96,80	100	98,77

The identification results of these twenty-five strains are shown in Table 5. The average percentage of correctly predicted spectra per strain was 98.58%. Concerning strains from the *Ascomycota* phylum, the spectra of sixteen out of twenty strains were identified with 100% accuracy at the genus and species level. The spectra of the 4 remaining strains were also correctly identified with a slightly lower accuracy comprised between 90 and 100 %. For *Penicillium brevicompactum* UBOCC-A-108093, the percentages of correctly predicted spectra at the genus and species levels were 100% and 93.3%, respectively. The 3 remaining strains *Aspergillus flavus* UBOCC-A-101060, *Penicillium nalgiovense* UBOCC-A-108109, and *Penicillium corylophilum* UBOCC-A-109224 were correctly identified at the genus level with 90.5%, 93.8% and 95.2% of correctly predicted spectra, respectively. In addition, the spectra correctly identified at the genus level were also accurately assigned (100 %) at the species level. Concerning the *Mucoromycotina* subphylum, the spectra of 4 out of five strains were correctly identified at the genus and species levels with 100% of correctly predicted spectra. For the remaining strain, 91.7% of the spectra of *Mucor circinelloides* UBOCC-A-108126 were correctly assigned at the genus level and therein these spectra were all correctly assigned at the species level.

Table 5: External validation of the calibration models and prediction results for twenty five fungal strains. The results obtained by DNA-sequencing, considered as the reference method for species identification, and the prediction results obtained by FTIR spectroscopy are compared. The results are expressed as a percentage of predicted spectra per strain and the average of well predicted spectra per strain is given.

Strain number	References species (sequencing)	Prediction	
		Predicted species (FTIR)	Percentage of predicted spectra per strain
UBOCC-A-101072	<i>A. niger</i>	<i>A. niger</i>	100
UBOCC-A-101060	<i>A. flavus</i>	<i>A. flavus</i> <i>P. roqueforti</i>	90,5 9,5
CBS 12135	<i>E. nidulans</i>	<i>E. nidulans</i>	100
CBS 109274	<i>A. versicolor</i>	<i>A. versicolor</i>	100
UBOCC-A-105014	<i>P. verrucosum</i>	<i>P. verrucosum</i>	100
UBOCC-A-110021	<i>P. expansum</i>	<i>P. expansum</i>	100
CBS 22130	<i>P. roqueforti</i>	<i>P. roqueforti</i>	100
UBOCC-A-106023	<i>P. chrysogenum</i>	<i>P. chrysogenum</i>	100
UBOCC-A-108109	<i>P. nalgiovense</i>	<i>P. nalgiovense</i> <i>E. chevalieri</i>	93,8 6,3
UBOCC-A-108093	<i>P. brevicompactum</i>	<i>P. brevicompactum</i> <i>P. verrucosum</i>	93,3 6,7
UBOCC-A-108107	<i>P. glabrum</i>	<i>P. glabrum</i>	100
UBOCC-A-109224	<i>P. corylophilum</i>	<i>P. corylophilum</i> <i>E. amstelodami</i>	95,2 4,8
UBOCC-A-102021	<i>P. oxalicum</i>	<i>P. oxalicum</i>	100
UBOCC-A-108082	<i>G. candidum</i>	<i>G. candidum</i>	100
UBOCC-A-101154	<i>F. oxysporum</i>	<i>F. oxysporum</i>	100
CBS 447.95	<i>F. verticillioides</i>	<i>F. verticillioides</i>	100
UBOCC-A-109085	<i>F. equiseti</i>	<i>F. equiseti</i>	100
UBOCC-A-101320	<i>V. lecanii</i>	<i>V. lecanii</i>	100
CBS 117143	<i>A. alternata</i>	<i>A. alternata</i>	100
UBOCC-A-101091	<i>A. pullulans</i>	<i>A. pullulans</i>	100
UBOCC-A-108126	<i>M. circinelloides</i>	<i>M. circinelloides</i> <i>C. elegans</i>	91,7 8,3
UBOCC-A-101364	<i>M. spinosus</i>	<i>M. spinosus</i>	100
UBOCC-A-109068	<i>M. racemosus</i>	<i>M. racemosus</i>	100
UBOCC-A-101369	<i>R. oryzae</i>	<i>R. oryzae</i>	100
UBOCC-A-103031	<i>L. corymbifera</i>	<i>L. corymbifera</i>	100
Number of strains			Average of well predicted spectra per strain (%)
25			98,58

4. Discussion

In this study, high-throughput FTIR spectroscopy analysis of mycelia was applied in order to discriminate and identify filamentous fungi. A rapid and reproducible method was developed for fungal cultivation and sample preparation. All strains were cultivated in the same culture medium, at the same temperature, and during an identical time period of 48 h. These standardized conditions and the use of a well-adapted culture medium permitted on the one hand to obtain a sufficient amount of mycelial biomass for all the strains. In addition, cultivation under agitation prevented spore production by the majority of strains, except for few species such as *Penicillium roqueforti* for which spore formation still occurred to a limited extent. However, we also observed that a few fungal species showed limited growth in the selected culture medium such as *Wallemia sebi* (data not shown). This is not surprising since this species is a xerophilic fungus meaning that it optimally grows at reduced water activity (aw). The use of another growth medium with a lower aw could be considered for such fungi. Nevertheless, the shorter cultivation time and the simplicity of the sample preparation protocol make the method used in this study more attractive than those developed in other studies based on FTIR analysis of spores. Thus, fungal mycelia were only subjected to two short preparation steps, e.g., dissociation and washing. Due to the inhomogeneous structure of mycelium, this dissociation process was necessary in order to obtain a suspension that could be easily deposited on the silicon plate, a material suitable for FTIR analysis. In contrast, preparation of standardized spore suspension is more laborious. Moreover, because fungal spores can be highly hydrophobic, chemical products such as ethanol or Tween may be required for the preparation of spore suspensions and for the deposition of reproducible deposits on the silicon plate (Fischer et al., 2006; Garon et al., 2010). The utilization of such chemical products can also alter the biological material, for example the fungal cell wall, and therefore lead to inconsistent FTIR spectra. Because spores have a spherical shape with sizes within the range of the mid-infrared wavelengths, physical effects like Mie-scattering can be introduced and may affect the biochemical information contained in the FTIR spectra (Kohler et al., 2009). Finally, filamentous fungi such as *Penicillium camemberti* and certain species of *Fusarium* and *Alternaria sp.* poorly sporulate under laboratory conditions which would render impossible the application of FTIR macrospectroscopy to such fungi.

As mentioned above, most studies dedicated to the identification of filamentous fungi with FTIR spectroscopy utilized spores as the basic biological material. In contrast, we used mycelium of 32 species also which was found to be a good biological material for inter-

species differentiation of filamentous fungi. Fungal mycelia have a more specific and complex biochemical composition than spores which are metabolically inactive. They are carriers of a large variety of biomolecules that may be more specific from one species to another and that may be used for identification at the species level. While taxonomic biomarkers of mycelium may be species-specific, they may also vary qualitatively and quantitatively depending on the fungal growth stage (Cantu, Greve, Labavitch, Powell, 2009). In this study, three independent cultures of each strain were made under the same conditions on three different days. The chemometric method used allowed us to group all or the majority of the spectra derived from these three cultures in a strain-specific cluster (data not shown). Only few spectra were misclassified and they originated from the three replicate cultures and never from a single culture. Hence, it can be concluded that FTIR measurements on mycelia were reproducible from one day to another.

Classification and identification of fungal strains were performed by spectral data analysis using the PLS-DA statistical method. In the first step, this statistical analysis was employed to optimize the calibration models using the cross-validation method. This method, applied on 106 fungal strains, allowed the internal validation of the calibration models and the optimization of the iteration number at each taxonomic rank. In our study, at this stage of the database construction, it was possible to keep the same optimized iteration number for all the models within a given taxonomic rank. However, it is possible that it might be necessary to optimize the iteration number for each calibration model with a more complex database comprising a higher number of genera and species.

In the second step, the data matrix was split into two sets of spectra, e.g., a calibration set containing 106 fungal strains and an external validation set of 25 fungal strains. Using PLS-DA, 98.97% (483 out of 488 spectra) and 98.77% (482 out of 488 spectra) were correctly assigned at the genus and species level, respectively. The main advantage of this method is that it is a supervised analysis method in contrast with the hierarchical cluster analysis (HCA) which was used as a preliminary approach. Except at high taxonomic ranks, classical HCA did not enable good identification of the fungal strains, e.g., at the genus and species levels. Indeed, using HCA, spectra from different species of the same genus generally did not cluster and strains from the same species did not always cluster together (data not shown). This phenomenon was also observed for several genera and species of yeasts in a previous report (Kummerle, Scherer, Seiler, 1998). In this study, the choice of a linear multivariate regression method over a non-linear method such as artificial neural network (ANN) was made because the infrared data showed fluctuations of intensity following the Beer-Lambert law which is

linear. This statistical method has been shown to be suitable for the analysis of FTIR spectral data (Gaydou, Kister, Dupuy, 2009). Nevertheless, a high number of references are required to get a robust database. Recently, Shapaval et al. (2013) used ANN to identify food spoilage fungi using FTIR spectroscopy. In the latter study, good predictions were achieved at the genus and species levels with 94% and 93.9% of correct predictions, respectively. However, it should be emphasized that these prediction results corresponded to an internal validation and not an external validation using other strains than those used to build the model.

In our calibration set, we utilized 38 *Penicillium* strains belonging to 9 species. The discrimination of these species by using 9 clusters did not give satisfactory results. The discrimination of *Penicillium* species was optimized by dividing the genus *Penicillium* into two groups, corresponding to the two subgenera *Penicillium* and *Aspergilloides*. Then, the results were equally optimized by dividing the sub-genus *Penicillium* into 5 sections: *Fasciculata*, *Penicillium*, *Roquefortorum*, *Chrysogena* and *Brevicompacta*. Given the high number of *Penicillium* species involved in food spoilage, e.g., more than 50 are listed in Pitt et al. (2009), it might be necessary to find specific discriminant spectral range(s) for the identification of the species comprised in each *Penicillium* section. Indeed, specific spectral ranges could be an additional parameter to optimize the models.

The PLS-DA results showed that only 6 out 488 spectra were not correctly identified although these spectra passed the quality test, indicating that these bad predictions were not related to a physical or measurement problem. One possible explanation is that these particular strains shared common biochemical features with strains from other species or genus. In our study, the analysis was performed on a large spectral range allowing us to better reflect the overall chemical composition of each sample. Once again, the search of more specific spectral regions may help to overcome this problem.

In agriculture and food industry, there is a need for rapid, easy-to-use, and affordable techniques to identify filamentous fungi for routine application in microbiology laboratories. As shown in the present work, high-throughput FTIR spectroscopy could constitute an alternative to currently available techniques without the need of thorough knowledge and expertise in mycology or molecular biology. In this study, fungal strains pertaining to 14 different genera and 32 species, which were correctly assigned at the species level using DNA sequencing, were used to build a FTIR spectral database. However, despite identification of the isolates was based on a polyphasic approach combining phenotypic and genotypic methods, misidentification of strains pertaining to closely related species is possible.

We showed that FTIR spectroscopy allowed discrimination of these strains at the genus and species levels and that, 25 strains absent from the database could be accurately identified using PLS-DA. Nevertheless, the database built is far from being exhaustive with respect to the diversity of fungal contaminants encountered in the food industry and agriculture. Based on the literature review of Pitt et al. (2009), we estimate that approximately 200 fungal species should be considered in order to provide a database as comprehensive as possible. Due to these promising preliminary results, a larger database totaling 850 strains is currently under development and will provide more in-depth information on the applicability and the limits of FTIR spectroscopy for rapid identification of filamentous fungi.

Acknowledgments

Aurélie Lecellier is thankful to the Conseil Régional de Champagne Ardenne for funding of her PhD. The “Pôle de compétitivité” VALORIAL, La Région Bretagne, La Région Champagne-Ardenne and the technological platform IBiSA “Imagerie Cellulaire et Tissulaire” are gratefully acknowledged. Financial support under project "Mycotech" of the European Union, the Région Bretagne and the Conseil Général du Finistère is also gratefully acknowledged. The authors are also grateful to Marie-Anne Le Bras and Valérie Vasseur for their expertise and help in fungal identification and to Amélie Weill and Olivia Le Bourhis for their excellent technical assistance.

References

- Bastert, J., Korting, H.C., Traenkle, P., Schmalreck, A.F. 1999. Identification of dermatophytes by Fourier transform infrared spectroscopy (FT-IR). *Mycoses* 42, 525-528.
- Bertrand, D., Dufour, E. 2006. *La spectroscopie infrarouge et ses applications analytiques*, 2nd edn. Technique et Documentation (ed) ed., Paris.
- Bylesjö M, Cloarec O, M., R. 2009. Normalization and Closure. *Compr. Chemometr* 2.07, 109-127.

Cantu, D., Greve, L.C., Labavitch, J.M., Powell, A.L. 2009. Characterization of the cell wall of the ubiquitous plant pathogen *Botrytis cinerea*. *Mycol Res* 113, 1396-1403.

Cassagne, C., Ranque, S., Normand, A.C., Fourquet, P., Thiebault, S., Planard, C., Hendrickx, M., Piarroux, R. 2011. Mould routine identification in the clinical laboratory by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *PLoS One* 6, e28425.

Chaiwun, B., Vanittanakom, N., Jiviriyawat, Y., Rojanasthien, S., Thorner, P. 2011. Investigation of dogs as a reservoir of *Penicillium marneffe* in northern Thailand. *Int J Infect Dis* 15, e236-239.

De Carolis, E., Posteraro, B., Lass-Flörl, C., Vella, A., Florio, A.R., Torelli, R., Girmenia, C., Colozza, C., Tortorano, A.M., Sanguinetti, M., Fadda, G. 2012. Species identification of *Aspergillus*, *Fusarium* and *Mucorales* with direct surface analysis by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Microbiol Infect* 18, 475-484.

Fischer, G., Braun, S., Thissen, R., Dott, W. 2006. FT-IR spectroscopy as a tool for rapid identification and intra-species characterization of airborne filamentous fungi. *J Microbiol Methods* 64, 63-77.

Garon, D., El Kaddoumi, A., Carayon, A., Amiel, C. 2010. FT-IR spectroscopy for rapid differentiation of *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus parasiticus* and characterization of aflatoxigenic isolates collected from agricultural environments. *Mycopathologia* 170, 131-142.

Gaydou, V., Kister, J., Dupuy, N. 2009. Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil. *Chemometrics and Intelligent Laboratory Systems* 106, 190-197.

Geiser, D.M., Gueidan, C., Miadlikowska, J., Lutzoni, F., Kauff, F., Hofstetter, V., Fraker, E., Schoch, C.L., Tibell, L., Untereiner, W.A., Aptroot, A. 2006. *Eurotiomycetes: Eurotiomycetidae and Chaetothyriomycetidae*. *Mycologia* 98, 1053-1064.

Glass, N.L., Donaldson, G.C. 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous *ascomycetes*. *Appl Environ Microbiol* 61, 1323-1330.

Helm, D., Labischinski, H., Naumann, D. 1991. Elaboration of a procedure for identification of bacteria using Fourier-Transform IR spectral libraries: a stepwise correlation approach. *J Microbiol Methods* 14, 127–142.

Hermet, A., Meheust, D., Mounier, J., Barbier, G., Jany, J.L. 2012. Molecular systematics in the genus *Mucor* with special regards to species encountered in cheese. *Fungal Biol* 116, 692-705.

Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T., Kirk, P.M., Lucking, R., Thorsten Lumbsch, H., Lutzoni, F., Matheny, P.B., McLaughlin, D.J., Powell, M.J., Redhead, S., Schoch, C.L., Spatafora, J.W., Stalpers, J.A., Vilgalys, R., Aime, M.C., Aptroot, A., Bauer, R., Begerow, D., Benny, G.L., Castlebury, L.A., Crous, P.W., Dai, Y.C., Gams, W., Geiser, D.M., Griffith, G.W., Gueidan, C., Hawksworth, D.L., Hestmark, G., Hosaka, K., Humber, R.A., Hyde, K.D., Ironside, J.E., Koljalg, U., Kurtzman, C.P., Larsson, K.H., Lichtwardt, R., Longcore, J., Miadlikowska, J., Miller, A., Moncalvo, J.M., Mozley-Standridge, S., Oberwinkler, F., Parmasto, E., Reeb, V., Rogers, J.D., Roux, C., Ryvarden, L., Sampaio, J.P., Schussler, A., Sugiyama, J., Thorn, R.G., Tibell, L., Untereiner, W.A., Walker, C., Wang, Z., Weir, A., Weiss, M., White, M.M., Winka, K., Yao, Y.J., Zhang, N. 2007. A higher-level phylogenetic classification of the Fungi. *Mycol Res* 111, 509-547.

Houbraken, J., Samson, R.A. 2011. Phylogeny of *Penicillium* and the segregation of *Trichocomaceae* into three families. *Stud Mycol* 70, 1-51.

Inderbitzin, P., Bostock, R.M., Davis, R.M., Usami, T., Platt, H.W., Subbarao, K.V. 2011. Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. *PLoS One* 6, e28341.

Kohler, A., Bocker, U., Warringer, J., Blomberg, A., Omholt, S.W., Stark, E., Martens, H. 2009. Reducing inter-replicate variation in fourier transform infrared spectroscopy by extended multiplicative signal correction. *Appl Spectrosc* 63, 296-305.

Kummerle, M., Scherer, S., Seiler, H. 1998. Rapid and reliable identification of food-borne yeasts by Fourier-transform infrared spectroscopy. *Appl Environ Microbiol* 64, 2207-2214.

Lay, J.O., Jr. 2001. MALDI-TOF mass spectrometry of bacteria. *Mass Spectrom Rev* 20, 172-194.

Liang, Y.Z., Kvalheim, O. 1996. Robust methods for multivariate analysis. *Chemom Intell Lab Syst* 32, 1-10.

Nie, M., Zhang, W.Q., Xiao, M., Luo, J.L., Bao, K., Chen, J.K., Li, B. 2007. FT-IR spectroscopy and artificial neural network identification of *Fusarium* species. *Journal of Phytopathology* 155, 364-367.

Nilsson, R.H., Ryberg, M., Abarenkov, K., Sjökvist, E., Kristiansson, E. 2009. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiol Lett* 296, 97-101.

Nucci, M., Anaissie, E. 2007. *Fusarium* infections in immunocompromised patients. *Clin Microbiol Rev* 20, 695-704.

O'Donnell, K., Kistler, H.C., Cigelnik, E., Ploetz, R.C. 1998. Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc Natl Acad Sci U S A* 95, 2044-2049.

Peterson, S.W. 2008. Phylogenetic analysis of *Aspergillus* species using DNA sequences from four loci. *Mycologia* 100, 205-226.

Pitt, J.I. 1994. The current role of *Aspergillus* and *Penicillium* in human and animal health. *J Med Vet Mycol* 32 Suppl 1, 17-32.

Pitt, J.I., Hocking, A.D. 2009. *Fungi and Food Spoilage*, 3rd edn. Pitt JI, Hocking AD ed. Springer, New York.

Rozynek, P., Gilges, S., Bruning, T., Wilhelm, M. 2004. Quality test of the MicroSeq D2 LSU Fungal Sequencing Kit for the identification of fungi. *Int J Hyg Environ Health* 207, 297-299.

Santos, C., Fraga, M.E., Kozakiewicz, Z., Lima, N. 2010a. Fourier transform infrared as a powerful technique for the identification and characterization of filamentous fungi and yeasts. *Res Microbiol* 161, 168-175.

Santos, C., Paterson, R.R., Venancio, A., Lima, N. 2010b. Filamentous fungal characterizations by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Appl Microbiol* 108, 375-385.

Savitzky, A., Golay, M.J.E. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* 36, 1627-1639.

Schmitt, I., Crespo, A., Divakar, P.K., Fankhauser, J.D., Herman-Sackett, E., Kalb, K., Nelsen, M.P., Nelson, N.A., Rivas-Plata, E., Shimp, A.D., Widhelm, T., Lumbsch, H.T. 2009. New primers for promising single-copy genes in fungal phylogenetics and systematics. *Persoonia* 23, 35-40.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A* 109, 6241-6246.

Shapaval, V., Moretro, T., Suso, H.P., Asli, A.W., Schmitt, J., Lillehaug, D., Martens, H., Bocker, U., Kohler, A. 2010. A high-throughput microcultivation protocol for FTIR spectroscopic characterization and identification of fungi. *J Biophotonics* 3, 512-521.

Shapaval, V., Schmitt, J., Moretro, T., Suso, H.P., Skaar, I., Asli, A.W., Lillehaug, D., Kohler, A. 2013. Characterization of food spoilage fungi by FTIR spectroscopy. *J Appl Microbiol* 114, 788-796.

Skiada, A., Pagano, L., Groll, A., Zimmerli, S., Dupont, B., Lagrou, K., Lass-Flörl, C., Bouza, E., Klimko, N., Gaustad, P., Richardson, M., Hamal, P., Akova, M., Meis, J.F., Rodriguez-Tudela, J.L., Roilides, E., Mitrousia-Ziouva, A., Petrikos, G. 2011. Zygomycosis in Europe: analysis of 230 cases accrued by the registry of the European Confederation of Medical Mycology (ECMM) Working Group on Zygomycosis between 2005 and 2007. *Clin Microbiol Infect* 17, 1859-1867.

Stone, M. 1974. Cross-validation choice and assessment of statistical predictions. *J R Stat Soc Series B Stat Methodol* 36, 111–147.

Suh, S.O., Blackwell, M., Kurtzman, C.P., Lachance, M.A. 2006. Phylogenetics of *Saccharomycetales*, the *ascomycete* yeasts. *Mycologia* 98, 1006-1017.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731-2739.

Tenenhaus, M. 1998. La régression PLS: théorie et pratique. Tenenhaus M ed. Technip, Paris.

Verscheure, M., Lognay, G., Marlier, M. 2002. Revue bibliographique: les méthodes chimiques d'identification et de classification des champignons. *Biotechnol Agron Soc Environ* 6, 131–142.

Vitale, R.G., de Hoog, G.S., Schwarz, P., Dannaoui, E., Deng, S., Machouart, M., Voigt, K., van de Sande, W.W., Dolatabadi, S., Meis, J.F., Walther, G. 2011. Antifungal susceptibility and phylogeny of opportunistic members of the order *mucorales*. *J Clin Microbiol* 50, 66-75.

Walker, D.M., Castlebury, L.A., Rossman, A.Y., White, J.F., Jr. 2012. New molecular markers for fungal phylogenetics: two genes for species-level systematics in the *Sordariomycetes* (*Ascomycota*). *Mol Phylogenet Evol* 64, 500-512.

Wang, Y., Geng, Y., Ma, J., Wang, Q., Zhang, X.G. 2012. *Sinomyces*: a new genus of anamorphic *Pleosporaceae*. *Fungal Biol* 115, 188-195.

White, T.J., Bruns, T., Lee, S., Taylor, J. 1990. PCR Protocols: A Guide to Methods and Applications. Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. ed. Academic Press, San Diego.

Wu, W., Guo, Q., Jouan-Rimbaud, D., Massart, D.L. 1999. Using contrasts as data pretreatment method in pattern recognition of multivariate data. *Chemometr Intell Lab* 45, 39-53.

Zhang, N., Castlebury, L.A., Miller, A.N., Huhndorf, S.M., Schoch, C.L., Seifert, K.A., Rossman, A.Y., Rogers, J.D., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Sung, G.H. 2006. An overview of the systematics of the *Sordariomycetes* based on a four-gene phylogeny. *Mycologia* 98, 1076-1087.

Zhang, Y., Crous, P.W., Schoch, C.L., Hyde, K.D. 2012. *Pleosporales*. *Fungal Divers* 53, 1-221.

III.3- Article 2

**Implementation of an FTIR spectral library
of 486 filamentous fungi strains for rapid
identification of molds**

Manuscrit soumis à *Food Microbiology*, Octobre 2013, actuellement en révision

- Préambule à l'article 2

Contexte de l'étude

Les champignons filamenteux sont des microorganismes ubiquitaires très importants pouvant jouer un rôle bénéfique ou néfaste. Certains d'entre eux sont employés pour produire des produits pharmaceutiques, des enzymes, des acides organiques ou des aliments. Le principal rôle des champignons dans la nature est le recyclage de la matière végétale organique. A l'inverse, certains champignons filamenteux produisent des mycotoxines qui sont la principale préoccupation de l'industrie agroalimentaire. Les méthodes conventionnelles d'identification des champignons filamenteux utilisées en routine reposent essentiellement sur l'analyse morphologique, sont chronophages et nécessitent une grande connaissance du domaine des micromycètes. Les méthodes moléculaires, utilisées en routine comme outils complémentaires, sont coûteuses et difficiles à mettre en œuvre. Dans ce contexte, il est nécessaire de développer des techniques simples, performantes, peu coûteuses et utilisables directement en industrie pour l'identification des champignons filamenteux. La spectroscopie IRTF à haut débit présente des qualités intéressantes et a fait ses preuves lors de nombreuses applications dans le domaine de la microbiologie en tant qu'outil de discrimination et d'identification des microorganismes.

Objectifs

Cette étude a plusieurs objectifs. Le premier objectif est tout d'abord d'utiliser un protocole simple et rapide pour l'identification des champignons filamenteux en utilisant la spectroscopie IRTF couplée à l'analyse discriminante par méthode des moindres carrés (PLS-DA). Cette dernière fait partie des méthodes chimiométriques supervisées d'analyse linéaire multivariée. Le deuxième objectif est de construire une banque de données spectrales pour l'identification des filamenteux. Ensuite, le troisième objectif est de valider la robustesse et la précision de la banque de données spectrales d'une part, en établissant un score et un seuil de prédiction pour la validation des résultats et d'autre part, en mettant en place une fonction de standardisation permettant la transférabilité de la méthode à un autre module IRTF que celui utilisé pour développer la base de données.

Matériels et méthodes

Un total de 486 souches (43 genres et 140 espèces), dont l'identification a été validée par séquençage moléculaire, ont été analysées à l'aide d'un premier spectromètre IRTF à haut débit. Les filaments ont été obtenus en cultivant les souches en milieu liquide (Chemboost YM, AES Chemunex/Biomérieux) pendant 48h. Chaque souche a fait l'objet de trois cultures indépendantes effectuées à des jours différents afin de vérifier la reproductibilité de la méthode. Des tests de qualité ont permis d'éliminer les spectres non conformes parmi les spectres enregistrés et des prétraitements mathématiques (correction de la ligne de base, dérivée seconde et normalisation vectorielle) ont été appliqués afin d'optimiser la matrice de données. La PLS-DA, méthode d'analyse statistique supervisée mettant en jeux des régressions PLS dites multivariées, a été utilisée comme méthode d'analyse chimiométrique dans les gammes spectrales 3200-2800 et 1800-800 cm^{-1} . Un premier jeu de spectres comprenant 288 souches (26 genres et 68 espèces) a été utilisé pour construire les différents modèles de calibration en cascade. Un deuxième jeu de spectres incluant 177 souches a été utilisé pour la validation des différents modèles de calibration et pour l'établissement d'un score et d'un seuil de validation des résultats de prédiction. Parmi ces 177 souches, seulement 105 sont représentées au sein de la banque de données, les souches restantes ne possèdent pas d'homologue au sein de la banque de données soit au niveau genre soit au niveau espèce. Enfin, un troisième jeu de spectres incluant 21 souches a été utilisé pour l'étude de la transférabilité de la méthode.

Résultats

La validation croisée des échantillons de calibration a permis d'optimiser le nombre de dimensions pour chaque modèle de calibration. L'identification de 105 souches de moisissures appartenant à 18 genres et 54 espèces, au niveau du genre et de l'espèce à respectivement 99.17% et à 92.3%, a permis la validation des différents modèles. L'établissement d'un score compris, entre 0 et 100, et d'un seuil de validation du résultat de prédiction fixé à 70 a permis de confirmer la validation de la banque de données spectrales. Ceci a été réalisé à l'aide des 105 souches représentées dans la banque de données et de 72 souches non représentées dont 27 non représentées au niveau du genre et 45 non représentées au niveau de l'espèce. Le pourcentage de résultats corrects obtenus en utilisant le score et le seuil est de 80.55%. L'analyse de 14 souches (3 genres et 7 espèces) sur deux modules IRTF

différents, l'un utilisé pour la construction de la banque de données spectrales (instrument 1) et l'autre situé sur un autre site et n'ayant pas servi à la construction des modèles de calibration (instrument 2), a permis de développer une fonction mathématique permettant la transférabilité de la méthode et donc de la banque de données à un autre appareil. Grâce à cette fonction, le pourcentage de spectres bien prédits de 7 autres souches (2 genres et 5 espèces), analysées sur l'instrument 2, à l'aide de la banque de données a été amélioré, passant de 72.15% à 89.13%.

Conclusion

Par sa simplicité de mise en œuvre, grâce au protocole développé, la spectroscopie IRTF couplée à une méthode d'analyse chimiométrique, constitue une réelle alternative aux autres méthodes de discrimination et d'identification des champignons filamenteux, de part sa rapidité et son faible coût. De plus, les résultats obtenus sont corrélés à l'identification moléculaire des souches étudiées. Une amélioration de la fonction de standardisation permettra de mener une étude multi-sites et de confronter les spectres obtenus avec le même protocole à une base de données centralisée.

Implementation of an FTIR spectral library of 486 filamentous fungi strains for rapid identification of molds

A. Lecellier^a, V. Gaydou^a, J. Mounier^b, A. Hermet^b, L. Castrec^b, G. Barbier^b, W. Ablain^c, M. Manfait^a, D. Toubas^{a, d}, G.D. Sockalingum^a

^a*MÉDIAN-Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, FRE CNRS 3481MEDyC, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 REIMS cedex, France*

^b*Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (EA3882), SFR148 ScInBioS, Université Européenne de Bretagne, Université de Brest, ESIAB, Technopôle de Brest Iroise, 29280 Plouzané, France*

^c*AES CHEMUNEX/BIOMERIEUX, Rue Maryse Bastié, CS17219 Ker Lann, 35172 Bruz cedex, France*

^d*Laboratoire de Parasitologie Mycologie, CHU de Reims, Hôpital Maison Blanche, 45 rue Cognacq Jay, 51092 Reims cedex, France*

Corresponding author:

Ganesh D Sockalingum

Université de Reims Champagne-Ardenne

Equipe MÉDIAN, Biophotonique et Technologies pour la Santé

Unité MEDyC, CNRS FRE3481

UFR Pharmacie, SFR CAP-Santé FED4231

51 rue Cognacq-Jay, Reims, France.

Tel: +33 (0)3 26 91 35 53

Fax: +33 (0)3 26 91 35 50

Email: ganesh.sockalingum@univ-reims.fr

Abstract

Filamentous fungi may cause food and feed spoilage and produce harmful metabolites to human and animal health such as mycotoxins. Identification of fungi using conventional phenotypic methods is time-consuming and molecular methods are still quite expensive and require specific laboratory skills. In the last two decades, it has been shown that Fourier transform infrared (FTIR) spectroscopy was an efficient tool for microorganism identification. The aims of this study were to use a simple protocol for the identification of filamentous fungi using FTIR spectroscopy coupled with a partial least squares discriminant analysis (PLS-DA), to implement a procedure to validate the obtained results, and to assess the transferability of the method and database. FTIR spectra of 486 strains (43 genera and 140 species) were recorded. An IR spectral database built with 288 strains was used to identify 105 different strains. It was found that 99.17% and 92.3% of spectra derived from these strains were correctly assigned at the genus and species levels, respectively. The establishment of a score and a threshold permitted to validate 80.79% of the results obtained. A standardization function (SF) was also implemented and tested on FTIR data from another instrument on a different site and permitted to increase the percentage of well predicted spectra for this set from 72.15% to 89.13%. This study confirms the good performance of high throughput FTIR spectroscopy for fungal identification using a spectral library of molds of industrial relevance.

Keywords: FTIR spectroscopy, Filamentous fungi, Identification, Food industry, PLS-DA, Standardization function.

1. Introduction

Microbiological hazards are a major issue for the food and feed industry and include contamination of food or feed with pathogenic and spoilage microorganisms. In the case of filamentous fungi, the main microbiological hazard is the production of specific secondary metabolites known as mycotoxins (Berthiller et al., 2013; Rodriguez, Rodriguez, Luque, Martin, & Cordoba, 2012) which are synthesized by many species belonging mainly to the *Aspergillus*, *Penicillium* and *Fusarium* genera and are considered harmful to human and animal health (Terra, Prado, Pereira, Ematne, & Batista, 2012). As a consequence, strict regulations are in place for the most toxic mycotoxins in most countries to manage the mycotoxin risk in food and feed (Moss, 2008). Moreover, regulatory pressure is increasingly important with the implementation of specific norms for the control of mycotoxins (Campagnoli et al., 2011; Fredlund et al., 2009). Hence, the identification of fungal contaminants represents a challenge to ensure and guarantee the sanitation and the control of food and feed products and their environments.

Currently, the routine identification of filamentous fungi is mainly based on phenotypic methods which often require an expertise in their morphological analysis (Marinach-Patrice et al., 2009). These methods are often time-consuming, laborious, and can lack sensitivity due to the vast diversity of the fungal species present in food or feed. The molecular methods currently available such as sequencing of the internal transcribed spacer region or other genes of interest are still costly and difficult to implement in routine laboratory practices (Nilsson, Ryberg, Abarenkov, Sjokvist, & Kristiansson, 2009). Consequently, the lack of adapted tools for the food and feed sectors may lead to a misidentification of fungi and generates an uncontrolled risk for these industries. In this context, there is a need for developing simple and rapid approaches adapted to industrial set-ups for the identification of filamentous fungi.

Recent developments concerning Fourier transform infrared (FTIR) spectroscopy have allowed implementing alternative identification methods adapted to a large range of microbiological samples. Infrared spectroscopy is a vibrational spectroscopic technique which is based on the measurement of the fundamental molecular bond vibrational modes. In this technique, a polychromatic infrared source (wavelengths 2.5 and 25 μm) interacts with the sample and the molecules therein can either absorb or reflect the radiation, whereby vibrational motions are stimulated. Only specific frequencies are absorbed corresponding to their molecular modes of vibration that are characteristic of their chemical bonding, composition, and structure. The changes in light absorption at specific frequencies allow

determining which molecular groups are present and how they are arranged or interacting. The result is given under the form of a FTIR spectrum. Thus, on this spectrum, each spectral band characterized by its frequency and intensity, reflects “the molecular fingerprint” of the characteristic molecules (Duygu, Baykal, Açikgöz, & Yildiz, 2009; Naumann, 2000). The spectral profile gives information about important macromolecules like proteins, lipids, nucleic acids and carbohydrates present in cells (Figure 1).

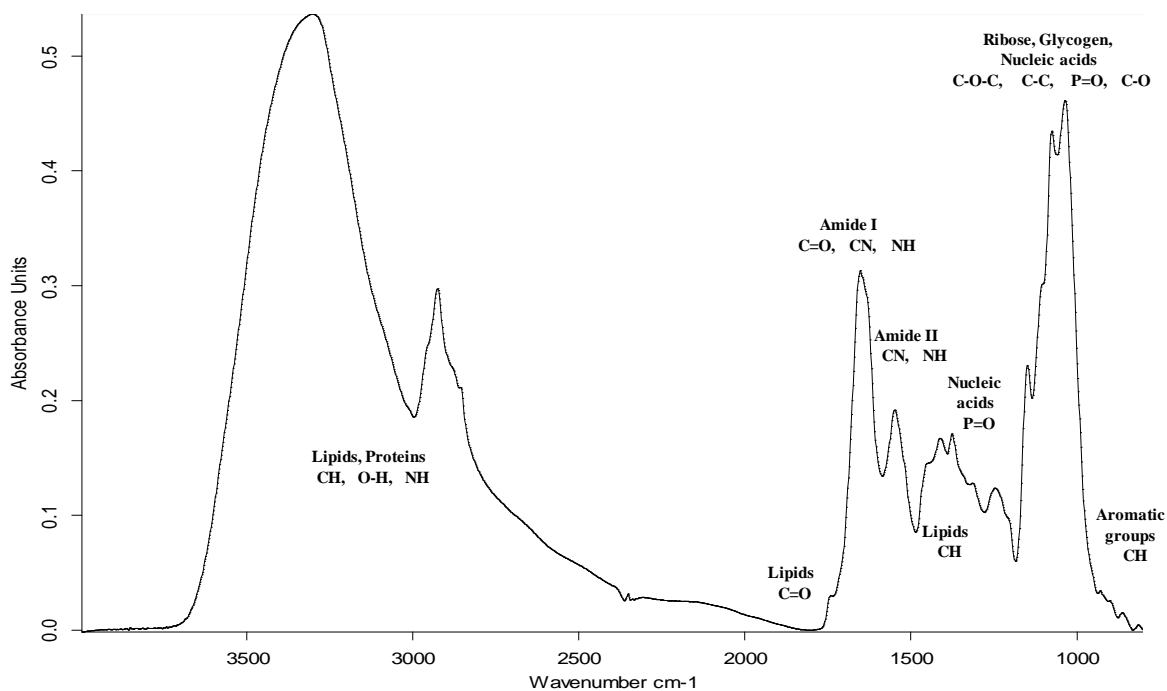


Figure 1: Characteristic raw infrared absorption spectrum of molds. The different molecular bonds are indicated with their biomolecular attribution (= stretching vibration, = bending vibration).

Concerning the filamentous fungi, this method has already been used for the differentiation and the classification of closely related species such as *Aspergillus fumigatus*, *Aspergillus flavus*, and *Aspergillus parasiticus* (Garon, El Kaddoumi, Carayon, & Amiel, 2010) and *Aspergillus niger*, *Aspergillus ochraceus*, and *Aspergillus westerdijkiae* (Tralamazza et al., 2013). Another recent study assessed the ability of FTIR spectroscopy for the differentiation and classification of clinically relevant *Trichophyton* species (Ergin et al., 2013). *Fusarium* species were also differentiated and discriminated using this method (Nie et al., 2007). In most studies related to the identification of filamentous fungi using FTIR, the focus was only made on one genus. Few studies relate to the ability of FTIR spectroscopy for discrimination

and identification of several fungal genera and species. FTIR spectroscopy was applied for the identification of airborne fungi belonging to the *Aspergillus*, *Emericella*, and *Penicillium* genera (Fischer, Braun, Thissen, & Dott, 2006). In recent studies, a high throughput protocol for FTIR spectroscopy was developed for the characterization and the identification of 11 strains (5 genera and 7 species) and this protocol was applied for the characterization of 59 food spoilage fungal strains belonging to 10 genera (Shapaval et al., 2010; Shapaval et al., 2013). The latter studies have shown that this method could be an alternative to the methods currently used for fungal identification as it was accurate and allowed a high throughput analysis using micro-culture.

Like for other emerging routine identification techniques such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), FTIR-based identification requires to build a spectral library specific to the filamentous fungal species of interest, e.g., the most frequently encountered species in the food and feed sectors. The performance of a library also relies on the choice of the statistical chemometric method used for developing calibration models. FTIR studies have shown that linear regression methods can be used to optimize models (Luna, da Silva, Ferre, & Boque, 2013; Navea, Tauler, Goormaghtigh, & de Juan, 2006). Recently, partial least squares discriminant analysis (PLS-DA) was compared to other chemometrics methods such as principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA) to assess FTIR spectroscopy based bacteria classification models (Preisner, Lopes, & Menezes, 2008). In the latter study, it was found that PLS-DA was more adequate than the other two methods in situation where spectral discrimination was difficult and this method appeared to be the most appropriate one for a classification going down to the species level.

In addition, one of the concerns to identify a microorganism at the species level using a spectral library is when the species to be predicted is not therein represented. In this case, it will be assigned to the closest species of the databank, leading to a wrong identification. Therefore, it is important to establish a score and a validation threshold in order to validate or invalidate the prediction result as developed for other identification methods based on spectral databanks such as MALDI-TOF MS (Khot, Couturier, Wilson, Croft, & Fisher, 2012; Marklein et al., 2009).

In IR spectroscopy, the spectral response of a given instrument requires a correction before its spectra could be compared with those of another instrument. Indeed, a spectral database built on a specific instrument and the different prediction models associated with this database can give good prediction results and a good identification of an unknown sample analyzed on the

same instrument. In the case where an unknown sample is analyzed on another instrument, the identification results can be biased. A standardization function from one instrument of a given laboratory to another instrument of a different laboratory is therefore necessary in order to use database with an unknown external sample (Zhang, Small, & Arnold, 2003).

In a recent publication (Lecellier et al., 2013), we developed a small database of FTIR spectra of food relevant molds. Identification levels of over 98% were achieved for the genus and species levels. The aims of this study were to expand the database developed previously for the filamentous fungi identification using a high-throughput FTIR spectroscopic analysis coupled with PLS-DA method and to test its robustness with 486 strains belonging to 43 genera and 140 species. We also established a score and threshold to validate prediction results and studied the transferability of the method to data acquired on another FTIR instrument.

2. Material and methods

2.1. Fungal strains

A total of 486 filamentous fungal strains (43 genera and 140 species), obtained from the Culture Collection of Université de Bretagne Occidentale (UBOCC, Plouzané, France) and the Culture Collection of Centraalbureau voor Schimmelcultures (CBS, Utrecht, The Netherlands), were used for this study in the form of mycelium plugs stored in 10% glycerol at -80°C. All strains were identified at the species level by morphological analysis based on macroscopic and microscopic features. The identification was then confirmed by DNA sequencing. Total genomic DNA was extracted using the 'FastDNA SPIN Kit' (MPBio, Illkirch, France) according to the manufacturer's instructions from mycelia grown in potato dextrose broth for 2 to 4 days at 25°C on a rotary shaker at 120 rpm. DNA amplification of 5 different regions depending on the fungal genus, e.g., the rDNA internal transcribed spacer (ITS) region including the 5.8S rRNA gene (all genera except *Fusarium spp.*), partial α -tubulin gene (*Penicillium* and *Aspergillus spp.*), partial translation elongation factor-1 alpha (TEF-1) gene (*Fusarium spp.*) and, partial mcm7 and partial tsr1 genes (*Mucor spp.*) was performed as described previously (Glass & Donaldson, 1995; Hermet, Meheust, Mounier, Barbier, & Jany, 2012; O'Donnell, Kistler, Cigelnik, & Ploetz, 1998; Schmitt et al., 2009; White, Bruns, Lee, & Taylor, 1990). After sequencing of the amplicons, the DNA sequences

were compared to the GenBank database using the Basic Local Alignment Search Tool (BLAST) to determine the taxonomic assignment of fungal isolates. Alignments of the resulting sequences and sequences from the NCBI database were performed using the MAFFT online server (MAFFT version 7) and the E-INS-i iterative refinement method. Phylogenetic trees were built in MEGA5 (Tamura et al., 2011) using the Neighbor-Joining method with 1000 bootstrapped data sets.

2.2. FTIR analysis

2.2.1. Culture conditions and sample preparation

For each cryopreserved strain, a sub-culture was realised on Sabouraud agar slants (Becton Dickinson, Le Pont de Claix, France) during 4 to 7 days at 25°C depending on the strain. The obtained spores were then inoculated in 20 ml of Chemboost YM broth (AES Chemunex, Bruz, France) and incubated for 48 h at 25°C under agitation (150 rpm). For each strain, 3 independent cultures were prepared on 3 different days (biological replicates). “M” tubes adapted to a gentleMACS Octo Dissociator (Miltenyi Biotec, Paris, France) were used to dissociate each culture (4000 rpm for 100 s) in order to obtain homogeneous suspensions for FTIR analysis. Two milliliters of each suspension, transferred into an eppendorf tube, were centrifuged (430 x g for 30 s) to eliminate the culture supernatant. The mycelia were washed using 1 ml of 0.9 % NaCl and the supernatant was discarded by centrifugation (430 x g for 30 s). The mycelial pellets were resuspended in approximately 300 µl of 0.9 % NaCl before FTIR spectroscopy analysis.

2.2.2. Spectral acquisition and pre-processing of FTIR spectra

Five microliters of each suspension were deposited on each position of a 384-well silicon plate in eight replicates to verify the measurement repeatability (instrumental replicates). The plate was then dried under mild vacuum for 1 h to avoid influence of water absorption on the infrared spectra. An FTIR high-throughput spectrometer HTS-XT Tensor 27 (Bruker Optics, Ettlingen, Germany) was used for spectral acquisition. The acquisition parameters were: 64 accumulations per well, spectral resolution of 4 cm⁻¹, spectral range of 4000-400 cm⁻¹. Before each sample measurement, the background spectrum of the silicon plate was recorded and subsequently removed from the sample signal. A quality test was performed on each raw

spectrum based on a previous report (Helm, Labischinski, & Naumann, 1991) using different parameters like signal intensity, noise, and water vapour signal in specific spectral regions. These parameters are summarized in Table 1. Raw spectra that did not passed the quality test were automatically eliminated. A fungal culture was validated when at least 5 out of 8 spectra corresponding to the 8 replicates passed this quality test. Pre-processing of infrared spectra included truncation in the region 4000-800 cm^{-1} ; a rubberband baseline correction with 64 baseline points (Wu, Guo, Jouan-Rimbaud, & Massart, 1999); a second derivation with Savitsky-Golay 9 smoothing points (Savitzky & Golay, 1964); and a vector normalization (Bylesjö M, Cloarec O, & Rantalainen M, 2009). All spectra were recorded and pre-processed with the OPUS software (version 5.5 and 6.6, Bruker Optics, Ettlingen, Germany).

Table 1: Parameters used for quality-testing applied on the raw spectra.

	X-ranges (cm^{-1})				
	2100-1600	1700-1600	1200-960	2100-2000	1847-1837
Absorbance (min-max)	0.17-1	-	-	-	-
Noise (max)	-	-	-	0.00016	-
Signal/Noise (min)	-	50	10	-	-
Water (max)	-	-	-	-	0.0003
Signal/Water (min)	-	20	4	-	-

2.3. Chemometric analysis

In this study, partial least square discriminant analysis (PLS-DA), developed in Matlab (version 7.2, Mathwork, USA) was used to classify the samples using their explained variables in the spectral ranges 3200-2800 and 1800-800 cm^{-1} . PLS-DA is a supervised linear analysis method based on the algorithm of multivariate PLS regression which was adapted as a method of classification (Tenenhaus, 1998) and was carried out using a binary code. Each class is coded by a combination of 0 and 1, based on whether or not the sample belongs to a given class. The PLS-DA algorithm, assisted by multivariate regression, permits to correlate explained data corresponding to the explained variables of the samples with a matrix of class properties corresponding to the different classes of the samples (Liang & Kvalheim, 1996). This study including 288 fungal strains belonging to 26 genera and 68 species (Table 2) and used for the calibration step were grouped into a pre-established class based on the molecular sequencing results and according to the current taxonomy. Models were built with these

strains for each of the taxonomic ranks from phylum/subphylum to species to which the fungal strains belonged. The quality of the calibration models and the choice of the best parameters for each model were evaluated by the partial leave-one-out cross-validation method (Stone, 1974). All replicate spectra from a same culture left-out and calibration models were developed with the remaining spectra. The accuracy of the different created models was verified using each spectrum of the removed culture. In this approach, all the spectra of the calibration set serve both for their construction and for the model optimization. The partial cross-validation was tested cumulatively from 1 to 35 dimensions in order to determine for each model the number of PLS-DA iterations which provided the best percentage of well classified spectra in the calibrations models. These numbers of iteration were used to construct the definitive calibration models. The validation was then performed with another spectral set obtained from 105 strains belonging to 18 genera and 54 species. All these strains belonged to species that were already represented in the database.

Table 2: List and the number of fungal species used to construct the spectral database.

Species	Number of strains	Species	Number of strains	Species	Number of strains	Species	Number of strains
<i>Absidia coerulea</i>	2	<i>Circinella sydowii</i>	2	<i>Fusarium sporotrichoides</i>	2	<i>Penicillium expansum</i>	7
<i>Actinomyces elegans</i>	3	<i>Cladosporium cladosporioides</i>	2	<i>Fusarium verticillioides</i>	8	<i>Penicillium chrysogenum</i>	6
<i>Alternaria alternata</i>	3	<i>Cladosporium ramotenellum</i>	2	<i>Geotrichum candidum</i>	5	<i>Penicillium citrinum</i>	2
<i>Aspergillus elegans</i>	2	<i>Cladosporium sphaerospermum</i>	3	<i>Geotrichum citri-aurantii</i>	2	<i>Penicillium commune</i>	4
<i>Aspergillus flavus</i>	9	<i>Cunninghamella elegans</i>	2	<i>Lichtheimia corymbifera</i>	3	<i>Penicillium crustosum</i>	2
<i>Aspergillus fumigatus</i>	15	<i>Emericella nidulans</i>	4	<i>Microdochium nivale</i>	3	<i>Penicillium glabrum</i>	5
<i>Aspergillus niger</i>	5	<i>Eurotium amstelodami</i>	2	<i>Mucor circinelloides</i>	10	<i>Penicillium nalgiovense</i>	2
<i>Aspergillus sclerotium</i>	2	<i>Eurotium chevalieri</i>	2	<i>Mucor hiemalis</i>	3	<i>Penicillium oxalicum</i>	4
<i>Aspergillus tamari</i>	3	<i>Eurotium rubrum</i>	2	<i>Mucor racemosus</i>	11	<i>Penicillium paneum</i>	3
<i>Aspergillus versicolor</i>	4	<i>Fusarium avenaceum</i>	10	<i>Mucor spinosus</i>	7	<i>Penicillium roqueforti</i>	6
<i>Aspergillus wentii</i>	3	<i>Fusarium culmorum</i>	6	<i>Mucor velutinosus</i>	2	<i>Penicillium verrucosum</i>	4
<i>Aspergillus sydowii</i>	3	<i>Fusarium equiseti</i>	15	<i>Paecilomyces variotii</i>	4	<i>Rhizopus oryzae</i>	6
<i>Aureobasidium pullulans</i>	4	<i>Fusarium graminearum</i>	12	<i>Paecilomyces lilacinus</i>	3	<i>Scopulariopsis fusca</i>	3
<i>Bionectria ochroleuca</i>	2	<i>Fusarium langsethiae</i>	2	<i>Penicillium brevicompactum</i>	4	<i>Stagonosporopsis valerianellae</i>	2
<i>Botrytis cinerea</i>	2	<i>Fusarium oxysporum</i>	10	<i>Penicillium camemberti</i>	3	<i>Umbelopsis isabellina</i>	2
<i>Ceratocystis paradoxa</i>	2	<i>Fusarium sambucinum</i>	4	<i>Penicillium carneum</i>	4	<i>Verticillium dahliae</i>	2
<i>Chaetomium globosum</i>	2	<i>Fusarium solani</i>	3	<i>Penicillium corylophilum</i>	3	<i>Verticillium lecanii</i>	2

2.4. Prediction score and validation threshold

In order to validate the prediction result corresponding to the predominantly predicted result obtained for all the spectra of one strain, a score (S) and a validation threshold were implemented. The score was calculated with Matlab version 7.2 (Mathwork, USA) using the Euclidian distances between each spectrum from the cluster corresponding to the predominantly predicted result and the spectrum to predict. The average of the scores for each

spectrum from a given strain to predict was calculated and the average of the scores obtained for all spectra to predict was determined. Then, the calculated scores were multiplied by h , corresponding to the homogeneity defined by the percentage of predominantly predicted spectra on the total of the spectra per strain, in order to add a weighting to the results (equation 2). To determine a score and a threshold, 3 spectral sets were used. The first set corresponded to the 105 strains (18 genera and 54 species) which were represented by strains from the same species in the database; the second set included 27 strains (17 genera and 27 species) not represented in the database at the genus level; and the third set included 45 strains (17 genera and 45 species) not represented at the species level in the database.

$$\mathbf{S} = (\mathbf{1-D}) \times \mathbf{h} \quad (\text{equation 2})$$

With: S = Score

D = Euclidian distance

h = homogeneity

2.5. Instrument to instrument standardization function

In order to verify the transferability of the identification method from an instrument to another, 14 strains were analyzed on two different instruments. Instrument 1 represents the instrument that was used to build the spectral library and instrument 2 was on a different geographic site and operated by a different operator. The samples were prepared using the same method and the spectra were recorded using the same spectral acquisition parameters and pre-processed as described above. A standardization function (SF) was calculated from the spectra of 14 strains (equation 1). Firstly, for each strain, the median of the derivative spectra corresponding to the result predominantly predicted was calculated. Then, the 14 median spectra from strains analyzed with instrument 1 were subtracted from the median spectra of the same 14 strains analyzed with instrument 2. Finally, the median of all median spectra was calculated. The quality of the calibration standardization function was tested by the total leave-one-out cross validation method (Stone, 1974). All spectra from one strain were left-out and the standardization function was calculated with the remaining strains. The accuracy of the different standardization functions were checked on each left-out strain. The standardization function was then applied to a spectral set to predict 7 other strains (2 genera

and 5 species), only analyzed with instrument 2, in order to validate the standardization function.

$$\mathbf{SF} = \mathbf{median (Inst1- Inst2)} \quad (\text{equation1})$$

With: Inst1= matrix including the median derivative spectra of each strain analyzed with instrument 1.

Inst2 = matrix including the median derivative spectra of each strain analyzed in the instrument 2.

3. Results

3.1. Database construction

The classification of 288 strains (26 genera and 68 species) using the PLS-DA algorithm applied to the second derivative IR spectra of these strains was undertaken. All these strains were used to build the calibration set composed of 6275 spectra. A total of 29 calibration models were constructed in cascade based on the current taxonomic classification of fungi using 9 taxonomic ranks from the subphylum to the species levels. The assignment of each strain to a taxonomic group was done on the basis of the results of the identification obtained by DNA sequencing. The partial leave-one-out cross validation method, used to test all the calibration models and to verify the accuracy of good prediction, also permitted to determine the number of iterations which provided the best percentage of sample prediction for each calibration model. The percentage of well classified spectra and the associated number of iterations for each model are shown in Table 3. It was found that the percentage of well classified spectra per model in the database was higher than 99% for all the models.

Table 3: Percentage of well classified spectra for the calibration model, with the corresponding taxonomic rank and the number of iterations associated.

Calibration model name	Taxonomic rank	Number of clusters	Number of spectra	Number of iterations	% of well classified spectra per model
<i>Micromycetes</i>	SUBPHYLUM	3	6275	22	99.98
<i>Pezizomycotina</i>	CLASS	4	5035	30	99.68
<i>Dothideomycetes</i>	ORDER	3	348	6	100
<i>Sordariomycetes</i>		5	2103	30	100
<i>Pleosporales</i>	FAMILY	2	110	2	100
<i>Mucorales</i>		3	1092	21	99.91
<i>Hypocreales</i>		4	1832	18	99.95
<i>Microascales</i>		2	104	2	100
<i>Lichtheimiaceae</i>	GENUS	2	86	2	100
<i>Mucoraceae</i>		4	931	14	100
<i>Trichocomaceae</i>		3	2536	25	99.53
<i>Fusarium</i>	SUBGENUS	6	1676	30	99.34
<i>Penicillium</i>		2	1241	25	100
<i>Aspergillus</i>	SECTION	7	1223	28	100
<i>Penicillium 1</i>		5	966	30	100
<i>Fasciculata</i>		2	276	5	99.64
<i>Cunninghamella</i>	SPECIES	2	75	2	100
<i>Nidulantes</i>		3	221	7	100
<i>Circumdati</i>		2	95	4	100
<i>Cremeri</i>		2	135	3	100
<i>Cladosporium</i>		3	156	8	99.36
<i>Eurotium</i>		3	132	9	100
<i>Sambucinum</i>		5	620	30	99.84
<i>Geotrichum</i>		2	148	2	100
<i>Mucor</i>		5	708	17	99.86
<i>Camemberti</i>		3	195	12	100
<i>Chrysogena</i>		2	162	4	100
<i>Roquefortorum</i>		3	318	30	100
<i>Aspergilloides</i>		4	275	4	100

3.2 Database validation

The database validation was carried out using a validation set composed of 2283 spectra corresponding to 105 different strains pertaining to 18 genera and 54 species. All these strains belonged to species that were represented in the database. The percentages of well predicted spectra for each taxonomic rank are shown in Figure 2. An average percentage of 91.56% of well predicted spectra per strain was obtained at the species level. The results also showed

that 99.17% and 92.30% of the FTIR spectra derived from these strains were correctly assigned at the genus and species level, respectively. The percentages of well predicted spectra per strain are illustrated in Table 4. The assignment of 101 out of 105 strains (96.2%) was correctly predicted at the species level. Concerning the misidentified strains, *Eurotium amstelodami* CBS 817.96 and *Eurotium rubrum* CBS 530.65 were misidentified as *Eurotium chevalieri* and *Eurotium amstelodami*, respectively while *Penicillium nalgiovense* UBOCC-A-101431 and *Penicillium citrinum* CBS 309.48 were misidentified as *Penicillium commune* and *Penicillium chrysogenum*, respectively. Although these 4 strains were not correctly assigned at the species level, their assignment at the genus level was correct.

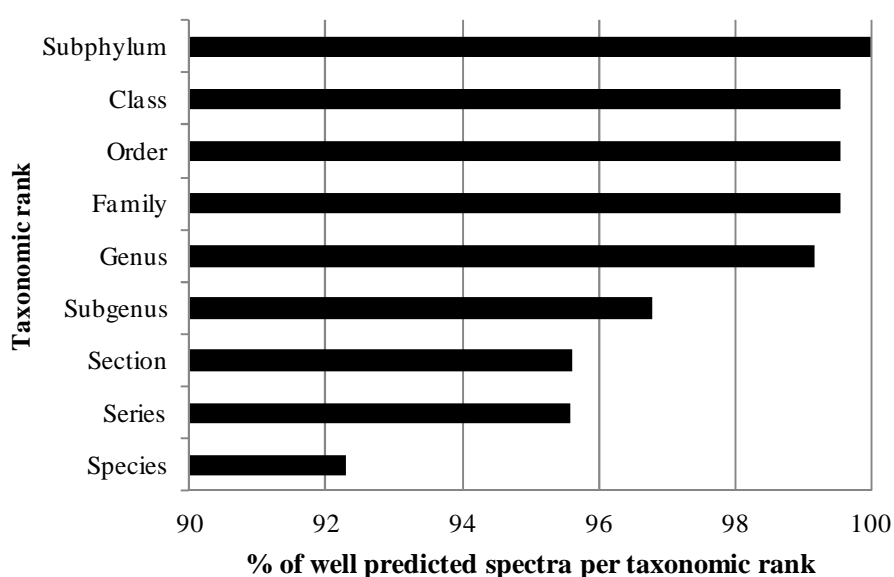


Figure 2: Histogram showing the percentage of well predicted spectra per taxonomic rank for the 105 strains used to validate the calibration models and represented in the spectral database.

3.3 Implementation of a score and validation threshold

The score established in order to validate the prediction results, was calculated for the 105 strains for which other strains from the same species were present in the database (Table 4) and for 72 other strains belonging to genus or species not represented in the database (Table 5). In our case, the scores varied between 0 and 100. With respect to the results obtained (identification using FTIR spectroscopy and DNA sequencing, and the calculated scores), a threshold of 70 was determined. If the score was greater than or equal to 70, the species

Table 4: Predicted species and the percentage of well predicted spectra and scores calculated per strain for the 105 strains used to validate the calibration models and represented in the spectral database.

Strains of the validation set	Predicted species	% of well predicted spectra per strain	Scores	Strains of the validation set	Predicted species	% of well predicted spectra per strain	Scores
^a A. flavus UB0CC-A-101060	^a A. flavus	100	83.31	^f F. graminearum CBS 447.95	^f F. graminearum	95.83	63.71
^a A. flavus UB0CC-A-106029	^a A. flavus	100	85.56	^f F. graminearum UB0CC-A-101143	^f F. graminearum	100	79.55
^a A. flavus UB0CC-A-106030	^a A. flavus	100	83.28	^f F. graminearum UB0CC-A-102016	^f F. graminearum	100	85.76
^a A. flavus UB0CC-A-106031	^a A. flavus	100	87.41	^f F. graminearum UB0CC-A-109011	^f F. graminearum	100	85.85
^a A. niger UB0CC-A-101072	^a A. niger	100	84.75	^f F. graminearum UB0CC-A-109032	^f F. graminearum	100	83.05
^a A. niger UB0CC-A-101073	^a A. niger	100	83.96	^f F. sambucinum UB0CC-A-109020	^f F. sambucinum	66.66	48.38
^a A. tamari CBS 129.49	^a A. tamari	100	78.58	^f F. culmorum UB0CC-A-101139	^f F. culmorum	58.33	42.84
^a A. wentii UB0CC-A-101090	^a A. wentii	100	76.14	^f F. culmorum UB0CC-A-109139	^f F. culmorum	100	74.14
^a A. fumigatus UB0CC-A-101066	^a A. fumigatus	100	80.27	^f F. culmorum UB0CC-A-109109	^f F. culmorum	100	75.48
^a A. fumigatus UB0CC-A-106001	^a A. fumigatus	100	89.86	^f F. langsethiae UB0CC-A-110061	^f F. langsethiae	95.83	66.24
^a A. fumigatus UB0CC-A-106002	^a A. fumigatus	100	91.05	^f F. avenaceum UB0CC-A-101137	^f F. avenaceum	73.33	57.54
^a A. fumigatus UB0CC-A-106003	^a A. fumigatus	100	88.75	^f F. avenaceum UB0CC-A-101136	^f F. avenaceum	87.5	70.18
^a A. fumigatus UB0CC-A-106004	^a A. fumigatus	100	88.33	^f F. avenaceum UB0CC-A-109048	^f F. avenaceum	100	76.62
^a A. fumigatus UB0CC-A-106005	^a A. fumigatus	100	88.41	^f F. avenaceum UB0CC-A-109033	^f F. avenaceum	100	80.97
^b E. nidulans CBS 589.65	^b E. nidulans	85.71	66.79	^f F. equiseti CBS 414.86	^f F. equiseti	80	70.07
^b E. nidulans CBS 492.65	^b E. nidulans	100	65.16	^f F. equiseti UB0CC-A-109007	^f F. equiseti	100	86.03
^a A. versicolor UB0CC-A-102012	^a A. versicolor	100	84.55	^f F. equiseti UB0CC-A-109015	^f F. equiseti	100	86.55
^a A. sydowii UB0CC-A-106017	^a A. sydowii	91.67	72.78	^f F. equiseti UB0CC-A-109016	^f F. equiseti	100	84.69
^a A. sclerotium UB0CC-A-105010	^a A. sclerotium	75	55.88	^f F. equiseti UB0CC-A-109029	^f F. equiseti	100	84.95
^e E. amstelodami CBS 817.96	^e E. chevalieri	0	69.56	^f F. equiseti UB0CC-A-109031	^f F. equiseti	100	86.32
^e E. chevalieri CBS 121704	^e E. chevalieri	65.28	37.39	^f F. equiseti UB0CC-A-109030	^f F. equiseti	100	84.67
^e E. rubrum CBS 530.65	^e E. amstelodami	12.5	27.84	^f F. solani CBS 128.29	^f F. solani	100	74.06
^d P. verrucosum UB0CC-A-105014	^d P. verrucosum	95.45	78.31	^e P. litacinus	^e P. litacinus	100	72.85
^d P. camemberti UB0CC-A-101455	^d P. camemberti	100	80.92	^e V. lecanii	^e V. lecanii	100	76.46
^d P. commune CBS 269.97	^d P. commune	100	79.21	^b M. nivale UB0CC-A-105026	^b M. nivale	100	76.18
^d P. commune UB0CC-A-108127	^d P. commune	72.41	60.96	^s S. fusca UB0CC-A-108119	^s S. fusca	100	83.61
^d P. expansum UB0CC-A-110021	^d P. expansum	81.82	71.11	^k A. alternata CBS 117143	^k A. alternata	100	76.04
^d P. expansum UB0CC-A-110024	^d P. expansum	94.44	81.42	^k A. pullulans UB0CC-A-101091	^k A. pullulans	95.83	65.53
^d P. expansum UB0CC-A-110023	^d P. expansum	100	88.41	^c Sphaerospermum UB0CC-A-101107	^c Sphaerospermum	75	49.53
^d P. roqueforti UB0CC-A-109090	^d P. roqueforti	100	72.71	^c C. ramotenellum UB0CC-A-108072	^c C. ramotenellum	100	78.5
^d P. roqueforti CBS 221.30	^d P. roqueforti	69.23	54.71	^m B. cinerea CBS 810.69	^m B. cinerea	100	78.55
^d P. carneum CBS 112297	^d P. carneum	100	88.98	ⁿ G. candidum UB0CC-A-101170	ⁿ G. candidum	100	88.53
^d P. carneum CBS 112489	^d P. carneum	100	87.95	ⁿ G. candidum UB0CC-A-101169	ⁿ G. candidum	100	76.45
^d P. paneum UB0CC-A-101448	^d P. paneum	84.62	66.25	ⁿ G. candidum UB0CC-A-108080	ⁿ G. candidum	100	82.53
^d P. chrysogenum CBS 111214	^d P. chrysogenum	100	86.94	^o M. circinelloides UB0CC-A-105016	^o M. circinelloides	100	86.73
^d P. chrysogenum CBS 478.84	^d P. chrysogenum	95.83	85.39	^o M. circinelloides UB0CC-A-105018	^o M. circinelloides	100	84.83
^d P. nalgiovensis UB0CC-A-101431	^d P. commune	0	77.33	^o M. circinelloides CBS 195.68	^o M. circinelloides	100	80.56
^d P. brevicompactum CBS 257.29	^d P. brevicompactum	91.3	73.37	^o M. circinelloides UB0CC-A-109066	^o M. circinelloides	100	83.12
^d P. brevicompactum UB0CC-A-108093	^d P. brevicompactum	80.00	63.93	^o M. velutinosis UB0CC-A-109075	^o M. velutinosis	72.73	63.76
^d P. glabrum UB0CC-A-108105	^d P. glabrum	95.65	78.54	^o M. spinosus UB0CC-A-101364	^o M. spinosus	95.83	77.42
^d P. glabrum UB0CC-A-109098	^d P. glabrum	100	81.4	^o M. spinosus CBS 246.58	^o M. spinosus	100	83.77
^d P. corylophilum UB0CC-A-109219	^d P. corylophilum	100	82.63	^o M. spinosus UB0CC-A-109061	^o M. spinosus	100	71.18
^d P. oxalicum UB0CC-A-101435	^d P. oxalicum	95.65	73.23	^o M. spinosus CBS 226.32	^o M. spinosus	65.22	52.82
^d P. oxalicum UB0CC-A-101436	^d P. oxalicum	100	85.02	^o M. racemosus UB0CC-A-109051	^o M. racemosus	100	73.86
^d P. citrinum CBS 309.48	^d P. chrysogenum	50	34.89	^o M. racemosus CBS 260.68	^o M. racemosus	87.5	67.08
^p Pa. variabilis UB0CC-A-103044	^p Pa. variabilis	83.33	65.7	^o M. racemosus UB0CC-A-108091	^o M. racemosus	100	79.71
^f F. oxysporum UB0CC-A-108128	^f F. oxysporum	100	86.45	^o M. racemosus UB0CC-A-109062	^o M. racemosus	100	79.12
^f F. oxysporum UB0CC-A-101157	^f F. oxysporum	86.96	71.46	^o M. hiemalis UB0CC-A-101360	^o M. hiemalis	100	71.66
^f F. oxysporum UB0CC-A-101151	^f F. oxysporum	100	85.68	^p Ac. elegans	^p Ac. elegans	100	74.98
^f F. oxysporum UB0CC-A-101152	^f F. oxysporum	100	84.27	^q R. oryzae CBS 112.07	^q R. oryzae	100	80.6
^f F. verticilloides CBS 119825	^f F. verticilloides	90.48	71.26	^q R. oryzae CBS 127.08	^q R. oryzae	100	78.42
^f F. verticilloides CBS 218.76	^f F. verticilloides	70.83	58.05	^r L. corymbifera UB0CC-A-101328	^r L. corymbifera	100	65.56
^f F. verticilloides UB0CC-A-110165	^f F. verticilloides	95.83	78.66				

^aAspergillus, ^bEmericella, ^cEurotium, ^dPenicillium, ^ePaecilomyces, ^fFusarium, ^gVerticillium, ^hMicrodochium, ⁱScopulariopsis, ^jAlternaria, ^kAureobasidium, ^lCladosporium, ^mBotrytis, ⁿGeotrichum, ^oMucor, ^pActinomyces, ^qRhizopus, ^rLichtheimia

Table 5: Predicted species and scores calculated per strain for the 72 strains used to validate the calibration models and not represented in the spectral database.

Strains of the validation set	Predicted species	% of predominantly predicted spectra	Scores	Strains of the validation set	Predicted species	% of predominantly predicted spectra	Scores
^a <i>P. brunneum</i> UBOCC-A-101391 ¹	^p <i>A. flavus</i>	70.59	49.96	^z <i>E. varicolor</i> UBOCC-A-101071 ²	^z <i>E. nidulans</i>	54.17	39.63
^b <i>T. flavus</i> UBOCC-A-101037 ¹	^r <i>P. lilacinus</i>	37.5	27.83	^p <i>A. westerdijkiae</i> UBOCC-A-101078 ²	^p <i>A. elegans</i>	100	80.09
^a <i>P. concavovrugulosum</i> UBOCC-A-101454 ¹	^s <i>B. ochroleuca</i>	62.5	44.83	^{aa} <i>E. repens</i> UBOCC-A-101079 ²	^{aa} <i>E. chevalieri</i>	55.56	39.69
^c <i>E. pinetorum</i> UBOCC-A-109223 ¹	^a <i>P. corylophilum</i>	66.67	50.39	^a <i>P. nordicum</i> CBS 323.92 ²	^a <i>P. verrucosum</i>	83.33	65.35
^d <i>T. aggressivum</i> CBS 101525 ¹	^r <i>F. graminearum</i>	62.5	43.01	^a <i>P. solitum</i> UBOCC-A-108113 ²	^a <i>P. verrucosum</i>	100	75.14
^d <i>T. harzianum</i> CBS 226.95 ¹	^r <i>P. lilacinus</i>	50	36.15	^a <i>P. viridicatum</i> UBOCC-A-108115 ²	^a <i>P. verrucosum</i>	80.95	63.33
^d <i>T. longibrachiatum</i> UBOCC-A-101290 ¹	^r <i>F. verticillioideis</i>	31.58	21.81	^a <i>P. aurantiogriseum</i> UBOCC-A-108092 ²	^a <i>P. verrucosum</i>	100	79.88
^d <i>T. viride</i> UBOCC-A-101288 ¹	^r <i>P. lilacinus</i>	56.52	44.4	^a <i>P. frei</i> CBS 477.84 ²	^a <i>P. verrucosum</i>	86.67	69.42
^e <i>H. virens</i> UBOCC-A-101176 ¹	^u <i>V. dahliae</i>	25	19.36	^a <i>P. palitans</i> CBS 311.48 ²	^a <i>P. roqueforti</i>	33.33	25.82
^f <i>M. cinctum</i> UBOCC-A-101201 ¹	^v <i>M. hiemalis</i>	69.57	49.81	^a <i>P. glandicola</i> UBOCC-A-101422 ²	^a <i>P. expansum</i>	60	48.5
^g <i>H. fuscoatra</i> UBOCC-A-101190 ¹	^w <i>C. globosum</i>	100	61.95	^a <i>P. raistrickii</i> UBOCC-A-101440 ²	^a <i>P. oxalicum</i>	47.62	34.95
<i>Gelasinospora</i> sp UBOCC-A-101018 ¹	^w <i>C. globosum</i>	83.33	54.26	^a <i>P. coralligerum</i> UBOCC-A-101404 ²	^a <i>P. brevicompactum</i>	43.48	33.02
^h <i>C. acutum</i> UBOCC-A-101180 ¹	^w <i>C. globosum</i>	47.37	35.25	^a <i>P. janthinellum</i> UBOCC-A-101428 ²	^a <i>P. chrysogenum</i>	43.48	33.79
^h <i>C. coccodes</i> UBOCC-A-101118 ¹	^w <i>C. globosum</i>	100	65.65	^a <i>P. rolfsii</i> UBOCC-A-101444 ²	^p <i>A. fumigatus</i>	37.5	28.94
<i>Pestalotiopsis</i> sp UBOCC-A-101216 ¹	^x <i>M. nivale</i>	34.78	24.72	^a <i>P. thomii</i> UBOCC-A-101463 ²	^a <i>P. glabrum</i>	45.83	36.7
ⁱ <i>K. pachypleura</i> UBOCC-A-101266 ¹	^y <i>S. fusca</i>	100	82.32	^a <i>P. spinulosum</i> UBOCC-A-101442 ²	^a <i>P. corylophilum</i>	27.27	20.04
<i>Papularia</i> sp UBOCC-A-101212 ¹	^w <i>C. globosum</i>	54.55	34.49	^a <i>P. fellutanum</i> CBS 172.44 ²	^a <i>P. chrysogenum</i>	60.87	46.82
^j <i>C. parasitica</i> UBOCC-A-101130 ¹	^z <i>E. nidulans</i>	73.33	48.06	^a <i>Pa. saturatus</i> UBOCC-A-101210 ²	^r <i>P. variotii</i>	80.95	62.61
<i>Phomopsis</i> sp UBOCC-A-101245 ¹	^{aa} <i>E. amstelodami</i>	21.74	17.04	^r <i>F. subglutinans</i> CBS 215.76 ²	^r <i>F. oxysporum</i>	75	57.2
^k <i>P. anserina</i> UBOCC-A-102026 ¹	^{ab} <i>A. alternata</i>	100	69.68	^r <i>F. temperatum</i> UBOCC-A-101148 ²	^r <i>F. verticillioideis</i>	62.5	43.15
^k <i>P. clade</i> UBOCC-A-101141 ¹	^{ab} <i>A. alternata</i>	65.22	50.60	^r <i>F. thapsinum</i> CBS 539.79 ²	^r <i>F. oxysporum</i>	100	82.56
^l <i>P. concavum</i> UBOCC-A-101181 ¹	^u <i>V. dahliae</i>	57.14	40.13	^r <i>F. proliferatum</i> UBOCC-A-109149 ²	^r <i>F. oxysporum</i>	45.83	40.28
^m <i>S. monosporum</i> UBOCC-A-101373 ¹	^{ac} <i>C. sydowii</i>	70.59	44.26	^s <i>B. aureofulvella</i> UBOCC-A-101174 ²	^s <i>B. ochroleuca</i>	82.61	64.16
^m <i>S. racemosum</i> UBOCC-A-101374 ¹	^{ad} <i>L. corymbifera</i>	52.17	35.06	^s <i>B. solani</i> UBOCC-A-102025 ²	^s <i>B. ochroleuca</i>	100	77.57
ⁿ <i>T. elegans</i> UBOCC-A-105020 ¹	^v <i>M. hiemalis</i>	95.83	69.7	^w <i>C. erectum</i> UBOCC-A-101010 ²	^w <i>C. globosum</i>	100	62.07
^o <i>M. zonata</i> UBOCC-A-101348 ¹	^y <i>M. circinnelloides</i>	47.62	33.68	^y <i>S. brevicaulis</i> UBOCC-A-101267 ²	^y <i>S. fusca</i>	100	82.37
^o <i>M. hyalina</i> UBOCC-A-101349 ¹	^{ad} <i>L. corymbifera</i>	54.17	37.44	^{ab} <i>A. chartarum</i> UBOCC-A-101045 ²	^{ab} <i>A. alternata</i>	100	75.57
^p <i>A. calidoustus</i> UBOCC-A-101086 ²	^p <i>A. versicolor</i>	79.17	63.1	^{ae} <i>C. brunhei</i> CBS 134.31 ²	^{ae} <i>C. ramotenellum</i>	100	82.4
^p <i>A. pseudoflectus</i> UBOCC-A-101085 ²	^p <i>A. sydowii</i>	60	42.89	^{ae} <i>C. herbarum</i> CBS 673.69 ²	^{ae} <i>C. ramotenellum</i>	100	82.36
^p <i>A. candidus</i> CBS 114985 ²	^p <i>A. elegans</i>	45.83	32.55	^{af} <i>G. silvicola</i> UBOCC-A-108083 ²	^{af} <i>G. candidum</i>	100	88.63
^p <i>A. clavati</i> UBOCC-A-101055 ²	^p <i>A. fumigatus</i>	62.5	38.43	^v <i>M. mucedo</i> UBOCC-A-101353 ²	^v <i>M. racemosus</i>	73.33	52.3
^q <i>N. fennelliae</i> CBS 584.90 ²	^p <i>A. fumigatus</i>	61.9	48.39	^v <i>M. fragilis</i> UBOCC-A-101356 ²	^v <i>M. circinnelloides</i>	81.25	61.1
^q <i>N. pseudofischeri</i> UBOCC-A-101204 ²	^p <i>A. fumigatus</i>	56.52	39.64	^{ag} <i>U. autotrophica</i> UBOCC-A-101347 ²	^{ag} <i>U. isabellina</i>	100	82.09
^q <i>N. fischeri</i> CBS 544.65 ²	^p <i>A. fumigatus</i>	87.5	69.34	^{ah} <i>C. binariae</i> UBOCC-A-101343 ²	^{ah} <i>C. elegans</i>	80	56.68
^q <i>N. hiratsukae</i> CBS 102802 ²	^p <i>A. versicolor</i>	75	49.81	^{ah} <i>C. blakesleeana</i> UBOCC-A-101341 ²	^{ah} <i>C. elegans</i>	63.64	42.1
^q <i>N. glabra</i> UBOCC-A-101203 ²	^p <i>A. fumigatus</i>	71.43	55	^{ai} <i>A. repens</i> UBOCC-A-101332 ²	^{ai} <i>A. coerulea</i>	64.29	23.7

¹not represented at the genus level in the database, ²not represented at the species level in the database

^a*Penicillium*, ^b*Talaromyces*, ^c*Eupenicillium*, ^d*Trichoderma*, ^e*Hypocrea*, ^f*Myrothecium*, ^g*Humicola*, ^h*Colletotrichum*, ⁱ*Kernia*, ^j*Cryphonectria*, ^k*Peyronellaea*, ^l*Pilidium*, ^m*Syncephalastrum*, ⁿ*Thamnidium*, ^o*Mortierella*, ^p*Aspergillus*, ^q*Neosartorya*, ^r*Paecilomyces*, ^s*Bionectria*, ^t*Fusarium*, ^u*Verticillium*, ^v*Mucor*, ^w*Chaetomium*, ^x*Microdochium*, ^y*Scopulariopsis*, ^z*Emericella*, ^{aa}*Eurotium*, ^{ab}*Alternaria*, ^{ac}*Circinnella*, ^{ad}*Lichtheimia*, ^{ae}*Cladosporium*, ^{af}*Geotrichum*, ^{ag}*Umbelopsis*, ^{ah}*Cunninghamella*, ^{ai}*Absidia*

prediction was validated while if the score was less than 70, the species prediction was not validated. The results showed that for the 105 strains represented in the database, 81 assignments were validated (77%) while the assignments of 24 strains were not validated (23%). Among the 81 strains, the identification of *Penicillium nalgiovense* UBOCC-A-101431 as *Penicillium commune* was validated although the predicted result was not correct, and among the 24 strains for which the species identification were not validated, the predicted results of *Eurotium amstelodami* CBS 817.96, *Eurotium rubrum* CBS 530.65 and *Penicillium citrinum* CBS 309.48 were not correct. The percentage of correct results corresponding to the correlated results between the prediction result and the calculated score for the 105 strains represented in the database was 79% (83 out of 105 strains). Concerning the 72 strains not represented in the database at the genus or species level, there were 12 strains for which species predictions were validated and 58 strains for which species predictions were not validated based on their respective scores. The percentage of correct results for these strains was 80.55%. For the strains not represented at the genus level in the database, 26 out of 27 predicted assignments were not validated while for the strains not represented at the species level, 34 out of 45 predicted assignments were not validated. The percentages of correct results were 96.29% and 75.56% for the strains not represented at the genus and species level, respectively. The strains that were not represented at the species level and for which the score was higher than 70, were correctly identified at the genus level. The average percentage of correct results for both sets of strains was 80.79%. These results are summarized in Figure 3.

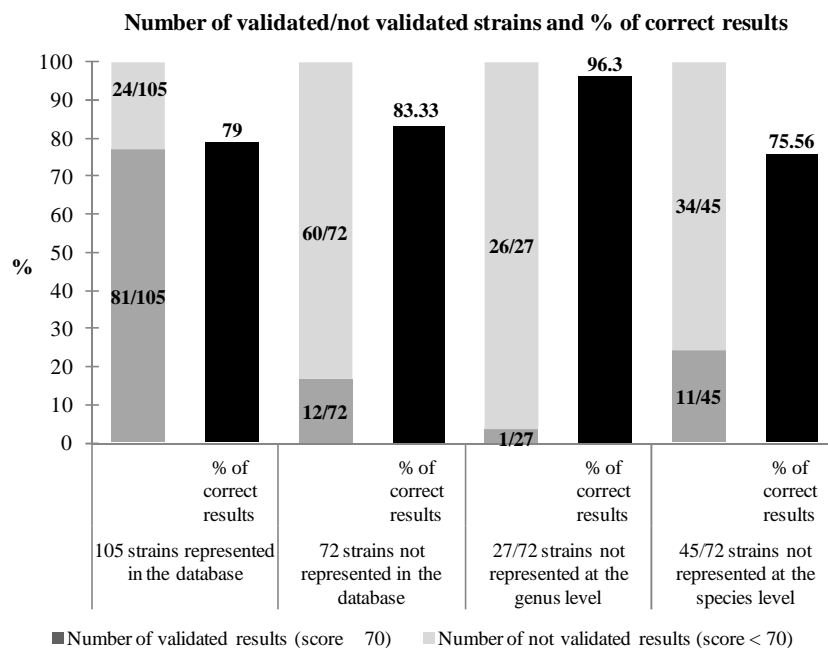


Figure 3: Summary of the percentage of correct results as a function of the threshold.

3.4 Implementation of a standardization function

The spectral library built using one instrument was then tested using two independent spectral data sets built from the same strains. To do that, 14 strains were analyzed using two different FTIR spectrometers located on two different sites and acquired by two different operators. The prediction result showed that the average of well predicted spectra per strain was 90.31% for the first spectral set recorded with instrument 1 (same instrument used for library implementation) and 65.29% for the second spectral set run on instrument 2 (instrument not used for library implementation). For each strain, the percentages of well predicted spectra for the two spectral sets are given in Table 6. All the strains of the first set were correctly predicted while 11 out of 14 strains were well predicted for the second spectral set. The bad predictions involved three strains of *Penicillium chrysogenum* that were misidentified as *Penicillium verrucosum* and *Penicillium commune*. The fact that the spectra of the second set of strains were recorded on a different instrument and confronted to the spectral library without any corrections, could be the reason for these wrong predictions. The second derivative spectra of these 14 strains were therefore used to calculate a standardization function (see equation 2) that was applied to a second spectral data set. Overall, the use of this function allowed to increase the average of well predicted spectra of data set 2 by 10% (from 65.10% to 75.35%). Interestingly, the three strains of *Penicillium chrysogenum* which were misidentified without the standardization function were correctly assigned after application of this function. The validation of the standardization function was then performed using the spectra of the 7 strains only recorded with the instrument 2. The results, illustrated in Table 7, showed that the application of the standardization function allowed increasing the average of well predicted spectra from 72.15% to 89.13%. The strain *Aspergillus versicolor* UBOCC-A-112085, predicted as *Paecilomyces variotii* before use of the standardization function, was correctly assigned as *Aspergillus versicolor* after use of this function. The results of the database validation using the spectra represented in the database and the spectra used to calculate the standardization function are summarized in Figure 4.

Table 6: Predicted species and the percentage of well predicted spectra per strain for the 14 strains used to calculate the standardization function and analysed on two different FTIR instruments.

Strains	Instrument 1		Instrument 2			
	Predicted species	% of well predicted spectra per strain	Predicted species	% of well predicted spectra per strain	Predicted species	% of well predicted spectra per strain
^a <i>A. flavus</i> UBOCC-A-101061	^a <i>A. flavus</i>	100	^a <i>A. flavus</i>	77.27	^a <i>A. flavus</i>	81.82
^a <i>A. niger</i> UBOCC-A-112080	^a <i>A. niger</i>	100	^a <i>A. niger</i>	86.67	^a <i>A. niger</i>	86.67
^a <i>A. niger</i> UBOCC-A-112082	^a <i>A. niger</i>	100	^a <i>A. niger</i>	68.75	^a <i>A. niger</i>	50
^b <i>P. roquefortii</i> UBOCC-A-112026	^b <i>P. roquefortii</i>	100	^b <i>P. roquefortii</i>	92.11	^b <i>P. roquefortii</i>	94.77
^b <i>P. paneum</i> UBOCC-A-111183	^b <i>P. paneum</i>	75	^b <i>P. paneum</i>	67.57	^b <i>P. paneum</i>	45.95
^b <i>P. chrysogenum</i> UBOCC-A-112065	^b <i>P. chrysogenum</i>	75	^b <i>P. verrucosum</i>	18.52	^b <i>P. chrysogenum</i>	70.37
^b <i>P. chrysogenum</i> UBOCC-A-112108	^b <i>P. chrysogenum</i>	91.67	^b <i>P. verrucosum</i>	0	^b <i>P. chrysogenum</i>	65.38
^b <i>P. chrysogenum</i> UBOCC-A-112077	^b <i>P. chrysogenum</i>	100	^b <i>P. commune</i>	16	^b <i>P. chrysogenum</i>	60
^b <i>P. brevicompactum</i> UBOCC-A-112048	^b <i>P. brevicompactum</i>	91.67	^b <i>P. brevicompactum</i>	59.26	^b <i>P. brevicompactum</i>	55.56
^b <i>P. corylophilum</i> UBOCC-A-112070	^b <i>P. corylophilum</i>	66.67	^b <i>P. corylophilum</i>	81.82	^b <i>P. corylophilum</i>	88.64
^b <i>P. corylophilum</i> UBOCC-A-112081	^b <i>P. corylophilum</i>	95.83	^b <i>P. corylophilum</i>	88.64	^b <i>P. corylophilum</i>	100
^c <i>F. oxysporum</i> UBOCC-A-112042	^c <i>F. oxysporum</i>	95.83	^c <i>F. oxysporum</i>	86.67	^c <i>F. oxysporum</i>	68.33
^d <i>G. candidum</i> UBOCC-A-103039	^d <i>G. candidum</i>	100	^d <i>G. candidum</i>	100	^d <i>G. candidum</i>	100
^e <i>M. circinelloides</i> CBS 223.56	^e <i>M. circinelloides</i>	72.72	^e <i>M. circinelloides</i>	70.83	^e <i>M. circinelloides</i>	87.5

^a*Aspergillus*, ^b*Penicillium*, ^c*Fusarium*, ^d*Geotrichum*, ^e*Mucor*

Table 7: Predicted species and the percentage of well predicted spectra per strain for the 7 strains used to validate the standardisation function.

Strains	Without standardization function		With standardization function	
	Predicted species	% of well predicted spectra per strain	Predicted species	% of well predicted spectra per strain
^a <i>A. niger</i> UBOCC-A-112064	^a <i>A. niger</i>	77.78	^a <i>A. niger</i>	97.22
^a <i>A. niger</i> UBOCC-A-112068	^a <i>A. niger</i>	66.67	^a <i>A. niger</i>	90
^a <i>A. versicolor</i> UBOCC-A-112085	^c <i>P. variotii</i>	50	^a <i>A. versicolor</i>	85.71
^b <i>P. chrysogenum</i> UBOCC-A-112074	^b <i>P. chrysogenum</i>	87.5	^b <i>P. chrysogenum</i>	93.75
^b <i>P. brevicompactum</i> UBOCC-A-112078	^b <i>P. brevicompactum</i>	79.17	^b <i>P. brevicompactum</i>	66.67
^b <i>P. corylophilum</i> UBOCC-A-112049	^b <i>P. corylophilum</i>	71.05	^b <i>P. corylophilum</i>	94.74
^b <i>P. corylophilum</i> UBOCC-A-112069	^b <i>P. corylophilum</i>	72.92	^b <i>P. corylophilum</i>	95.83

^a*Aspergillus*, ^b*Penicillium*, ^c*Paecilomyces*

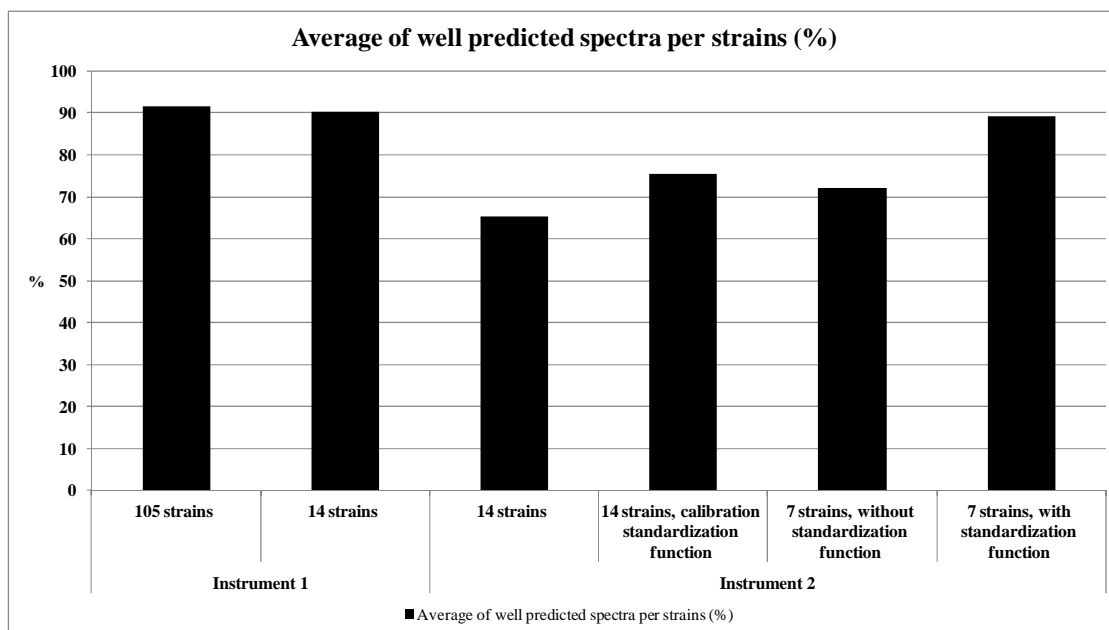


Figure 4: Summary of the average percentages of well predicted spectra per strain for all the validation sets.

4. Discussion

This study showed that with the use of a standardized protocol for fungal culture, sample preparation and spectral data acquisition parameters, FTIR spectroscopy can be applied as an alternative method for rapid and accurate identification of filamentous fungi. Our results clearly demonstrated that it was possible to identify filamentous fungi at the genus (99.17%) and species (92.3%) levels and also illustrated the capacity of FTIR spectroscopy to correctly identify fungal strains in a blind manner (101 out to 105 strains were correctly assigned) using the supervised PLS-DA chemometric method. These findings reinforce our preliminary study on a limited library of 106 strains (Lecellier et al., 2013). Indeed, the FTIR-based identification of the filamentous fungi used in this study was in very good agreement with those obtained using culture-based and molecular methods, thus highlighting the robustness of this approach. Nevertheless, molecular identification of filamentous fungi at the species level can be difficult and misclassification may occur. In fact, it is possible that a strain may be misclassified at the species level, especially regarding closely-related species.

The FTIR spectroscopy represents a real progress towards effective mold identification techniques as it is cost-effective and time-saving (2 days instead of 5) over conventional methods. In fact, the filamentous fungi identification based currently on the morphological features analysis is difficult because of the very high phenotypic biodiversity. Furthermore,

the conventional methods require a good knowledge of the fungal strains and are time-consuming. The use of molecular approaches based on DNA sequencing for filamentous fungi identification presents some limits like the cost and the application constraints (Alexander, 2002). The most recent promising method is the use of MALDI-TOF mass spectrometry for routine fungal identification (Del Chierico et al., 2012; Normand et al., 2013). This method is rapid, reliable, involves low labor and consumable cost but the equipment is quite expensive as compared to a high-throughput FTIR system (Santos, Paterson, Venancio, & Lima, 2010).

In this study, FTIR coupled with PLS-DA allowed to highlight the species-specific features from the global features of the strains and rendered discrimination and identification of filamentous fungi possible.

Different chemometric methods have been utilized for the FTIR-based analysis and classification of microorganisms. Among the mostly used, PLS-DA (Wenning & Scherer, 2013) presents several advantages such as the linearity relationship, put forward by Beer-Lambert law, linking chemical concentration and absorption, and implying that the linear approach appears appropriate. PLS-DA models also exhibit a good predictive power and these models can deal with large training datasets. They also have the ability to circumvent the “curse of dimensionality” and a high ease to set up the classifier (Trevisan, Angelov, Carmichael, Scott, & Martin, 2012).

For the 177 strains not included in the FTIR database and used for database validation and implementation of a score and validation threshold, 80.79% were correctly assigned. The score and the validation threshold of the prediction result permitted to prevent misidentification in situations where the species to predict was not represented in the reference database. Library-based identification requires a certain robustness and accuracy to classify an unknown sample. These can be improved by increasing the number of strains per species present in the database. The impact of such improvement was demonstrated in a recent study on the application of MALDI-TOF MS for the identification of filamentous fungi (Normand et al., 2013). As in the present study, a score was used to validate the prediction results. Nevertheless, it is estimated that out of the 1.5×10^6 species of fungi, only 5% have been formally identified (Hawksworth, 2001). Therefore, even if small and highly specialized libraries are well suited for specific applications where only a limited number of species are expected to be present (Wenning & Scherer, 2013), a library will never be exhaustive.

The application of a standardization function, on spectra obtained from another FTIR instrument not used to construct the spectral library, improved the percentage of well

predicted spectra by 17% (72.15% to 89.13%). It also allowed evaluating whether the transfer of the database from an instrument to another was feasible. Despite the same method was used for sample preparation and spectral acquisition, and the same strains were analyzed, there were differences in the percentage of well predicted spectra (90.31% and 65.29% with instruments 1 and 2 respectively). These differences can be explained by inter-instrument and atmospheric condition variabilities. Indeed, the characteristics of a given instrument can vary over time and the spectra obtained on different instruments of the same type can be different depending on the conditions in which the measurements are performed, irrespective of the auto-controls and auto-calibrations made (Chamrad et al., 2003). In IR spectroscopy, the stability of the energy, the reproducibility of the wavenumbers, the variations of pressure and temperature, and the noise, can influence the spectral quality. The method developed in the present study was based on the analysis on both apparatus of several isolates considered as reference strains but not included in the database. The two spectral sets obtained were then used to calculate the standardization function. Such a standardization function has already been applied in other analytical fields. For example, reference fungal strains were used as an internal calibration set for identification of fungi using MALDI-TOF MS (Del Chierico et al., 2012). A study on the analysis of tablet samples by near infrared spectroscopy showed that the use of a calibration standardization on spectral data could be used to modify or standardize spectral data to predict, instead of transforming, the calibration model. In fact, this method offers the advantage of maintaining a single database (Cogdill, Anderson, & Drennen, 2005). Nevertheless, in the present study, the transferability of the method was only studied using a low number of strains. It is planned to verify the transferability with a larger number of strains, and we also envisage to extend to more than two FTIR instruments.

5. Conclusions

FTIR spectroscopy coupled with chemometric method represents a promising technique to discriminate and identify filamentous fungi in routine laboratories. Moreover, this technique is rapid, easy to use, and less expensive compared to other methods such as molecular methods. It is envisaged to further extend this spectral database with other filamentous fungal species frequently encountered in the food, pharmaceutical and cosmetic industries, and in the health sector. It would then be possible to adapt the technology for a standardized routine use in an industrial environment permitting rapid investigations in case of crisis management so as to implement adapted corrective actions to the contaminants involved and to secure

industrial processes. This method may show an additional interest in the microbiology field, not only for quality control but also for research and development. Nevertheless, this technique should not be considered as an alternative to existing methods, but as a supplementary and complementary tool that allows a rapid identification of fungal isolates and can also be used as a dereplication tool for subsequent analyses.

Acknowledgements

Aurélie Lecellier is thankful to the Conseil Régional de Champagne Ardenne for funding of her PhD. The “Pôle de compétitivité” VALORIAL, La Région Bretagne, La Région Champagne-Ardenne and the technological platform IBiSA “Imagerie Cellulaire et Tissulaire” are gratefully acknowledged. Financial support under project "Mycotech" of the European Union, the Région Bretagne and the Conseil Général du Finistère are also gratefully acknowledged. The authors are also grateful to Marie-Anne Le Bras and Valérie Vasseur for their expertise and help in fungal identification and to Amélie Weill and Olivia Le Bourhis for their excellent technical assistance.

References

- Alexander, B. D. (2002). Diagnosis of fungal infection: new technologies for the mycology laboratory. *Transpl Infect Dis*, 4 Suppl 3, 32-37.
- Berthiller, F., Crews, C., Dall'Asta, C., Saeger, S. D., Haesaert, G., Karlovsky, P., et al. (2013). Masked mycotoxins: a review. *Mol Nutr Food Res*, 57(1), 165-186.
- Bylesjö M, Cloarec O, & Rantalainen M. (2009). Normalization and Closure. *Compr. Chemometr*, 2.07, 109-127.
- Campagnoli, A., Cheli, F., Polidori, C., Zaninelli, M., Zecca, O., Savoini, G., et al. (2011). Use of the electronic nose as a screening tool for the recognition of durum wheat naturally contaminated by deoxynivalenol: a preliminary approach. *Sensors (Basel)*, 11(5), 4899-4916.

Chamrad, D. C., Koerting, G., Gobom, J., Thiele, H., Klose, J., Meyer, H. E., et al. (2003). Interpretation of mass spectrometry data for high-throughput proteomics. *Anal Bioanal Chem*, 376(7), 1014-1022.

Cogdill, R. P., Anderson, C. A., & Drennen, J. K., 3rd. (2005). Process analytical technology case study, part III: calibration monitoring and transfer. *AAPS PharmSciTech*, 6(2), E284-297.

Del Chierico, F., Masotti, A., Onori, M., Fiscarelli, E., Mancinelli, L., Ricciotti, G., et al. (2012). MALDI-TOF MS proteomic phenotyping of filamentous and other fungi from clinical origin. *J Proteomics*, 75(11), 3314-3330.

Duygu, D., Baykal, T., Açikgöz, D., & Yildiz, K. (2009). Fourier Transform Infrared (FT-IR) Spectroscopy for Biological Studies. *Journal of science*, 22(3), 117-121.

Ergin, C., Ilkit, M., Gok, Y., Ozel, M. Z., Con, A. H., Kabay, N., et al. (2013). Fourier transform infrared spectral evaluation for the differentiation of clinically relevant *Trichophyton* species. *J Microbiol Methods*, 93(3), 218-223.

Fischer, G., Braun, S., Thissen, R., & Dott, W. (2006). FT-IR spectroscopy as a tool for rapid identification and intra-species characterization of airborne filamentous fungi. *J Microbiol Methods*, 64(1), 63-77.

Fredlund, E., Thim, A. M., Gidlund, A., Brostedt, S., Nyberg, M., & Olsen, M. (2009). Moulds and mycotoxins in rice from the Swedish retail market. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*, 26(4), 527-533.

Garon, D., El Kaddoumi, A., Carayon, A., & Amiel, C. (2010). FT-IR spectroscopy for rapid differentiation of *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus parasiticus* and characterization of aflatoxigenic isolates collected from agricultural environments. *Mycopathologia*, 170(2), 131-142.

Glass, N. L., & Donaldson, G. C. (1995). Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous *ascomycetes*. *Appl Environ Microbiol*, 61(4), 1323-1330.

Hawksworth, D. L. (2001). The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research*, 105, 1422-1432.

Helm, D., Labischinski, H., & Naumann, D. (1991). Elaboration of a procedure for identification of bacteria using Fourier-Transform IR spectral libraries: a stepwise correlation approach. *J Microbiol Methods*, 14, 127–142.

Hermet, A., Meheust, D., Mounier, J., Barbier, G., & Jany, J. L. (2012). Molecular systematics in the genus *Mucor* with special regards to species encountered in cheese. *Fungal Biol*, 116(6), 692-705.

Khot, P. D., Couturier, M. R., Wilson, A., Croft, A., & Fisher, M. A. (2012). Optimization of matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis for bacterial identification. *J Clin Microbiol*, 50(12), 3845-3852.

Lecellier, A., Mounier, J., Gaydou, V., Castrec, L., Barbier, G., Ablain, W., et al. (2013). Differentiation and identification of filamentous fungi by high-throughput FTIR spectroscopic analysis of mycelia. *Int J Food Microbiol*, In Press.

Liang, Y. Z., & Kvalheim, O. (1996). Robust methods for multivariate analysis. *Chemom Intell Lab Syst*, 32, 1–10.

Luna, A. S., da Silva, A. P., Ferre, J., & Boque, R. (2013). Classification of edible oils and modeling of their physico-chemical properties by chemometric methods using mid-IR spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc*, 100, 109-114.

Marinach-Patrice, C., Lethuillier, A., Marly, A., Brossas, J. Y., Gene, J., Symoens, F., et al. (2009). Use of mass spectrometry to identify clinical *Fusarium* isolates. *Clin Microbiol Infect*, 15(7), 634-642.

Marklein, G., Josten, M., Klanke, U., Muller, E., Horre, R., Maier, T., et al. (2009). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for fast and reliable identification of clinical yeast isolates. *J Clin Microbiol*, 47(9), 2912-2917.

Moss, M. O. (2008). Fungi, quality and safety issues in fresh fruits and vegetables. *J Appl Microbiol*, 104(5), 1239-1243.

Naumann, D. (2000). *Infrared spectroscopy in microbiology* (R.A. Meyers ed.). Chichester: John Wiley and Sons Ltd.

Navea, S., Tauler, R., Goormaghtigh, E., & de Juan, A. (2006). Chemometric tools for classification and elucidation of protein secondary structure from infrared and circular dichroism spectroscopic measurements. *Proteins*, 63(3), 527-541.

Nie, M., Zhang, W. Q., Xiao, M., Luo, J. L., Bao, K., Chen, J. K., et al. (2007). FT-IR spectroscopy and artificial neural network identification of *Fusarium* species. *Journal of Phytopathology*, 155(6), 364-367.

Nilsson, R. H., Ryberg, M., Abarenkov, K., Sjökvist, E., & Kristiansson, E. (2009). The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiol Lett*, 296(1), 97-101.

Normand, A. C., Cassagne, C., Ranque, S., L'Ollivier, C., Fourquet, P., Roesems, S., et al. (2013). Assessment of various parameters to improve MALDI-TOF MS reference spectra libraries constructed for the routine identification of filamentous fungi. *BMC Microbiol*, 13, 76.

O'Donnell, K., Kistler, H. C., Cigelnik, E., & Ploetz, R. C. (1998). Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc Natl Acad Sci U S A*, 95(5), 2044-2049.

Preisner, O., Lopes, J. A., & Menezes, J. C. (2008). Uncertainty assessment in FT-IR spectroscopy based bacteria classification models. *Chemometrics and Intelligent Laboratory Systems*, 94(1), 33-42.

Rodriguez, A., Rodriguez, M., Luque, M. I., Martin, A., & Cordoba, J. J. (2012). Real-time PCR assays for detection and quantification of aflatoxin-producing molds in foods. *Food Microbiol*, 31(1), 89-99.

Santos, C., Paterson, R. R. M., Venancio, A., & Lima, N. (2010). Filamentous fungal characterizations by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Applied Microbiology*, 108(2), 375-385.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem*, 36, 1627-1639.

Schmitt, I., Crespo, A., Divakar, P. K., Fankhauser, J. D., Herman-Sackett, E., Kalb, K., et al. (2009). New primers for promising single-copy genes in fungal phylogenetics and systematics. *Persoonia*, 23, 35-40.

Shapaval, V., Moretro, T., Suso, H. P., Asli, A. W., Schmitt, J., Lillehaug, D., et al. (2010). A high-throughput microcultivation protocol for FTIR spectroscopic characterization and identification of fungi. *J Biophotonics*, 3(8-9), 512-521.

Shapaval, V., Schmitt, J., Moretro, T., Suso, H. P., Skaar, I., Asli, A. W., et al. (2013). Characterization of food spoilage fungi by FTIR spectroscopy. *J Appl Microbiol*, 114(3), 788-796.

Stone, M. (1974). Cross-validation choice and assessment of statistical predictions. *J R Stat Soc Series B Stat Methodol*, 36, 111-147.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, 28(10), 2731-2739.

Tenenhaus, M. (1998). *L'algorithme de régression PLS1* (Tenenhaus M ed.). Paris: Technip.

Terra, M. F., Prado, G., Pereira, G. E., Ematne, H. J., & Batista, L. R. (2012). Detection of ochratoxin A in tropical wine and grape juice from Brazil. *J Sci Food Agric*, 93(4), 890-894.

Tralamazza, S. M., Bozza, A., Destro, J. G., Rodriguez, J. I., do Rocio Dalzoto, P., & Pimentel, I. C. (2013). Potential of Fourier transform infrared spectroscopy (FT-IR) to differentiate environmental *Aspergillus* fungi species *A. niger*, *A. ochraceus*, and *A. westerdijkiae* using two different methodologies. *Appl Spectrosc*, 67(3), 274-278.

Trevisan, J., Angelov, P. P., Carmichael, P. L., Scott, A. D., & Martin, F. L. (2012). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst*, 137(14), 3202-3215.

Wenning, M., & Scherer, S. (2013). Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method. *Appl Microbiol Biotechnol*, 97(16), 7111-7120.

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics (Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. ed.). San Diego: Academic Press.

Wu, W., Guo, Q., Jouan-Rimbaud, D., & Massart, D. L. (1999). Using contrasts as data pretreatment method in pattern recognition of multivariate data. *Chemometr Intell Lab*, 45, 39-53.

Zhang, L., Small, G. W., & Arnold, M. A. (2003). Multivariate calibration standardization across instruments for the determination of glucose by Fourier transform near-infrared spectrometry. *Anal Chem*, 75(21), 5905-5915.

Chapitre IV : Travaux supplémentaires

IV.1.Article 3

Assessing the potential of linear and non-linear supervised discrimination chemometrics methods on various filamentous fungi FTIR spectral database

En finalisation, soumission prévue à « *Analytical Chemistry* »

- Préambule à l'article 3

Contexte de l'étude

Les progrès en informatique ont permis de développer des méthodes statistiques sophistiquées permettant de traiter des ensembles de données complexes. Ces méthodes ont été appliquées dans de nombreux domaines scientifiques, comme la physique, la chimie et la biologie. La chimiométrie a émergé de ces méthodes comme approche puissante pour la compréhension et l'interprétation des données. Dans le cadre de données spectrales infrarouge, ces méthodes peuvent être utilisées pour l'extraction d'informations moléculaires pertinentes permettant une discrimination et une classification des spectres. Suivant la nature du problème à étudier, le choix de la méthode chimiométrique la plus appropriée est une étape essentielle.

Objectif

L'objectif de cette étude est de comparer et d'évaluer le potentiel discriminant de 8 méthodes chimiométriques linéaires et non-linéaires (impliquant 11 algorithmes de calcul) sur la même base de données spectrales. Cette base est composée de 5960 spectres infrarouge, réalisés (avec le protocole établi précédemment) à partir de 277 souches de champignons filamenteux appartenant à 14 genres et 36 espèces, et dont l'identification par séquençage est connue. À notre connaissance, c'est la première fois qu'une telle étude a été menée.

Matériels et méthodes

La base de données spectrales a été construite et modélisée afin de prédire l'identification d'une espèce de champignon filamenteux inconnue en se basant sur son spectre infrarouge. Parmi les 277 souches étudiées, 194 souches ont été utilisées pour l'optimisation et l'étape de calibration des modèles de discrimination et de classification des champignons filamenteux et 83 souches ont été utilisées pour l'étape de validation. Ainsi, 20 modèles de calibration ont été construits en cascade et de manière supervisée basée sur les différents rangs taxonomiques allant de la division jusqu'à l'espèce. Huit méthodes de classification ont été utilisées pour la construction des modèles dont 4 méthodes

chimiométriques linéaires : la LDA (Linear Discriminant Analysis), la FDA (Factorial Discriminant Analysis), la SIMCA (Soft Independent Modeling of Class Analogy) et la PLS-DA (Partial Least Square Discriminant Analysis), et 4 méthodes chimiométriques non linéaires : la QDA (Quadratic Discriminant Analysis), la KNN (k-Nearest Neighbor), la PNN (Probabilistic Neural Network) et la SVM (Support Vector Machine).

Résultats

En ce qui concerne les méthodes linéaires, les meilleurs résultats de prédiction de spectres des champignons filamenteux ont été obtenus en utilisant la méthode chimiométrique PLS-DA. Le pourcentage de bonne prédiction des spectres est de 98,9% et 93,2% au niveau genre et espèce respectivement. En ce qui concerne les méthodes non linéaires, la KNN a permis d'obtenir les meilleurs résultats de prédiction avec 90,4% et 78,2% de spectres bien prédits respectivement au niveau genre et espèce. Les résultats obtenus en couplant deux méthodes linéaires (SVM et PLS-DA) dans le modèle en cascade, a sensiblement amélioré les taux d'identification passant à 99,9% et 94,2% au niveau genre et espèce respectivement. Ces résultats montrent que la SVM serait plus appropriée pour les hauts rangs taxonomiques (sous-division au sous-genre) alors que la PLS-DA semble plus appropriée pour les rangs plus spécifiques et plus difficiles à différencier (de la section à l'espèce).

Conclusion

La spectroscopie IRTF couplée à la PLS-DA a permis d'implémenter une méthode d'identification des champignons filamenteux. Les résultats mettent en évidence la supériorité de la méthode chimiométrique PLS-DA, méthode d'analyse statistique multivariée linéaire, en comparaison aux autres méthodes utilisées dans cette étude. Le couplage de la PLS-DA avec la SVM améliore sensiblement les taux d'identification.

Assessing the discrimination potential of linear and non-linear supervised chemometrics methods on a filamentous fungi FTIR spectral database

V. Gaydou¹, A. Lecellier¹, D. Toubas^{1, 2}, J. Mounier³, L. Castrec³, G. Barbier³, W. Ablain⁴, M. Manfait¹, G.D. Sockalingum^{1*}

¹*MéDIAN-Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, FRE CNRS 3481-MEDyC, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 REIMS cedex, France*

²*Laboratoire de Parasitologie Mycologie, CHU de Reims, Hôpital Maison Blanche, 45 rue Cognacq-Jay, 51092 Reims cedex, France*

³*Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (EA3882), SFR148 ScInBioS, Université Européenne de Bretagne, Université de Brest, ESIAB, Technopôle de Brest Iroise, 29280 Plouzané, France*

⁴*AES CHEMUNEX/BIOMERIEUX, Rue Maryse Bastié, CS17219 Ker Lann, 35172 Bruz cedex, France*

*Corresponding author:

Ganesh D. Sockalingum

Université de Reims Champagne-Ardenne

MéDIAN, Biophotonique et Technologies pour la Santé

Unité MEDyC, CNRS FRE3481

UFR Pharmacie, SFR CAP-Santé FED4231

51 rue Cognacq-Jay, Reims, France.

Tel: +33 3 26 91 35 53

Fax: +33 3 26 91 35 50

ganesh.sockalingum@univ-reims.fr

Abstract

This study proposes a comparative investigation of different linear and non-linear chemometrics methods for filamentous fungi discrimination and identification applied to the same database of infrared spectra. The latter concerned 277 strains, (14 genus, 36 species), identified and validated by DNA sequencing, analyzed by high throughput FTIR spectroscopy. A cascade of 20 supervised models based on taxonomic ranks was defined to predict spectra until the species taxonomic rank. The cascade modeling and FTIR spectra, acquired from the mycelia, were used to test 11 algorithms of supervised classification methods. Among these, 5 algorithms were linear (LDA, FDA, SIMCA, PLS-DA, and SVM) and 6 non-linear (QDA, KNN, PNN, and SVM with RBF (Radial Basis Function), sigmoid and polynomial kernel function). To assess these algorithms, indicators of classification rates and McNemar's tests were defined and applied in same way to each of them. Concerning linear algorithms, the PLS-DA showed the best classification potential and for non-linear algorithms, the KNN method gave the best classification. It is noteworthy that the performances of SVM and PLS-DA algorithms are almost equivalent and highlights a tendency to possible complementarities between the two methods.

Keywords: Supervised discrimination, Cascade models, Chemometrics methods, Linear, Non-linear, FTIR spectral database, Filamentous fungi.

1. Introduction

The technological progress in computing has allowed developing sophisticated statistical methods enabling to process complex data sets. These methods are applied in various scientific domains, such as in physics, chemistry and biology (1-3). Chemometrics has emerged from these methods as a powerful approach for data mining, interpretation, and understanding, specifically for extracting relevant molecular information in different fields of spectroscopy.

These methods were developed to answer specific problems. So, for a given study, the methodology generally concerns only one or two chemometrics methods, usually in perfect adequacy with the problematic and the nature of the data. For example, it is well known that the SVM method is one of the most appropriate for character recognition (4, 5).

Also, very often an algorithm can be slightly modified and optimized as a function of the study requirements and depending on the knowledge, skills and know-how of research group. Indeed, IR spectral data has been used to set up models of identification of fungi involved in food spoilage (6). However, few studies have assessed the performances of linear and non-linear supervised classification algorithms on the same spectral library (7, 8).

The aim of this study was to experimentally compare the discriminating potential of 11 algorithms grouped into 8 supervised linear and non-linear chemometrics methods on the same dataset composed of 5960 Fourier transform infrared (FTIR) fungal spectra. The chemometrics question raised in this study was to select the most appropriate statistical method able to discriminate and identify an unknown strain of filamentous fungi from its FTIR spectrum and using a spectral library. To do so, a non-exhaustive spectral database was constituted using 277 fungal strains, belonging to 14 genera and 36 species. Among these, 194 strains (4159 spectra) were used for the optimization and calibration steps and 83 strains (1801 spectra) were used for the validation step.

The assessed methods were all supervised discrimination methods requiring a calibration step and grouped in two categories. The linear methods included 5 methods: LDA (Linear Discriminant Analysis), FDA (Factorial Discriminant Analysis), SIMCA (Soft Independent Modeling of Class Analogies), PLS-DA (Partial Least Square - Discriminant Analysis) and SVM (Support Vector Machine) used with linear Kernel function. The non-linear methods comprised QDA (Quantitative Descriptive Analysis), KNN (K Nearest Neighbors), PNN (Probabilistic Neural Network), and SVM (Support Vector Machine) with RBF (Radial Basis Function), Sigmoid and polynomial Kernel function (13-44). To assess these supervised

discrimination methods in terms of statistical significance, indicators of classification rates and McNemar's tests were defined and applied in same way to each of the studied algorithm.

2. Materials and Methods

Two hundred and seventy-seven strains of fungi (14 genera and 36 species yielding 6648 spectra) from the following culture collections (Université de Bretagne Occidentale and Centraalbureau voor Schimmelcultures) were sub-cultured on inclined Sabouraud (Becton Dickinson, Le pont de Claix, France) agar at 25°C for 4 to 7 days. The strains were identified by sequencing of specific DNA region like the rDNA internal transcribed spacer (ITS) region. The cultures were dissociated using a gentleMACS Octo Dissociator (Miltenyi Biotec, Paris, France) in order to obtain a homogenous suspension suitable for depositing and spectral acquisition. Dissociated mycelia suspensions were then transferred into an Eppendorf tube, centrifuged, and pellets were resuspended in 300 µl of 0.9 % physiological saline water.

Samples were then deposited on a 384-well silicon plate in 8 instrumental replicates in order to appraise the instrumental repeatability and the silicon plate was dried under mild vacuum during one hour. The spectral acquisition was performed using a FTIR high-throughput system composed of a spectrometer (Tensor 27, Bruker Optics, Etlingen, Germany) coupled to a high-throughput module (HTS-Xt, Bruker Optics). The spectrometer was driven by OPUS 6.5 software (Bruker Optics) and the acquisition parameters were 64 accumulations per well with a spectral resolution of 4 cm⁻¹, a spectral range of 4000-400 cm⁻¹, and a zero filling factor of 2. The background spectrum of the blank silicon plate was recorded in the same conditions before each sample measurement.

2.1 Spectral data pre-processing

The FTIR spectra were compiled in a two dimensions (900 x 6648) data matrix. The first dimension represents the absorption intensities and in the above experimental conditions, 900 absorbance values were recorded per spectrum. The second dimension represents all the 6648 analyzed spectra corresponding to the 277 fungal strains. Following this, the data matrix was subjected to a series of procedures as outlined below.

Spectral Quality Test

The quality test (QT) developed for this study and adapted for fungi was based on that reported for microbiological studies by Helm *et al.* (7). For a spectrum to pass the quality test, the following conditions must be satisfied:

- the absorbance in the region 1600-2100 cm^{-1} must be included between 0.17 and 1 arbitrary unit.
- the noise signal (N value) defined in the region 2000-2100 cm^{-1} , where there is no absorption peak, must be less than 0.00016.
- the residual water signal (W value) included between 1837 and 1847 cm^{-1} should be less than 0.0003.
- $S1/N > 50$, $S2/N > 10$, $S1/W > 20$, and $S2/W > 4$, where S1 corresponds to the highest absorbance included between 1600 and 1700 cm^{-1} and S2 to the highest absorbance included between 960 and 1260 cm^{-1} .

Any spectrum not satisfying the defined quality test conditions was automatically removed from the data matrix. The QT allows selecting, in an automatic way, the spectra presenting defects due to the sample preparation protocol or the experimental conditions. Approximately 10 % (688) of spectra were excluded by the quality test. So, 5960 spectra were conserved in the data matrix.

Mathematical preprocessing

Preprocessing is commonly used for signal improvement and is composed of several steps. It also allows to improve the accuracy of the models built preliminarily. For FTIR spectra of moulds, the preprocessing procedures used include baseline correction, second derivative, and normalization. These 3 mathematical transformations allowed building the best preliminary models (figure 1). The selection of variables, 800 to 1800 cm^{-1} and 2800 to 3200 cm^{-1} was realized using literature data on similar IR spectroscopy studies of *fungi* (6). Each of these steps was computed in the following order: QT, baseline correction, derivation, vector normalization, and variable selection. The quality test and mathematical preprocessing were performed using the OPUS 5.5 software (Bruker Optics).

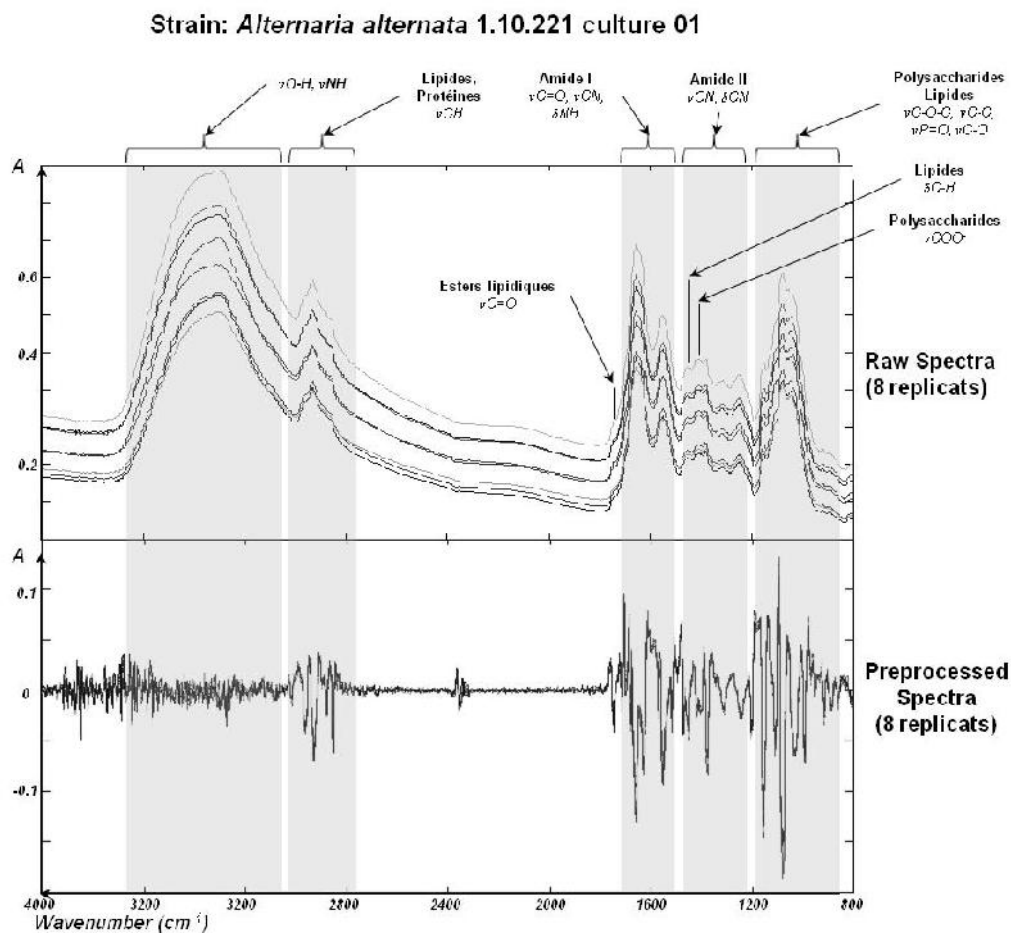


Figure 1: Raw and preprocessed FTIR spectra of *Alternaria alternata* culture with a tentative band assignment of major macromolecules.

Reference data and cascade modeling

The aim of the modeling is to predict the membership of species from the IR spectrum of an unknown filamentous fungus. The number of fungal species present in the data matrix is equal to 36 (5960 spectra). However, for IR spectroscopic data, the establishment of a single model of discrimination, parameterized by more than around thirty clusters is not feasible at the moment. Such one-model procedure is difficult to implement since the zones of variance and covariance overlap and become inconsistent with the number of clusters. For this reason, a modeling called “in cascade” has been developed (3) to circumvent this problem in this study (figure 2). The particularity of the cascade modeling is that it is parameterized from a reference arborescence, and for the study presented here, it is the taxonomic classification of fungi that is used in this respect. At every taxonomic rank, samples were distributed in subphylum, class, order, family, genus, subgenus, section, serial, and species. In so doing,

several “subgroups” were established at every rank and for each model the number of clusters was around 3 and so on, until the last rank called “species” rank is reached. The taxonomic tree is thus used to structure the data matrix in a subgroup and cluster cascade. We call "taxonomic nodes" the subgroups highlighted by the taxonomic tree. For every taxonomic node, a discrimination model was built. So, this technique allowed constructing the discrimination model in cascades including not less than 20 models with a maximum of 7 models required to reach the species taxonomic rank as regards to *Camemberti* serial.

The main advantage of the cascade modeling is that it allows obtaining a strong method of discrimination although the final number of clusters is high. On the other hand, this method requires the elaboration of numerous models which are interlocked, requiring therefore, a meticulous and delicate optimization. Furthermore, this method is completely parameterized and thus totally dependent on the cascade reference to which it is associated. Yet, the fungal taxonomy is in constant evolution and consequently training variation on taxonomic nodes can influence the outcome in a significant way.

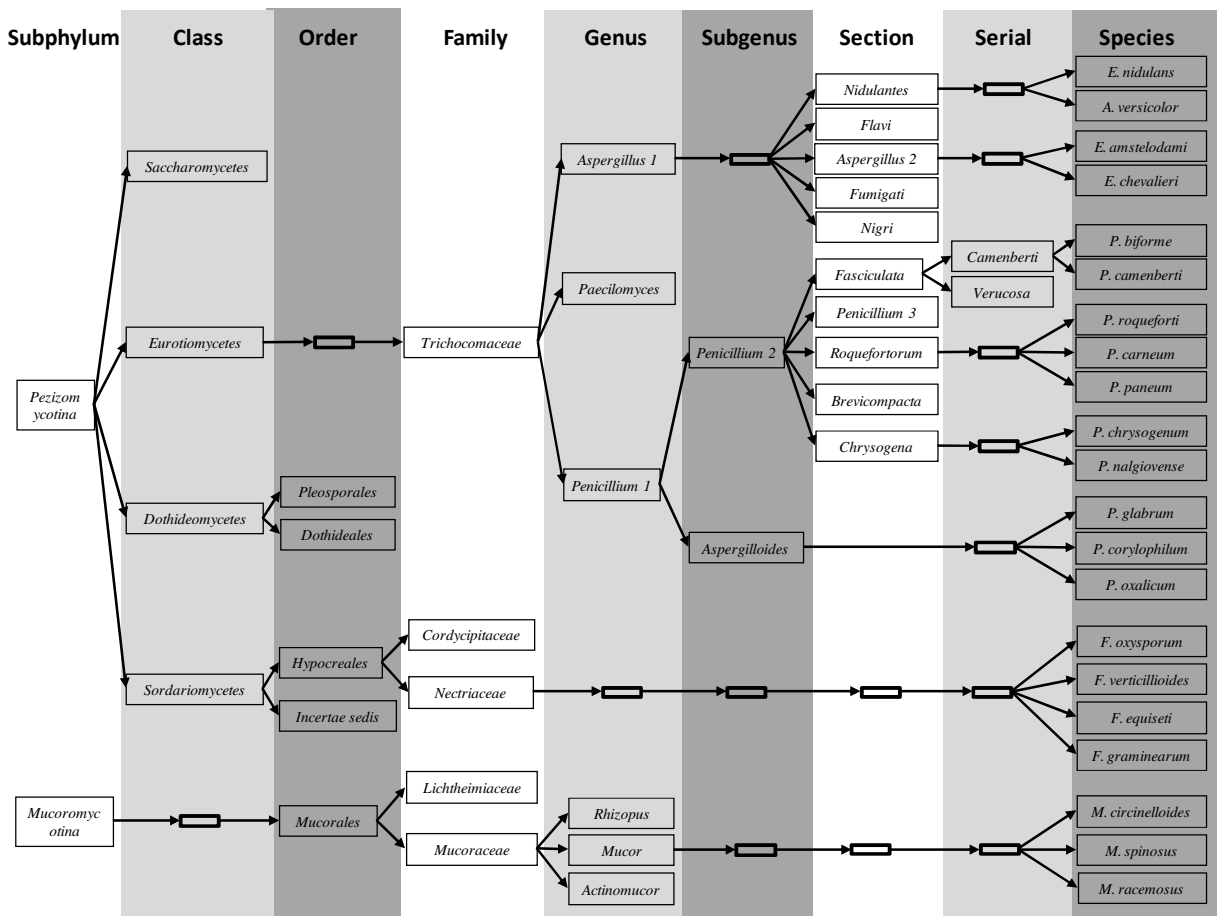


Figure 2: Organigram of the modeling cascade based on the current mould taxonomy.

Building of the calibration and validation sets

In a conventional way and more particularly within the framework of a theoretical study on the various methods of regression, the models are presented in two steps by means of two sample sets: the calibration set and the validation set. The calibration set allows building the model that is a parameterized mathematical algorithm associating the “explained variables” to the “explanatory variables”. The validation set allows estimating by an external way the models built. It is important that the explained variables of the calibration set are distributed in a homogeneous way between both extremes of the variables of all samples. Thus, the constructed model will be more robust towards the explained variables of validation set.

The data matrix was split into 2 sets; about two-thirds of samples (4159 spectra) were attributed to the calibration set and the rest (1801 spectra) to the validation set. Because the distribution of samples was made in a random way, the homogeneity of the calibration set variables was inspected and corrected when necessary. Depending on the study, a reallocation of some spectra could be required (most generally to ensure that the relative variance of the validation set is inferior to that of the calibration set) (8).

For this study, a random selection will most likely end on the impossibility to build all the models. The discrimination tree would then have missing branches, and the concerned samples of validation would be impossible to predict (at least up to the foliage of the discrimination tree). For a species represented in the data matrix by 3 different strains, only one of these three strains will be randomly chosen for the validation set and the two others for the calibration set.

If one species is represented by only 2 strains, no strain will be selected for the validation set, because the construction of a calibration model based on the variance of one strain cannot during the validation step, highlight the variance connected with species and that connected with strains. The models may be specific but not at all robust. The method of semi random (or half random) selection also presents the advantage to validate in a rather homogeneous way all the constructed models. Naturally, when the number of strains which represents a species is higher than five, 2 strains will be randomly selected among six of the same species (and so on, for every multiple of 3).

On the other hand, the bank of moulds was developed to maximize the number of presented species. Furthermore, to represent each of these species, at least three moulds per species were selected and analyzed (within the limit of the bank of moulds available to us).

Cross validation for parameter optimisation

Fundamentally, cross validation was developed for chemometrics experiments with a low sample population (9). Because of this low population it is impossible to split the data matrix into calibration and validation sets while keeping a representative sample set. Thus, the cross validation allows to estimate the accuracy and robustness only with one sample set. For the present study, the calibration set was checked with cross validation and several chemometrics parameters were optimized for all the studied chemometrics methods (10). The major interest of cross validation is that all of the individual samples of the calibration set were used at the same time for the calibration and for the validation. Several kinds of cross validation can be used. In the total cross validation, all samples are removed one by one and in the partial cross validation, all samples are removed group by group. Every sample or group of samples is alternately excluded and a regression model is elaborated with the remaining individuals. The established model is then tested by the sample or the group of samples which was left out. In this way, it is possible to know at the end of the calibration step, the superior average potential of the obtained model.

In this study, a large number of spectra is available. However, although the number of samples is quite high, the proportion of the number of species and that of the number of strains is close to two (that is only 2 strains per species for the calibration set). That is, each of the species present in the data matrix is represented on average by only 2 different strains. Therefore, the use of cross validation is justified. The cascade structure of all models is complex and the data matrix is constituted of biological and technical replicates. These two features must be taken into account in the implementation of the cross validation. So, three algorithms of cross validation were developed and tested: total cross validation, partial cross validation by strain, and partial cross validation by culture.

- Total cross validation

Total cross validation consists of testing, according to the chemometrics parameters, one by one every spectrum of the data matrix (protocol requiring the highest computing resources). This method has the merit to claim the maximal precision of the calibration models. However, it does not allow appreciating the intra-species covariance but rather the spectral covariance. In other words, the spectra of the same strain and the same culture are going to be classified very easily in the good category probably due to the specific variances in the culture or in the strain (and not in the species). The results of the total leave-one-out cross validation are

particularly good (very close to 100 % for accuracy) and this is why it was decided to set up a partial cross validation

- Partial cross validation by strain

The partial cross validation was scripted to remove from every validation all spectra belonging to the same strain. So, during every validation, all spectra of the same strain are tested. This method allows emphasizing the inter-species and inter-strain covariances. However, this method cannot report on the inter-culture covariance. Furthermore, if there are only two strains to represent a species, the model to associate with this species cannot test the covariance intra species (indeed because for a given species, there will be only one strain in calibration and another one in validation, thus, it was not possible any more to distinguish the covariance intra strain and covariance intra species). The results of the partial cross validation by moulds were not satisfying since it gave only 60 % accuracy.

- Partial cross validation by culture

The partial cross validation by culture was then worked out such that in every iteration all spectra associated to the culture of a strain were removed, then tested in validation phase. By applying this algorithm, it was able to test all cultures of the calibration set. The partial cross validation by culture allows estimating (partially) the intra-species, intra-strains, and intra-cultures covariances. Further, concerning the species represented by only 2 strains, this cross validation algorithm was more stable and allowed the observation of intra-species covariance. The obtained results are encouraging and hold promises for robust and accurate models since good prediction about 95 % for accuracy are expected. The various chemometrics methods were optimized by means of this partial cross validation.

Chemometrics parameters optimisation, Percentage of Good Prediction (PGP), and McNemar (McN) test

Each of the studied methods is inclined to variable chemometrics parameters which can have a significant influence on the developed prediction models and it is particularly important to pay a lot of attention to these chemometrics parameters. These latter were presented in table 1, with the exploration ranges applied on each parameter.

Table 1: Optimized parameters used for the different chemometrics methods.

Chemometrics methods	Used parameters	Chemometrics methods	Used parameters
<i>Linear methods</i>		<i>Non-linear methods</i>	
LDA	Kdim (positive integer included in 1 to 35): size of eigenvalues matrix	QDA	Kdim (positive integer included in 1 to 35): size of eigenvalues matrix
FDA	maxscore (integer included in 1 to 35): size of PCA-score matrix allowed to the model (PCA step) Kdim (positive integer included in 1 to 35): size of eigenvalues matrix	KNN	NumNeighbors (positive integer included in 1 to 30) : specifying the number of nearest neighbors in calibration data to find for classifying each point when predicting Metric choice: function use to specify the distance metric between neighbors (among 11 distances metric process)
SIMCA	maxscores (positive integer included in 1 to 35): size of PCA-score matrix allowed to each clusters (PCA step)	PNN	FN (positive integer included in 1 to 35): the number of computed iterations 2 (positive real included in 0 to +): "smoothing parameter" of the probability function estimator
PLS-DA	IN (positive integer included in 1 to 35) is the Iteration Number : the number of computed regression vector	SVM	Kernel function choice (among 3 K-functions: RBF , Sigmoid and polynomial) (positive real included in 0 to 1): "level of detail" or hyperplan resolution
SVM	Linear Kernel function choice (positive real included in 0 to 1): "level of detail" or hyperplan resolution		(positive real included in 0 to +): selected value of in Kerne function (RBF, sigmoid and polynomial choice) coef0 (positive real included in 0 to +): selected value of coef0 in Kernel function (RBF and sigmoid choice) d (positive integer included in 1 to 5): selected degree in kernel function (polynomial choice)

Every method has its own optimized chemometrics parameters, obtained and controlled by means of these three following steps: cross validation to optimize parameters, calibration to build discrimination models with optimized parameters, and validation to evaluate the power of the optimized methods.

The statistical indices called here the Percentage of Good Prediction (PGP) were calculated at the end of the cross validation, the calibration, and the validation steps. These indices were calculated by ratioing the number of good predicted spectra over the total number of spectra to predict. They allowed estimating the accuracy then the robustness of the discrimination model.

The optimized models (computing while the calibration step) are then validated while the validation step by means of validation data set.

The validation step allows estimating the real power of prediction of the given model by the use of an external sample set. To compare the various chemometrics methods, the samples of calibration and validation sets were kept identical for each tested methods.

McNemar's test (11, 12) is a statistical procedure able to estimate if the prediction power of two methods is significantly different. This test is based on a χ^2 with one degree of freedom

because the sample's number of each model is always higher than twenty. The χ^2 critical value with a 5% level of significance (: type I error), written $\chi^2_{(1,0.95)}$ is equal to 3.8414. McNemar's values (McN) were computed by means of equation 1. Two algorithms A and B are trained and validated with the same sets:

$$\text{McN} = \frac{(|n_A - n_B| - 1)^2}{n_A + n_B} \quad (\text{equation 1})$$

n_A : number of misclassified samples for algorithm A at the validation step,

n_B : number of misclassified samples for algorithm B at the validation step.

If McN value is less than $\chi^2_{(1,0.95)}$, the null hypothesis is valid with more than 95% of probability and the two algorithms are not significantly different. If the McN value is greater than $\chi^2_{(1,0.95)}$, the null hypothesis has 95% of probability to be false and the two algorithms are significantly different.

2.2 Linear and non-linear chemometrics methods

The methodological rules of these two categories are entirely different and the data are not visualized in the same way. For linear methods, the variance of the explanatory variables is considered as linear and a proportionality relationship between them and the explained variables is assumed. Non-linear methods take into account two types of variances, the global variance of the explanatory and explained variables, and then try to correlate these by means of a non-linear function such as the polynomial Kernel function for SVM algorithm. Also, for these two categories of algorithms, chemometrics models were not built around the same statistical rules. For supervised discrimination studies, the variety of chemometrics methods available is quite diverse. The linear methods are generally the most used with spectroscopic data. Indeed, the linearity relationship, put forward by Beer-Lambert, linking concentration and absorption implies that the linear approach appears better (13). However, the evolution of non-linear methods has allowed the elaboration of effective approaches such as SVM (Support Vector Machine) or Neural Network, which have been successfully applied in numerous experimental cases, including complex biological spectral data (14).

In order to optimize data mining and improve the understanding of biological phenomena from spectral results, it becomes essential to evaluate both linear and non-linear methods. Many of these linear and non-linear algorithms were declined in various specific algorithms, e.g., for the PLS algorithm, it was “declined” in robust or double PLS, quadratic PLS, splines function PLS or GIFI-PLS and many algorithms were combined such as the neural networks PLS or the least square SVM (15). For this study, only the “classical” (not “declined”) algorithms were used in order to assess the fundamental computing methodology of each of the following described algorithms.

2.2.1 Linear chemometrics methods

Linear Discriminant Analysis (LDA)

LDA is a linear method of supervised discrimination that can improve the spreading of the sample distribution (16). The aims of this method are to maximize the ratio of the inter- to intra-class distances and to find a linear transformation allowing to achieve the maximum class discrimination. The classical LDA tries to find an optimal discriminating subspace (spanned by the column vectors of a projection matrix) to maximize the inter-class separability and the intra-class compactness of the data samples in a low-dimensional vector space (17). The ideal discriminating subspace can be obtained by performing the eigenvalue decomposition on the inter- and intra-class scatter matrices. However, for the classical LDA the scatter matrices must be non-singular, which is well-known as the under sampling problem. To get round this problem many solutions exist. One of them is to precede LDA by a Principal Component Analysis (PCA) in order to extract the discriminant information. Nevertheless, PCA-LDA may lead to a loss of discriminant information during the PCA step (18). So, for this study only the classical LDA was tested.

Factorial Discriminant Analysis (FDA)

FDA aims at finding the subspace of the original variable space that best separates clusters by maximizing the inter-class variance with regard to the total variance (19, 20). This descriptive analysis builds a discriminant model to determine which cluster a new sample belongs to. This is simply done by projecting this sample onto the eigenvectors space and by selecting the nearest cluster. Several distances can be used for this decision. The two most commonly used distances are the simple Euclidean distance from the mass centers of the clusters and the Mahalanobis distance, which takes into account the shape of the clusters. In this study,

because of the small number of samples in each cluster and the resulting difficulty to assess shape of the clusters, the Euclidean distance was preferred.

Soft Independent Modeling of Class Analogy (SIMCA)

Wold and Sjöström were the first to describe the SIMCA chemometrics method (21). It is a supervised classification method which considers every “cluster of samples” or “groups” separately. This method is very useful for classifying high-dimensional observations because it incorporates PCA for dimension reduction. So for every cluster, decomposition into principal components (PC) is carried out providing a matrix of scores and loadings for each. The most practical interest of this analysis is that each cluster can be reduced to a set of PCs (22) and the optimal PCs are determined during the calibration step by estimating the variance explained in function of PCs calculated. After PCA, the discrimination models are built using Euclidean distance between clusters and PCA subspaces, taking into account the information and properties of clusters. Nowadays, many deviations of the SIMCA algorithm have been developed to improve this classification method (23). The common modifications concern the methods of distance calculation. For this study only the Euclidean distance was used due to the same reason as explained above.

Partial Least Square Discriminant Analysis (PLS-DA)

PLS-DA is a supervised classification method based on the PLS regression algorithm. The PLS regression is a model that links a property variable, such as concentration, to a set of explanatory variables, numerical or categorical (non quantitative) (24, 25). The algorithm is based on the ability to mathematically maximize the variance-covariance between the explanatory variable matrix and the property variable matrix. To maximize this covariance, the PLS algorithm builds by iterating many regression vectors (orthogonal to each other) that can explain the sample properties (in the vectorial space) with the minimum of approximation. This approximation represents the distance or error between reference property matrix and prediction of regression vectors. PLS-DA applies the PLS algorithm to establish discrimination rules by means of a binary matrix. Thus, for each cluster of the sample set, a binary code is defined and the bit number of each code corresponds to the maximum number of clusters (26).

To date, the PLS-DA method is used in various applications without much fundamental algorithm modification. However, this discrimination method can take advantage of any PLS

improvement (27). For this study, the classical PLS-DA was applied after building the binary property variable matrix.

2.2.2 Non-linear chemometrics methods

Quadratic Discriminant Analysis (QDA)

QDA is non-linear algorithm because it is based on a quadratic function but it is not very much different from LDA except that it is assumed that the covariance matrix can be different for each cluster, where it is estimated separately as a Gaussian distribution. The Gaussian parameters for each cluster are computed from training points with maximum likelihood estimation (28). This discriminant function is a quadratic function with second order terms and the classification rule is to find the cluster which maximizes the quadratic discriminant function (29). Because it allows for more flexibility for the covariance matrix, QDA tends to fit the data better than LDA, but however with a more significant number of parameters to estimate. For this study, this method was applied on the PCA scores of the data matrix.

k-Nearest Neighbor (KNN)

KNN techniques were developed to answer problematics about density estimation and pattern classification (30, 31). These methods are commonly employed for analyzing data sets which cannot be assumed to follow a normal distribution and have been concretely introduced by J. H. Friedman in 1975 (32). This algorithm can be used in a supervised and non supervised way with continuous or categorical variables to predict. In our study, we used the supervised way with categorical variable to predict.

Processing of this algorithm consists of basically ordering the training samples in a d-dimensional unit hypercube by means of a metrics distance measure. Then, for each tested sample, the training matrix is examined in the order of their projected distance from the tested sample on the sorted coordinate. Various metrics could be used for distance calculation: *cityblock distance*, *Chebychev distance* (maximum coordinate difference), *correlation* (one minus the sample linear correlation between observations), *cosine* (one minus the cosine of the included angle between observations), *Euclidean distance*, *Hamming distance* (percentage of coordinates that differs), *Jaccard* (the percentage of nonzero coordinates that differs), *Mahalanobis distance*, *Minkowski distance*, *standardized Euclidean distance*, and *Spearman* (one minus the sample Spearman's rank correlation between observations). During the validation step, the distances between each unknown sample and the training samples are

computed. The prediction of the unknown sample is determined by the most representative cluster of the K nearest neighbors (33). To optimize the training model, the K integer and the metrics of distance can be adjusted.

Probabilistic Neural Network (PNN)

Neural networks were successfully used to solve complicated pattern recognition and classification problems in different domains. The probabilistic neural networks (PNN) method presents a few advantages over the conventional neural network (34, 35). It provides a robust classification with noisy data. PNN operates on the concept of the Parzen windows classifier using Bayesian statistics (36). The Bayes decision rule (37) was proposed a few decades ago, but it stayed just a theoretical method until powerful computers became available. PNN combines different concepts: neural computing, Bayes classification rule, and non parametric estimation of the probability density function. The network contains an input layer which has as many elements as there are separable parameters needed to describe the objects to be classified. It has a pattern layer which organizes the training set such that an individual processing element represents each input vector. And finally, the network contains an output layer, called the summation layer, which has as many processing elements as there are classes to be recognized (38). In this study, the PNN method was employed on the eigenvalues of the data matrix, after PCA preprocessing, and the Mahalanobis method was used for distance computing.

Support Vector Machine (SVM)

SVM is a supervised method originally proposed by Vapnik *et al.* in 1963 (39). Fifty years later, many publications reporting on SVM and its extensions as a multiclass classification method can be found in literature (40). SVM is an attractive method because of its property to condense information in the training step and to provide a sparse representation by using a very small number of data points (41). These properties allow SVM to be used in a wide field of applications. Generally, classification methods try to minimize the error of prediction while SVM aims at maximizing the margin between the separating hyperplane and the data. The SVM algorithm classifies data by finding the best hyperplane that separates all data points of one class from the others classes. The best hyperplane for an SVM corresponds to the one with the largest margin between the two classes. In fact, the model is optimized in the space of Lagrangian multipliers by using the Kernel Function. It is a convex and deterministic process, and is guaranteed to converge to a single global minimum (42).

In this study, The nu-SVM algorithm was employed and this algorithm could be used with many Kernel Functions such as the Linear function, the Radial Basis Function (RBF), the Polynomial and Sigmoid functions (43). These four Kernel Functions were used with several chemometrics parameters depending on the Kernel Function chosen (table 1).

When the SVM was used with a Linear Kernel Function, this algorithm was considered as linear method, and for this study, the results of SVM with a Linear Kernel Function were associated with the results obtained by other linear methods.

All the chemometrics analyses were performed with Matlab R2013a (32-bit) (Mathwork, USA) and was used to classify the samples using their explanatory variables. The used algorithms for LDA, QDA and KNN were developed by MathWorks, USA. The used algorithms for SIMCA were developed by Cleiton A. Nunes, Brazil (available on Mathwork/matlabcentral). The used algorithms for SVM called lib-SVM were developed by Chih-Chung Chang and Chih-Jen Lin, China (43). The used algorithms for FDA, PLS-DA and PNN were developed by Dominique Bertrand and Christophe Cordella, INRA, France (44).

3. Results

For all linear and non-linear methods described previously, the results obtained take into account the three following steps: optimization of the chemometric parameters, model computing in cascade, and validation.

Optimization of chemometric parameters

The various methods of discrimination compared in this work require parameter optimization. These parameters are different from each other and are directly associated to the chemometrics method used (table 1). They have naturally a strong influence on the final results. It is thus essential that these parameters should be optimized in the most rigorous way. For this purpose, the method of partial cross validation by culture was employed. The parameters are optimized by means of the calibration sample set only and for each method, the influence of the variability of the various parameters (or combination of parameters such as for SVM methods) was tested by cross validation.

The details of the models in cascade are presented in table 2 and each model was built taking into account the taxonomic tree as described in figure 2. In total, 20 models were built to complete the cascade. Table 2 shows the model names associated with their cluster names.

Table 2: Details of the 20 discrimination models tested.

Taxonomic rank	Model name	Corresponding clusters names	Number of spectra (number of strains)	
			Optimization & calibration	External validation
Subphylum	<i>Micromycetes</i>	<i>Pezizomycotina</i> <i>Mucoromycotina</i>	4159 (194)	1801 (83)
Class	<i>Pezizomycotina</i>	<i>Eurotiomycetes</i> <i>Sordariomycetes</i> <i>Saccharomycetes</i> <i>Dothideomycetes</i>	3302 (154)	1430 (66)
Order	<i>Dothideomycetes</i>	<i>Dothideales</i> <i>Pleosporales</i>	144 (7)	39 (2)
	<i>Sordariomycetes</i>	<i>Xylariales</i> <i>Hypocreales</i>	1087 (51)	433 (19)
Family	<i>Mucorales</i>	<i>Mucoraceae</i> <i>Lichtheimiaceae</i>	857 (42)	371 (17)
	<i>Hypocreales</i>	<i>Cordycipitaceae</i> <i>Nectriaceae</i>	1046 (45)	412 (18)
Genus	<i>Mucoraceae</i>	<i>Rhizopus</i> <i>Mucor</i> <i>Actinomucor</i>	822 (40)	355 (16)
	<i>Trichocomaceae</i>	<i>Paecilomyces</i> <i>Penicillium 1</i> <i>Aspergillus 1</i>	1948 (91)	886 (42)
Subgenus	<i>Penicillium 1</i>	<i>Penicillium 2</i> <i>Aspergilloides</i>	1043 (49)	441 (22)
Section	<i>Aspergillus 1</i>	<i>Flavi</i> <i>Fumigati</i> <i>Nidulantes</i> <i>Nigri</i> <i>Aspergillus 2</i>	851 (39)	427 (19)
	<i>Penicillium 2</i>	<i>Brevicompacta</i> <i>Fasciculata</i> <i>Chrysogena</i> <i>Roquefortorum</i> <i>Penicillium 3</i>	808 (37)	337 (17)
Serial	<i>Fasciculata</i>	<i>Camemberti</i> <i>Verrucosa</i>	190 (9)	45 (2)
Species	<i>Nidulantes</i>	<i>E. nidulans</i> <i>A. versicolor</i>	147 (7)	57 (3)
	<i>Aspergillus 2</i>	<i>E. chevalieri</i> <i>E. amstelodami</i>	110 (5)	21 (1)
	<i>Fusarium</i>	<i>F. equiseti</i> <i>F. graminearum</i> <i>F. verticillioides</i> <i>F. oxysporum</i>	1001 (43)	392 (17)
	<i>Mucor</i>	<i>M. circinelloides</i> <i>M. racemosus</i> <i>M. spinosus</i>	647 (31)	320 (13)
	<i>Camemberti</i>	<i>P. biforme</i> <i>P. camemberti</i>	108 (5)	24 (1)
	<i>Chrysogena</i>	<i>P. nalgiovense</i> <i>P. chrysogenum</i>	141 (7)	84 (4)
	<i>Roquefortorum</i>	<i>P. roqueforti</i> <i>P. carneum</i> <i>P. paneum</i>	250 (12)	119 (6)
	<i>Aspergilloides</i>	<i>P. corylophilum</i> <i>P. glabrum</i> <i>P. oxalicum</i>	235 (11)	104 (5)

These 20 models were built independently during the optimization and calibration steps but during the validation step, these 20 models were interlocked into each other, taxonomic rank by taxonomic rank. Table 2 shows also the details of the spectral population for each model at the optimization/calibration and validation steps. The *micromycetes* model (first row) had the maximum population taking into account all the spectra of the calibration and validation sets. Then, the model population decreased when the studied rank approached the species rank; with the limit of one strain per species model for validation steps (see rows of *Aspergillus 2* and *Camenberti* models). These observations underlined that the robustness, the sensitivity, and the accuracy could equally decrease when the studied rank approached the species rank, particularly during the validation step (because of the interlocked models built in cascade).

In the scope of this study, it is not possible to describe all the chemometric parameters for all methods and all models (tables 1 and 2). We have therefore chosen to illustrate for two methods: PLS-DA and SVM.

Concerning the PLS-DA methods, the parameter to optimize is the iteration number (IN) of each model in the cascade (table 2). In this study, the IN parameter varies from 1 to 35. The limit of 35 was chosen principally for computing reasons. These parameters were optimized by partial cross validation and each IN was tested with each culture. The average of Percentage of Good Prediction (PGP) as a function of the IN was computed and plotted highlighting a maximum of PGP (figure 3). The IN corresponding to this maximum was taken as the best parameter. In fact, the IN optimization is important to minimize the under- and over-fitting (3). All of the IN parameters were defined for each model of the cascade, and each model was built with its appropriate IN as presented in Table 3 (see the IN column). It was possible to observe that the IN parameters highlighted the complexity of the model, because more the model took into account a high number of clusters, the higher is the IN parameter.

Concerning the SVM, there are numerous parameters to be optimized and by means of the cross validation and the calibration sample set, a very large number of parameter combinations could be assessed. The same protocol as for PLS-DA optimization was repeated for each Kernel function and each combination of parameters. The association between the Kernel function and parameter combination giving the best PGP was selected for the calibration step. The values of “ ” parameters are positive and real numbers from 0 to 1, and they operate on "level of detail" also called “hyperplan resolution”. The tested values of for each model were 0.00000001, 0.000001, 0.00001, 0.0001, 0.0005, 0.001, 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.4 and 0.8. The optimization results of SVM, with

linear Kernel Function, are presented in table 3 (see the column). Further, the partial cross validation by culture could be used to control the homogeneity between the 3 biological replicates and, in case of inhomogeneous results, samples could be re-prepared and re-analyzed by FTIR spectroscopy.

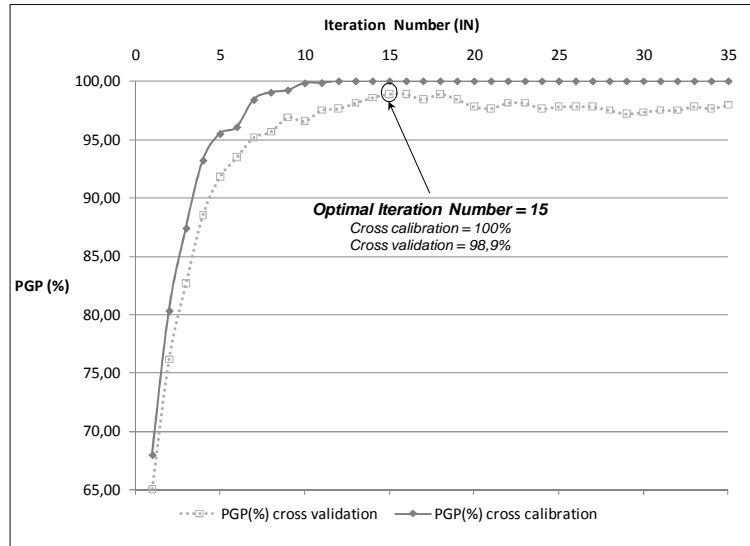


Figure 3: IN optimization (PLS-DA methods) for the *Mucor* species model with 3 clusters (*M. racemosus*, *M. circinelloides*, *M. spinosus*).

Table 3: Comparison of calibration and validation PGP's between PLS-DA and SVM models.

Taxonomic rank	Model name	PLS-DA models			SVM models with linear kernel		
		IN	PGP Calibration	PGP Validation	γ	PGP Calibration	PGP Validation
Subphylum	<i>Micromycetes</i>	7	100	100	0,001	100	100
Class	<i>Pezizomycotina</i>	12	99,9	99,2	0,01	99,8	100
Order	<i>Dothideomycetes</i>	5	100	100	0,01	100	100
	<i>Sordariomycetes</i>	10	100	99,6	0,005	99,3	100
Family	<i>Mucorales</i>	16	100	100	0,005	98,4	100
	<i>Hypocreales</i>	5	100	99,5	0,001	100	100
Genus	<i>Mucoraceae</i>	20	100	100	0,005	99,1	100
	<i>Trichocomaceae</i>	25	99,9	98,8	0,05	95,7	99,9
Subgenus	<i>Penicillium 1</i>	13	100	99,6	0,2	98,7	95,7
Section	<i>Aspergillus 1</i>	20	100	97,7	0,05	99,2	99,5
	<i>Penicillium 2</i>	15	100	95,4	0,2	94,7	84,7
Serial	<i>Fasciculata</i>	5	100	100	0,2	100	100
Species	<i>Nidulantes</i>	6	100	100	0,01	95,8	100
	<i>Aspergillus 2</i>	5	100	85,7	0,8	100	0,0
	<i>Fusarium</i>	20	99,1	95,5	0,2	96,1	94,5
	<i>Mucor</i>	14	100	99,0	0,2	96,6	95,3
	<i>Camemberti</i>	10	100	83,3	0,1	93,3	37,5
	<i>Chrysogena</i>	5	100	86,5	0,1	95,0	75,0
	<i>Roquefortorum</i>	20	100	66,0	0,01	100	59,0
	<i>Aspergilloides</i>	5	100	100	0,01	100	100

Model computing

After parameters optimization, the models were built in cascade taking into account the current taxonomy and the optimized parameters. During the model computing step, the discrimination abilities of each model and each method was evaluated by means of the PGP of calibration. The calibration results concerning PLS-DA and SVM (with linear Kernel function) algorithms are presented in table 3, but for the other algorithms, in order to clearly present the results, it was chosen to present only the validation results (table 4). In table 3, the mean of PGP of calibration for PLS-DA algorithm was equal to 99.94%. This good result allowed expecting a PGP for validation near to 100%. Concerning linear Kernel function SMV algorithm, the mean of PGP of calibration was close to 98%. This result was good but less than for PLS-DA, and implied that the validation result could be good but less than for the PLS-DA algorithm.

Table 4: Percentage of good prediction for validation test (1801 spectra) as a function of the taxonomic rank.

Chemometrics methods	Subphylum	Class	Order	Family	Genus	Subgenus	Section	Serial	Species
<i>linear methods</i>									
LDA (Linear Discriminant Analysis)	100	99,9	99,2	99,2	96,4	96,1	94,4	94,4	89,6
FDA (Factorial discriminante Analysis)	100	99,9	99,6	99,5	95,0	93,7	91,1	91,1	85,8
SIMCA (Soft Independent Modeling of Class Analogy)	97,7	89,6	88,9	87,8	66,5	61,6	52,0	51,7	35,4
PLS-DA (Partial Least Square Discriminant Analysis)	100	99,2	99,2	99,2	98,9	98,9	97,7	97,7	93,2
SVM (Support Vector Machine) Linear Kernel function	100	100	100	100	99,9	99,0	97,2	97,2	90,1
<i>non-linear methods</i>									
QDA (Quadratic Discriminant Analysis)	100	96,9	96,8	94,6	90,0	88,3	83,1	83,1	71,5
PNN (probalistic Neural Network)	65,2	49,3	49,3	48,3	39,8	39,6	36,1	36,1	26,9
KNN (k-Nearest Neighbor)	100	99,3	99,2	98,7	90,4	90,2	85,0	85,0	78,2
SVM (Support Vector Machine) RBF Kernel function	100	99,9	99,9	99,6	91,9	91,8	75,5	75,4	42,2
SVM (Support Vector Machine) Sigmoid	100	99,9	99,9	99,8	82,0	80,8	74,8	73,7	50,7
SVM (Support Vector Machine) Polynomial	99,5	90,3	89,4	81,2	43,1	41,4	34,5	34,5	24,7

Validation step

The prediction capacity of all the classification models was evaluated by means of a blind or external sample set, i.e., a validation sample set. This validation step allows observing, in real conditions, the behavior of the various models tested in this investigation. For a proper comparison, the validation set used for each method was kept identical. The obtained results are shown in the table 4 and a global view is presented in figure 4a (linear methods) and 4b (non-linear methods). Table 4 presents the PGP of validation spectra at each taxonomic rank and for each tested algorithm. Figures 4a and 4b show the broken curves corresponding to the PGP of validation spectra of each tested algorithm versus the taxonomic rank. For figure 4a, concerning the linear algorithm, the broken curve of SIMCA is not shown in order to have the best scale for PGP.

For the linear methods, the best results were obtained with the PLS-DA method. This method allowed reaching a PGP of 98.9% for the genus taxonomic rank and 93.2% for the species taxonomic rank. The LDA and FDA methods respectively gave a PGP around 3% and 6% less than the PLS-DA method, with 96.4% and 95% for the genus taxonomic rank and 89.6% and 85.8% for the species taxonomic rank. The SVM algorithm with the linear Kernel function, gave very good results with a PGP of 99.8% for the genus taxonomic rank and 91.3% for the species taxonomic rank. The SIMCA method showed the worst results with PGP of 66.5% for the genus taxonomic rank and less than 50% for the species taxonomic rank.

Concerning the non-linear methods, the best result was obtained with the KNN algorithm and gave a PGP of 90.4% and 78.2% respectively for genus and species taxonomic rank. The PGP of this algorithm was close to 100% up to the family taxonomic rank and from the genus to the species rank, the PGP was around 15% less than the PLS-DA algorithm. The second best non-linear algorithm was the QDA algorithm. This algorithm gave PGP values close to the KNN algorithm, nearly 5% less, with a PGP of 71.5% for the species taxonomic rank. The SVM algorithms used with the 3 non-linear Kernel functions (RBF, sigmoid, and polynomial) showed PGP values near to 100% until the family taxonomic rank, but for the ranks following, these PGP decreased strongly, at the genus taxonomic rank (92%, 82% and 43%) and at species taxonomic rank (42%, 51% and 25%), respectively for RBF, sigmoid and polynomial Kernel function. Finally, the PNN algorithm gave the worst results, similar to SVM with the polynomial Kernel function, with PGP values around 50% from the class taxonomic rank, which then decreased substantially until the species taxonomic rank.

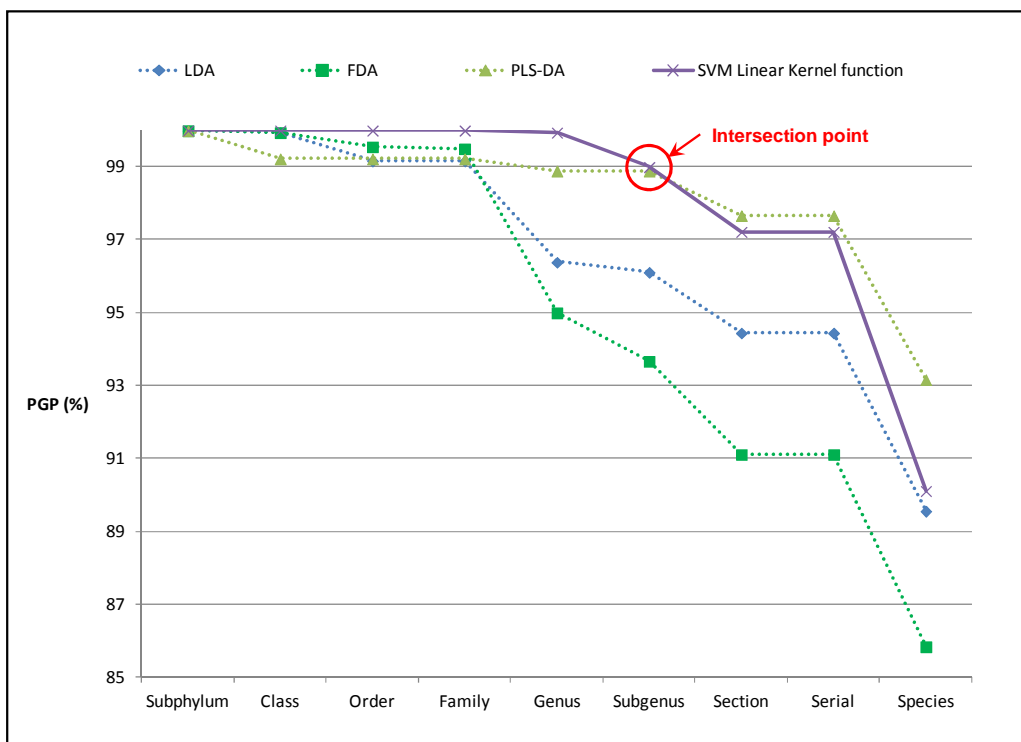


Figure 4a: Comparison of validation PGPs by taxonomic rank between linear chemometrics methods.

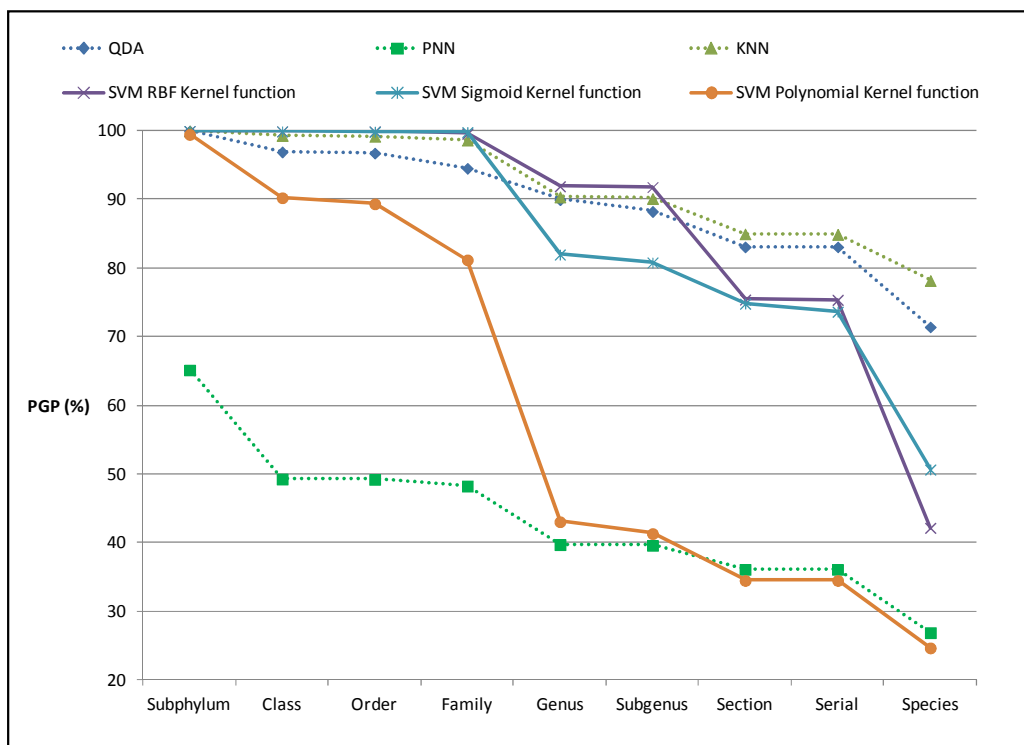


Figure 4b: Comparison of validation PGPs by taxonomic rank between non-linear chemometrics methods.

In order to evaluate if the prediction power of two methods was significantly different, the McNemar's test was applied to each investigated method. The results are displayed in table 5 in the form of a two-dimension correlation matrix. These tests were computed for the species taxonomic rank.

Concerning the linear methods, all the McNemar's values were greater than $\chi^2_{(1,0.95)}$ (equal to 3.8414), except for LDA and SVM with a linear Kernel function. These two algorithms therefore had 95% of probability to be equivalent to each other.

Concerning the non-linear methods, only the PNN coupled with SVM's polynomial Kernel function showed McNemar's value inferior than $\chi^2_{(1,0.95)}$. This result underlines the prediction power proximity between PNN and SVM's polynomial Kernel function. These two methods were not significantly different.

The McNemar's tests concerning the mix of linear and non-linear algorithms are also presented in Table 5 and all the McNemar's values were greater than $\chi^2_{(1,0.95)}$. This could mean that all investigated linear algorithms had more than 95% probability to be significantly different from each other non-linear algorithm at the species rank.

Table 5: Correlation matrix of McNemar's test, presenting McNemar's value for each pair of chemometrics methods at the species level.

Misclassified sample's number (species level)	Tested chemometrics algorithms	linear methods					non-linear methods					
		LDA	FDA	SIMCA	PLS-DA	SVM Linear	QDA	PNN	KNN	SVM RBF	SVM Sigmoid	SVM Polynomial
188	LDA	0	10	703	13	0,2	150	845	71	591	454	882
255	FDA	10	0	581	45	13	87	715	29	475	349	751
1164	SIMCA	703	581	0	840	723	251	9,2	382	6,8	37	14
123	PLS-DA	13	45	840	0	10	239	987	139	722	577	1026
178	SVM Linear	0,2	13	723	10	0	162	865	80	610	472	903
514	QDA	150	87	251	239	162	0	351	16	178	99	378
1316	PNN	845	715	9,2	987	865	351	0	499	32	83	0,6
392	KNN	71	29	382	139	80	16	499	0	293	191	531
1041	SVM RBF	591	475	6,8	722	610	178	32	293	0	12	41
888	SVM Sigmoid	454	349	37	577	472	99	83	191	12	0	97
1356	SVM Polynomial	882	751	14	1026	903	378	0,6	531	41	97	0

A complementarity between PLS-DA and SVM's linear Kernel function algorithms was suggested in figure 4a and table 4. At the subgenus rank, PLS-DA and SVM- linear Kernel function gave similar accuracy with respectively PGP of 98.9 and 99.0%. But, at the genus rank, the PGP of these two algorithms were 98.9 and 99.9% respectively for PLS-DA and SVM's linear Kernel function; and at the species rank, respectively 93.2 and 90.1%. The remarkable intersection point of these two broken curves was pointed by a circle in figure 4a. Due to this singularity between both algorithms, a "combined cascade" was built. SVM algorithm with linear Kernel function was use to elaborate the 8 models from the subphylum to the genus taxonomic rank, and PLS-DA algorithm was used from the sub-genus to species taxonomic rank (12 models). The validation results of this combined cascade are presented in figure 5.

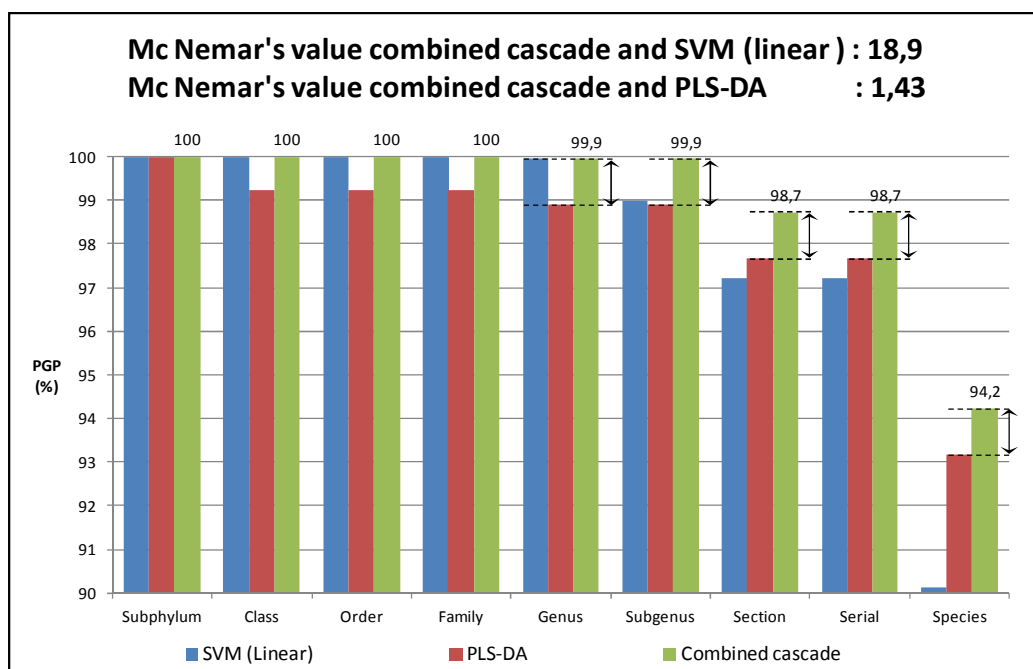


Figure 5: Comparison between the combined SVM (linear Kernel function) and PLS-DA cascade with the SVM and PLS-DA cascades taken independently. With ϵ representing the delta between SVM and PLS-DA PGPs at subgenus rank.

This particular cascade, called "combined cascade" showed the best performances compared to all the other "regular cascades", with 94.2% of PGP at species taxonomic rank. The table 6 shows details of validation step at subgenus taxonomic rank, for PLS-DA and SVM's kernel function algorithms. At this taxonomic rank, only one percent of the 1801 spectra selected for the validation set were misclassified. Twenty spectra were misclassified when the PLS-DA

algorithm was used. Strains corresponding to these misclassified spectra are shown in dark grey. For the SVM-linear algorithm with Kernel function, eighteen spectra were misclassified, and the corresponding strains are highlighted in light grey.

Table 6: Number of misclassified spectra by strains at the subgenus taxonomic rank for PLS-DA and SVM (linear Kernel function) chemometrics methods.

Strains ID	Number of spectra by strains	Taxonomic references	Number of misclassified spectra by strains at subgenus taxonomic rank	
			PLS-DA	SVM
1.01.073	24	<i>A. niger</i>	7	0
1.01.074	22	<i>A. niger</i>	5	0
1.03.044	18	<i>P. variotii</i>	6	0
1.01.143	24	<i>F. graminearum</i>	2	0
1.06.031	24	<i>A. flavus</i>	0	1
1.08.109	16	<i>P. nalgiovense</i>	0	10
1.10.065	20	<i>P. brevicompactum</i>	0	2
1.10.028	18	<i>P. expansum</i>	0	1
1.01.397	24	<i>P. chrysogenum</i>	0	4

4. Discussion

This study compares various linear and non linear chemometrics methods on a single FTIR spectral data set of mould species. The results presented highlighted the interest of the cascade modeling based on taxonomy because of the size and nature of the data set (figure 2). Indeed, extending a complex discrimination problem into several steps allowed to distribute the studied variance on several models and so to target, on every taxonomic node, the adequate variance. Further, the supervised cascade model amplifies the discrimination effect of each tested algorithm by means of the interlocked models. This observation is supported by table 4 and the McNemar's values matrix. In fact if the McNemar test were computed at the subphylum taxonomic rank, probably many algorithms would not be significantly different. For the linear and non-linear methods (figures 4a and 4b), all the presented curves were decreasing. This decrease could be correlated with the variance sought at each taxonomic

rank. In fact, during the validation step, while the prediction was at the level of the species taxonomic rank, the predicted spectra were more and more similar to calibration spectra of applied models. We noticed that the change from the genus to the subgenus and from the section to the taxonomic species rank induced complications. These may be associated firstly, to the morphological proximity at the subgenus rank and to the closeness of the biochemical structures at the species rank. These observations seemed to converge with difficulties of the morphological identification, in particular for *Aspergillus* 2, *Camemberti*, *Chrysogena* and *Roquefortorum* species models (table 3) (45- 47). This was comforting but it also showed the dependence due to the taxonomic references and consequently the limits of the supervised cascade computing for the use of spectral analysis.

Apart from the decreasing tendency, the profiles of the broken curves exhibited by linear and by non-linear algorithms were strongly different. Firstly, the mean of PGP at species taxonomic rank were 78.8% for linear algorithms and to 44.6% for non-linear algorithms. Secondly, the rank by rank decrement coefficients were more constant for linear than for non-linear algorithms because the linear broken curves (figure 4a) were nearly parallel and rectilinear while the non-linear broken curves (figure 4b) presented numerous intersection points and irregular diminutions. These observations highlight the suitability of the linear algorithms for, at least, the concerned problematic.

Concerning linear methods, the PLS-DA algorithm showed superior ability than the other linear methods. It is that LDA and FDA methods use mathematical algorithm based on variance maximization of the spectral data matrix, whereas the PLS-DA algorithm is based on the covariance maximization between the spectral data matrix and binary references matrix. In light of the results displayed in Table 4, it is clear that the multivariate PLS algorithm seemed more adapted than the other linear algorithms. On the opposite, the SIMCA method uses the internal variance of each cluster separately. The pertinent variance searched for each model becomes finer when the modelisation is near to the taxonomic species rank. When the SIMCA method estimates the internal variance of each cluster, the pertinent variance is probably masked by other internal variances caused by the physical and biological replicate effects.

The four Kernel functions used with the SVM algorithm (table 4) gave good and similar results up to the taxonomic family rank. After this rank, a decrease in performance could be observed, in particular for the RBF. The SVM-linear Kernel function was able to reach the second best PGP for the last taxonomic rank. The three other SVM algorithms (RBF, sigmoid and polynomial Kernel function) are non-linear and require optimization of a high number of parameters (table 1). The optimization step could, after a long calculating time, induce an

over-fitted model. This could explain the low SVM performance obtained with the polynomial, sigmoid, and RBF algorithms. In addition, the interlocked cascade models could probably exacerbate this effect. These results and considerations about SVM algorithms underline the increased adaptability (by the Kernel function choice), and also the fitting constraints of SVM algorithms.

For the non-linear methods (QDA, KNN and PNN), the superiority of KNN algorithm is demonstrated while PNN was found to be the least efficient. This could be explained by the singular skills of the PNN. In fact, the neural networks were developed for discrimination studies concerning data sets with strong and non-linear variabilities and seems, given this study's problematic, completely unsuitable.

The McNemar's values were computed to verify if the difference between tested methods is statistically significant at the species taxonomic rank. All of tested algorithms were significantly different except for the following pairs: LDA and SVM (linear Kernel function) and, PNN and SVM (polynomial Kernel function). This could probably be explained by the implementation of the cascade modeling because the interlocked models amplified the discrimination effects as mentioned earlier.

Some proximity between the two best linear methods, PLS-DA and SVM-linear Kernel function, are observed (table 3 and figure 5). Up to the genus taxonomic rank, the discriminating abilities of SVM was higher than PLS-DA. SVM was also more robust because the PGP values of the concerned models were all near 100 %, whereas those of the PLS-DA decreased regularly to reach 98.9% at the genus taxonomic rank. Beyond the subgenus taxonomic rank, the performance of the SVM models was below that of PLS-DA. In particular, for the models of *Penicillium* 1 and 2 presenting respectively validation PGP values of 95.7 and 84.7% and concerning about a quarter of the spectra of validation set (tables 2 and 3). The SVM algorithm therefore seems less appropriate at this level. The difficulties encountered by the SVM could be explained by the complexity of taxonomic attribution after the genus taxonomic rank. This complexity could be correlated to the high number of clusters per model and the low sample population associated that could induce a risk of over-fitting (particularly during the SVM optimization step). On other side, the identification based on morphological features using microscopic methods presented a relative feasibility until the genus taxonomic rank, but after this level, the identification requires a more extended and complex expertise inside a specific genus.

For the species models, table 3 underlines some problematic models for both SVM and PLS-DA algorithms. The concerned models were principally *Aspergillus* 2, *Camemberti*,

Chrysogena, and particularly *Roquefortorum*, presenting respectively validation PGP values of 85.7%, 83.3% and 61.0% for PLS-DA and 0%, 37.5 and 59.0% for SVM. These 4 models gave, during the calibration step, a PGP of 100% (or close to 100%) but their validation was low and illustrated a lack of accuracy. These inconveniencies could be explained, for *Aspergillus 2*, *Chrysogena* and *Camemberti* models, by the low population of strains and by the high number of needed taxonomic nodes (table 2 and figure 2). For the *Roquefortorum* model, the taxonomic references of strains concerned by this model are in evolution (45) and the difficulties to discriminate by sequencing and the genetic proximity of these concerned strains were real and correlated by the outcome of our chemometrics models. In other words, these observations underline the limits of spectral discrimination due to the biochemical proximity between the strains of each model of species taxonomic rank. The superior performance of the PLS-DA algorithm for the species taxonomic rank is thus clearly highlighted (table 3).

The combined cascade model with SVM-linear Kernel function from the subphylum to the genus taxonomic ranks and PLS-DA from subgenus to species taxonomic ranks, is quite successful since it was possible to observe that spectra that were wrongly predicted by PLS-DA were then correctly predicted by SVM at genus taxonomic rank, and were almost all correctly indentified until the species taxonomic rank. The gain of one percent (illustrated by in figure 5), due to the combined cascade, was maintained from the genus down to the species taxonomic ranks.

The McNemar's values computed between the combined cascade, SVM and PLS-DA cascades, showed that the former was significantly different from SVM with linear Kernel function but not from PLS-DA algorithms (with a confidence interval of 95%). Although not significantly different from the PLS-DA cascade, it however gave an improvement tendency and holds promises for a pertinent complementarity between SVM and PLS-DA.

On the one hand, the SVM methods used with linear Kernel function is the most pertinent method to discriminate fungi strains until the genus taxonomic rank, which means linear SVM could be better adapted than PLS-DA algorithm for voluminous sample sets. On the other side, the PLS-DA method is the most pertinent method to identify fungi strain at species taxonomic rank, which suggests that PLS-DA could be better adapted than the linear SVM algorithm for reduced and complex sample sets. Furthermore, the combination of both methods tends to indicate an improvement of the identification capacity. The results shown in table 6 illustrate the complementarities between SVM and PLS-DA algorithms. It highlights the interest to use both algorithms in tandem.

5. Conclusion

The choice of the supervised cascade to develop chemometrics discrimination methods was found to be appropriate for the fungi spectral data bank giving high prediction accuracy at each taxonomic rank from the subphylum to the species. For the linear algorithms, the PLS-DA method gave the best identification performance over SVM-linear Kernel function with PGP values of 98.9% and 93.2% for the genus and species taxonomic rank respectively. Concerning the non-linear algorithms, the best performance was reached with the KNN method giving a PGP values of 90.4% and 78.2% for genus and species taxonomic rank respectively. The McNemar's test showed that all these methods were significantly different except for LDA *versus* SVM (linear Kernel function) and PNN, *versus* SVM (polynomial Kernel function).

Combining the two well suited models, PLS-DA and SVM (linear Kernel function) indicate an improvement of the identification accuracy from the subphylum to the species taxonomic ranks.

Furthermore, the PLS-DA regression vectors could very interesting to study spectroscopic markers for each model in order to link the spectral and biochemical information of fungi. Thus, the identification accuracy could be improved by the development of an adjusted "combined cascade" modeling.

Acknowledgements

The "Pôle de compétitivité" Valorial, La Région Bretagne, La Région Champagne-Ardenne and the technological platform PICT-IBiSA "Imagerie Cellulaire et Tissulaire" are gratefully acknowledged. Financial support under the MOLDID project, project "Mycotech" of the European Union, the Région Bretagne and the Conseil Général du Finistère is also gratefully acknowledged. The authors are also grateful to Marie-Anne Le Bras and Valérie Vasseur for their expertise and help in fungal identification, to Amélie Weill and Olivia Le Bourhis for their excellent technical assistance and to Cyril Gobinet for reading the chemometrics section.

References

- 1 **B.R. Kowalski,**
Chemometrics : view and proposition,
J. Chem Inf. Comput. Sci., 15 (1975) pp. 201-203.

- 2 **A. Höskuldsson,**
Prediction methods in science and technology,
Basic Theory vol. 1, Thor Publishing, Copenhagen, Denmark (1996) pp. 245.

- 3 **D. Bertrand, E. Dufour,**
Identification et caractérisation des microorganismes, La spectroscopie infrarouge et ses
applications analytiques,
2nd ed. Lavoisier, Paris (2006) pp. 561-581.

- 4 **C. Cortes, V. Vapnik,**
Support-vector networks Machine Learn.,
20(3) (1995), pp. 273-297.

- 5 **B. Schölkopf, C. Burges, A. Smola,**
Introduction to support vector learning, B. Schölkopf, C. Burges, A. Smola (Eds.),
Advances in Kernel Methods—Support Vector Learning, MIT Press (1999), pp. 1-15.

- 6 **V. Shapaval, J. Schmitt, T. Moretro, H.P. Suso, I. Skaar, A.W. Asli, et al,**
Characterization of food spoilage fungi by FTIR spectroscopy. J Appl Microbiol,
114 (2013) 788-796.

- 7 **D. Helm, H. Labischinski, D. Naumann,**
Elaboration of a procedure for identification of bacteria using Fourier-Transform IR spectral
libraries: a stepwise correlation approach,
J Microbiol Methods, 14 (1991) pp. 127-142.

8 **L.J. Tashman,**

Out-of-sample tests of forecasting accuracy: an analysis and review,
International Journal of Forecasting, 16(4) (2000) pp. 437-450.

9 **S. Arlot, A. Celisse,**

A survey of cross-validation procedures for model selection,
Statistics Surveys, 4 (2010) pp. 40-79.

10 **M. Stone,**

Cross-validation choice and assessment of statistical predictions.
J R Stat Soc Series B Stat Methodol. 36 (1974) pp. 111-147.

11 **T.G. Dietterich,**

Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms
Department of Computer Science, Oregon State University, Corvallis, OR 97331, U.S.A.
October 1, 10(7) (1998) pp. 1895-1923. Posted Online March 13, 2006.

12 **Y. Roggo, L. Duponchel, C. Ruckebusch, J.P. Huvenne,**

Statistical tests for comparison of quantitative and qualitative models developed with near
infrared spectral data,
Volume 654, (1–3) 25 June 2003, pp. 253-262.

13 **J. Workman,**

Review of Chemometrics Applied to Spectroscopy: Quantitative and Qualitative Analysis,
The Handbook of Organic Compounds, NIR, IR, Raman, and UV-Vis Spectra Featuring
Polymers and Surfactants,
1(3) (2001) pp. 301-326.

14 **F. Chauchard, R. Cogdill, S. Roussel, J.M. Roger, V. Bellon-Maurel,**

Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a
robust and portable sensor for acidity prediction in grapes, Chemometrics and Intelligent
Laboratory systems, 71 (2004), pp. 141-150.

- 15 **M. Mörtzell, M. Gulliksson,**
An overview of some non-linear techniques in Chemometrics,
Rapportserie FSCN - ISSN 1650-5387 2001:6, Mid-Sweden University, (2001).
- 16 **D.H. Moore,**
Combining linear and quadratic discriminants,
Comput. Biomed. Res., 6 (1973) pp. 422-429.
- 17 **J.M. Geoffrey,**
Discriminant Analysis and Statistical Pattern Recognition,
Wiley, New York (1992).
- 18 **D.L. Swets, J. Weng,**
Using discriminant eigenfeatures for image retrieval,
IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (1996) 831-836.
- 19 **R. Fisher,**
The use of multiple measurements in taxonomic problems,
Ann. Eugenics, 7 (1936) pp. 179-188.
- 20 **J.M. Romeder,**
Méthodes et Programmes d'Analyse Discriminante,
Dunod, Paris, France (1973).
- 21 **S. Wold, M. Sjostrom,**
SIMCA: A method for analyzing chemical data in terms of similarity and analogy,
Chemometrics, theory and application, 52 (1977) pp. 243.
- 22 **M.A. Sharaf, D.L. Illman, B.R. Kowalski,**
Chemometrics,
Wiley, New York, (1986).

- 23 **K. Vanden Branden, M. Hubert,**
Robust Classification in High Dimensions, based on the SIMCA Method,
Chemometrics and Intelligent Laboratory Systems, 79 (1-2) 2005 pp. 10-21.
- 24 **S. Wold, H. Martens, H. Wold,**
The multivariate calibration problem in chemistry solved by the PLS method,
In: Ruhe, A., Kastrom, B. (Eds.), Proceedings of the Conference Matrix Pencils, March 1982,
Lecture Notes in Mathematics, Springer, Heidelberg, (1983) pp. 286-293.
- 25 **P.H. Garthwaite,**
An interpretation of partial least squares,
J. Amer. Statist. Assoc. 89 (425) (1994) 122-127.
- 26 **M. Tenenhaus,**
L'algorithme de régression PLS1,
In Tenenhaus M (ed), La régression PLS: théorie et pratique. Technip, Paris (1998) pp. 75-77.
- 27 **P. Bastien, V.E. Vinzi, M. Tenenhaus,**
PLS generalised linear regression,
Computational Statistics & Data Analysis 48 (2005) pp. 17-46.
- 28 **T. Hastie, R. Tibshirani, J. Friedman,**
The Elements of Statistical Learning,
Springer-Verlag, New York, 2001.
- 29 **S. Srivastava, M.R. Gupta, B.A. Frigyik,**
Bayesian Quadratic Discriminant Analysis,
Journal of Machine Learning Research 8 (2007) pp. 1277-1305.
- 30 **D.O. Loftsgaarden, C.P. Quesenberry,**
A nonparametric density function,
Ann. Math Statist., 36 pp. (1965) pp. 1049-1051.

- 31 **T.J. Wagner,**
Convergence of the nearest neighbor rule,
IEEE Trans. Inform. Theory, 17 pp. (1971) 566-571.
- 32 **J.H. Friedman, F. Baskett, L.J. Shustek,**
An algorithm for finding nearest neighbors,
IEEE Trans. Comput., 24 (1975), pp. 1000-1006.
- 33 **L. Labart, A. Morineau, N. Tabart,**
Technique de la description statistique, Méthodes et logiciels pour l'analyse des grands tableaux,
Ed. Dunod, Paris, France (1987).
- 34 **D. Specht,**
Probabilistic neural networks,
Neural networks, 3(1) (1990) pp. 110-118.
- 35 **P. Wasserman,**
Advanced methods in neural networks,
Van Nostrand Reinhold, New York, USA (1993).
- 36 **A. Gelman, J. Carlin, H. Stern, D. Rubin,**
Bayesian data analysis,
Boca Raton, FL: CRC Press (2003).
- 37 **A.M. Mood, F.A. Graybill,**
Introduction to the theory of statistics,
Macmillan, New York, (1962).
- 38 **Y. Shana, R. Zhaoa, G. Xua, H.M Liebichb, Y. Zhanga,**
Application of probabilistic neural network in the clinical diagnosis of cancers based on clinical chemistry data,
Analytica Chimica Acta, 471(1), 23 October 2002 pp 77-86.

- 39 **V. Vapnik, A. Lerner,**
Pattern recognition using generalized portrait method,
Automation and Remote Control 24 (1963) pp. 774-780.
- 40 **C.C. Chang, C.J. Lin,**
LIBSVM : a library for support vector machines,
ACM Transactions on Intelligent Systems and Technology, 2(27) (2011) pp. 1-27,
Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- 41 **F. Girosi,**
An equivalence between sparse approximation and support vector machines,
Neural Computation, 20 (1998) pp. 1455–1480.
- 42 **B. Boser, I. Guyon, V. Vapnik,**
A training algorithm for optimal, margin classifiers,
Fifth Annual Workshop on Computational Learning, Theory, New York: ACM Press, USA
(1992).
- 43 **J.A.K. Suykens, J. Vandewalle,**
Nonlinear Modeling: Advanced Black-Box Techniques,
Kluwer Academic Publishing, Boston, (1998) pp. 1-274.
- 44 http://www.chimie.ucl.ac.be/~saisir_conceptors.html.
- 45 **M. Boysen, P. Skouboe, J. Frisvad, L. Rossen,**
Reclassification of the *Penicillium roqueforti* group into three species on the basis of
molecular genetic and biochemical profiles,
Microbiology. 142(3) (1996) pp. 541-9.
- 46 **F. Giraud, T. Giraud, G. Aguilera, E. Fournier, R. Samson, C. Cruaud, and al.,**
Microsatellite loci to recognize species for the cheese starter and contaminating strains
associated with cheese manufacturing,
Int J Food Microbiol. 137(2-3) (2010) pp. 204-13.

47 **V. Hubka, M. Kolarik, A. Kubatova, S.W. Peterson,**
Taxonomic revision of *Eurotium* and transfer of species to *Aspergillus*,
Mycologia, 105(4) (2013) Jul-Aug pp. 912-37.

Chapitre V : Conclusion générale

Conclusion générale

La connaissance des contaminants fongiques est un enjeu majeur pour garantir la maîtrise sanitaire des environnements alimentaires, pharmaceutiques, cosmétiques et médicaux. Devant le constat de faiblesse analytique dans le domaine de l'identification des moisissures en milieu industriel, une approche biophysique reposant sur la spectroscopie IRTF a été évaluée au cours de ce travail. Cette approche biophotonique, basée sur l'interaction onde-matière, est capable d'identifier les souches étudiées par les variations spectrales liées aux modifications des différents constituants moléculaires (lipides, polysaccharides, acides nucléiques, protéines...).

Dans le cadre de ce travail, nous avons pu mettre en évidence le potentiel de la spectroscopie IRTF couplée à la méthode chimiométrique PLS-DA comme approche prometteuse et complémentaire aux méthodes d'identification conventionnelles et moléculaires des champignons filamenteux. Les avantages de cette méthode sont sa simplicité de mise en œuvre, sa rapidité et son faible coût.

- La première étape a porté sur la mise au point d'un protocole standardisé et simplifié de préparation des échantillons ne nécessitant, après culture, qu'une étape de broyage du mycélium suivi d'un lavage par centrifugation. La seconde étape a consisté en la mise au point d'une procédure d'analyse spectrale à haut débit, ne nécessitant qu'un simple dépôt des échantillons sur une plaque de silicium multipuits. Cette démarche a permis la discrimination et l'identification des souches de champignons filamenteux au niveau de l'espèce.

- L'identification des moisissures après seulement 48 h de culture permet de réduire considérablement le temps de l'analyse par rapport aux méthodes conventionnelles, ce qui rend cette technique particulièrement attractive dans un contexte industriel et/ou médical.

- La méthode mise au point représente un avantage économique certain par rapport aux autres méthodes car elle ne nécessite aucun réactif, les plaques de silicium sont réutilisables et la maintenance de l'appareil est quasi-nulle. De plus, le coût de l'appareil n'est pas excessif et s'élève à 50 keuros.

La mise en œuvre de la méthode d'identification a nécessité la construction d'une bibliothèque de spectres IRTF restreinte, à partir de 288 souches provenant du secteur agroalimentaire et appartenant à 28 genres et 68 espèces. Des modèles de classifieur ont été construits sur la base de méthodes chimiométriques linéaires et non linéaires. Le classifieur est composé de modèles en cascade pour prendre en compte la complexité des moisissures ainsi que leur taxonomie allant de la sous-division à l'espèce. Dans ces méthodes, 11 algorithmes de calcul ont été testés. Les résultats de ces tests montrent que la méthode PLS-DA est légèrement supérieure à la méthode SVM, en ce qui concerne la capacité d'identification au niveau de l'espèce. Nos résultats montrent également que le couplage entre deux méthodes linéaires comme la PLS-DA et la SVM peut améliorer le pouvoir d'identification. En effet, la SVM semble être adaptée pour traiter un volume de données conséquent (de la sous-division au sous-genre) tandis que la PLS-DA serait plus efficace pour traiter des données complexes (de la section à l'espèce) et déceler des variations minimales.

Une étude préliminaire sur la transférabilité de la base de données à un autre appareil IRTF, situé sur un autre site a pu être réalisée par l'implémentation d'une fonction de standardisation. Le pourcentage de spectres bien prédits s'améliore après l'application de cette fonction, passant de 72,15% à 89,13%. Cependant, seule une optimisation de cette fonction et la confrontation de spectres obtenus avec le même protocole sur des sites multiples, avec la base de données spectrales, permettra de valider la transférabilité de la méthode à différents appareils.

La méthode développée au cours de cette étude est en cours d'évolution. Ces résultats prometteurs nous engagent à poursuivre cette étude afin d'élargir la base de données initiale à un nombre de souches plus important et appartenant à plus de 200 espèces des moisissures les plus rencontrées dans l'industrie agroalimentaire, pharmaceutique, cosmétique et dans le domaine de la santé publique.

L'une des perspectives de cette étude est d'évaluer de manière externe le potentiel de la méthodologie développée et des bibliothèques de spectres, à identifier rapidement les moisissures par différents partenaires industriels. Après cette étape d'évaluation, une adaptation de la technologie pour une utilisation standardisée en routine et applicable dans un contexte industriel et/ou médical pourra être envisagée.

Enfin, il serait possible de mettre en place une identification à distance, c'est à dire de créer une banque de données spectrales infrarouge centralisée, à laquelle seraient confrontés les spectres d'échantillons à prédire acquis sur différents sites.

Références bibliographiques

1. Bowman SM, Free SJ. The structure and synthesis of the fungal cell wall. *Bioessays*. 2006 Aug;28(8):799-808.
2. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol*. 2005 Sep-Oct;52(5):399-451.
3. Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, et al. A higher-level phylogenetic classification of the Fungi. *Mycol Res*. 2007 May;111(Pt 5):509-47.
4. Ropars J, Cruaud C, Lacoste S, Dupont J. A taxonomic and ecological overview of cheese fungi. *Int J Food Microbiol*. 2012 Apr 16;155(3):199-210.
5. Karaffa L, Sandor E, Fekete E, Szentirmai A. The biochemistry of citric acid accumulation by *Aspergillus niger*. *Acta Microbiol Immunol Hung*. 2001;48(3-4):429-40.
6. Lv XC, Huang ZQ, Zhang W, Rao PF, Ni L. Identification and characterization of filamentous fungi isolated from fermentation starters for Hong Qu glutinous rice wine brewing. *J Gen Appl Microbiol*. 2012;58(1):33-42.
7. Rice LG, Ross PF. Methods for detection and quantitation of fumonisins in corn, cereal products and animal excreta. *Journal of Food Protection*. 1994 Jun;57(6):536-40.
8. Marin S, Ramos AJ, Cano-Sancho G, Sanchis V. Mycotoxins: Occurrence, toxicology, and exposure assessment. *Food Chem Toxicol*. 2013 Jul 29;60C:218-37.
9. Terra MF, Prado G, Pereira GE, Ematne HJ, Batista LR. Detection of ochratoxin A in tropical wine and grape juice from Brazil. *J Sci Food Agric*. 2012 Mar 15;93(4):890-4.
10. Moss MO. Fungi, quality and safety issues in fresh fruits and vegetables. *J Appl Microbiol*. 2008 May;104(5):1239-43.
11. Li FQ, Yoshizawa T, Kawamura O, Luo XY, Li YW. Aflatoxins and fumonisins in corn from the high-incidence area for human hepatocellular carcinoma in Guangxi, China. *J Agric Food Chem*. 2001 Aug;49(8):4122-6.
12. De Lucca AJ. Harmful fungi in both agriculture and medicine. *Rev Iberoam Micol*. 2007 Mar;24(1):3-13.
13. Pagano L, Caira M, Candoni A, Offidani M, Martino B, Specchia G, et al. Invasive aspergillosis in patients with acute myeloid leukemia: a SEIFEM-2008 registry study. *Haematologica*. 2009 Apr;95(4):644-50.

14. Baddley JW, Stroud TP, Salzman D, Pappas PG. Invasive mold infections in allogeneic bone marrow transplant recipients. *Clin Infect Dis*. 2001 May 1;32(9):1319-24.
15. Nucci M, Garnica M, Gloria AB, Lehugeur DS, Dias VC, Palma LC, et al. Invasive fungal diseases in haematopoietic cell transplant recipients and in patients with acute myeloid leukaemia or myelodysplasia in Brazil. *Clin Microbiol Infect*. 2001 Aug;19(8):745-51.
16. Eucker J, Sezer O, Graf B, Possinger K. Mucormycoses. *Mycoses*. 2001;44(7-8):253-60.
17. Muhammed M, Coleman JJ, Carneiro HA, Mylonakis E. The challenge of managing fusariosis. *Virulence*. 2011 Mar-Apr;2(2):91-6.
18. Troke P, Aguirrebengoa K, Arteaga C, Ellis D, Heath CH, Lutsar I, et al. Treatment of scedosporiosis with voriconazole: clinical experience with 107 patients. *Antimicrob Agents Chemother*. 2008 May;52(5):1743-50.
19. Petrikkos G, Drogari-Apiranthitou M. Zygomycosis in Immunocompromised non-Haematological Patients. *Mediterr J Hematol Infect Dis*. 2011;3(1):e2011012.
20. Liu JC, Modha DE, Gaillard EA. What is the clinical significance of filamentous fungi positive sputum cultures in patients with cystic fibrosis? *J Cyst Fibros*. 2013 May;12(3):187-93.
21. Poisson DM, Da Silva NJ, Rousseau D, Esteve E. Tinea corporis gladiatorum: Specificity and epidemiology. *Journal De Mycologie Medicale*. 2007 Sep;17(3):177-82.
22. Verscheure M, Lognay G, Marlier M. Revue bibliographique: les méthodes chimiques d'identification et de classification des champignons. *Biotechnol Agron Soc Environ*. 2002;6:131-42.
23. Bougnoux ME, Espinasse F. Nouvelles applications des techniques de biologie moléculaire en mycologie médicale. *Revue française des laboratoires*. 2003;2003(351):67-71.
24. de Valk HA, Klaassen CH, Meis JF. Molecular typing of *Aspergillus* species. *Mycoses*. 2008 Nov;51(6):463-76.
25. Geisen R, Cantor MD, Hansen TK, Holzappel WH, Jakobsen M. Characterization of *Penicillium roqueforti* strains used as cheese starter cultures by RAPD typing. *Int J Food Microbiol*. 2001 May 10;65(3):183-91.
26. Perrone G, Susca A, Epifani F, Mule G. AFLP characterization of Southern Europe population of *Aspergillus* Section *Nigri* from grapes. *Int J Food Microbiol*. 2006 Sep 1;111 Suppl 1:S22-7.
27. Marvin LF, Roberts MA, Fay LB. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clin Chim Acta*. 2003 Nov;337(1-2):11-21.

28. Cassagne C, Ranque S, Normand AC, Fourquet P, Thiebault S, Planard C, et al. Mould routine identification in the clinical laboratory by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *PLoS One*. 2011;6(12):e28425.
29. De Carolis E, Posteraro B, Lass-Flörl C, Vella A, Florio AR, Torelli R, et al. Species identification of *Aspergillus*, *Fusarium* and *Mucorales* with direct surface analysis by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Microbiol Infect*. 2012 May;18(5):475-84.
30. Normand AC, Cassagne C, Ranque S, L'Ollivier C, Fourquet P, Roesems S, et al. Assessment of various parameters to improve MALDI-TOF MS reference spectra libraries constructed for the routine identification of filamentous fungi. *BMC Microbiol*. 2013;13:76.
31. Duygu D, Baykal T, Açıkgöz D, Yildiz K. Fourier Transform Infrared (FT-IR) Spectroscopy for Biological Studies. *Journal of science*. 2009;22(3):117-21.
32. Naumann D. Infrared spectroscopy in microbiology. R.A. Meyers (ed) ed. Chichester: John Wiley and Sons Ltd; 2000.
33. Mariey L, Signolle J, Amiel C, Travert J. Discrimination, classification, identification of microorganisms using spectroscopy and chemometrics. *Vib Spectrosc*. 2001;26:151-9.
34. Wenning M, Scherer S. Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method. *Appl Microbiol Biotechnol*. 2013 Aug;97(16):7111-20.
35. Amiel C, Mariey L, Curk-Daubié M-C, Pichon P, Travert J. Potentiality of Fourier transform infrared spectroscopy (FTIR) for discrimination and identification of dairy lactic acid bacteria. *Lait*. 2000;80:445-59.
36. Guibet F, Amiel C, Cadot P, Cordevant C, Desmots MH, Lange M, et al. Discrimination and classification of *Enterococci* by Fourier transform infrared (FT-IR) spectroscopy. *Vib Spectrosc*. 2003;33:133-42.
37. Rebuffo CA, Schmitt J, Wenning M, von Stetten F, Scherer S. Reliable and rapid identification of *Listeria monocytogenes* and *Listeria* species by artificial neural network-based Fourier transform infrared spectroscopy. *Appl Environ Microbiol*. 2006 Feb;72(2):994-1000.
38. Mouwen DJ, Capita R, Alonso-Calleja C, Prieto-Gomez J, Prieto M. Artificial neural network based identification of *Campylobacter* species by Fourier transform infrared spectroscopy. *J Microbiol Methods*. 2006 Oct;67(1):131-40.

39. Amiali NM, Mulvey MR, Sedman J, Louie M, Simor AE, Ismail AA. Rapid identification of coagulase-negative *staphylococci* by Fourier transform infrared spectroscopy. *J Microbiol Methods*. 2007 Feb;68(2):236-42.
40. Maquelin K, Kirschner C, Choo-Smith LP, van den Braak N, Endtz HP, Naumann D, et al. Identification of medically relevant microorganisms by vibrational spectroscopy. *J Microbiol Methods*. 2002 Nov;51(3):255-71.
41. Sandt C, Madoulet C, Kohler A, Allouch P, De Champs C, Manfait M, et al. FT-IR microspectroscopy for early identification of some clinically relevant pathogens. *J Appl Microbiol*. 2006 Oct;101(4):785-97.
42. Rubio C, Ott C, Amiel C, Dupont-Moral I, Travert J, Mariey L. Sulfato/thiosulfato reducing bacteria characterization by FT-IR spectroscopy: A new approach to biocorrosion control. *J Microbiol Methods*. 2006;64:287–96.
43. Boudaud N, Coton M, Coton E, Pineau S, Travert J, Amiel C. Biodiversity analysis by polyphasic study of marine bacteria associated with biocorrosion phenomena. *J Appl Microbiol*. 2009 Jul;109(1):166-79.
44. Haag H, Gremlich H-G, Bergmann R, J-J S. Characterization and identification of *actinomycetes* by FT-IR spectroscopy. *J Microbiol Methods*. 1996;27:157–63.
45. Wenning M, Seiler H, Scherer S. Fourier-transform infrared microspectroscopy, a novel and rapid tool for identification of yeasts. *Appl Environ Microbiol*. 2002 Oct;68(10):4717-21.
46. Kummerle M, Scherer S, Seiler H. Rapid and reliable identification of food-borne yeasts by Fourier-transform infrared spectroscopy. *Appl Environ Microbiol*. 1998 Jun;64(6):2207-14.
47. Essendoubi M, Toubas D, Bouzaggou M, Pinon JM, Manfait M, Sockalingum GD. Rapid identification of *Candida* species by FT-IR microspectroscopy. *Biochim Biophys Acta*. 2005 Aug 5;1724(3):239-47.
48. Essendoubi M, Toubas D, Lepouse C, Leon A, Bourgeade F, Pinon JM, et al. Epidemiological investigation and typing of *Candida glabrata* clinical isolates by FTIR spectroscopy. *J Microbiol Methods*. 2007 Dec;71(3):325-31.
49. Sandt C, Sockalingum GD, Aubert D, Lepad H, Lepouse C, Jaussaud M, et al. Use of Fourier-transform infrared spectroscopy for typing of *Candida albicans* strains isolated in intensive care units. *J Clin Microbiol*. 2003 Mar;41(3):954-9.

50. Toubas D, Essendoubi M, Adt I, Pinon JM, Manfait M, Sockalingum GD. FTIR spectroscopy in medical mycology: applications to the differentiation and typing of *Candida*. *Anal Bioanal Chem*. 2007 Mar;387(5):1729-37.
51. Beekes M, Lasch P, Naumann D. Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research. *Vet Microbiol*. 2007 Aug 31;123(4):305-19.
52. Erukhimovitch V, Karpasasa M, Huleihel M. Spectroscopic detection and identification of infected cells with herpes viruses. *Biopolymers*. 2009 Jan;91(1):61-7.
53. Bastert J, Korting HC, Traenkle P, Schmalreck AF. Identification of dermatophytes by Fourier transform infrared spectroscopy (FT-IR). *Mycoses*. 1999;42(9-10):525-8.
54. Ergin C, Ilkit M, Gok Y, Ozel MZ, Con AH, Kabay N, et al. Fourier transform infrared spectral evaluation for the differentiation of clinically relevant *Trichophyton* species. *J Microbiol Methods*. 2013 Jun;93(3):218-23.
55. Garon D, El Kaddoumi A, Carayon A, Amiel C. FT-IR spectroscopy for rapid differentiation of *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus parasiticus* and characterization of aflatoxigenic isolates collected from agricultural environments. *Mycopathologia*. 2010 Aug;170(2):131-42.
56. Tralamazza SM, Bozza A, Destro JG, Rodriguez JI, do Rocio Dalzoto P, Pimentel IC. Potential of Fourier transform infrared spectroscopy (FT-IR) to differentiate environmental *Aspergillus* fungi species *A. niger*, *A. ochraceus*, and *A. westerdijkiae* using two different methodologies. *Appl Spectrosc*. 2013 Mar;67(3):274-8.
57. Nie M, Zhang WQ, Xiao M, Luo JL, Bao K, Chen JK, et al. FT-IR spectroscopy and artificial neural network identification of *Fusarium* species. *Journal of Phytopathology*. 2007 Jun;155(6):364-7.
58. Fischer G, Braun S, Thissen R, Dott W. FT-IR spectroscopy as a tool for rapid identification and intra-species characterization of airborne filamentous fungi. *J Microbiol Methods*. 2006 Jan;64(1):63-77.
59. Shapaval V, Moretro T, Suso HP, Asli AW, Schmitt J, Lillehaug D, et al. A high-throughput microcultivation protocol for FTIR spectroscopic characterization and identification of fungi. *J Biophotonics*. 2010 Aug;3(8-9):512-21.
60. Shapaval V, Schmitt J, Moretro T, Suso HP, Skaar I, Asli AW, et al. Characterization of food spoilage fungi by FTIR spectroscopy. *J Appl Microbiol*. 2013 Mar;114(3):788-96.

61. Santos C, Fraga ME, Kozakiewicz Z, Lima N. Fourier transform infrared as a powerful technique for the identification and characterization of filamentous fungi and yeasts. *Res Microbiol.* 2010 Mar;161(2):168-75.
62. Bertrand D, Dufour E. Identification et caractérisation des microorganismes. Lavoisier, editor. Paris; 2006.
63. Szeghalmi A, Kaminskyj S, Gough KM. A synchrotron FTIR microspectroscopy investigation of fungal hyphae grown under optimal and stressed conditions. *Anal Bioanal Chem.* 2007 Mar;387(5):1779-89.
64. Jilkine K, Gough KM, Julian R, Kaminskyj SG. A sensitive method for examining whole-cell biochemical composition in single cells of filamentous fungi using synchrotron FTIR spectromicroscopy. *J Inorg Biochem.* 2008 Mar;102(3):540-6.
65. Khot PD, Couturier MR, Wilson A, Croft A, Fisher MA. Optimization of matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis for bacterial identification. *J Clin Microbiol.* 2012 Dec;50(12):3845-52.
66. Marklein G, Josten M, Klanke U, Muller E, Horre R, Maier T, et al. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for fast and reliable identification of clinical yeast isolates. *J Clin Microbiol.* 2009 Sep;47(9):2912-7.
67. Bouveresse E, Massart DL. Standardisation of near-infrared spectrometric instruments: A review. *Vibrational Spectroscopy.* 1996 Mar;11(1):3-15.
68. Chamrad DC, Koerting G, Gobom J, Thiele H, Klose J, Meyer HE, et al. Interpretation of mass spectrometry data for high-throughput proteomics. *Anal Bioanal Chem.* 2003 Aug;376(7):1014-22.
69. Chen T, Martin E. The impact of temperature variations on spectroscopic calibration modelling: a comparative study. *Journal of Chemometrics.* 2007 May-Jun;21(5-6):198-207.
70. Zhang L, Small GW, Arnold MA. Multivariate calibration standardization across instruments for the determination of glucose by Fourier transform near-infrared spectrometry. *Analytical Chemistry.* 2003 Nov;75(21):5905-15.
71. Trevisan J, Angelov PP, Carmichael PL, Scott AD, Martin FL. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst.* 2013 Jul 21;137(14):3202-15.
72. Hibbett DS, Nilsson RH, Snyder M, Fonseca M, Costanzo J, Shonfeld M. Automated phylogenetic taxonomy: an example in the *homobasidiomycetes* (mushroom-forming fungi). *Syst Biol.* 2005 Aug;54(4):660-8.

73. Samson RA, Seifert KA, Kuijpers AFA, Houbraken J, Frisvad JC. Phylogenetic analysis of *Penicillium* subgenus *Penicillium* using partial P-tubulin sequences. *Studies in Mycology*. 2004(49):175-200.
74. Kristensen R, Torp M, Kosiak B, Holst-Jensen A. Phylogeny and toxigenic potential is correlated in *Fusarium* species as revealed by partial translation elongation factor 1 alpha gene sequences. *Mycol Res*. 2005 Feb;109(Pt 2):173-86.
75. Kearsley SE, Labib K. MCM proteins: evolution, properties, and role in DNA replication. *Biochim Biophys Acta*. 1998 Jun 16;1398(2):113-36.
76. Gelperin D, Horton L, Beckman J, Hensold J, Lemmon SK. Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. *RNA*. 2001 Sep;7(9):1268-83.
77. White TJ, Bruns T, Lee S, Taylor J. *PCR Protocols: A Guide to Methods and Applications*. Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. ed. San Diego: Academic Press; 1990.
78. Glass NL, Donaldson GC. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous *ascomycetes*. *Appl Environ Microbiol*. 1995 Apr;61(4):1323-30.
79. O'Donnell K, Kistler HC, Cigelnik E, Ploetz RC. Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc Natl Acad Sci U S A*. 1998 Mar 3;95(5):2044-9.
80. Schmitt I, Crespo A, Divakar PK, Fankhauser JD, Herman-Sackett E, Kalb K, et al. New primers for promising single-copy genes in fungal phylogenetics and systematics. *Persoonia*. 2009 Dec;23:35-40.
81. Hermet A, Meheust D, Mounier J, Barbier G, Jany JL. Molecular systematics in the genus *Mucor* with special regards to species encountered in cheese. *Fungal Biol*. 2012 Jun;116(6):692-705.
82. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011 Oct;28(10):2731-9.
83. Helm D, Labischinski H, Naumann D. Elaboration of a procedure for identification of bacteria using Fourier-Transform IR spectral libraries: a stepwise correlation approach. *J Microbiol Methods*. 1991;14:127-42.
84. Wu W, Guo Q, Jouan-Rimbaud D, Massart DL. Using contrasts as data pretreatment method in pattern recognition of multivariate data. *Chemometr Intell Lab*. 1999;45:39-53.

85. Levillain P, Fompeydie D. Derivative spectrophotometry – Principales, advantages and limitations. *Appl Anal.* 1986;14:1-20.
86. Arakaki LSL, Burns DH. Multispectral analysis for quantitative measurements of myoglobin oxygen fractional saturation in the presence of hemoglobin interference. *Appl Spectrosc.* 1992;46:1919-28.
87. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem.* 1964;36:1627-39.
88. Bylesjö M, Cloarec O, M. R. Normalization and Closure. *Compr Chemometr.* 2009;2.07:109-27.
89. Tenenhaus M. La régression PLS: théorie et pratique. Tenenhaus M ed. Paris: Technip; 1998.
90. Liang YZ, Kvalheim O. Robust methods for multivariate analysis. *Chemom Intell Lab Syst.* 1996;32:1–10.
91. Stone M. Cross-validation choice and assessment of statistical predictions. *J R Stat Soc Series B Stat Methodol.* 1974;36:111–47.

Publications et communications

Publications internationales

1) Differentiation and identification of filamentous fungi by high-throughput FTIR spectroscopic analysis of mycelia

A. Lecellier, J. Mounier, V. Gaydou, L. Castrec, G. Barbier, W. Ablain, M. Manfait, D. Toubas, G.D. Sockalingum

International Journal of Food Microbiology, 168-169 (2014), pp. 32-41

2) Implementation of an FTIR spectral library of 486 filamentous fungi strains for rapid identification of molds

A. Lecellier, V. Gaydou, J. Mounier, A. Hermet, L. Castrec, G. Barbier, W. Ablain, M. Manfait, D. Toubas, G.D. Sockalingum

Soumise dans le journal « *Food Microbiology* », Octobre 2013, actuellement en révision

3) Assessing the discrimination potential of linear and non-linear supervised chemometrics methods on a filamentous fungi FTIR spectral database

V. Gaydou, A. Lecellier, D. Toubas, J. Mounier, L. Castrec, G. Barbier, W. Ablain, M. Manfait, G.D. Sockalingum

En finalisation, soumission prévue dans le journal « *Analytical Chemistry* »

Communications orales

- Internationales

Mould identification: Infrared spectroscopy as a high-throughput method

Lecellier A., Mounier J., Gaydou V., Castrec L., Barbier G., Huet S., Ablain W., Manfait M., Toubas D., Sockalingum G.D.

Microbial Spoilers in Food, Quimper, France, July 1-3, 2013

- Nationales

1) Discrimination et identification des champignons filamenteux par analyse du mycélium par spectroscopie infrarouge à haut débit

Lecellier Aurélie, Gaydou Vincent, Mounier Jérôme, Toubas Dominique, Le Bras Marie-Anne, Barbier Georges, Leden Nadia, Huet Stéphane, Ablain Wilfried, Manfait Michel, Sockalingum Ganesh
Journée des jeunes chercheurs de la SFR CAP-Santé, Reims, le 28 Mars 2013

2) Analyse du mycélium par spectroscopie infrarouge à transformée de Fourier pour l'identification des champignons filamenteux

A. Lecellier, V. Gaydou, J. Mounier, L. Castrec, G. Barbier, N. Leden, S. Huet, W. Ablain, M. Manfait, G.D. Sockalingum, D. Toubas
SFMM : Société Française de Mycologie Médicale, Dijon, 15-17 Mai 2013

Communications affichées

- Internationales

1) Comparative FTIR spectroscopic analysis of spores and mycelia for differentiating filamentous fungi

A. Lecellier, J. Mounier, D. Toubas, A. Kerviel, M. Le Bras, G. Barbier, N. Leden, S. Huet, M. Manfait, G.D. Sockalingum
4th Congress of European Microbiologists, FEMS 2011, Geneva, Switzerland, June 26-30, 2011

2) FTIR spectroscopic analysis of spores and mycelia: a comparative study for the identification of filamentous fungi

A. Lecellier, J. Mounier, D. Toubas, A. Kerviel, M. Le Bras, G. Barbier, N. Leden, S. Huet, M. Manfait, G.D. Sockalingum
14th European Conference on the Spectroscopy of Biological Molecules, ECSBM 2011, Portugal, Coimbra University, 29th august to 3rd September 2011

3) Rapid FTIR spectroscopic analysis of mycelia for the identification of filamentous fungi

A. Lecellier, V. Gaydou, J. Mounier, L. Castrec, G. Barbier, N. Leden, S. Huet, W. Ablain, M. Manfait, G.D. Sockalingum, **D. Toubas**

ECCMID: European Society of Clinical Microbiology and Infectious Diseases, Berlin, April 27-30, 2013

- Nationales

1) FTIR spectroscopic analysis of spores and mycelia: a comparative study for the identification of filamentous fungi

A. Lecellier, J. Mounier, D. Toubas, A. Kerviel, M. Le Bras, G. Barbier, N. Leden, S. Huet, M. Manfait, G.D. Sockalingum

CRP Santé Luxembourg – SFR CAP Santé Reims, 28 Novembre 2011, Reims

2) Différenciation et identification des champignons filamenteux par analyse du mycélium par spectroscopie infrarouge à haut débit

LECELLIER Aurélie, MOUNIER Jérôme, TOUBAS Dominique, LE BRAS Marie-Anne, BARBIER Georges, LEDEN Nadia, HUET Stéphane, GOBINET Cyril, MANFAIT Michel, SOCKALINGUM Ganesh

Journée des jeunes chercheurs de la SFR CAP-Santé, 7 juin 2012, Amiens

Annexes

Annexe1 : Liste des souches fongiques utilisées dans cette étude (Collection UBOCC : Université de Bretagne Occidentale, Collection CBS : Centraalbureau voor Schimmelcultures).

<i>Absidia coerulea</i>	UBOCC-A-101326	<i>Aspergillus fumigatus</i>	UBOCC-A-106012
<i>Absidia coerulea</i>	UBOCC-A-101327	<i>Aspergillus fumigatus</i>	UBOCC-A-106013
<i>Absidia repens</i>	UBOCC-A-101332	<i>Aspergillus fumigatus</i>	UBOCC-A-106014
<i>Actinomucor elegans</i>	CBS 153.86	<i>Aspergillus fumigatus</i>	UBOCC-A-106015
<i>Actinomucor elegans</i>	UBOCC-A-102005	<i>Aspergillus fumigatus</i>	UBOCC-A-106016
<i>Actinomucor elegans</i>	UBOCC-A-106035	<i>Aspergillus fumigatus</i>	UBOCC-A-106018
<i>Actinomucor elegans</i>	UBOCC-A-101333	<i>Aspergillus fumigatus</i>	UBOCC-A-110089
<i>Alternaria alternata</i>	CBS 116329	<i>Aspergillus niger</i>	CBS 554.65
<i>Alternaria alternata</i>	CBS 117143	<i>Aspergillus niger</i>	UBOCC-A-101072
<i>Alternaria alternata</i>	CBS 916.96	<i>Aspergillus niger</i>	UBOCC-A-101073
<i>Alternaria alternata</i>	UBOCC-A-111005	<i>Aspergillus niger</i>	UBOCC-A-101074
<i>Alternaria chartarum</i>	UBOCC-A-101045	<i>Aspergillus niger</i>	UBOCC-A-101075
<i>Aspergillus calidoustus</i>	UBOCC-A-101086	<i>Aspergillus niger</i>	UBOCC-A-101076
<i>Aspergillus candidus</i>	CBS 114985	<i>Aspergillus niger</i>	UBOCC-A-101089
<i>Aspergillus clavati</i>	UBOCC-A-101055	<i>Aspergillus niger</i>	UBOCC-A-112064
<i>Aspergillus elegans</i>	CBS 108.08	<i>Aspergillus niger</i>	UBOCC-A-112068
<i>Aspergillus elegans</i>	UBOCC-A-105015	<i>Aspergillus niger</i>	UBOCC-A-112080
<i>Aspergillus flavus</i>	CBS 100927	<i>Aspergillus niger</i>	UBOCC-A-112082
<i>Aspergillus flavus</i>	UBOCC-A-101060	<i>Aspergillus pseudoflectus</i>	UBOCC-A-101085
<i>Aspergillus flavus</i>	UBOCC-A-101061	<i>Aspergillus sclerotium</i>	UBOCC-A-105001
<i>Aspergillus flavus</i>	UBOCC-A-101063	<i>Aspergillus sclerotium</i>	UBOCC-A-105010
<i>Aspergillus flavus</i>	UBOCC-A-106025	<i>Aspergillus sclerotium</i>	UBOCC-A-105013
<i>Aspergillus flavus</i>	UBOCC-A-106026	<i>Aspergillus sydowii</i>	UBOCC-A-108050
<i>Aspergillus flavus</i>	UBOCC-A-106028	<i>Aspergillus sydowii</i>	UBOCC-A-108052
<i>Aspergillus flavus</i>	UBOCC-A-106029	<i>Aspergillus sydowii</i>	UBOCC-A-108053
<i>Aspergillus flavus</i>	UBOCC-A-106030	<i>Aspergillus sydowii</i>	UBOCC-A-106017
<i>Aspergillus flavus</i>	UBOCC-A-106031	<i>Aspergillus tamari</i>	CBS 104.13
<i>Aspergillus flavus</i>	UBOCC-A-106032	<i>Aspergillus tamari</i>	CBS 104.14
<i>Aspergillus flavus</i>	UBOCC-A-106033	<i>Aspergillus tamari</i>	CBS 590.68
<i>Aspergillus flavus</i>	UBOCC-A-108066	<i>Aspergillus tamari</i>	CBS 129.49
<i>Aspergillus flavus</i>	UBOCC-A-108068	<i>Aspergillus versicolor</i>	CBS 109274
<i>Aspergillus fumigatus</i>	CBS 121719	<i>Aspergillus versicolor</i>	CBS 583.65
<i>Aspergillus fumigatus</i>	UBOCC-A-101065	<i>Aspergillus versicolor</i>	UBOCC-A-101087
<i>Aspergillus fumigatus</i>	UBOCC-A-101066	<i>Aspergillus versicolor</i>	UBOCC-A-101088
<i>Aspergillus fumigatus</i>	UBOCC-A-106001	<i>Aspergillus versicolor</i>	UBOCC-A-102012
<i>Aspergillus fumigatus</i>	UBOCC-A-106002	<i>Aspergillus versicolor</i>	UBOCC-A-112085
<i>Aspergillus fumigatus</i>	UBOCC-A-106003	<i>Aspergillus wentii</i>	CBS 104.07
<i>Aspergillus fumigatus</i>	UBOCC-A-106004	<i>Aspergillus wentii</i>	CBS 127.28
<i>Aspergillus fumigatus</i>	UBOCC-A-106005	<i>Aspergillus wentii</i>	CBS 229.67
<i>Aspergillus fumigatus</i>	UBOCC-A-106006	<i>Aspergillus wentii</i>	UBOCC-A-101090
<i>Aspergillus fumigatus</i>	UBOCC-A-106007	<i>Aspergillus westerdijkiae</i>	UBOCC-A-101078
<i>Aspergillus fumigatus</i>	UBOCC-A-106008	<i>Aureobasidium pullulans</i>	UBOCC-A-101091
<i>Aspergillus fumigatus</i>	UBOCC-A-106009	<i>Aureobasidium pullulans</i>	UBOCC-A-101092
<i>Aspergillus fumigatus</i>	UBOCC-A-106010	<i>Aureobasidium pullulans</i>	UBOCC-A-108047
<i>Aspergillus fumigatus</i>	UBOCC-A-106011	<i>Aureobasidium pullulans</i>	UBOCC-A-108056

<i>Aureobasidium pullulans</i>	UBOCC-A-108057	<i>Eurotium rubrum</i>	UBOCC-A-108078
<i>Bionectria aureofulvella</i>	UBOCC-A-101174	<i>Eurotium rubrum</i>	CBS 530.65
<i>Bionectria ochroleuca</i>	UBOCC-A-101319	<i>Fusarium avenaceum</i>	CBS 408.86
<i>Bionectria ochroleuca</i>	UBOCC-A-105019	<i>Fusarium avenaceum</i>	UBOCC-A-101136
<i>Bionectria solani</i>	UBOCC-A-102025	<i>Fusarium avenaceum</i>	UBOCC-A-101137
<i>Botrytis cinerae</i>	CBS 810.69	<i>Fusarium avenaceum</i>	UBOCC-A-109018
<i>Botrytis cinerae</i>	UBOCC-A-101099	<i>Fusarium avenaceum</i>	UBOCC-A-109033
<i>Botrytis cinerae</i>	UBOCC-A-101100	<i>Fusarium avenaceum</i>	UBOCC-A-109035
<i>Ceratocystis paradoxa</i>	UBOCC-A-101283	<i>Fusarium avenaceum</i>	UBOCC-A-109048
<i>Ceratocystis paradoxa</i>	UBOCC-A-101285	<i>Fusarium avenaceum</i>	UBOCC-A-109096
<i>Chaetomium erectum</i>	UBOCC-A-101010	<i>Fusarium avenaceum</i>	UBOCC-A-109138
<i>Chaetomium globosum</i>	CBS 107.14	<i>Fusarium avenaceum</i>	UBOCC-A-109139
<i>Chaetomium globosum</i>	CBS 148.51	<i>Fusarium avenaceum</i>	UBOCC-A-109140
<i>Circinella sydowii</i>	UBOCC-A-101338	<i>Fusarium avenaceum</i>	UBOCC-A-109141
<i>Circinella sydowii</i>	UBOCC-A-101339	<i>Fusarium avenaceum</i>	UBOCC-A-109142
<i>Cladosporium brunhei</i>	CBS 134.31	<i>Fusarium avenaceum</i>	UBOCC-A-109143
<i>Cladosporium cladosporioides</i>	CBS 109.21	<i>Fusarium culmorum</i>	UBOCC-A-101139
<i>Cladosporium cladosporioides</i>	UBOCC-A-101114	<i>Fusarium culmorum</i>	UBOCC-A-107001
<i>Cladosporium herbarum</i>	CBS 673.69	<i>Fusarium culmorum</i>	UBOCC-A-109109
<i>Cladosporium ramotenellum</i>	CBS 109031	<i>Fusarium culmorum</i>	UBOCC-A-109110
<i>Cladosporium ramotenellum</i>	UBOCC-A-108072	<i>Fusarium culmorum</i>	UBOCC-A-109124
<i>Cladosporium ramotenellum</i>	UBOCC-A-108073	<i>Fusarium culmorum</i>	UBOCC-A-109125
<i>Cladosporium sphaerospermum</i>	UBOCC-A-101107	<i>Fusarium culmorum</i>	UBOCC-A-109126
<i>Cladosporium sphaerospermum</i>	UBOCC-A-101110	<i>Fusarium culmorum</i>	UBOCC-A-109128
<i>Cladosporium sphaerospermum</i>	UBOCC-A-101111	<i>Fusarium culmorum</i>	UBOCC-A-109139
<i>Cladosporium sphaerospermum</i>	UBOCC-A-108054	<i>Fusarium equiseti</i>	CBS 123566
<i>Colletotrichum acutatum</i>	UBOCC-A-101180	<i>Fusarium equiseti</i>	CBS 163.57
<i>Colletotrichum coccodes</i>	UBOCC-A-101118	<i>Fusarium equiseti</i>	CBS 414.86
<i>Cryphonectria parasitica</i>	UBOCC-A-101130	<i>Fusarium equiseti</i>	CBS 791.70
<i>Cunninghamella binariae</i>	UBOCC-A-101343	<i>Fusarium equiseti</i>	UBOCC-A-109007
<i>Cunninghamella blakesleeana</i>	UBOCC-A-101341	<i>Fusarium equiseti</i>	UBOCC-A-109015
<i>Cunninghamella elegans</i>	UBOCC-A-101342	<i>Fusarium equiseti</i>	UBOCC-A-109016
<i>Cunninghamella elegans</i>	UBOCC-A-102008	<i>Fusarium equiseti</i>	UBOCC-A-109017
<i>Emericella nidulans</i>	CBS 121.35	<i>Fusarium equiseti</i>	UBOCC-A-109029
<i>Emericella nidulans</i>	CBS 589.65	<i>Fusarium equiseti</i>	UBOCC-A-109030
<i>Emericella nidulans</i>	CBS 119.55	<i>Fusarium equiseti</i>	UBOCC-A-109031
<i>Emericella nidulans</i>	CBS 465.65	<i>Fusarium equiseti</i>	UBOCC-A-109037
<i>Emericella nidulans</i>	CBS 492.65	<i>Fusarium equiseti</i>	UBOCC-A-109038
<i>Emericella nidulans</i>	UBOCC-A-101069	<i>Fusarium equiseti</i>	UBOCC-A-109039
<i>Emericella nidulans</i>	UBOCC-A-110152	<i>Fusarium equiseti</i>	UBOCC-A-109040
<i>Emericella variegata</i>	UBOCC-A-101071	<i>Fusarium equiseti</i>	UBOCC-A-109041
<i>Eupenicillium pinetorum</i>	UBOCC-A-109223	<i>Fusarium equiseti</i>	UBOCC-A-109042
<i>Eurotium amstelodami</i>	CBS 117323	<i>Fusarium equiseti</i>	UBOCC-A-109043
<i>Eurotium amstelodami</i>	CBS 119376	<i>Fusarium equiseti</i>	UBOCC-A-109044
<i>Eurotium amstelodami</i>	CBS 817.96	<i>Fusarium equiseti</i>	UBOCC-A-109045
<i>Eurotium chevalieri</i>	CBS 129.54	<i>Fusarium equiseti</i>	UBOCC-A-109046
<i>Eurotium chevalieri</i>	CBS 522.65	<i>Fusarium equiseti</i>	UBOCC-A-109047
<i>Eurotium chevalieri</i>	CBS 121704	<i>Fusarium equiseti</i>	UBOCC-A-109085
<i>Eurotium repens</i>	UBOCC-A-101079	<i>Fusarium graminearum</i>	CBS 110266
<i>Eurotium rubrum</i>	CBS 104.18	<i>Fusarium graminearum</i>	CBS 447.95

<i>Fusarium graminearum</i>	UBOCC-A-101142	<i>Fusarium verticillioides</i>	CBS 218.76
<i>Fusarium graminearum</i>	UBOCC-A-101143	<i>Fusarium verticillioides</i>	UBOCC-A-101150
<i>Fusarium graminearum</i>	UBOCC-A-101144	<i>Fusarium verticillioides</i>	UBOCC-A-101165
<i>Fusarium graminearum</i>	UBOCC-A-101145	<i>Fusarium verticillioides</i>	UBOCC-A-109118
<i>Fusarium graminearum</i>	UBOCC-A-102016	<i>Fusarium verticillioides</i>	UBOCC-A-109119
<i>Fusarium graminearum</i>	UBOCC-A-109002	<i>Fusarium verticillioides</i>	UBOCC-A-109121
<i>Fusarium graminearum</i>	UBOCC-A-109010	<i>Fusarium verticillioides</i>	UBOCC-A-109122
<i>Fusarium graminearum</i>	UBOCC-A-109011	<i>Fusarium verticillioides</i>	UBOCC-A-109144
<i>Fusarium graminearum</i>	UBOCC-A-109023	<i>Fusarium verticillioides</i>	UBOCC-A-109145
<i>Fusarium graminearum</i>	UBOCC-A-109032	<i>Fusarium verticillioides</i>	UBOCC-A-110165
<i>Fusarium graminearum</i>	UBOCC-A-109050	<i>Gelasinospora</i> sp	UBOCC-A-101018
<i>Fusarium graminearum</i>	UBOCC-A-109105	<i>Geotrichum candidum</i>	UBOCC-A-101167
<i>Fusarium graminearum</i>	UBOCC-A-109106	<i>Geotrichum candidum</i>	UBOCC-A-101169
<i>Fusarium graminearum</i>	UBOCC-A-109107	<i>Geotrichum candidum</i>	UBOCC-A-101170
<i>Fusarium graminearum</i>	UBOCC-A-109129	<i>Geotrichum candidum</i>	UBOCC-A-101172
<i>Fusarium graminearum</i>	UBOCC-A-109130	<i>Geotrichum candidum</i>	UBOCC-A-101173
<i>Fusarium langsethiae</i>	UBOCC-A-109148	<i>Geotrichum candidum</i>	UBOCC-A-103039
<i>Fusarium langsethiae</i>	UBOCC-A-110061	<i>Geotrichum candidum</i>	UBOCC-A-108080
<i>Fusarium langsethiae</i>	UBOCC-A-110063	<i>Geotrichum candidum</i>	UBOCC-A-108081
<i>Fusarium oxysporum</i>	CBS 221.49	<i>Geotrichum candidum</i>	UBOCC-A-108082
<i>Fusarium oxysporum</i>	UBOCC-A-101135	<i>Geotrichum citri-aurantii</i>	UBOCC-A-101171
<i>Fusarium oxysporum</i>	UBOCC-A-101138	<i>Geotrichum citri-aurantii</i>	UBOCC-A-101206
<i>Fusarium oxysporum</i>	UBOCC-A-101140	<i>Geotrichum silvicola</i>	UBOCC-A-108083
<i>Fusarium oxysporum</i>	UBOCC-A-101151	<i>Humicola fuscoatra</i>	UBOCC-A-101190
<i>Fusarium oxysporum</i>	UBOCC-A-101152	<i>Hypocrea virens</i>	UBOCC-A-101176
<i>Fusarium oxysporum</i>	UBOCC-A-101154	<i>Kernia pachypleura</i>	UBOCC-A-101266
<i>Fusarium oxysporum</i>	UBOCC-A-101155	<i>Lichtheimia corymbifera</i>	UBOCC-A-101328
<i>Fusarium oxysporum</i>	UBOCC-A-101156	<i>Lichtheimia corymbifera</i>	UBOCC-A-101331
<i>Fusarium oxysporum</i>	UBOCC-A-101157	<i>Lichtheimia corymbifera</i>	UBOCC-A-102023
<i>Fusarium oxysporum</i>	UBOCC-A-101158	<i>Lichtheimia corymbifera</i>	UBOCC-A-103031
<i>Fusarium oxysporum</i>	UBOCC-A-108128	<i>Microdochium nivale</i>	UBOCC-A-102027
<i>Fusarium oxysporum</i>	UBOCC-A-109102	<i>Microdochium nivale</i>	UBOCC-A-105025
<i>Fusarium oxysporum</i>	UBOCC-A-110142	<i>Microdochium nivale</i>	UBOCC-A-105026
<i>Fusarium oxysporum</i>	UBOCC-A-112042	<i>Microdochium nivale</i>	UBOCC-A-105027
<i>Fusarium proliferatum</i>	UBOCC-A-109149	<i>Mortierella hyalina</i>	UBOCC-A-101349
<i>Fusarium sambucinum</i>	UBOCC-A-109006	<i>Mortierella zonata</i>	UBOCC-A-101348
<i>Fusarium sambucinum</i>	UBOCC-A-109019	<i>Mucor circinelloides</i>	CBS 195.68
<i>Fusarium sambucinum</i>	UBOCC-A-109024	<i>Mucor circinelloides</i>	CBS 205.68
<i>Fusarium sambucinum</i>	UBOCC-A-109027	<i>Mucor circinelloides</i>	CBS 223.56
<i>Fusarium sambucinum</i>	UBOCC-A-109020	<i>Mucor circinelloides</i>	UBOCC-A-101354
<i>Fusarium solani</i>	CBS 128.29	<i>Mucor circinelloides</i>	UBOCC-A-102003
<i>Fusarium solani</i>	UBOCC-A-101164	<i>Mucor circinelloides</i>	UBOCC-A-102017
<i>Fusarium solani</i>	UBOCC-A-109146	<i>Mucor circinelloides</i>	UBOCC-A-105016
<i>Fusarium solani</i>	UBOCC-A-110136	<i>Mucor circinelloides</i>	UBOCC-A-105017
<i>Fusarium sporotrichoides</i>	UBOCC-A-102015	<i>Mucor circinelloides</i>	UBOCC-A-105018
<i>Fusarium sporotrichoides</i>	UBOCC-A-109116	<i>Mucor circinelloides</i>	UBOCC-A-108126
<i>Fusarium subglutinans</i>	CBS 215.76	<i>Mucor circinelloides</i>	UBOCC-A-109066
<i>Fusarium temperatum</i>	UBOCC-A-101148	<i>Mucor circinelloides</i>	UBOCC-A-109067
<i>Fusarium thapsinum</i>	CBS 539.79	<i>Mucor circinelloides</i>	UBOCC-A-109072
<i>Fusarium verticillioides</i>	CBS 119825	<i>Mucor circinelloides</i>	UBOCC-A-109073

<i>Mucor circinelloides</i>	UBOCC-A-109082	<i>Paecilomyces lilacinus</i>	UBOCC-A-108027
<i>Mucor circinelloides</i>	UBOCC-A-109084	<i>Paecilomyces lilacinus</i>	UBOCC-A-108030
<i>Mucor circinelloides</i>	UBOCC-A-110124	<i>Paecilomyces saturatus</i>	UBOCC-A-101210
<i>Mucor circinelloides</i>	UBOCC-A-110127	<i>Paecilomyces variotii</i>	CBS 101032
<i>Mucor fragilis</i>	UBOCC-A-101356	<i>Paecilomyces variotii</i>	UBOCC-A-101209
<i>Mucor hiemalis</i>	UBOCC-A-101359	<i>Paecilomyces variotii</i>	UBOCC-A-103043
<i>Mucor hiemalis</i>	UBOCC-A-101360	<i>Paecilomyces variotii</i>	UBOCC-A-103044
<i>Mucor hiemalis</i>	UBOCC-A-109054	<i>Paecilomyces variotii</i>	UBOCC-A-106027
<i>Mucor hiemalis</i>	UBOCC-A-109059	<i>Paecilomyces variotii</i>	UBOCC-A-110151
<i>Mucor mucedo</i>	UBOCC-A-101353	<i>Papularia sp</i>	UBOCC-A-101212
<i>Mucor racemosus</i>	CBS 113.08	<i>Penicillium aurantiogriseum</i>	UBOCC-A-108092
<i>Mucor racemosus</i>	CBS 115.08	<i>Penicillium brevicompactum</i>	CBS 257.29
<i>Mucor racemosus</i>	CBS 260.68	<i>Penicillium brevicompactum</i>	UBOCC-A-101198
<i>Mucor racemosus</i>	CBS 636.67	<i>Penicillium brevicompactum</i>	UBOCC-A-108093
<i>Mucor racemosus</i>	UBOCC-A-101352	<i>Penicillium brevicompactum</i>	UBOCC-A-108094
<i>Mucor racemosus</i>	UBOCC-A-101366	<i>Penicillium brevicompactum</i>	UBOCC-A-108095
<i>Mucor racemosus</i>	UBOCC-A-102007	<i>Penicillium brevicompactum</i>	UBOCC-A-110065
<i>Mucor racemosus</i>	UBOCC-A-108091	<i>Penicillium brevicompactum</i>	UBOCC-A-112048
<i>Mucor racemosus</i>	UBOCC-A-109051	<i>Penicillium brevicompactum</i>	UBOCC-A-112078
<i>Mucor racemosus</i>	UBOCC-A-109056	<i>Penicillium brunneum</i>	UBOCC-A-101391
<i>Mucor racemosus</i>	UBOCC-A-109062	<i>Penicillium camenberti</i>	CBS 299.48
<i>Mucor racemosus</i>	UBOCC-A-109063	<i>Penicillium camenberti</i>	UBOCC-A-101398
<i>Mucor racemosus</i>	UBOCC-A-109068	<i>Penicillium camenberti</i>	UBOCC-A-101455
<i>Mucor racemosus</i>	UBOCC-A-109069	<i>Penicillium camenberti</i>	UBOCC-A-108097
<i>Mucor racemosus</i>	UBOCC-A-109078	<i>Penicillium carneum</i>	CBS 100539
<i>Mucor racemosus</i>	UBOCC-A-109083	<i>Penicillium carneum</i>	CBS 302.97
<i>Mucor spinosus</i>	CBS 226.32	<i>Penicillium carneum</i>	CBS 466.95
<i>Mucor spinosus</i>	CBS 226.32	<i>Penicillium carneum</i>	CBS 468.95
<i>Mucor spinosus</i>	CBS 246.58	<i>Penicillium carneum</i>	CBS 112297
<i>Mucor spinosus</i>	UBOCC-A-101363	<i>Penicillium carneum</i>	CBS 112489
<i>Mucor spinosus</i>	UBOCC-A-101364	<i>Penicillium chrysogenum</i>	CBS 111214
<i>Mucor spinosus</i>	UBOCC-A-102004	<i>Penicillium chrysogenum</i>	CBS 478.84
<i>Mucor spinosus</i>	UBOCC-A-103032	<i>Penicillium chrysogenum</i>	UBOCC-A-101393
<i>Mucor spinosus</i>	UBOCC-A-108086	<i>Penicillium chrysogenum</i>	UBOCC-A-101399
<i>Mucor spinosus</i>	UBOCC-A-109052	<i>Penicillium chrysogenum</i>	UBOCC-A-101400
<i>Mucor spinosus</i>	UBOCC-A-109053	<i>Penicillium chrysogenum</i>	UBOCC-A-101458
<i>Mucor spinosus</i>	UBOCC-A-109061	<i>Penicillium chrysogenum</i>	UBOCC-A-102022
<i>Mucor spinosus</i>	UBOCC-A-109094	<i>Penicillium chrysogenum</i>	UBOCC-A-106023
<i>Mucor spinosus</i>	UBOCC-A-110133	<i>Penicillium chrysogenum</i>	UBOCC-A-106024
<i>Mucor velutinosus</i>	UBOCC-A-103030	<i>Penicillium chrysogenum</i>	UBOCC-A-110067
<i>Mucor velutinosus</i>	UBOCC-A-109070	<i>Penicillium chrysogenum</i>	UBOCC-A-112065
<i>Mucor velutinosus</i>	UBOCC-A-109075	<i>Penicillium chrysogenum</i>	UBOCC-A-112074
<i>Myrothecium cinctum</i>	UBOCC-A-101201	<i>Penicillium chrysogenum</i>	UBOCC-A-112077
<i>Neosartorya fennelliae</i>	CBS 584.90	<i>Penicillium chrysogenum</i>	UBOCC-A-112108
<i>Neosartorya fischeri</i>	CBS 544.65	<i>Penicillium citrinum</i>	CBS 252.55
<i>Neosartorya glabra</i>	UBOCC-A-101203	<i>Penicillium citrinum</i>	CBS 309.48
<i>Neosartorya hiratsukae</i>	CBS 102802	<i>Penicillium citrinum</i>	UBOCC-A-110169
<i>Neosartorya pseudofischeri</i>	UBOCC-A-101204	<i>Penicillium commune</i>	CBS 261.29
<i>Paecilomyces lilacinus</i>	UBOCC-A-101208	<i>Penicillium commune</i>	CBS 269.97
<i>Paecilomyces lilacinus</i>	UBOCC-A-108014	<i>Penicillium commune</i>	UBOCC-A-101403

<i>Penicillium commune</i>	UBOCC-A-108098	<i>Penicillium paneum</i>	UBOCC-A-109218
<i>Penicillium commune</i>	UBOCC-A-108127	<i>Penicillium paneum</i>	UBOCC-A-111183
<i>Penicillium commune</i>	UBOCC-A-110150	<i>Penicillium raistrickii</i>	UBOCC-A-101440
<i>Penicillium concavovrugulosum</i>	UBOCC-A-101454	<i>Penicillium rolfsii</i>	UBOCC-A-101444
<i>Penicillium coralligerum</i>	UBOCC-A-101404	<i>Penicillium roqueforti</i>	CBS 221.30
<i>Penicillium corylophilum</i>	UBOCC-A-101405	<i>Penicillium roqueforti</i>	UBOCC-A-101445
<i>Penicillium corylophilum</i>	UBOCC-A-109219	<i>Penicillium roqueforti</i>	UBOCC-A-101449
<i>Penicillium corylophilum</i>	UBOCC-A-109222	<i>Penicillium roqueforti</i>	UBOCC-A-108110
<i>Penicillium corylophilum</i>	UBOCC-A-109224	<i>Penicillium roqueforti</i>	UBOCC-A-108111
<i>Penicillium corylophilum</i>	UBOCC-A-112049	<i>Penicillium roqueforti</i>	UBOCC-A-108112
<i>Penicillium corylophilum</i>	UBOCC-A-112069	<i>Penicillium roqueforti</i>	UBOCC-A-109090
<i>Penicillium corylophilum</i>	UBOCC-A-112070	<i>Penicillium roqueforti</i>	UBOCC-A-110052
<i>Penicillium corylophilum</i>	UBOCC-A-112081	<i>Penicillium roqueforti</i>	UBOCC-A-112026
<i>Penicillium crustosum</i>	UBOCC-A-101434	<i>Penicillium solitum</i>	UBOCC-A-108113
<i>Penicillium crustosum</i>	UBOCC-A-110068	<i>Penicillium spinulosum</i>	UBOCC-A-101442
<i>Penicillium expansum</i>	UBOCC-A-108102	<i>Penicillium thomii</i>	UBOCC-A-101463
<i>Penicillium expansum</i>	UBOCC-A-110021	<i>Penicillium verrucosum</i>	CBS 115508
<i>Penicillium expansum</i>	UBOCC-A-110023	<i>Penicillium verrucosum</i>	UBOCC-A-105004
<i>Penicillium expansum</i>	UBOCC-A-110024	<i>Penicillium verrucosum</i>	UBOCC-A-105007
<i>Penicillium expansum</i>	UBOCC-A-110026	<i>Penicillium verrucosum</i>	UBOCC-A-105014
<i>Penicillium expansum</i>	UBOCC-A-110028	<i>Penicillium verrucosum</i>	UBOCC-A-109221
<i>Penicillium expansum</i>	UBOCC-A-110030	<i>Penicillium viridicatum</i>	UBOCC-A-108115
<i>Penicillium expansum</i>	UBOCC-A-110032	<i>Pestalotiopsis sp</i>	UBOCC-A-101216
<i>Penicillium expansum</i>	UBOCC-A-110034	<i>Peyronellaea anserina</i>	UBOCC-A-102026
<i>Penicillium expansum</i>	UBOCC-A-110070	<i>Peyronellaea clade</i>	UBOCC-A-101141
<i>Penicillium fellutanum</i>	CBS 172.44	<i>Phomopsis sp</i>	UBOCC-A-101245
<i>Penicillium freii</i>	CBS 477.84	<i>Pilidium concavum</i>	UBOCC-A-101181
<i>Penicillium glabrum</i>	UBOCC-A-108105	<i>Rhizopus oryzae</i>	CBS 112.07
<i>Penicillium glabrum</i>	UBOCC-A-108106	<i>Rhizopus oryzae</i>	CBS 127.08
<i>Penicillium glabrum</i>	UBOCC-A-108107	<i>Rhizopus oryzae</i>	CBS 146.90
<i>Penicillium glabrum</i>	UBOCC-A-108114	<i>Rhizopus oryzae</i>	CBS 278.38
<i>Penicillium glabrum</i>	UBOCC-A-109089	<i>Rhizopus oryzae</i>	CBS 515.94
<i>Penicillium glabrum</i>	UBOCC-A-109098	<i>Rhizopus oryzae</i>	UBOCC-A-101369
<i>Penicillium glabrum</i>	UBOCC-A-110122	<i>Rhizopus oryzae</i>	UBOCC-A-101371
<i>Penicillium glandicola</i>	UBOCC-A-101422	<i>Rhizopus oryzae</i>	UBOCC-A-101372
<i>Penicillium janthinellum</i>	UBOCC-A-101428	<i>Scopulariopsis brevicaulis</i>	UBOCC-A-101267
<i>Penicillium nalgiovense</i>	UBOCC-A-101430	<i>Scopulariopsis fusca</i>	UBOCC-A-101271
<i>Penicillium nalgiovense</i>	UBOCC-A-101431	<i>Scopulariopsis fusca</i>	UBOCC-A-101272
<i>Penicillium nalgiovense</i>	UBOCC-A-108109	<i>Scopulariopsis fusca</i>	UBOCC-A-108119
<i>Penicillium nordicum</i>	CBS 323.92	<i>Scopulariopsis fusca</i>	UBOCC-A-108120
<i>Penicillium oxalicum</i>	CBS 301.97	<i>Stagonosporopsis valerianellae</i>	UBOCC-A-101243
<i>Penicillium oxalicum</i>	UBOCC-A-101435	<i>Stagonosporopsis valerianellae</i>	UBOCC-A-101244
<i>Penicillium oxalicum</i>	UBOCC-A-101436	<i>Syncephalastrum monosporum</i>	UBOCC-A-101373
<i>Penicillium oxalicum</i>	UBOCC-A-101437	<i>Syncephalastrum racemosum</i>	UBOCC-A-101374
<i>Penicillium oxalicum</i>	UBOCC-A-101438	<i>Talaromyces flavus</i>	UBOCC-A-101037
<i>Penicillium oxalicum</i>	UBOCC-A-102021	<i>Thamnidium elegans</i>	UBOCC-A-105020
<i>Penicillium palitans</i>	CBS 311.48	<i>Trichoderma aggressivum</i>	CBS 101525
<i>Penicillium paneum</i>	CBS 303.97	<i>Trichoderma harzianum</i>	CBS 226.95
<i>Penicillium paneum</i>	CBS 464.95	<i>Trichoderma longibrachiatum</i>	UBOCC-A-101290
<i>Penicillium paneum</i>	UBOCC-A-101448	<i>Trichoderma viride</i>	UBOCC-A-101288

<i>Umbelopsis autotrophica</i>	UBOCC-A-101347
<i>Umbelopsis isabellina</i>	UBOCC-A-101350
<i>Umbelopsis isabellina</i>	UBOCC-A-101351
<i>Verticillium dahliae</i>	UBOCC-A-101313
<i>Verticillium dahliae</i>	UBOCC-A-101314
<i>Verticillium lecanii</i>	UBOCC-A-101320
<i>Verticillium lecanii</i>	UBOCC-A-108019
<i>Verticillium lecanii</i>	UBOCC-A-108023

Annexe 2 : Embranchement taxonomique des souches fongiques utilisées dans cette étude.

Subdivision	Classe	Ordre	Famille	Genre	Sous Genre	Section	Serie	Espèce		
<i>Pezizomycotina</i>	<i>Eurotiomycetes</i>	<i>Eurotiales</i>	<i>Trichocomaceae</i>	<i>Aspergillus</i>		<i>Flavi</i>		<i>A. flavus</i>		
						<i>Nigri</i>		<i>A. niger</i>		
						<i>Cremeri</i>		<i>A. tamari</i>		
								<i>A. wentii</i>		
						<i>Usti</i>		<i>A. calidoustus</i>		
								<i>A. pseudoflectus</i>		
						<i>Candidi</i>		<i>A. candidus</i>		
						<i>Clavati</i>		<i>A. clavati</i>		
						<i>Fumigati</i>		<i>A. fumigatus</i>		
								<i>N. pseudofischeri</i>		
								<i>N. fischeri</i>		
								<i>N. fennelliae</i>		
								<i>N. hiratsukae</i>		
						<i>Nidulantes</i>		<i>N. glabra</i>		
								<i>E. nidulans</i>		
								<i>E. varicolor</i>		
								<i>A. versicolor</i>		
						<i>Circumdati</i>		<i>A. sydowii</i>		
								<i>A. sclerotium</i>		
								<i>A. elegans</i>		
						<i>Eurotium</i>		<i>A. westerdijkiae</i>		
								<i>E. amstelodami</i>		
								<i>E. chevalieri</i>		
								<i>E. rubrum</i>		
						<i>Penicillium</i>	<i>Penicillium</i>	<i>Fasciculata</i>	<i>Verrucosa</i>	<i>P. verrucosum</i>
										<i>P. nordicum</i>
										<i>Solita</i>
				<i>Viridicata</i>	<i>P. viridicatum</i>					
					<i>P. aurantiogriseum</i>					
				<i>Camemberti</i>	<i>P. freii</i>					
					<i>P. camemberti</i>					
					<i>P. commune</i>					
					<i>P. palitans</i>					
					<i>P. crustosum</i>					
				<i>Penicillium</i>	<i>Expansa</i>					<i>P. expansum</i>
					<i>Claviforma</i>					<i>P. glandicola</i>
				<i>Roquefortorum</i>						<i>P. roqueforti</i>
										<i>P. carneum</i>
										<i>P. paneum</i>
				<i>Chrysogena</i>		<i>P. chrysogenum</i>				
						<i>P. nalgiovense</i>				
				<i>Brevicompacta</i>		<i>P. brevicompactum</i>				
				<i>Ramosa</i>		<i>P. raistrickii</i>				
<i>Canescentia</i>		<i>P. coralligerum</i>								
<i>Aspergilloides</i>				<i>P. glabrum</i>						
				<i>P. corylophilum</i>						
				<i>P. oxalicum</i>						
				<i>P. citrinum</i>						
				<i>P. janthinellum</i>						
				<i>P. rolfsii</i>						
				<i>P. thomii</i>						
				<i>P. spinulosum</i>						
				<i>P. fellutanum</i>						
				<i>P. saturatus</i>						
<i>P. variotii</i>										
<i>Paecilomyces I</i>				<i>P. brunneum</i>						
<i>Talaromyces</i>				<i>T. flavus</i>						
				<i>P. concavorugulosum</i>						
<i>Eupenicillium</i>				<i>E. pinetorum</i>						

Sordariomycetes	Hypocreales	Nectriaceae	Fusarium	Oxysporum	F. oxysporum	
				Fujikuroi	F. verticillioides	
					F. subglutinans	
					F. temperatum	
					F. thapsinum	
					F. proliferatum	
				Sambucinum	F. graminearum	
					F. sambucinum	
					F. culmorum	
					F. sporotrichoides	
					F. langsethiae	
				Tricinatum	F. avenaceum	
				Incarnatum-equiseti	F. equiseti	
				Solani	F. solani	
		Bionectriaceae	Bionectria	B. aureofulvella		
				B. ochroleuca		
				B. solani		
		Ophiocordycipitaceae	Paecilomyces 2	P. lilacinus		
		Cordycipitaceae	Verticillium 1	V. lecanii		
		Hypocreaceae	Trichoderma	T. aggressivum		
				T. harzianum		
				T. longibrachiatum		
				T. viride		
			Myrothecium	M. cinctum		
		Sordariales	Chaetomiaceae	Chaetomium		
				Humicola		
			Sordariaceae	Gelasinospora	Gelasinospora sp	
		Glomerales	Plectosphaerelleceae	Verticillium 2	V. dahliae	
			Glomerellaceae	Colletotrichum	C. acutatum	
					C. coccodes	
		Xylariales	Incertae sedis	Microdochium	M. nivale	
			Amphisphaeriaceae	Pestalotiopsis	Pestalotiopsis sp	
		Microascales	Ceratocystidaceae	Ceratocystis	C. paradoxa	
			Microasaceae	Scopulariopsis	S. brevicaulis	
					S. fusca	
			Kernia	K. pachypleura		
		Incertae sedis	Apiosporaceae	Papularia	Papularia sp	
		Diaporthales	Cryphonectriaceae	Cryphonectria	C. parasitica	
			Valsaceae	Phomopsis	Phomopsis sp	
		Dothideomycetes	Pleosporales	Pleosporaceae	Alternaria	A. alternata
						A. chartarum
						P. anserina
Didymellaceae	Peyronellaea			P. clade		
	Stagonosporopsis			S. valerianellae		
Dothideales	Dothioraceae		Aureobasidium	A. pullulans		
Capnodiales	Davidiellaceae		Cladosporium	C. sphaerospermum		
				C. ramotenellum		
				C. cladosporioides		
				C. brunhei		
			C. herbarum			
			B. cineræ			
Leotiomyces	Helotiales	Sclerotiniaceae	Botrytis	P. concavum		
		Incertae sedis	Pilidium	G. candidum		
Saccharomycotina	Saccharomycetes	Saccharomycetales	Endomycetaceae	Geotrichum	G. silvicola	
					G. citri-aurantii	
					M. circinelloides	
					M. velutinosus	
Mucoromycotina	Incertae sedis	Mucorales	Mucoraceae	Mucor	M. spinosus	
					M. racemosus	
					M. hiemalis	
					M. mucedo	
					M. fragilis	
					Actinomucor	Ac. elegans
					Rhizopus	R. oryzae
					Umbelopsis	U. isabellina
						U. autotrophica

			<i>Lichtheimiaceae</i>	<i>Lichtheimia</i>				<i>L. corymbifera</i>
				<i>Circinella</i>				<i>C. sydowii</i>
			<i>Syncephalastraceae</i>	<i>Syncephalastrum</i>				<i>S. monosporum</i>
								<i>S. racemosum</i>
			<i>Cunninghamellaceae</i>	<i>Cunninghamella</i>				<i>C. elegans</i>
								<i>C. binariae</i>
								<i>C. blakesleeana</i>
								<i>A. coerulea</i>
								<i>A. repens</i>
			<i>Thamniaceae</i>	<i>Thamnidium</i>				<i>T. elegans</i>
<i>Mortierellomycotina</i>	<i>Incertea sedis</i>	<i>Mortierellales</i>	<i>Mortierellaceae</i>	<i>Mortierella</i>				<i>M. zonata</i>
								<i>M. hyalina</i>

RESUME en français

Les contaminations par les moisissures représentent un problème majeur au sein de l'industrie agroalimentaire, pharmaceutique, cosmétique, et dans le secteur médical. Actuellement, l'identification des champignons filamenteux est basée sur l'analyse des caractéristiques phénotypiques, nécessitant une expertise et pouvant manquer de précision, ou sur les méthodes moléculaires, coûteuses et fastidieuses. Dans ce contexte, l'objectif de cette étude a consisté à développer un protocole simple et standardisé à l'aide de la spectroscopie infrarouge à transformée de Fourier (IRTF) combinée à une méthode d'analyse chimiométrique, proposant une méthode alternative pour l'identification rapide des moisissures. Au total, 498 souches de champignons filamenteux (45 genres et 140 espèces) ont été analysées à l'aide d'un spectromètre IRTF à haut débit. L'analyse discriminante des moindres carrés partiels (PLS -DA), méthode chimiométrique supervisée, a été appliquée à chaque spectre dans les gammes spectrales 3200-2800 et 1800-800 cm^{-1} . Différents modèles de calibration ont été construits à partir de 288 souches, ceci en cascade de la sous-division jusqu'à l'espèce en se basant sur la taxonomie actuelle. La prédiction des spectres en aveugle, obtenus à partir de 105 souches, au niveau du genre et de l'espèce est respectivement de 99,17 % et 92,3 %. La mise en place d'un score de prédiction et d'un seuil a permis de valider 80,22 % des résultats. L'implémentation d'une fonction de standardisation (SF) a permis d'augmenter le pourcentage de spectres bien prédits, acquis sur un autre instrument, de 72,15 % (sans fonction) à 89,13 %, validant la transférabilité de la méthode. Puisqu'une biomasse mycélienne suffisante peut être obtenue après 48h de culture et que la préparation des échantillons implique l'utilisation d'un protocole simple, la spectroscopie IRTF combinée à la PLS-DA apparaît comme une méthode rapide et peu coûteuse, ce qui la rend particulièrement attractive pour l'identification des champignons filamenteux au niveau industriel. Les résultats obtenus placent la spectroscopie IRTF parmi les méthodes analytiques prometteuses et avant-gardistes, possédant un haut pouvoir discriminant et une forte capacité d'identification, en comparaison avec les techniques conventionnelles.

TITRE en anglais

Characterization and identification of filamentous fungi by vibrational spectroscopy

RESUME en anglais

Mold contaminants represent a major problem in various areas such as food and agriculture, pharmaceuticals, cosmetics and health. Currently, molds identification is based either on phenotypic characteristics, requiring an expertise and can lack accuracy, or on molecular methods, which are quite expensive and fastidious. In this context, the objective was to develop a simple and standardized protocol using Fourier transform infrared (FTIR) spectroscopy combined with a chemometric analysis, allowing to implement an alternative method for rapid identification of molds. In total, 498 fungal strains (45 genera and 140 species) were analyzed using a high-throughput FTIR spectrometer. Partial Least Squares Discriminant Analysis (PLS-DA), a supervised chemometrics method, was applied to each spectrum in the spectral ranges 3200-2800 and 1800-800 cm^{-1} for the identification process. Using 288 strains, different calibration models were constructed in cascade and following the current taxonomy, from the subphylum to the species level. Blind prediction of spectra from 105 strains at the genus and species levels was achieved at 99.17 % and 92.3% respectively. The establishment of a prediction score and a threshold permitted to validate 80.22% of the obtained results. The implementation of a standardization function (SF) permitted to increase the percentage of well predicted spectra from strains analyzed using another instrument from 72.15% (without SF) to 89.13% and permitted to verify the transferability of the method. Since sufficient mycelial biomass can be obtained at 48h culture and sample preparation involved a simple protocol, FTIR spectroscopy combined with PLS-DA is a very rapid and cost effective method, which could be particularly attractive for the identification of moulds at the industrial level. The results obtained places FTIR spectroscopy among the avant-garde promising analytical approaches, with high discriminant power and identification capacity, compared to conventional techniques.

DISCIPLINE

Biologie-Biophysique

MOTS-CLES

Champignons filamenteux, Identification, Spectroscopie IRTF à haut débit, Chimiométrie

KEYWORDS

Filamentous fungi, Identification, Highthroughput FTIR spectroscopy, Chemometrics

INTITULE ET ADRESSE DE L'UNITE DE RECHERCHE

MéDIAN-Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, FRE CNRS 3481MEDyC, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 REIMS cedex, France.