



HAL
open science

Etude et application de nouvelles techniques d'imagerie et d'intelligence artificielle pour l'identification bactérienne

Julien Peyroux

► **To cite this version:**

Julien Peyroux. Etude et application de nouvelles techniques d'imagerie et d'intelligence artificielle pour l'identification bactérienne. Médecine humaine et pathologie. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALS026 . tel-03922686

HAL Id: tel-03922686

<https://theses.hal.science/tel-03922686v1>

Submitted on 4 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : BIS - Biotechnologie, instrumentation, signal et imagerie pour la biologie, la médecine et l'environnement

Arrêté ministériel : 25 mai 2016

Présentée par

Julien PEYROUX

Thèse dirigée par **Max MAURIN**, Université Grenoble Alpes
et codirigée par **Ahmed LBATH**, Université Grenoble Alpes
et co-encadrée par **Yvan CASPAR**, Université Grenoble Alpes

préparée au sein des **Laboratoire Translational Innovation in Medicine and Complexity** et **Laboratoire d'informatique de Grenoble**
dans **l'École Doctorale Ingénierie pour la Santé la Cognition et l'Environnement**

**Etude et application de nouvelles techniques
d'imagerie et d'intelligence artificielle pour
l'identification bactérienne**

**Study and application of new imaging techniques
and artificial intelligence for bacterial identification**

Thèse soutenue publiquement le **28 octobre 2022**, devant
le jury composé de :

Madame Catherine Neuwirth

PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER,
Université de Dijon, Présidente du jury

Madame Sophie Jarraud

PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER,
Université de Lyon, Rapporteur

Monsieur Christophe Marquette

DIRECTEUR DE RECHERCHE, Université de Lyon, Rapporteur

Madame Murielle Cornet

PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER,
Université Grenoble Alpes, Examinatrice

Monsieur Said Oulibouch

DOCTEUR EN SCIENCES, Revartis, Membre invité

Monsieur Max MAURIN

PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER,
Université Grenoble Alpes, Directeur de thèse

Monsieur Ahmed Lbath

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes,
Co-directeur de thèse

Monsieur Yvan Caspar

MAITRE DE CONFERENCE DES UNIVERSITES - PRATICIEN
HOSPITALIER, Université Grenoble Alpes, Co-encadrant de thèse

REMERCIEMENTS

Je souhaite tout d'abord remercier les membres du jury pour avoir accepté d'évaluer ce travail. Professeur Sophie Jarraud et Docteur Christophe Marquette, merci d'avoir accepté d'être également rapporteurs. Merci au Professeur Catherine Neuwirth d'avoir accepté de présider le jury. Enfin, je remercie le Professeur Murielle Cornet et Saïd Oualibouch, examinateurs dans le jury. Je remercie également mes co-directeurs de thèse, les Professeurs Ahmed Lbath et Max Maurin pour m'avoir fait confiance sur ce projet. Enfin je tiens à remercier mon co-encadrant Dr Yvan Caspar. Merci pour ton aide sur l'écriture, pour les relectures parfois juste avant de partir en vacances et pour tes conseils.

Je remercie également Ahmed Alsoffi pour avoir mis en place ce projet au travers de la société Biokube Automation. Merci également à l'équipe de Floralis, en particulier Mariana Tsymbrovska, Clémentine Ferrand et Philippe Beliard sans qui la mise en place et le déroulement de ce projet aurait été impossible.

Ce projet ayant donné lieu à une collaboration avec la société Revartis, je remercie aussi Saïd Oualibouch pour sa participation aux comités de suivi de thèse et au jury de soutenance, mais également pour son implication profonde dans le projet de thèse. Je remercie également Nebojsa Davidovic, et Edin Redjovic qui ont été d'une aide précieuse. Merci pour avoir toujours été disponible pour travailler sur les logiciels, corriger les erreurs, faire avec mes changements d'idées de dernière minute et ce bien au-delà des clauses de la collaboration.

Je remercie également le Docteur Eric Faudry, pour avoir accepté de présider mon comité de suivi de thèse et pour des conseils plus que précieux.

Ayant fait ce doctorat au sein du laboratoire de Bactériologie du CHU de Grenoble, je remercie toute l'équipe de bactériologie pour son accueil. Merci à Thomas Girard pour avoir été toujours présent depuis les 10 ans que nous nous connaissons, pour m'avoir encouragé à venir au CHU et pour avoir été un soutien tout au long de cette thèse. Merci à Camille Brunet pour les relectures, les encouragements, les cartes reconnaissances éternelles et les idées de scénarios pour la série TV. Merci à tous les deux, pour ces 4 années passées ensemble, la fine équipe n'a pas besoin de travailler au même endroit pour faire des étincelles. Merci à Léa Pondérand pour

sa bonne humeur. Merci à Isabelle Pelloux pour la colocation du bureau, les discussions sur tous les sujets et surtout pour avoir toujours répondu à mes questions incessantes de néophyte en bactériologie cliniques. Merci à Aurélie Hennebique pour avoir été Barnabé à contre cœur. Merci à Christine et Sandrine pour les connaissances que vous m'avez transmises. Je remercie également Kevin José Sánchez García pour son aide sur la constitution des banques de données. Merci aux jeunes du DAI, pour la bonne humeur, les soirées et l'accueil du « gars de la recherche ». Merci à tous les techniciens du laboratoire pour la collecte des souches, leurs conseils et leur accueil.

Je tiens enfin à remercier mes proches, famille et amis, qui m'ont encouragé et soutenu, m'ont vu disparaître pendant la rédaction de la thèse tout en sachant quand me faire sortir de mon antre.

Merci à toutes les personnes qui ont participé de près ou de loin à ce projet de thèse car il n'aurait pas été réalisable sans vous.

RESUME

Depuis la découverte des bactéries et de leur implication dans de nombreuses pathologies humaines, leur détection et identification est un pilier central du diagnostic et du traitement des infections bactériennes. L'évolution des connaissances médicales a établi que la rapidité et la précision de l'identification bactérienne ont un impact direct sur l'amélioration de la prise en charge médicale des patients, l'accélération de leur guérison et la réduction de la mortalité associée aux infections. De nos jours, la détermination des profils protéomique et/ou moléculaire d'une souche à identifier, a émergé comme gold-standard pour l'identification de micro-organismes, bien que des techniques plus anciennes comme la coloration de Gram, l'utilisation des tests biochimiques ou la croissance sur milieu gélosé complètent toujours l'arsenal diagnostique utilisé.

L'apparition de nouvelles techniques n'est pas le seul point d'évolution de la bactériologie moderne. Le nombre de prélèvements à visée diagnostique reçus dans les laboratoires de bactériologie est en croissance constante, dû en partie au vieillissement de la population mais également à la flambée des résistances bactériennes aux antibiotiques. Les aspects médico-économiques des infections bactériennes poussent en permanence les laboratoires de biologie médicale à chercher à raccourcir le temps d'analyse pour l'identification bactérienne tout en conservant un coût des analyses aussi bas que possible. L'apparition des premiers systèmes automatisés, basés sur l'identification biochimique, dans les années 1990, puis l'utilisation de la spectrométrie de masse en routine une décennie plus tard, ont marqué l'essor de l'automatisation des laboratoires de bactériologie médicale. Actuellement, l'automatisation des techniques de bactériologie tend à devenir prépondérante vis-à-vis des techniques manuelles.

Le besoin de techniques d'identification fiable, peu coûteuse, toujours plus rapide et capable de gérer un volume conséquent d'analyses, fait de la bactériologie un domaine où l'évolution des techniques d'identification est permanente, soit à partir des techniques actuelles, soit par le développement de nouvelles approches innovantes. L'analyse de pattern de diffraction LASER de colonies bactériennes utilise le phénomène de diffraction de Fresnel. Lorsqu'il est dirigé sur une colonie bactérienne, un LASER se retrouve diffracté en un pattern de diffraction.

Il a été démontré que ces patterns de diffraction sont spécifiques de la forme et de la conformation de la colonie bactérienne et donc possiblement de l'espèce bactérienne. Cette approche peut ainsi être applicable à l'identification bactérienne. C'est d'ailleurs l'objectif de la société BioKube Automation qui développe un automate basé sur ce principe et qui a financé ce projet de thèse.

L'objectif de ce travail de thèse était donc d'établir une preuve de concept pour l'utilisation de pattern de diffraction de colonies bactériennes dans l'optique d'obtenir une identification de l'espèce bactérienne ainsi que de développer un automate permettant ces identifications. Ainsi, le développement du prototype de l'automate BioKube a été mené en parallèle du développement de son logiciel et de l'intégration d'un réseau neuronal convolutif permettant la classification des patterns de diffraction en partenariat avec la société Revertis. Ce travail a abouti à l'obtention d'un prototype fonctionnel permettant la détection de colonies, l'acquisition de leur pattern de diffraction ainsi que leur classification.

Pour permettre une identification bactérienne grâce au réseau neuronal convolutif, ce dernier doit être entraîné sur des bases de données. Ainsi, pour cette preuve de concept, deux bases de données ont été développées en incluant 6 espèces bactériennes. Les patterns de diffraction constituant ces bases de données ont été acquis à deux temps différents d'incubation (12 et 16 heures) pour étudier les délais de croissance bactérienne nécessaires à l'obtention d'une identification par cette nouvelle approche. Les modèles de classification développés ont permis une identification des espèces bactériennes avec 95,4 % et 94 % de classifications correctes après respectivement 12 et 16 heures d'incubation. Obtenus à partir d'une base de données restreinte incluant seulement 6 espèces bactériennes, ces résultats établissent une preuve de concept pour cette nouvelle approche d'identification bactérienne mais nécessitent des études complémentaires sur des bases de données contenant un nombre bien plus important d'espèces bactériennes. Ceci dans le but de valider l'utilisation de cette nouvelle technique pour l'identification bactérienne.

ABSTRACT

Since the discovery of bacteria and of their involvement in numerous human pathologies, their detection and identification are at the center of the diagnosis and treatment of bacterial infections. The evolution of medical knowledge has established that the rapidity and accuracy of bacterial identification have a direct impact on patient care with a quicker recovery and a diminution of mortality. Nowadays, the determination of proteomic and/or molecular profiles of a strain has emerged as gold standard for the identification of pathogenic microorganisms, even if older techniques like Gram staining, biochemical assays or culture on agar media are still present in the panel of diagnostic techniques.

The appearance of new techniques is not the single point of evolution of modern bacteriology. The number of samples received for diagnosis in bacteriology laboratories is in constant increase, partly due to the ageing of population but also to the burden of antimicrobial resistance. The economic and medical aspects of bacterial infections are pushing clinical microbiology laboratories to shorten times of analysis for bacterial identification while keeping the cost of the analysis as low as possible. The arrival of the first automated systems, based on biochemical identification in the 1990s, then the use of mass spectrometry a decade later, have marked the rise of automation in bacteriology laboratories. Nowadays, the automatization of bacteriology techniques tends to become predominant in regards of manual techniques.

The need of identification techniques that are reliable, low cost, faster and able to deal with large volume of analysis have made bacteriology a field in which identification techniques evolution is permanent, either from currently used techniques or by the development of new innovative approaches. The analysis of diffraction pattern from bacterial colonies is based on the phenomenon of Fresnel diffraction. When a LASER is directed to a bacterial colony, it is diffracted in a specific diffraction pattern. It has been demonstrated that such diffraction patterns are specific of the shape and conformation of the bacterial colony and thus possibly of the bacterial species. This approach is therefore potentially applicable to bacterial identification. It is the objective of the Biokube Automation company which is developing a device based on this principle and has funded the PhD project presented here.

The aim of the PhD work was to establish the proof of concept of the possible use of diffraction pattern from colonies for bacterial identification and to develop a device able to realize those identification. The development of the BioKube prototype has been done in parallel with the development of a software and the implementation of a convolutional neural network allowing the classification of diffraction patterns in partnership with the company Revertis. This work led to a functional prototype allowing the detection of colonies, the acquisition of their diffraction pattern as well as their classification.

To perform bacterial identification using convolutional neural network, the training of the network on databases is required. Therefore, for this proof of concept, two databases including 6 bacterial species have been built. The diffraction patterns constituting these databases were acquired at two different incubation times (12 and 16 hours) to study the growth delay required for identification by this new approach. Classification models developed from the databases have allowed an identification of bacterial species with 95,4% and 94% of correct classification respectively after 12 and 16 hours of incubation. Established from a restricted database including only 6 different bacterial species, these results give a proof of concept of this new approach of bacterial identification but required complementary studies to validate the use of this new technique for bacterial identification.

TABLE DES MATIERES

Remerciements.....	II
Résumé.....	IV
Abstract.....	VI
Table des matières.....	VIII
Liste des figures et tableaux.....	XIII
Liste des abreviations.....	XIX
Avant propos.....	1
I. Introduction	3
I.A. Techniques actuelles d'identification bactérienne	9
I.A.1. Identification bactérienne après culture des bactéries en milieu liquide ou gélifié	11
I.A.1.a. Types d'échantillons et flux d'analyse.....	11
I.A.1.b. Cytologie.....	11
I.A.1.c. Coloration de Gram.....	12
I.A.1.d. Culture bactérienne	14
I.A.1.e. Techniques d'identification bactérienne sur colonies	18
I.A.2. Identification bactérienne sans étape de culture : techniques moléculaires .	25
I.A.2.a. Hybridation <i>In Situ</i> par Fluorescence (FISH)	26
I.A.2.b. PCR et PCR en temps réel	27
I.A.2.c. PCR multiplexe et approche syndromique.....	31
I.A.2.d. PCR et séquençage.....	32
I.A.2.e. Séquençage de nouvelle génération (NGS).....	34

I.A.3.	Approches innovantes.....	36
I.A.3.a.	Automatisation en bactériologie.....	36
I.A.3.b.	MALDI TOF MS sur échantillon natif	39
I.A.3.c.	PCR digitale	42
I.A.3.d.	Techniques d’amplification moléculaire isothermale	44
I.A.4.	Technologies émergentes.....	47
I.A.4.a.	Analyse des composés volatiles microbiens	47
I.A.4.b.	Approches spectroscopiques et optiques.....	48
I.A.4.c.	Autres approches	50
I.B.	Intelligence artificielle et réseaux neuronaux convolutifs.....	51
I.B.1.	Principe de fonctionnement d’un réseau neuronal.....	51
I.B.2.	Apprentissage supervisé et non supervisé.....	54
I.B.3.	Réseau neuronal convolutif.....	55
I.B.3.a.	Principe.....	55
I.B.3.b.	Entraînement d’un réseau neuronal convolutif	57
I.B.4.	Evaluation de la performance d’un modèle de classification	59
I.B.5.	Application en bactériologie	61
I.C.	Diffraction de Fresnel.....	63
II.	Identification automatisée en non destructive par diffraction LASER (BioKube)...	65
II.A.	Contexte et objectifs de la thèse	66
II.A.1.	Contexte du projet de thèse	67
II.A.2.	Objectifs de la thèse	68
II.B.	Matériel et Méthodes	69
II.B.1.	Prototype BioKube en développement	69
II.B.2.	Architecture du réseau neuronal convolutif.....	72

II.B.3. Souches bactériennes utilisées	73
II.B.4. Identification des espèces du complexe <i>Enterobacter cloacae</i>	75
II.B.5. Préparation des géloses et acquisition des patterns de diffraction LASER des colonies	76
II.B.6. Validation de la fonction de compensation	76
II.B.7. Construction des bases de données et développement des modèles de classification	77
II.B.8. Classification des patterns de diffraction.....	78
II.B.9. Représentation graphique des résultats.....	78
II.C. Résultats.....	79
II.C.1. Analyse du prototype initial	79
II.C.2. Flux de fonctionnement de l'automate BioKube.....	84
II.C.3. Développement logiciel.....	87
II.C.3.a. Programme de calibration	87
II.C.3.b. Programme de compensation de la visée du LASER	90
II.C.3.c. Développement d'une nouvelle interface graphique.....	95
II.C.4. Acquisition des patterns.....	98
II.C.4.a. Choix des temps d'incubation pour l'établissement des bases de données de spectre de diffraction.....	98
II.C.4.b. Critères de sélection des patterns pour la constitution des bases de données	100
II.C.5. Classification à partir de souches de référence.....	103
II.C.5.a. Création des bases de données à 12h et 16h d'incubation à partir de souches de référence	103
II.C.5.b. Détermination du nombre d'epochs pour l'apprentissage du réseau neuronal convolutif	104

II.C.5.c. Performances des modèles de classification établis à partir des bases de données de patterns de souches de référence	106
II.C.5.d. Identification de patterns de souches cliniques par les modèles de classification basés sur des souches de référence	110
II.C.6. Bases de données basées sur souches cliniques.....	118
II.C.6.a. Développement de modèles de classification basés sur 15 souches cliniques par espèce	119
II.C.6.b. Performances des modèles de classification établis à partir des bases de données de patterns de souches cliniques	121
II.C.6.c. Identification de patterns des souches cliniques par les modèles de classification basés sur des souches cliniques	125
II.C.7. Evaluation des modalités d'acquisition et de classification des patterns ..	130
II.C.7.a. Impact du type de gélose sur la classification	130
II.C.7.b. Seuil de classification	132
II.D. Discussion et perspectives	135
II.D.1. Evaluation et adaptation du prototype initial	135
II.D.2. Identification des souches bactériennes	136
II.D.2.a. Performance de classification des modèles basés sur souches cliniques	138
II.D.2.b. Souches de référence et souches cliniques	139
II.D.3. Diffraction LASER et espèces bactériennes	140
II.D.3.a. Causes biologiques des changements de formes de patterns de diffraction	140
II.D.3.b. Sérotypes et sérovars	142
II.D.3.c. Acquisition des patterns.....	144
II.D.4. Utilisation d'un réseau neuronal convolutif pour la classification de pattern de diffraction	147

II.D.5. Perspectives d'optimisation de l'automate et de son logiciel.....	148
III. Conclusion et perspectives.....	151
Références.....	156
Publications et communications	178
ANNEXE	181

LISTE DES FIGURES ET TABLEAUX

Figure 1 : Evolution des techniques d'identification.....	5
Figure 2 : Exemple de protocoles d'identification bactérienne en fonction du contexte clinique ..	8
Figure 3: Principales techniques d'identification : temps d'analyse versus coût.....	10
Figure 4 : Images de colorations de Gram.	13
Figure 5 : Exemples de cultures sur géloses chromogéniques ou sélectives.	17
Figure 6 : Identification bactérienne par spectrométrie de masse MALDI-TOF.....	24
Figure 7: Principe d'amplification en chaînes par polymérase d'après Sreejith, 2018 (67).	29
Figure 8 : Exemple de TLA, Chaîne BD Kiestra™	38
Figure 9: Protocole d'identification rapide MALDI-TOF Sepsityper d'après Ponderand et al, 2020 (38).....	40
Figure 10 : Principe de PCR digitale d'après Quan et al, 2018 (153).....	43
Figure 11 : Principe d'amplification isothermale médiée par boucle (LAMP) d'après Tomita et al, 2008 (199).....	46
Figure 12 : Principe de régression logistique.....	51
Figure 13 : Architecture d'un réseau neuronal	53
Figure 14 : Illustration des paramètres de convolution.....	56
Figure 15 : Exemples de pattern de diffraction issus de la diffraction d'une onde lumineuse au travers d'ouverture de forme circulaire, triangulaire ou rectangulaire	63
Figure 16 : Automate Biokube et station d'analyse	69
Figure 17 : Flux de fonctionnement d'acquisition de patterns	71
Figure 18 : Illustration de la zone de non détection des colonies due au manque d'homogénéité de l'éclairage LED.	79

Figure 19 : Illustration de la rotation des gélules durant leur transfert entre la station LED (A) et la station LASER (B).	80
Figure 20 : Estimation du ratio pixels par millimètres à partir d'une image de papier gradué. ...	81
Figure 21 : Synthèse d'analyse du prototype initial et répartition des contributions à son amélioration.	83
Figure 22 : Processus de détection des colonies	86
Figure 23 : Plan et simulation 3D de la plaque de calibration de l'automate BioKube	88
Figure 24 : Evolution du pattern de diffraction d'une colonie bactérienne en fonction de la position de LASER par rapport au centre de la colonie observée.	91
Figure 25 : Répartition des circularités des patterns de diffraction avant et après compensation de la position du LASER par rapport au centre des colonies observées.	94
Figure 26 : Ancienne interface utilisateur du logiciel BioKube	96
Figure 27 : Fenêtre de gestion des analyses en cours de la nouvelle interface utilisateur	96
Figure 28 : Nouvelle fenêtre d'accueil du logiciel BioKube	97
Figure 29 : Fenêtre de résultats d'analyse de la nouvelle interface utilisateur.....	97
Figure 30 : Evolution de pattern de diffraction au cours de 24 heures d'incubation.....	99
Figure 31 : Images de patterns avec un artefact causé par la présence proche d'une seconde colonie.....	101
Figure 32 : Images de patterns ayant des formes aberrantes.	102
Figure 33 : Images de la classes "aucun pattern détecté"	103
Figure 34: Répartition des pourcentages d'identifications correctes des patterns par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 12 heures d'incubation.	114
Figure 35 : Répartition des pourcentages d'identifications correctes des patterns par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 16h d'incubation.	115

Figure 36 : Exemple d'hétérogénéité de patterns de diffraction pour <i>Enterococcus faecium</i> après 12 heures d'incubation.	116
Figure 37: Exemple d'hétérogénéité de patterns de diffraction pour <i>Proteus vulgaris</i> après 16 heures d'incubation.	116
Figure 38 : Pourcentages d'identifications correctes par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 12h d'incubation.	128
Figure 39 : Répartition des pourcentages d'identifications correctes par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 16h d'incubation.	129
Figure 40 : Patterns de diffraction de colonies sur 3 milieux gélosés différents : BHI, TSA et MHE.	131
Figure 41 : Patterns de diffraction pour les souches de références d' <i>Escherichia coli</i>	143
Figure 42 : Patterns de diffraction de sérovars de <i>Salmonella</i> sp.	144
Figure 43 : Acquisition de la plaque de calibration.	150
Figure 44 : Automate ScanStation 100 (Interscience)	185
Figure 45 : Exemple du délai entre la détection des colonies et leur identification dans la catégorie de couleur correcte pour un dépistage de SARM dans un prélèvement nasal.	188
Figure 46 : Exemple de résultats faux positifs et faux négatifs observés durant l'analyse ScanStation.	190

Tableau 1 : Discrimination de 5 espèces bactériennes par 20 critères biochimiques inclus dans les galeries API 20E. D'après Fiche technique API 20E (biomérieux)	19
Tableau 2 : Liste et origine des différentes souches bactériennes utilisées.	74
Tableau 3 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches de références après 12h d'incubation	105
Tableau 4 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches de références après 16h d'incubation.	105
Tableau 5 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de 5 souches de référence après 12 heures d'incubation.	107
Tableau 6 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns de 5 souches de référence après 12 heures d'incubation.	107
Tableau 7 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de 6 souches de référence après 16 heures d'incubation..	108
Tableau 8 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns de 6 souches de référence après 16 heures d'incubation.	109
Tableau 9 : Matrice de confusion des résultats de classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 5 souches de références après 12 heures d'incubation	111
Tableau 10 : Sensibilités et spécificités d'identification par classe de la classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 5 souches de références après 12 heures d'incubation	111
Tableau 11 : Matrice de confusion des résultats de classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 16 heures d'incubation.	

La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 6 souches de références après 16 heures d'incubation.	112
Tableau 12 : Sensibilités et spécificités d'identification par classe de la classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 16 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 6 souches de références après 16 heures d'incubation	112
Tableau 13 : Identification des souches cliniques appartenant au complexe <i>Enterobacter cloacae</i>	118
Tableau 14 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches cliniques après 12 heures d'incubation.....	120
Tableau 15 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches de cliniques après 16h d'incubation.	120
Tableau 16 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de souches cliniques après 12h d'incubation.....	121
Tableau 17 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns de souches cliniques après 12 heures d'incubation	122
Tableau 18 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de souches cliniques après 16h d'incubation.....	123
Tableau 19 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns après 16h d'incubation	123
Tableau 20 : Matrice de confusion des résultats de classification des patterns de souches cliniques après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 12 heures d'incubation..	125
Tableau 21 : Sensibilités et spécificités d'identification par classe de la classification des patterns de souches cliniques après 12 heures d'incubation. La classification a été réalisée en utilisant le	

modèle de classification basé sur la base de données de souches cliniques après 12 heures d'incubation	126
Tableau 22 : Matrice de confusion des résultats de classification des patterns de souches cliniques après 16 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 16 heures d'incubation..	126
Tableau 23 : Sensibilités et spécificités d'identification par classe de la classification des patterns de souches cliniques après 16 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 16 heures d'incubation	127
Tableau 24 : Proportion d'identifications correctes et erronées ayant un score de pertinence supérieur au seuil fixé.	133
Tableau 25 : Concordances positives et négatives et valeurs prédictives du dépistage de bactéries multirésistantes par l'automate ScanStation sur milieux gélosés chromogéniques CHROMagar™	191
Tableau 26 : Corrélation entre l'observation visuelle et les résultats du ScanStation pour l'analyse de <i>Staphylococcus aureus</i> résistants à la méticilline.	193
Tableau 27 : Corrélation entre l'observation visuelle et les résultats du ScanStation pour l'analyse de bactéries productrices de carbapénémases.	195
Tableau 28 : Corrélation entre l'observation visuelle et les résultats du ScanStation pour l'analyse de bactéries productrices de β -lactamases à spectre étendu.	197
Tableau 29 : Temps moyen de détection et d'identification correcte de la couleur lors des analyses ScanStation pour le dépistage de SARM, de BPC et de BP-BLSE	198
Tableau 30 : Corrélation entre l'observation visuelle et les résultats ScanStation pour les prélèvements rectaux avec des populations mixtes sur milieu mSuperCARBA	204
Tableau 31 : Corrélation entre l'observation visuelle et les résultats ScanStation pour les prélèvements rectaux avec des populations mixtes sur milieu ESBL.....	205

LISTE DES ABREVIATIONS

ADN : Acide désoxyribonucléique

BLSE : β lactamase à spectre étendu

BP-BLSE : Bactéries productrices de β lactamases à spectre étendu

BPC : Bactérie productrices de carbapénémases

dNTP : désoxyriboNucléotide TriPhosphate

dPCR : PCR digitale

IA : Artificial Intelligence (Intelligence artificielle)

MALDI-TOF MS : Matrix Assisted Laser Ionisation Desorption by Time of Flight Mass Spectrometry (spectrométrie de masse par temps de vol après ionisation et désorption laser assistée par matrice)

NPA : Negative Percentage Agreement (Concordance négative)

NPV : Negative Predictive Value (Valeur prédictive négative)

NGS : Next Generation Sequencing (Séquençage de nouvelle génération)

OMS : Organisation Mondiale de la Santé

PCR : Polymerase Chain Reaction (Réaction de polymérase en chaînes)

PPA : Positive Percentage Agreement (Concordance positive)

PPV : Positive Predictive Value (Valeur prédictive positive)

qPCR : PCR quantitative

SARM : *Staphylococcus aureus* Résistant à la Méricilline

ssDNA : single stranded DNA (ADN simple brin)

TLA : Total Lab Automation (Automatisation de laboratoire complète)

UFC : Unité Formant Colonie

AVANT PROPOS

L'identification rapide des bactéries pathogènes est indispensable dans les domaines de la santé et de l'agro-alimentaire. Actuellement, cette identification se fait principalement par spectrométrie de masse MALDI TOF (MALDI TOF MS). Ce procédé très performant permet en quelques minutes, à partir de colonies bactériennes isolées sur gélose, d'obtenir une identification de la plupart des bactéries pathogènes chez les humains et les animaux. Toutefois, le temps d'incubation nécessaire aux cultures bactériennes pour obtenir suffisamment de biomasse pour réaliser cette analyse est de 24h, voire plus pour des bactéries de croissance lente. De plus, certaines espèces ne peuvent pas être différenciées alors que leur pouvoir pathogène est parfois très différent (ex : *E. coli* et *Shigella spp.* ; *Streptococcus pneumoniae* par rapport aux autres espèces du groupe *Streptococcus mitis/oralis*, espèces du complexe *Enterobacter cloacae*...). La MALDI TOF MS ne permet pas non plus d'identifier différents sérotypes bactériens, chose nécessaire en santé humaine pour différencier certaines souches au sein d'espèces bactériennes comme c'est le cas pour les bactéries du genre *Salmonella*.

Lorsqu'un LASER rencontre un objet circulaire d'une taille similaire, il est diffracté selon le principe de la diffraction de Fresnel. Cette diffraction peut être observée par une caméra située derrière l'objet, l'image en résultant est appelée un pattern de diffraction de Fresnel. Il a été observé que les patterns de diffraction de colonies bactériennes étaient spécifiques de l'espèce bactérienne. C'est ce constat qui a amené au principe de l'automate BioKube qui utilise l'analyse de patterns de Fresnel suite à la diffraction d'un LASER sur une colonie bactérienne afin d'obtenir son identification.

C'est dans ce contexte que s'est situé ce projet de thèse. Il a émergé d'une collaboration entre la société BioKube et 2 laboratoires universitaires : le laboratoire de bactériologie - hygiène hospitalière du CHU Grenoble Alpes et le Laboratoire d'Informatique de Grenoble (LIG).

I. Introduction

Depuis la première observation au microscope de bactéries par Antoine van Leeuwenhoek en 1676 et la découverte de leur implication dans les pathologies humaines par Louis Pasteur en 1879, l'identification des bactéries pathogènes est devenue un pilier central du diagnostic et du traitement des patients souffrant d'infection par ces micro-organismes. Les premières méthodes de classification utilisaient les caractéristiques macroscopiques des colonies bactériennes poussant sur des géloses d'agar comme leur forme, leur couleur et même leur odeur. L'évolution des connaissances médicales a par la suite prouvé que la rapidité ainsi que la précision de l'identification bactérienne avaient un impact direct sur le rétablissement des patients infectés. Par la suite, ces méthodes ancestrales d'identification devinrent insuffisantes et obsolètes face à une taxonomie bactérienne de plus en plus complexe (1). Les techniques d'identification furent ainsi appelées à évoluer pour améliorer la précision et la rapidité de l'identification des pathogènes (Figure 1).

Une des premières évolutions majeures en microbiologie fut l'observation microscopique des formes des bactéries après une coloration : la coloration de Gram. Cette coloration au violet de gentiane (ou Crystal violet) et à la fuschine permet, après observation au microscope, de distinguer les bactéries en deux classes suivant l'épaisseur de leur paroi externe : les bactéries à Gram positif et celles à Gram négatif. Par la suite, l'observation microscopique ainsi que l'analyse des profils phénotypiques et biochimiques des bactéries furent utilisés pour l'identification bactérienne (2,3). Pendant des décennies, les techniques de bactériologie étaient réalisées principalement manuellement et requéraient de nombreuses étapes de manipulations. L'arrivée des premiers systèmes automatisés à la fin des années 1990 pour l'identification des bactéries, à l'aide notamment de systèmes d'identification biochimiques miniaturisées (comme les systèmes automatisés BD Phoenix ou Vitek), marqua une autre évolution majeure en bactériologie (4–6). Une dizaine d'année plus tard, l'essor de l'identification bactérienne par spectrométrie de masse réalisant un profil protéique des bactéries par Temps de vol après Ionisation et Désorption Laser Assistée par Matrice (MALDI-TOF) en routine de laboratoire (7,8) changea drastiquement la bactériologie moderne. C'est actuellement la technique principalement utilisée en laboratoire.

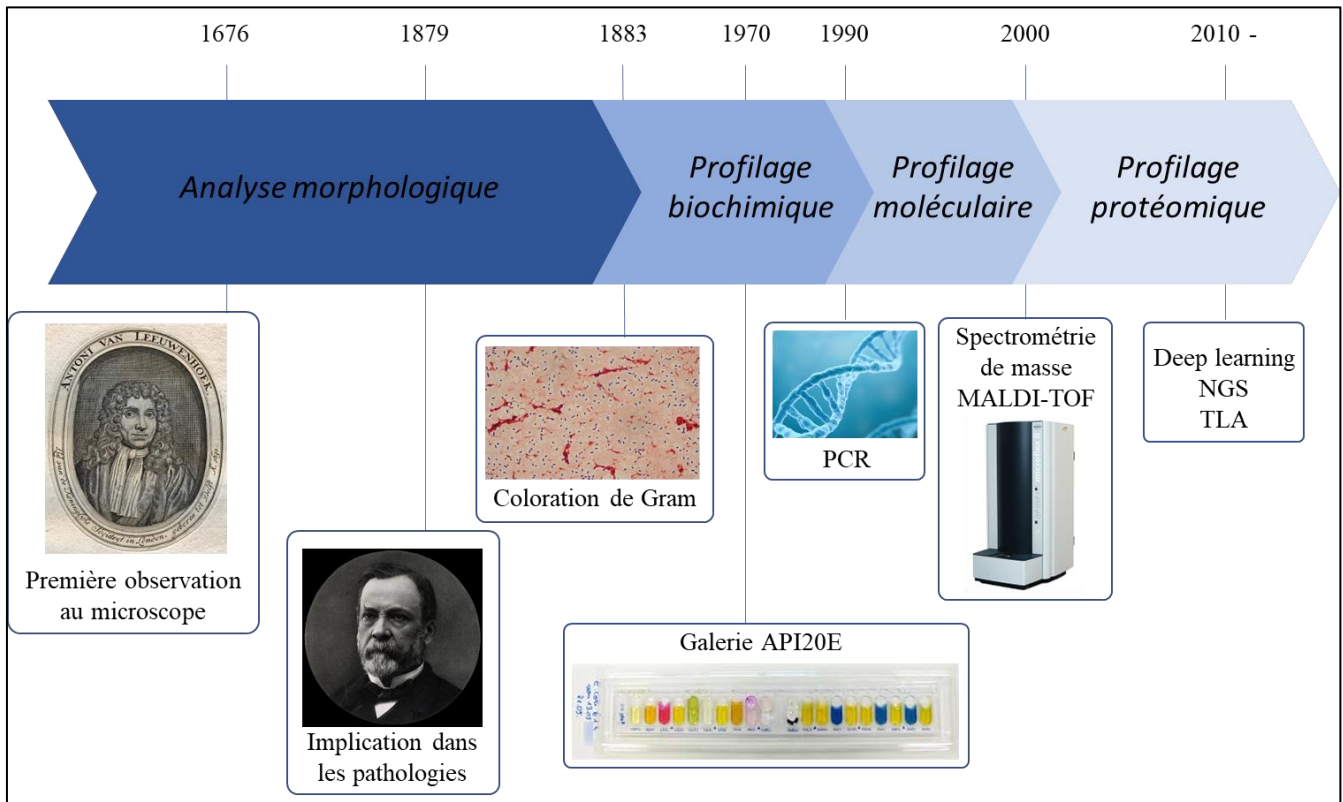


Figure 1 : Evolution des techniques d'identification

De nos jours, identifier les espèces bactériennes est un besoin clinique pour le diagnostic de patients atteints d'infection mais aussi sanitaire et de surveillance agro-alimentaire. La complexité du domaine bactérien fait de la bactériologie une discipline à part entière. En effet, plus de 2000 espèces bactériennes sont actuellement connues pour être pathogènes chez les humains. Cette diversité complique bien entendu les processus d'identification. Bien que la grande majorité des bactéries pathogènes pour les humains puissent être cultivées sur des milieux artificiels gélosés ou liquides à une température de $36 \pm 1 \text{ }^\circ\text{C}$ (9), certaines peuvent nécessiter une température plus basse ($25\text{-}30^\circ\text{C}$ pour *Yersinia spp* par exemple). Un changement dans la composition de l'atmosphère de croissance peut également être nécessaire (en absence d'oxygène ; avec seulement 5% d'oxygène ; ou avec un enrichissement en gaz carbonique). D'autres bactéries pathogènes ne se développent qu'à l'intérieur des cellules de mammifères

qu'elles ont infectées (*Coxiella spp.*) rendant leur culture sur milieu gélosé impossible. Enfin certaines bactéries causant des infections graves voir létales chez les humains sont classées agents pathogènes de classe 3 (*Bacillus anthracis*, *Mycobacterium tuberculosis*). Leur manipulation en laboratoire est règlementée et nécessite des infrastructures spécifiques comme un Laboratoire de Sécurité Biologique de Niveau 3 (LSBN 3).

Chaque espèce bactérienne possède des caractéristiques phénotypiques et/ou génétiques qui leur permettent d'être différenciées mais peuvent également rendre certaines techniques d'identification inefficaces. De plus, une variabilité au sein d'une même espèce peut exister nécessitant de caractériser plus finement les souches d'une espèce pour identifier celles pathogènes pour les humains, notamment par la distinction de différents sérotypes ou la recherche de sécrétion de toxines. Certaines souches bactériennes dites toxigènes ont la plupart du temps une pathogénicité augmentée, il est donc nécessaire de les identifier (*Escherichia coli* producteur de toxine *shiga*-like (10) ou *Clostridium difficile* (11) par exemple). D'autres espèces bactériennes sont sous-divisées en sérotypes qui nécessitent également une analyse approfondie (par exemple le sérotypage de *Salmonella* (12)).

Ainsi les techniques d'identification bactérienne actuelles sont multiples, complémentaires et nécessitent de fortes connaissances des différentes bactéries pour pouvoir être utilisées correctement. De plus, le nombre d'échantillons reçus en laboratoire de biologie médicale est en constante augmentation en lien avec le vieillissement de la population et à la place centrale des analyses de biologie médicale dans l'établissement d'un diagnostic. Dans le même temps, les laboratoires de biologie médicale sont constamment en recherche de diminution du temps nécessaire à l'identification des pathogènes, puisque cette rapidité a été corrélée à un meilleur pronostic chez le patient (13). Cela en essayant de garder au minimum les couts engendrés par l'augmentation de ces analyses, car le critère financier est également un des facteurs limitants principaux de la mise en place d'innovations diagnostiques aujourd'hui à l'hôpital. Dans ce contexte, l'automatisation des laboratoires est un levier permettant de pallier à ces augmentations d'activité. Cette automatisation est en constant essor et a drastiquement changé l'organisation et le flux de travail au sein des laboratoires de microbiologie.

Ainsi, l'identification de bactéries par culture sur milieu gélosé requiert actuellement plusieurs étapes manuelles ainsi que des temps de croissance de 24 à 72 heures pour la plupart des bactéries. Le développement des techniques d'identification bactérienne reste nécessaire pour trouver de nouvelles approches permettant de remplir ces critères de performances, de rapidité d'analyse et financiers. L'identification non destructive de colonies bactériennes par analyse de leur pattern de diffraction LASER peut répondre à ces critères et fait l'objet du travail de thèse ici présenté. Il est cependant important dans un premier temps de détailler en introduction les principales techniques d'identification bactériennes utilisées en routine, celles actuellement en développement ainsi que la généralisation de l'automatisation des laboratoires de microbiologie à visée médicale.

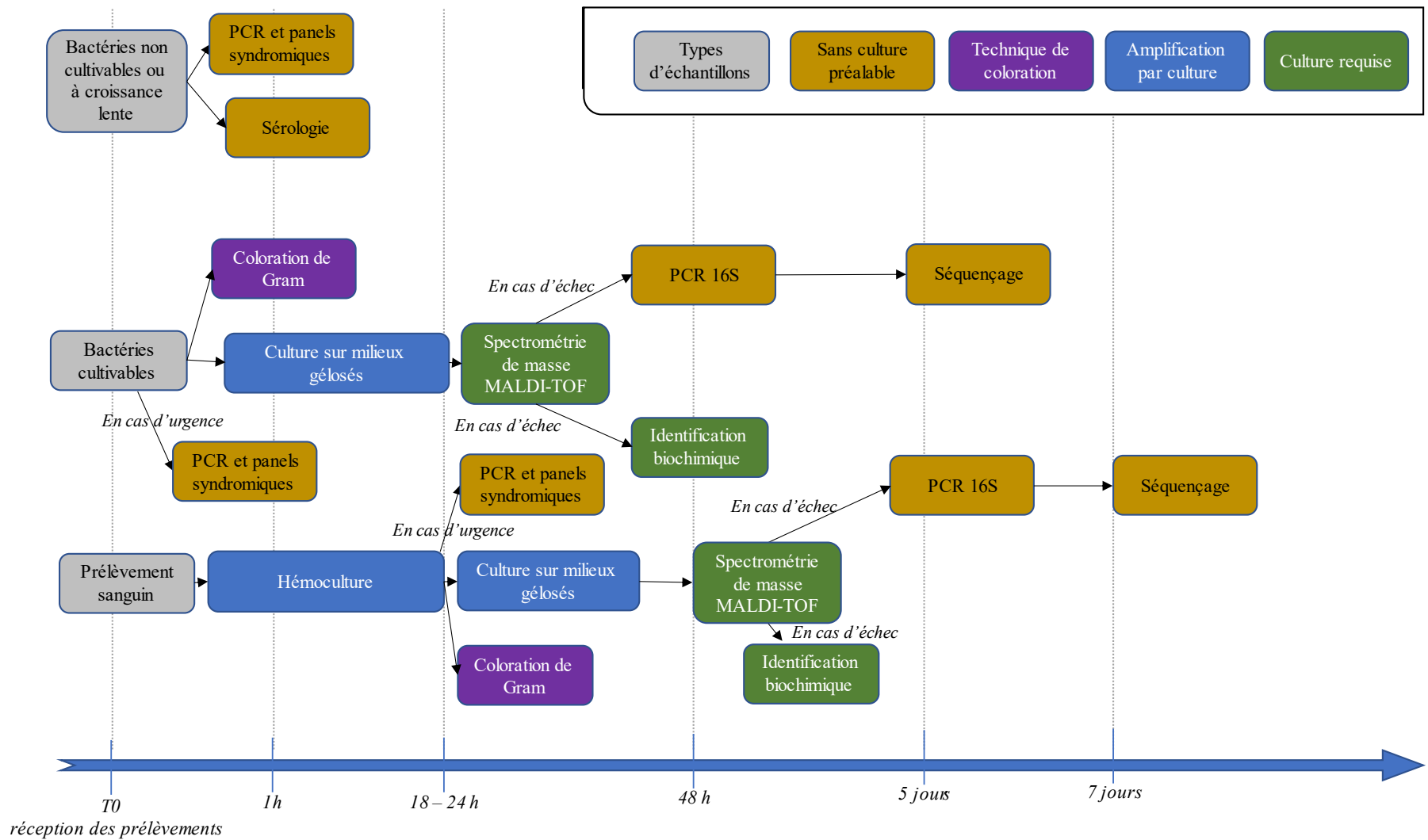


Figure 2 : Exemple de protocoles d'identification bactérienne en fonction du contexte clinique

I.A. Techniques actuelles d'identification bactérienne

Les techniques d'identification actuelles étant nombreuses, les laboratoires de biologie médicale doivent choisir lesquelles utiliser et dans quel contexte les mettre en place. Il existe plusieurs moyens de classer les techniques d'identification bactérienne. Une des premières différences notoires entre les techniques est le fait de nécessiter ou non une première étape de culture avant analyse. Cette étape va fatalement influencer le temps global d'analyse puisque les cultures sur milieux gélosés prennent entre 24 à 72 heures pour la plupart des bactéries. Ces délais de culture peuvent être plus longs pour certaines espèces bactériennes : entre 3 et 5 jours pour les *Legionella spp* (14), et entre 10 et 30 jours pour les mycobactéries à croissance lente (15). En revanche, cette étape de culture a également des avantages qui seront détaillés plus loin. A l'inverse, les techniques ne nécessitant pas d'étape de culture peuvent être appliquées directement sur échantillons cliniques et donc sont souvent plus rapide en termes d'identification.

Une autre façon de distinguer les techniques d'identification est d'évaluer le spectre d'espèces identifiables. En effet, certaines techniques permettent d'identifier un large panel d'espèces bactériennes avec un protocole unique, d'autres doivent avoir leur protocole adapté à l'espèce recherchée et enfin certaines permettent uniquement de différencier quelques espèces entre elles. Il est donc nécessaire en laboratoire de connaître toutes ces techniques ainsi que leurs champ d'application et coût respectif pour pouvoir choisir les techniques d'identification les plus appropriées mais également définir vers quelles évolutions les futurs développements doivent être menés. Les principales techniques d'identification bactériennes utilisées actuellement en laboratoire de biologie médicale ainsi que certaines innovations en cours de développement seront détaillées dans la suite de ce manuscrit.

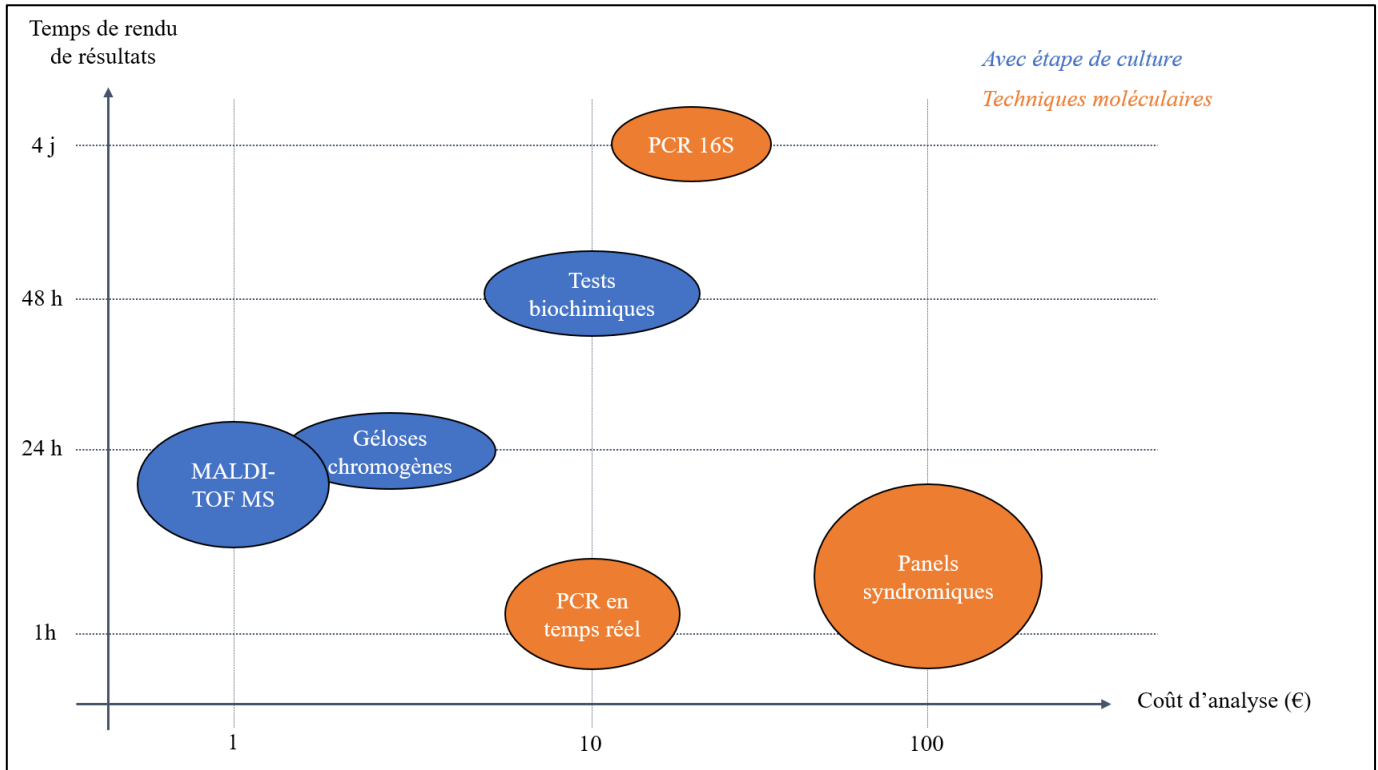


Figure 3: Principales techniques d'identification : temps d'analyse versus coût

I.A.1. Identification bactérienne après culture des bactéries en milieu liquide ou gélosé

I.A.1.a. Types d'échantillons et flux d'analyse

Les échantillons cliniques reçus en laboratoire de bactériologie médicale pour analyse peuvent être répartis en 3 catégories selon leurs origines et la façon dont ils seront traités : les échantillons solides, les échantillons liquides (dont urine, sang, liquide de ponction divers...) et les écouvillonnages (rectaux, nasaux, vaginaux, plaie...). Les techniques de sérologie, qui permettent une identification indirecte du pathogène dans les échantillons sanguins, ne seront pas traitées dans ce travail. Certains des échantillons permettant une identification bactérienne directe passeront par une analyse par coloration de Gram mais presque tous seront ensemencés pour culture sur milieux gélosés et/ou en milieu liquide.

I.A.1.b. Cytologie

Pour les urines et certains autres prélèvements comme les liquides de ponction (liquide pleural, liquide articulaire, liquide péricardique, ...), une numération des leucocytes est généralement pratiquée. Si cette dernière est en faveur d'une suspicion d'infection, elle peut s'accompagner de la réalisation d'une formule leucocytaire. Le compte cellulaire peut être réalisé manuellement (avec un hémocytomètre Malassez par exemple) ou en utilisant un système automatisé comme les cytomètres en flux Sysmex UF-1000i[®] et UF 4000[®] (Sysmex) (16,17) ou le microscope automatisé analyseur de sédiments SediMax[®] (Menarini Diagnostics) (18,19). La charge bactérienne de l'échantillon peut être également évaluée par certains de ces automates. Ces systèmes ont été utilisés pour la réalisation d'analyses cytologiques sur plusieurs types d'échantillons liquides (urines, liquide cébrospinal et liquide articulaire) mais jusqu'ici seulement sur les urines en terme de quantification bactérienne (20–22). Ils ne permettent pas actuellement l'identification des bactéries.

I.A.1.c. Coloration de Gram

La coloration de Gram est la principale technique de coloration utilisée dans les laboratoires de bactériologie depuis sa description en 1883 par le médecin danois Hans Christian Gram (23). Bien que la sensibilité analytique de cette technique soit limitée (avec une limite de détection autour de 10^4 - 10^5 UFC/mL (24,25)), elle reste un test fréquemment utilisé car peu coûteuse, rapide et utile pour rendre un premier résultat rapide au clinicien. Premièrement, la coloration de Gram peut donner, en quelques minutes, une première idée du pathogène impliqué (coque, bacille, Gram positif ou négatif, Figure 4), permettant ainsi au clinicien de décider d'une thérapie antibiotique empirique appropriée ou de réadapter le traitement antibiotique si celui-ci a déjà débuté. Deuxièmement, cette technique permet aussi au laboratoire d'évaluer la qualité de certains prélèvements et d'ajouter des milieux de cultures spécifiques à certaines bactéries en cas de besoin.

Concernant l'évaluation de la qualité du prélèvement, à titre d'exemple, un fort nombre des cellules épithéliales buccales dans une expectoration est synonyme d'une contamination de l'échantillon par de la flore salivaire et entraîne le refus du prélèvement pour non-conformité associé à une demande de re-prélèvement du patient. D'autre part, l'observation microscopique permet de choisir les milieux de culture les plus adaptés si ceux-ci ne sont pas ensemencés de façon standard sur le type d'échantillon concerné : par exemple, l'observation de bactéries à Gram positif pouvant faire suspecter la présence de *Corynebacterium urealyticum* sur un prélèvement urinaire pourra conduire à l'ensemencement d'un milieu gélosé au sang pour permettre la potentiel détection de ce pathogène. Dans certains rares cas, la coloration de Gram peut permettre une identification présomptive de la bactérie ou affirmer un diagnostic, tels que l'identification d'*Helicobacter pylori* dans les biopsies gastroduodénales, ou le diagnostic d'angine de Vincent (présence combinée de *Spirochetes spp* et de *Fusobacterium spp*) ou de vaginoses bactériennes (absence de Lactobacilles avec prolifération de bactéries de type *Gardnerella*, *Bacteroides* et/ou *Mobiluncus spp*). Dans la majorité des cas, l'observation au microscope après coloration de Gram permet uniquement d'obtenir des informations partielles sur le type de bactérie impliquée et doit être suivie d'analyses supplémentaires pour obtenir l'identification des pathogènes présents dans le prélèvement.

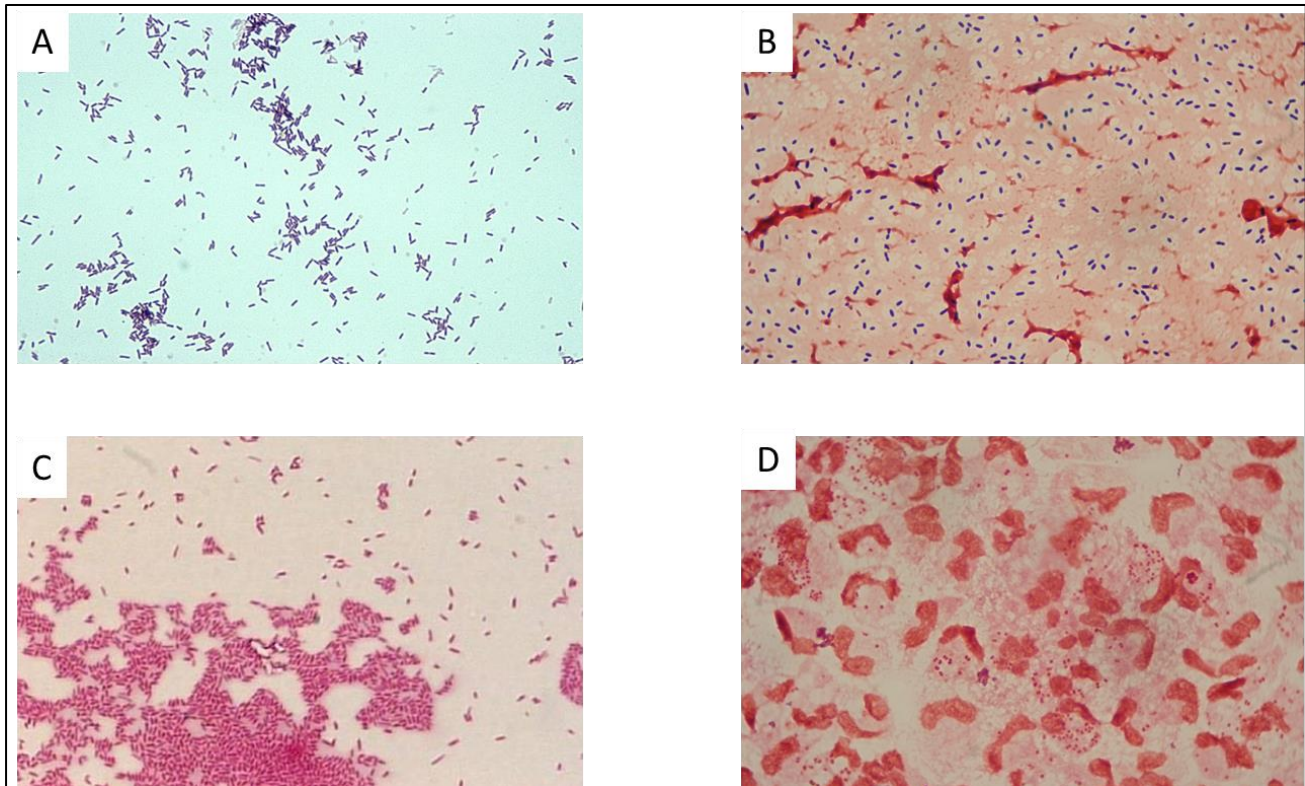


Figure 4 : Images de colorations de Gram. A : Bacilles Gram positifs. B : Coques Gram Positifs. C : Bacille Gram négatifs. D : Coques Gram négatifs.

Il existe d'autres méthodes de coloration disponible mais aucune ne permet d'obtenir une identification précise des pathogènes. La coloration de May-Grünwald-Giemsa (MGG) permet une meilleure différenciation des polynucléaires, lymphocytes et monocytes (tous étant des marqueurs biologiques d'infection) en comparaison avec la coloration de Gram et peut être utilisée pour la réalisation de formule leucocytaire sur des prélèvements liquides tels que les liquides pleuraux, synoviaux ou d'ascite. Les mycobactéries, quant à elles, ne sont pas détectable avec une coloration de Gram alors qu'elles sont révélées en tant que bacille acide-alcollo résistant par les colorations de Ziehl-Neelsen, à l'auramine ou de Kinyoun (26,27). Des solutions automatisées pour ces colorations existent (28,29) et sont commercialisées pour une utilisation en routine.

I.A.1.d. Culture bactérienne

L'isolement de bactérie sur milieu de culture est actuellement la méthode la moins coûteuse pour rapidement diagnostiquer des infections causées par des bactéries non fastidieuses. On entend par bactéries fastidieuses des espèces bactériennes nécessitant des suppléments nutritionnels et des conditions de croissance particulières (besoin de cystéine pour la culture de *Francisella spp* par exemple). De plus, ces bactéries ne croissent pas, très difficilement et/ou très lentement sur des milieux de culture classiquement utilisés en routine (3 à 5 jours pour *Francisella spp*, jusqu'à 14 jours pour *Brucella spp* et 30 jours pour *Leptospira spp* (30)).

Dans le cas de bactéries non fastidieuses, des colonies bactériennes sont généralement obtenues en 24 à 72 heures pour la plupart des bactéries pathogènes chez les humains. L'ensemencement de prélèvements sur différents milieux de culture sélectifs peut également permettre l'isolement des différentes souches bactériennes dans le cas d'infection polymicrobienne. Cette étape de culture permet une amplification de la biomasse bactérienne rendant possible la réalisation de tests complémentaires comme l'identification de la bactérie et l'évaluation de sa sensibilité aux antibiotiques. Les conditions de culture peuvent être adaptées suivant le contexte clinique, la nature du prélèvement ainsi que suivant les résultats de la coloration de Gram. Ces informations préliminaires permettent aux cliniciens et biologistes de faire des suppositions sur le pathogène incriminé.

Les milieux de culture sont incubés sous des atmosphères spécifiques : aérobie (air ambiant), anaérobie (absence d'oxygène), microaérophilie (pression d'oxygène de 5%) ou enrichie avec 5% de CO₂. Les géloses sont généralement incubées à 36 ± 1 °C mais certaines espèces (*Yersinia spp* par exemple) ont une meilleure croissance à des températures plus basses (25 à 30 °C). Il est important de noter qu'actuellement seules les atmosphères aérobie et enrichie avec 5% de CO₂ peuvent être générées dans des incubateurs. Les conditions d'anaérobie et de microaérophilie sont produites dans des jarres ou des sachets spécifiques incubés dans des enceintes chauffées classiques (31). Actuellement, on observe également l'essor d'incubateurs dits « intelligents » équipés notamment de bras robotisés permettant un suivi précis de la

croissance des pathogènes par imagerie des géloses (32). Leurs intérêts et utilisation seront détaillés dans un paragraphe spécifique.

- Cas particuliers des hémocultures

En cas de bactériémie (présence d'agents bactériens circulant dans le sang), les concentrations bactériennes sont en général très faibles (33). Une étape de culture en milieu liquide de l'échantillon sanguin permettant l'amplification de la biomasse bactérienne est nécessaire. Cette étape de culture liquide se fait dans des flacons d'hémocultures aérobie et anaérobie pendant un minimum de 5 jours (jusqu'à 15 jours dans le cas de suspicion d'espèces à croissance lente, ou par exemple en cas d'endocardite). Cette étape est généralement réalisée avec des systèmes automatisés qui permettent la détection automatique des flacons d'hémocultures positifs (cultures ayant une croissance bactérienne). Plusieurs systèmes d'hémocultures sont actuellement commercialisés comme le BD Bactec™ Plus (BD Diagnostics, Sparks, USA), le BacT/Alert® (bioMérieux, Durham, NC) ou le VersaTREK™ (TREK Diagnostics Systems, Cleveland, USA) (34). Les hémocultures positives rejoignent ensuite le flux d'analyse classique du laboratoire, à savoir une analyse par coloration de Gram puis un repiquage des flacons d'hémoculture sur milieux gélosés.

- Milieux gélosés chromogéniques

La culture sur milieu gélosé chromogénique peut être utilisée pour isoler, différencier voire même identifier des micro-organismes spécifiques. Ces milieux contiennent un composé chromogénique qui permet de mettre en évidence une activité enzymatique spécifique. En effet, le clivage du chromogène par son enzyme cible permet la libération d'un chromophore qui, en précipitant, donnera une coloration spécifique à la colonie bactérienne. Il est donc possible en ciblant des activités enzymatiques propres à une espèce bactérienne, ou à un sous-type d'espèce, de mettre en évidence, par coloration, la présence de cette dernière au sein d'un prélèvement. C'est le cas, par exemple, de la différenciation des pathogènes majeurs d'infection urinaires avec l'utilisation de substrat ciblant l'activité β -glucuronidase d'*Escherichia coli*, l'activité β -

glucosidase des entérocoques et du groupe *Klebsiella/Enterobacter/Serratia* et la présence de tryptophane désaminase chez *Proteus*, *Providencia* et *Morganella* (35). Il peut également être cité l'utilisation de 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) pour la détection de l'activité glycosyltransférase de souches de *Clostridium difficile* toxigènes de types A et B (36). Ces milieux chromogènes sont généralement également supplémentés avec des antibiotiques spécifiques pour augmenter leur sélectivité. Deux compagnies sont actuellement leaders dans ce domaine : CHROMagar™ et bioMérieux. Il existe actuellement une trentaine de milieux chromogènes disponibles pour l'analyse microbiologique en laboratoire.

Les limitations principales de cette technologie sont le coût élevé des géloses chromogènes en comparaison des géloses plus classiques (entre 2 et 3,5€ pour une gélose chromogénique contre moins de 1€ pour les milieux de culture gélosés conventionnels) ainsi que des spécificités souvent trop basses pour une identification certaine de l'espèce dans le cadre d'un usage médical (37). La plupart des laboratoires n'utilisent donc pas systématiquement les milieux gélosés chromogènes mais les réservent à certaines analyses principalement pour le dépistage de certaines bactéries au sein d'échantillons cliniques (38). Dans la plupart des laboratoires de microbiologie, les milieux chromogènes sont utilisés pour la détection et l'identification rapide des *Escherichia coli* dans les urines. Ils peuvent aussi permettre l'orientation du diagnostic vers la présence de *Proteus*, d'*Enterococcus* ou d'Entérobactéries appartenant aux genres *Klebsiella*, *Enterobacter* ou *Serratia* sans pour autant donner une identification précise au niveau de l'espèce (Figure 5). D'autres situations peuvent justifier l'utilisation de milieux chromogènes comme le dépistage de *Staphylococcus aureus* résistants à la méticilline (SARM) (Figure 5A), la recherche de *Salmonella* spp, *Yersinia enterocolitica*, *Streptococcus agalactiae* (Streptocoque du groupe B), d'*Enterococcus* résistants à la vancomycine ou d'Entérobactéries productrices de carbapénémases (39). Cependant, les milieux chromogènes ont souvent une spécificité faible. Ces milieux ne sont donc généralement pas utilisés comme techniques de diagnostics (exception faite des *Escherichia coli* dans les urines) mais comme des approches de screening nécessitant une analyse de confirmation comme l'identification par spectrométrie de masse MALDI-TOF.

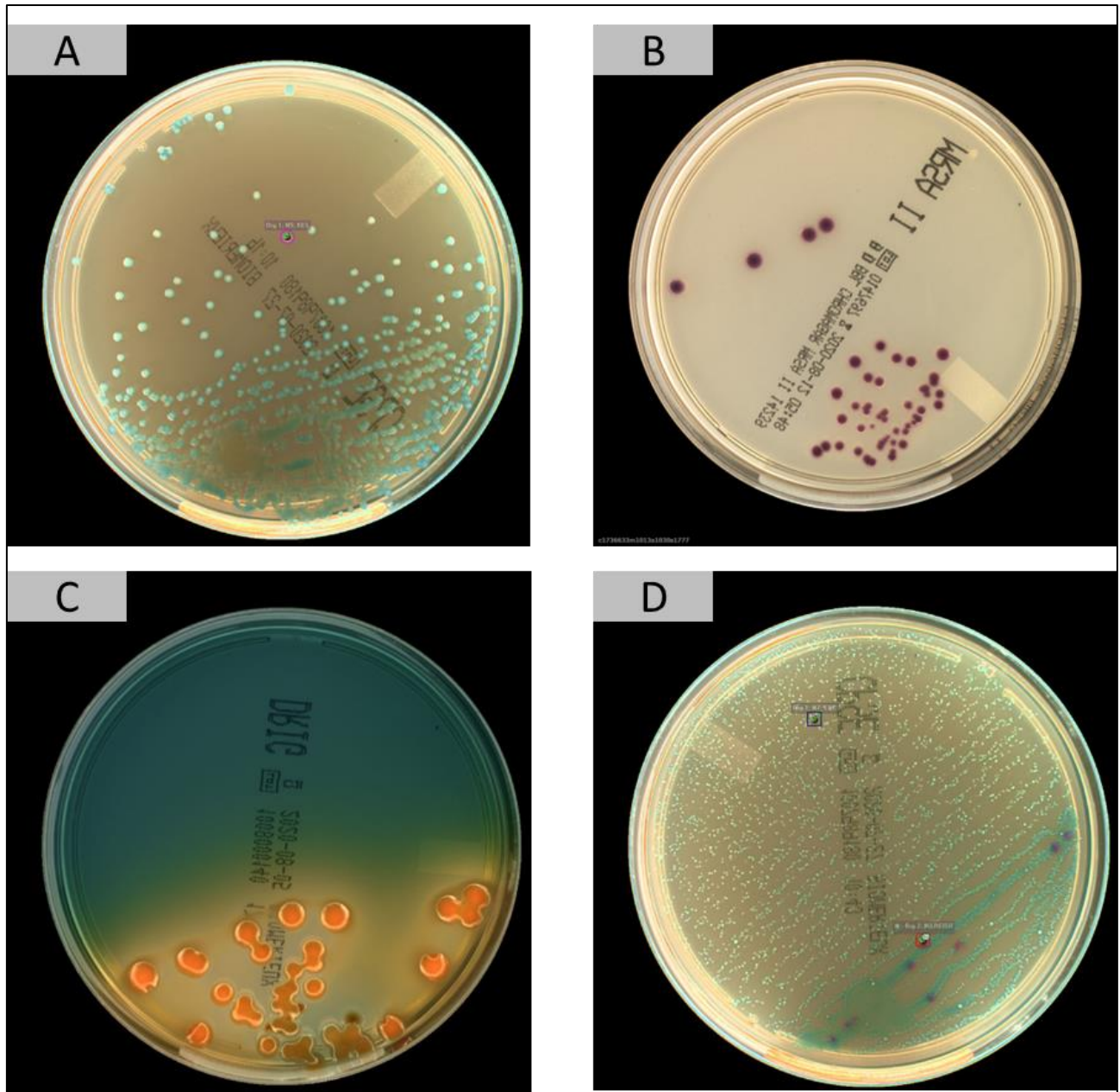


Figure 5 : Exemples de cultures sur géloses chromogéniques ou sélectives. A : *Klebsiella pneumoniae* (bleu clair) sur gélose CPSE (bioMerieux). B : *Staphylococcus aureus* résistant à la méticilline (rose) sur gélose MRSA (CHROMagar). C : *Klebsiella pneumoniae* (orange) et *Proteus mirabilis* (marron) sur gélose Drigalski (bioMerieux). D : *Enterococcus faecalis* (bleu) et *Escherichia coli* (Rose) sur gélose CPSE (bioMerieux)

Il existe également des bouillons d'enrichissement (comme le milieu Schaedler ou bouillon Cœur-Cervelle) qui peuvent être utilisés pour l'isolement spécifique de bactéries fastidieuses ou anaérobies (en milieu semi-gélosé). Ils peuvent être également utilisés pour amplifier la charge bactérienne d'un prélèvement avant ensemencement sur milieu gélosé, soit dans le cas d'échantillons cliniques avec très peu de bactéries, soit pour des échantillons cliniques solides qui ne peuvent pas être ensemencés sur gélose (matériel prosthétique par exemple). Au contraire des milieux gélosés classiques, qui se déshydratent au bout de 5 à 10 jours d'incubation, les bouillons d'enrichissement peuvent être incubés pendant de longues périodes (jusqu'à plusieurs semaines). Un des principaux inconvénients de ce type de milieu est les potentielles interactions délétères entre plusieurs espèces bactériennes dans un échantillon polymicrobien. En effet, il est possible dans ces milieux riches qu'une espèce bactérienne se développe plus rapidement que les autres et cause la disparition des autres espèces présentes initialement dans le prélèvement. Il est donc possible, et nécessaire de savoir, qu'après une subculture du bouillon d'enrichissement sur milieu gélosé, les colonies bactériennes obtenues ne sont pas forcément le reflet de la totalité de la population bactérienne initiale du prélèvement.

I.A.1.e. Techniques d'identification bactérienne sur colonies

- Tests biochimiques

Les tests biochimiques ont été utilisés pendant des décennies pour identifier des souches bactériennes (40). Ces tests sont basés sur l'analyse des propriétés biochimiques des microorganismes liées à la présence d'enzymes spécifiques permettant l'assimilation ou la dégradation de carbohydrates, lipides ou peptides spécifiques. Tout d'abord réalisés manuellement, ces protocoles ont été miniaturisés puis automatisés dans les années 90. On peut notamment citer les galeries API, très largement utilisées à cette époque. Elles consistent à coupler un ensemble de tests biochimiques pour déterminer le profil biochimique d'un pathogène permettant ainsi son identification. Un exemple de discrimination d'espèce bactérienne par galerie API 20E est représenté dans le tableau 1. L'automatisation de cette méthode a mené à la large utilisation de systèmes tels que le BD Phoenix™ Automated Microbiology system (BD Diagnostics) ou le Vitek® 2 System (bioMérieux).

Espèce bactérienne	β-D-galactosidase	Arginine Dihydrolase	Lysine Décarboxylase	Ornithine Décarboxylase	Utilisation du citrate	Production de sulfure d'hydrogène	Uréase	Tryptophane désaminase	Production d'indole	Production d'acétoïne	Gélatinase	Fermentation – oxydation (Glucose)	Fermentation – oxydation (Mannitol)	Fermentation – oxydation (Inositol)	Fermentation – oxydation (Sorbitol)	Fermentation – oxydation (Rhamnose)	Fermentation – oxydation (Saccharose)	Fermentation – oxydation (Melibiose)	Fermentation – oxydation (Amygdaline)	Fermentation – oxydation (Arabinose)
<i>Proteus mirabilis</i>	-	-	-	+	V	+	+	+	-	-	V	+	-	-	-	-	V	-	-	-
<i>Stenotrophomonas maltophilia</i>	+	-	V	-	V	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
<i>Enterobacter cloacae</i>	+	+	-	V	+	-	-	-	-	+	-	+	+	V	+	+	+	+	+	+
<i>Escherichia coli</i>	+	-	+	+	-	-	-	-	+	-	-	+	+	-	+	+	-	+	-	+
<i>Klebsiella pneumoniae ssp pneumoniae</i>	+	-	+	-	+	-	V	-	-	V	-	+	+	+	+	+	+	+	+	+

Tableau 1 : Discrimination de 5 espèces bactériennes par 20 critères biochimiques inclus dans les galeries API 20E. D'après Fiche technique API 20E (bioMérieux) (41). + : réaction positive, - : réaction négative, V : résultat variable

Ces deux systèmes se basent sur des tests biochimiques miniaturisés regroupés sous la forme de différents panels permettant l'identification d'espèces bactériennes proches morphologiquement (ex : panels Gram négatifs, Gram positifs, anaérobies, *Neisseria/Haemophilus*, *Streptococcus*, etc...). Chaque carte contient un ensemble d'une quarantaine de test biochimiques (42). Avec l'ensemble des panels, ces systèmes permettent l'identification d'approximativement 2000 phénotypes de microorganismes avec une précision allant jusqu'à 93 % d'identification correcte (43) après un temps d'incubation de 8 à 16 heures (après l'étape d'isolement). Depuis que la classification bactérienne a évolué vers des méthodes basées sur des analyses moléculaires, beaucoup de nouvelles espèces bactériennes ont été caractérisées. Les techniques d'identification basées sur les tests biochimiques sont alors devenues obsolètes à cause d'un manque de discrimination de ces nouvelles espèces. Au niveau des analyses de routine en laboratoire, elles ont été remplacées par la Spectrométrie de Masse MALDI-TOF (MALDI-TOF MS) présentée plus loin. Depuis, dans les pays développés, les automates BD Phoenix™ (BD Diagnostic) et Vitek® 2 (bioMérieux) sont maintenant principalement utilisés pour l'analyse de la sensibilité aux antibiotiques des souches pathogènes. Cependant, certains tests phénotypiques restent utilisés dans quelques situations, principalement quand l'identification par MALDI-TOF MS est prise en défaut. Par exemple, le test à l'indole est utilisé en complément du MALDI-TOF MS pour différencier les espèces du genre *Proteus* (*vulgaris*, *penneri* et *hauseri*). Le test à l'optochine est utilisé pour faire la distinction entre *Streptococcus pneumoniae* et les autres espèces du groupe *Streptococcus mitis/oralis*. Le test d'hydrolyse de l'hippurate permet quant à lui de différencier *Streptococcus agalactiae* des autres streptocoques β -hémolytique. Enfin, les galeries API 20E® (bioMérieux) sont encore utilisées pour distinguer des souches d'*Escherichia coli* (lactose négatives) des souches du genre *Shigella* ou encore différencier les sérotypes de *Salmonella enterica* typhoïdique des non typhoïdiques. D'autres tests comme ceux de la catalase (*Staphylococcus/Streptococcus*), l'oxidase (*Moraxella*, *Campylobacter*, *Pasteurella spp*), la coagulase (*Staphylococcus* à coagulase négatifs vs *Staphylococcus aureus*) ou les galeries API NH® (pour *Neisseria* et *Haemophilus spp*) sont également encore utilisés ponctuellement (2).

- Spectrométrie de masse MALDI TOF

La spectrométrie de masse a été proposée comme outil d'identification bactérienne dès 1975 (44). Cependant, c'est seulement dans les années 1990 avec le développement de l'ionisation-désorption douce et l'utilisation de matrice organique que la technique a atteint la précision requise pour l'analyse des macromolécules organiques (45). Les évolutions en automatisation et informatique dans les années 2000, avec notamment l'essor de larges bases de données validées, ont permis à la Spectrométrie de Masse MALDI-TOF (Matrix-Assisted Laser Desorption Ionisation Time Of Flight) d'être utilisée comme méthodologie d'identification bactérienne (46–49). Le développement d'un protocole universel de préparation des échantillons avec un prétraitement utilisant une matrice d'acide α -Cyano-4-HydroxyCinnamique (CHCA), les larges bases de données d'identification ainsi que la rapidité de l'identification ont permis l'utilisation du MALDI-TOF MS en tant qu'outil universel et haut-débit pour l'identification bactérienne de routine (7).

La MALDI TOF MS est maintenant largement utilisée dans les laboratoires de microbiologie modernes et est considérée comme technique de référence pour l'identification bactérienne. Cette dernière se fait à partir de colonies bactériennes en quelques minutes avec un protocole unique pouvant identifier plus de 2000 espèces et genres bactériens. Après le dépôt de la colonie à analyser sur la plaque MALDI-TOF, l'ajout de la matrice permet une cristallisation avec les composés organiques de l'analyte. Après insertion dans l'automate, un tir avec un LASER permet un transfert d'énergie aux analytes et leur ionisation positive. Ces ions sont alors vaporisés, accélérés par un champ magnétique et envoyés au travers du tube de vol. L'ionisation étant faite de manière homogène, le nombre de charge de chaque analyte est proportionnel à sa taille (critère masse-charge m/Z en spectrométrie de masse). Le tube de vol étant sous vide, la vitesse de traversée est uniquement influencée par l'accélération initiale créée par le champ magnétique. Plus une molécule aura une grosse taille, plus elle se déplacera lentement dans le tube de vol. Un détecteur d'ions, situé en bout du tube de vol, détectera le passage de chaque ion permettant ainsi de créer un spectre où chaque pic représente la quantité d'ions détectés en fonction de leur rapport masse/charge. Ce spectre représente la composition moléculaire de l'échantillon analysé. Dans le cas de bactéries, certains composants moléculaires (particulièrement protéique pour l'identification bactérienne) sont spécifiques du genre et/ou de

l'espèce bactérienne. La comparaison des spectres à une base de données permet ainsi la détection de pics spécifiques et l'identification de l'espèce bactérienne de la colonie analysée (Figure 6). Bien que la technologie en elle-même soit assez complexe, la préparation de l'échantillon et son analyse ne requièrent qu'un logiciel expert et peu de formation. Un technicien de laboratoire est donc capable en une heure de formation d'utiliser cette technique d'identification de manière autonome (8). Des colonies bactériennes entières peuvent être déposées directement sur une plaque MALDI pour être identifiées en très peu de temps de manipulation (moins de deux minutes par dépôt)(50). Le procédé entier prend approximativement 6 minutes (51) et une plaque MALDI peut contenir 95 échantillons différents qui seront analysés consécutivement. Les colonies peuvent, si besoin, être prétraitées pour augmenter l'extraction protéique et ainsi améliorer l'identification MALDI TOF MS. Par exemple, un traitement des colonies avec de l'acide formique directement sur la plaque MALDI permet d'augmenter les scores d'identification au niveau de l'espèce bactérienne de plus de 20% pour les bactéries à Gram positifs (52,53). Plusieurs études ont fait état d'une précision d'identification supérieure à 95 % au niveau de l'espèce bactérienne pour la MALDI TOF MS (50–52,54). De plus, l'identification par MALDI TOF MS est peu coûteuse (moins d'un euro par analyse) avec une réduction de coûts d'identification évaluée à 75% par rapport à une identification phénotypique (50).

Ces différents avantages de la MALDI TOF MS en comparaison des techniques conventionnelles expliquent la très rapide diffusion de cette méthode dans les laboratoires de microbiologie clinique. Cependant, la MALDI TOF MS possède des inconvénients. Tout d'abord, les bactéries ont besoin d'être mise en culture pour obtenir des colonies isolées et suffisamment de biomasse pour permettre leur dépôt sur la plaque MALDI puis leur identification. Une quantité de 10^4 UFC est requise pour une identification précise (8). L'obtention d'une biomasse suffisante dépend de la vitesse de croissance de chaque espèce bactérienne et est classiquement de 18 à 24 heures. Des protocoles plus rapides ont été développés pour l'identification des bactéries présentes dans les flacons d'hémoculture positifs (55). Par ailleurs, la spectrométrie de masse MALDI TOF MS n'est pas capable d'identifier un mélange polymicrobien et nécessite donc l'isolement des pathogènes avant analyse. L'identification directe de bactéries dans des échantillons cliniques complexes n'est donc pas applicable actuellement en analyse de routine (56). La MALDI TOF MS manque également de

spécificité pour différencier certaines espèces ou genres bactériens avec des profils spectraux proches. Par exemple, cette méthode ne peut faire la différence entre *Escherichia coli* et *Shigella spp* (57), entre les espèces du complexe *Enterobacter cloacae* (58,59) ou celles que complexe *Burkholderia cepacia* (60). La différenciation entre *Streptococcus pneumoniae* et les espèces du groupe *Streptococcus mitis/oralis* ou entre les espèces *Proteus vulgaris*, *penneri* et *hauseri* est également impossible. Enfin, il est important de noter les forts coûts d'installation et d'entretien des automates MALDI TOF MS, bien que les faibles coûts d'analyse, de consommables, et de temps techniques en font une technique rentable sur le moyen terme. Enfin, la spectrométrie de masse MALDI-TOF est une technique dite destructive. La colonie analysée est détruite et ne peut pas être réutilisée pour d'autres tests d'identification bactérienne ou de sensibilité aux antibiotiques. Il est donc nécessaire de faire ces tests supplémentaires sur des colonies jugées *de visu* comme identiques de la gélose ce qui peut induire un risque d'erreur dans l'analyse de l'échantillon.

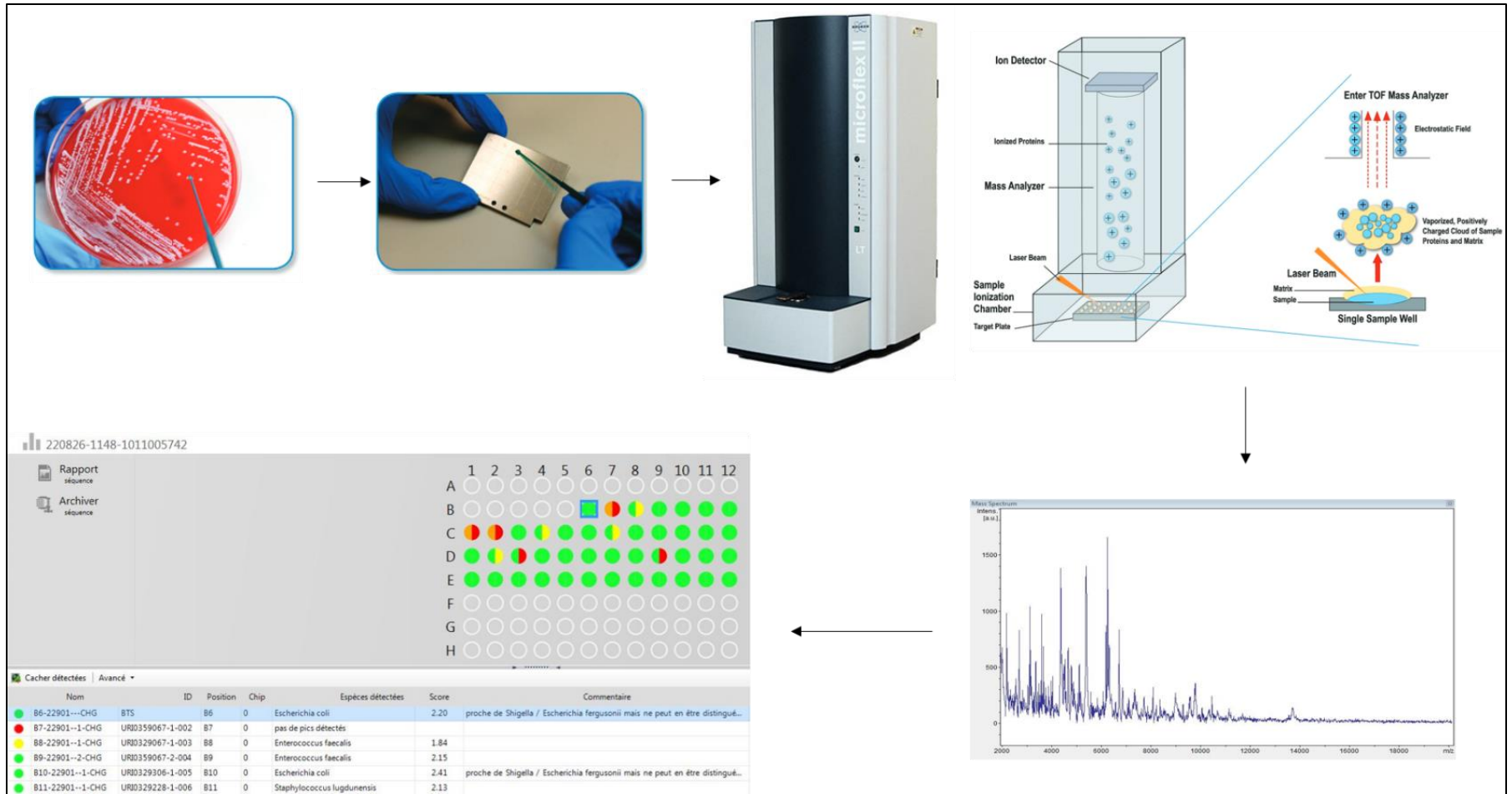


Figure 6 : Identification bactérienne par spectrométrie de masse MALDI-TOF

I.A.2. Identification bactérienne sans étape de culture : techniques moléculaires

Les techniques d'identification moléculaires sont basées sur l'analyse des différences génétiques entre espèces bactériennes. Ces approches ont des avantages et inconvénients communs à la grande majorité des techniques détaillées ci-dessous. Les techniques moléculaires sont généralement plus discriminantes que la technique de spectrométrie de masse MALDI-TOF. Elles permettent notamment de définir une sous-espèce ou un génotype (identification d'espèces au sein du complexe *Enterobacter cloacae* par exemple (61)). Elles permettent une identification directe et rapide des microorganismes à partir des échantillons cliniques car elles ne nécessitent pas d'amplification de la biomasse bactérienne. Ce court délai d'identification est d'un intérêt majeur dans les situations d'urgences comme les septicémies ou les méningites bactériennes. De plus, ne nécessitant pas de culture, les approches moléculaires sont utilisables pour la détection rapide de bactéries non cultivables ou fastidieuses. Enfin, grâce à la grande stabilité de l'ADN, il est possible de détecter la présence de microorganismes jusqu'à deux semaines après qu'un traitement antibiotique ait été commencé, alors que les cultures sont généralement négative à ce moment (62).

Cependant, plusieurs inconvénients sont partagés par les techniques moléculaires. De par leur besoin de cibler une séquence génétique spécifique, ces approches ne peuvent cibler qu'une seule espèce bactérienne en même temps, quelques-unes tout au plus dans le cas d'un multiplexage de la technique (63). Ce spectre étroit d'identification pose plusieurs problèmes. Tout d'abord, l'utilisation de ces approches requiert la suspicion d'une espèce bactérienne en particulier. Si cette suspicion est fautive, le résultat de l'analyse ne donnera aucune information complémentaire. Les techniques moléculaires répondent donc plus à la question : « Est-ce que cette espèce bactérienne est présente dans l'échantillon ? », à la différence du MALDI TOF MS qui répond plus à la question : « Quelle est la bactérie présente dans l'échantillon ? » grâce à son protocole universel. De par les consommables qu'elles nécessitent (amorces d'ADN, sondes...), l'un des inconvénients majeurs des techniques moléculaires est leur coût bien plus important que celui d'une analyse de spectrométrie de masse MALDI-TOF (environ 40 € pour une PCR contre moins de 1€ pour une analyse MADLI-TOF). Ensuite, la présence, dans d'autres microorganismes (du même genre bactérien que la cible par exemple), de séquences génétiques

proches et semblables à la cible de la technique peut mener à des amplifications croisées et des résultats faussement positifs. Enfin, les techniques moléculaires utilisées en diagnostic microbiologique ne donnent pas d'informations quant à la viabilité des bactéries identifiées et peuvent donc détecter la présence d'ADN résiduel en l'absence d'infection active. Par exemple, de l'ADN bactérien peut être détecté sur des valves cardiaques des années après une endocardite traitée (64) rendant ainsi difficile l'interprétation d'un résultat positif en cas d'un nouvel épisode infectieux (65).

I.A.2.a. Hybridation *In Situ* par Fluorescence (FISH)

La technique FISH est basée sur la liaison d'un oligonucléotide couplé avec un fluorochrome à une région d'ADN ou d'ARN complémentaire. Elle a été décrite pour la première fois dans les années 1980 (66). En ciblant une séquence génétique spécifique à une espèce bactérienne, la FISH peut être utilisée pour l'identification bactérienne (67). De nos jours, les techniques FISH utilisent principalement des sondes d'acides nucléo-peptidiques (PNA-FISH). Les complexes d'hybridation ARN-PNA ont une très bonne stabilité thermique permettant d'améliorer la spécificité de la technique (68). De plus, les PNA ont un squelette carboné non chargé et peuvent ainsi, au contraire des sondes conventionnelles, diffuser au travers des membranes bactériennes (69). Les techniques PNA-FISH ont un temps d'analyse court (60 à 90 minutes) avec très peu de temps de manipulation (5 à 15 minutes) ce qui les rend particulièrement intéressantes dans le contexte d'urgence médicale où la rapidité de diagnostic a un impact sur le succès des traitements (70). La PNA-FISH peut être utilisée pour l'identification rapide d'une bactérie directement à partir d'hémocultures pour des infections du sang avec des spécificités supérieures à 99% (71).

Comme d'autres techniques moléculaires, la PNA-FISH peut être utilisée pour identifier des bactéries directement à partir des prélèvements cliniques avec des résultats obtenus en 90 minutes (72). La FISH peut être également appliquée dans le cadre d'infections multi-microbiennes en combinant plusieurs sondes (73). Comparée aux techniques de PCR, la FISH ne

nécessite que peu d'équipements techniques, pas de pièces dédiées et du matériel peu coûteux à acquérir (74). Cependant, la FISH demeure une technique microscopique très peu standardisée et non automatisée. Elle requiert des techniciens formés et expérimentés pour l'interprétation des résultats. L'automatisation des techniques FISH, notamment avec l'utilisation de systèmes micro-fluidiques, est actuellement à l'étude (75). La sensibilité de la FISH est inférieure aux techniques de PCR, particulièrement quand elle est utilisée directement sur échantillons cliniques (76). Cependant, les sensibilités et spécificités de cette technique sont extrêmement dépendantes des sondes utilisées et de la bactérie ciblée (99% de sensibilité pour les *Streptococci* β -hémolytiques (77), 81,5% pour les mycobactéries non tuberculeuses ou atypiques (78)). Ainsi, la FISH est de moins en moins utilisée en routine dans les laboratoires de microbiologie.

Ces dernières années, un système automatisé utilisant la technique FISH a été commercialisé : Accelerate Pheno[®] (Accelerate Diagnostics). Il consiste en une analyse FISH multiplexée permettant la détection et l'identification de 16 espèces bactériennes au maximum en 90 minutes. Les espèces identifiables sont regroupées en panels syndromiques (panels respiratoires, méningo-encéphalite, gastro-intestinal). Le concept de panels syndromiques est abordé plus en détail dans le paragraphe « PCR multiplexe et approche syndromique » (I.A.2.c, page 28). Actuellement, seul un panel pour les infections du sang est commercialisé. Le système Accelerate Pheno[®] donne un résultat avec une sensibilité similaire aux autres techniques disponibles (95,6 % de résultats concordants avec le MALDI-TOF MS) mais avec un temps d'analyse et un coût supérieurs (79,80).

I.A.2.b. PCR et PCR en temps réel

Les tests d'amplification en chaîne par polymérase (PCR) sont utilisés depuis des dizaines d'années pour la détection de microorganismes à la fois à des fins de recherche et de diagnostic. Les cibles sont généralement des séquences d'ADN spécifiques d'une espèce bactérienne ou d'un groupe d'espèces génétiquement proches (comme le complexe *Mycobacterium tuberculosis*). De nos jours, la PCR en temps réel (ou PCR quantitative, qPCR) a

remplacé la PCR conventionnelle pour la plupart des tests diagnostiques. En utilisant des sondes ou intercalants d'ADN fluorescent pour détecter et quantifier en temps réel les amplicons créés, la qPCR est plus rapide et souvent plus sensible que la PCR traditionnelle. Les résultats sont ainsi disponibles directement à la fin des cycles d'amplification sans avoir besoin d'une étape supplémentaire pour l'identification des produits de PCR (81) (par gel d'électrophorèse ou Southern Blot par exemple). Fortement automatisée ces dernières années, la qPCR, dans le cadre de tests diagnostiques, a de très bonnes sensibilités et spécificités avec un temps de résultats court (souvent moins d'une heure) (82). La qPCR permet également l'analyse des courbes de fusion pour distinguer les amplicons d'intérêt des produits non spécifiques ou des dimères d'amorces, ce qui améliore la fiabilité du résultat (83). L'analyse des courbes de fusion peut également permettre dans certains cas de différencier des espèces bactériennes ou d'identifier des mutations particulières impliquées dans des résistances antibiotiques comme par exemple avec le kit LightCycler® SeptiFast (Roche) (84,85). De plus, la qPCR a une grande plage dynamique de linéarité quantitative (6 logarithmes décimaux contre 3 pour la PCR conventionnelle) ce qui est important pour la détection, l'identification et la quantification des pathogènes cliniques (82). Cependant, la qPCR ne peut amplifier qu'un plus petit nombre de paires de bases par rapport à la PCR classique (86).

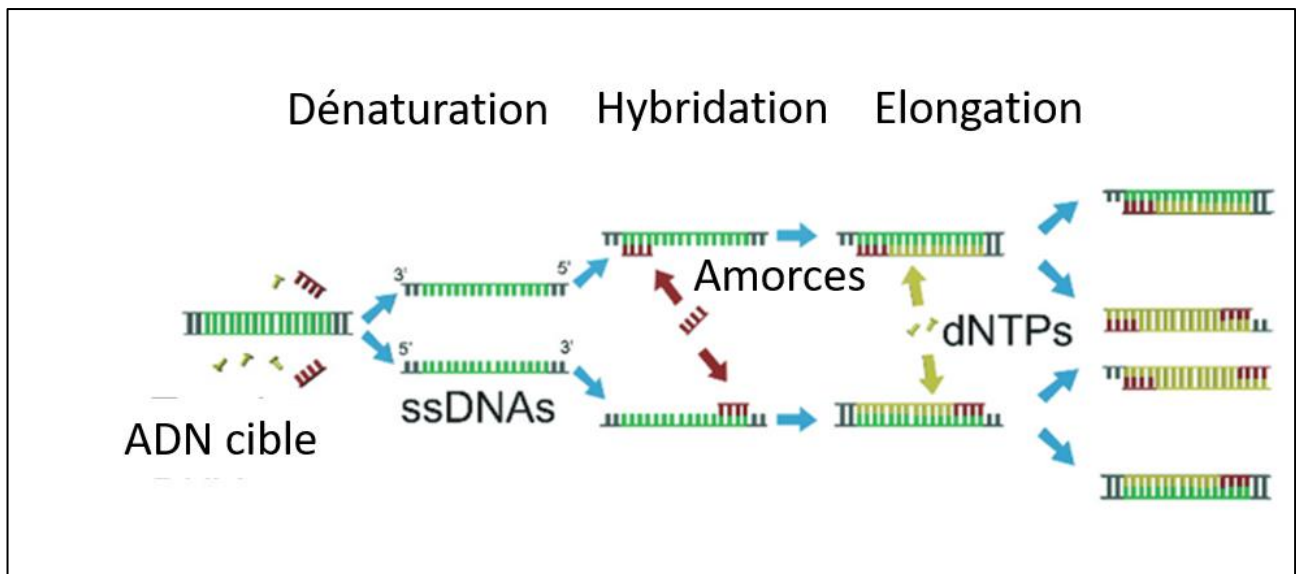


Figure 7: Principe d'amplification en chaînes par polymérase d'après Sreejith, 2018 (67).

ssDNA : ADN simple brin. *dNTPs* : désoxyribonucléotide triphosphate

Les techniques de PCR ont plusieurs inconvénients majeurs. Elles nécessitent d'avoir un *a priori* sur le pathogène en cause car seule une séquence d'ADN spécifique (et donc une seule espèce bactérienne) peut être ciblée à chaque analyse. Des PCR multiplexes ont été développées pour essayer de résoudre en partie ce problème (décrites plus loin). Deuxièmement, ne passant pas par une étape de culture, un antibiogramme ne peut pas être réalisé à la suite d'une PCR. Enfin, les méthodes de PCR sont très sensibles aux contaminations par de l'ADN précédemment amplifié dans la même pièce, même en l'absence d'étapes post-PCR. La qPCR a un risque inférieur de résultats faussement positifs, grâce à la mesure en temps réel des amplicons en comparaison aux PCR conventionnelles (87). Pour éviter ces contaminations, des précautions sont à prendre comme l'utilisation de réactifs ADN-free et la séparation dans des pièces de laboratoires différentes avec accès restreint pour les étapes pré- et post-PCR. L'automatisation (d'abord grâce aux thermocycleurs puis avec des solutions complètement automatisées) des qPCR aide également à réduire les risques de contaminations. Ces systèmes permettent aussi une meilleure reproductibilité (notamment au niveau de la gestion des cycles d'amplification), une meilleure standardisation ainsi qu'un temps d'analyse réduit. Des systèmes tels que le BDmax™

(BD Diagnostics), le GeneXpert® (Cepheid) ou le LightCycler® (Roche) permettent des qPCR semi ou complètement automatisées en utilisant des kits commerciaux. Ces systèmes peuvent, dans certains cas, permettre l'utilisation de protocoles « maison » (82).

Toutefois, la qPCR, reste une méthode d'identification plus coûteuse que les techniques basées sur la culture des bactéries, particulièrement en utilisant des kits commerciaux (88). La limite de détection des techniques de PCR est bien souvent assez faible (10 à 100 bactéries/mL). Cependant, à partir d'un échantillon clinique, cette limite de détection est quelques fois plus élevée en raison de la présence d'inhibiteurs de PCR, de difficulté d'extraction ou d'une dilution de l'ADN bactérien dans de l'ADN humain présent en plus grande quantité. Dans ces cas-là, la sensibilité analytique, en partant de l'échantillon clinique, est quelques fois moins bonne pour les PCR faites directement sur échantillons cliniques en comparaison d'une amplification des bactéries par culture puis d'une identification par MALDI TOF MS par exemple. En effet, dans le cas d'échantillons cliniques avec une faible charge bactérienne, il est possible que l'ADN bactérien ne soit pas en quantité suffisante dans le faible volume analysé en PCR pour être détectable. En revanche, une seule bactérie déposée sur un milieu de culture donnera, 24 heures plus tard, suffisamment de biomasse pour une identification par MALDI TOF MS (89).

En raison de son coût élevé, l'identification bactérienne par qPCR est généralement utilisée uniquement dans des cas spécifiques : pour des diagnostics d'urgences (méningites, septicémies, encéphalites ou suspicion de bioterrorisme) ou dans certaines situations particulières : en cas de suspicion de bactéries fastidieuses ou ne se développant pas sur milieu de culture (*Mycobacterium leprae*, *Mycoplasma genitalium*), en cas d'endocardites avec des hémocultures négatives (la PCR pouvant être faite directement sur prélèvements de sang ou sur valves réséquées) ou si le patient a déjà reçu un traitement antibiotique avant l'analyse microbiologique. Il est important de noter que, de par ces cibles uniques, la qPCR (mais également les PCR conventionnelles) n'est utilisée que lorsqu'une ou quelques espèces bactériennes sont fortement suspectées à partir des contextes cliniques et microbiologiques. Le design et le choix des sondes et amorces mais aussi de l'ADN polymérase ont bien évidemment un impact majeur sur la sensibilité et la spécificité de la technique PCR (90–92).

I.A.2.c. PCR multiplexe et approche syndromique

Pour résoudre le problème du ciblage unique des PCR et qPCR, la méthode d'amplification multiplexée (PCR multiplexe) a été développée (93). Elle consiste en l'amplification et la détection de deux ou plusieurs séquences nucléiques spécifiques en une seule réaction de PCR grâce à la combinaison de jeux spécifiques d'amorces et de sondes. Plus récemment, l'utilisation de technologies microfluidiques a permis de paralléliser plusieurs réactions de PCR dans des cartouches miniaturisées. La PCR multiplexe peut être basée sur la PCR conventionnelle (93) ou sur la qPCR (94). En comparaison des PCR simples, la PCR multiplexe a un coût, un temps d'analyse et un volume d'échantillon nécessaire qui sont inférieurs (par rapport au nombre de PCR simples nécessaire pour le même résultat). Le faible volume d'échantillon requis peut être d'un grand intérêt dans le cas de prélèvements cliniques précieux disponibles seulement en petite quantité (liquide céphalo-rachidien par exemple) (63). La PCR multiplexe a les mêmes avantages que les PCR simples et peut être, dans certains cas, plus sensible que les techniques basées sur la culture pour la détection de pathogènes, comme pour *Yersinia pestis* par exemple (95).

Cependant, les tests de PCR multiplexe sont plus complexes à développer que ceux de PCR simple de par la difficulté à combiner plusieurs jeux d'amorces et de sondes sans réactivité croisée. Ils restent, par ailleurs, moins sensibles de manière générale que les PCR simples (96). Les PCR multiplexes sont utilisées pour l'identification bactérienne depuis les années 90 (97). Cependant, le nombre de cibles au sein d'une même PCR reste limité du fait du nombre maximum de sondes spectralement discriminables disponibles (98). Grâce aux évolutions de la technique et à son automatisation, il est aujourd'hui possible de multiplexer une vingtaine de cibles au sein d'un même test (84,99). Cela en combinant notamment plusieurs cibles par PCR multiplexe, en réalisant plusieurs PCR par test en simultané ou en utilisant d'autres moyens de détection que la fluorescence (comme avec le système ePlex (Genmark) et l'hybridation de l'ADN compétitive et la détection électrochimique). Des solutions automatisées sont commercialisées et disponibles pour l'identification bactérienne en routine de laboratoire comme l'ePlex® (Genmark), le FilmArray® (bioMérieux), le Verigene® (Nanosphere) et le GeneXpert® (Cepheid).

Tous ces systèmes proposent des tests PCR basés sur des panels syndromiques. Ces panels sont adaptés spécialement pour la détection rapide de multiples pathogènes dans un contexte clinique spécifique. Ils permettent de cibler en même temps tous les pathogènes responsables en majorité d'un syndrome clinique. Il existe des panels respiratoires (100,101), gastro-intestinaux (102,103), pour les bactériémies (104–106), pour les infections génitales (107), pour les pathogènes responsables de méningites et encéphalites (108,109) mais également des panels pour la détection d'agents potentiels de bioterrorisme (110). Ces tests de panels syndromiques sont forcément automatisés, miniaturisés avec de la micro-fluidique, avec très peu de temps de manipulation et un temps d'analyse court (entre 1 et 5 heures). Cependant, ils sont également très coûteux (entre 100 et 150 euros par analyse) et peuvent quelques fois manquer d'identifier tous les pathogènes présents dans une infection multiple (même si tous sont présents dans le panel) (111). De plus, bien que le nombre de pathogènes détectables soit conséquent, il est toujours possible d'être face à un pathogène rare qui n'est pas pris en compte dans ces tests et qui restera donc non détecté. Les PCR multiplexe basées sur panels syndromiques sont donc, à cause de leur fort coût, utilisées uniquement pour des situations d'urgence ou quand elles sont les seules techniques utilisables.

I.A.2.d. PCR et séquençage

La PCR 16S ARNr (ou PCR 16s) est une méthode basée sur le fait que le gène codant la sous-unité 16S de l'ARN ribosomal (ARNr) est hautement conservé au travers de tout le domaine bactérien (112). Dès lors, il peut être ciblé et amplifié chez toutes les bactéries avec un jeu d'amorces unique. Au milieu des régions conservées, le gène codant l'ARNr 16S possède des régions variables dont les séquences nucléiques sont propres à chaque espèce bactérienne. On peut donc, à partir de ces séquences, faire de l'identification bactérienne. Pour ce faire, le gène codant l'ARNr 16s est tout d'abord amplifié par PCR. Le produit de PCR est ensuite séquencé pour pouvoir, au sein la séquence nucléique du gène amplifié, identifier les changements nucléotidiques spécifiques menant à l'identification bactérienne par comparaison avec les bases de données génétiques existantes (113). La force de la PCR 16S couplée au séquençage (parfois appelée PCR universelle) est sa relative absence de sélectivité au moment de la PCR permettant de détecter et d'identifier presque toutes les espèces bactériennes (114,115). La PCR 16S fait

partie de méthodes souvent appelées PCR à champ large (broad-range PCR en anglais) qui ciblent des gènes conservés dans tout le domaine bactérien pour donner une identification à l'espèce après séquençage. Parmi ces gènes, les plus souvent utilisés pour les PCR à champ large sont les gènes codant les ARN ribosomiaux, dont celui codant l'ARNr 16S. Les ARNr sont souvent comparés à des horloges moléculaires car ils reflètent par leur mutation les changements évolutifs, sans que ces changements ne soient trop grands à cause du rôle fonctionnel vital des ARNr (116). Le gène codant la sous-unité 23S de l'ARNr est aussi une cible potentielle pour les PCR à champ large mais l'ARNr 16S reste majoritairement utilisé car c'est sur ce gène que les bases de données disponibles sont les plus complètes. Les gènes *tuf*, *hsp65*, *gyrB* et *rpoB* font également partie des gènes pouvant être cibler par PCR à champ large mais n'étant pas conservés dans l'entièreté du domaine bactérien, leur spectre d'espèces bactériennes identifiables est plus restreint (117,118).

Un des avantages de la PCR 16S est que, même si aucune identification ne peut être obtenue, la bactérie inconnue sera quand même positionnée taxonomiquement (89). La PCR 16S est d'ailleurs l'une des premières techniques utilisées lors de l'établissement d'une nouvelle espèce bactérienne, car elle permet de mettre en évidence une espèce s'éloignant génétiquement de toutes celles existantes. Cependant, cette technique n'est pas parfaite. Certaines bactéries peuvent avoir plusieurs copies du gène codant l'ARNr 16S et des changements de paires de bases entre ces copies peuvent rendre l'interprétation complexe à cause de l'incertitude de séquence que cela cause (113). De plus, il est important de noter que certaines espèces proches ont des séquences du gène codant l'ARNr 16S similaires voir identiques et sont donc indifférenciables par PCR 16S. C'est le cas entre autre de *Bacillus cereus* et *Bacillus anthracis*, des espèces du groupe *Streptococcus mitis*, d'*Escherichia coli* et *Shigella spp.*, de *Neisseria meningitidis* et *Neisseria cinereal*, de *Bordetella pertussis*, *Bordetella parapertussis* et *Bordetella bronchiseptica* ainsi que des espèces du complexe *Mycobacterium tuberculosis* (113,115).

La PCR 16S n'est pas une technique rapide d'identification bactérienne puisqu'elle requiert le séquençage du produit de PCR ce qui peut prendre quelques jours (suivant l'équipement du laboratoire ou si le séquençage est externalisé). La PCR 16S est donc généralement utilisée lorsqu'une bactérie fastidieuse est suspectée, si les résultats de la culture

sont négatifs alors qu'il existe une forte suspicion clinique et biologique d'une infection bactérienne ou lorsqu'un traitement antibiotique a été initié avant la réalisation du prélèvement. Enfin, la PCR 16S est également une alternative pour l'identification des souches cliniques isolées pour lesquelles les méthodes d'identification par culture ne peuvent fournir un résultat probant (114,119). Cependant, le manque d'automatisation de la PCR 16S est un frein majeur à son utilisation (119). La PCR et le séquençage peuvent être automatisés séparément mais aucun système automatisé ne combine ces deux étapes. D'autres facteurs limitants de la technique sont un coût élevé, principalement dû aux séquenceurs automatisés (96), un temps d'analyse long, ainsi que la nécessité d'être réalisée dans des laboratoires adaptés (120) avec du matériel complexe (121). Ce problème peut être partiellement résolu en externalisant le séquençage à des compagnies qui reçoivent les produits de PCR et renvoient les séquences d'ADN. La limite de détection de la PCR 16S est de manière générale plus élevée que celle d'autres PCR spécifiques (1000 bactéries/mL). La PCR 16S est également très sensible aux contaminations. En effet, la nature non-sélective inhérente à la technique fait qu'elle est hautement sensible à toute contamination par de l'ADN bactérien (122) amené soit par un réactif soit par contamination d'un automate lors de précédentes analyses. C'est pour cela que des mesures de prévention et de stérilité doivent être prises en laboratoire et que la PCR 16S est recommandée uniquement sur des échantillons cliniques issus de compartiments corporels « stériles ». La sensibilité à la contamination explique en partie les taux de résultats faussement positifs pour la PCR 16S (123,124). Même si encore peu utilisé en routine, il est important de noter que, associée avec des bibliothèques de clones et du séquençage NGS (125), la PCR 16S peut être utilisée dans le cadre d'infections polymicrobienne comme la parodontite (126), la prostatite (127), les infections du canal dentaire (128) ou la kératite ulcéreuse (129).

I.A.2.e. Séquençage de nouvelle génération (NGS)

L'apparition du séquençage à haut-débit, aussi appelé séquençage de nouvelle génération (NGS), a conduit à une baisse du coût « par base » du séquençage génomique et à une augmentation du volume de données extraites (130). Plusieurs séquenceurs sont maintenant

commercialisés et les revues suivantes donnent une excellente analyse et comparaison de ces derniers (131,132). Ces avancées technologiques ainsi que la petite taille du génome bactérien donnent alors la possibilité d'utiliser le séquençage de génome complet (Whole Genome Sequencing ou WGS) pour l'identification bactérienne (133). Le WGS permet, en plus d'identifier l'espèce de la bactérie, d'avoir des informations sur son génotype, sa virulence et ses résistances antibiotiques (134). Cependant, aux vues du coût encore élevé du séquençage et du temps nécessaire pour analyser les données entières d'un génome complet, les NGS ne sont pas encore prêts à remplacer les méthodes d'identification actuelles et sont encore rarement utilisés de nos jours en routine (131).

I.A.3. Approches innovantes

Les techniques actuellement disponibles pour l'identification bactérienne possèdent toutes des inconvénients. Ils peuvent impacter la qualité des résultats (sensibilité ou spécificité trop basses), le délai nécessaire à l'obtention du résultat ou les coûts, rendant ces techniques difficilement utilisables à grande échelle. Pour remédier à ces problèmes, des évolutions de ces techniques sont constamment publiées. Ce chapitre a pour but de détailler des évolutions récentes proposées par la communauté scientifique. Certaines ne sont qu'au stade de preuve de concept, d'autres à un stade plus avancé de validation de procédé. Certaines de ces approches récentes ont commencées à être commercialisées depuis peu et à être utilisées dans les laboratoires de microbiologie.

I.A.3.a. Automatisation en bactériologie

Un changement technologique majeur des dernières décennies en bactériologie est l'automatisation des techniques d'ensemencement et d'analyses. L'augmentation constante du volume de prélèvements à analyser, la nécessité d'accélération des délais de rendu des résultats (ayant un impact direct sur la morbi-mortalité des patients ainsi que l'évolution de leur état de santé (135)) ainsi que le besoin de traçabilité et de surveillance de la qualité des analyses (136) ont fait de l'automatisation d'une technique un critère primordial pour son intégration en routine dans un laboratoire de bactériologie médicale. Cela se traduit par le développement de solutions automatisées pour les techniques d'identification décrites précédemment (137,138) mais également par l'émergence du concept d'automatisation totale de laboratoire (TLA) (139,140). Il consiste en une solution complètement automatisée de la réception du prélèvement jusqu'à l'analyse des résultats (Figure 8) dans laquelle techniciens et biologistes décident des protocoles et techniques à appliquer sans jamais (ou presque) manipuler l'échantillon (141). Ce concept de TLA est déjà présent sur le marché avec des solutions complètement ou partiellement

automatisées telles que les chaînes BD Kiestra™ (BD Diagnostics) et WASPLAB® (Copan) ou l'ensemenceur PreLUD® (I2a Diagnostics).

Disponibles depuis moins de 10 ans, ces chaînes automatisées d'ensemencement sont en train de modifier considérablement les laboratoires de bactériologie qui restaient parmi les laboratoires de biologie médicale avec le plus de tâches manuelles. Les bénéfices de l'automatisation sont nombreux. Elle permet tout d'abord un traitement haut-débit et continu du flux de prélèvement, apportant une diminution du délai global du temps d'analyse (142). L'automatisation permet également une traçabilité et une reproductibilité bien supérieures aux techniques manuelles. Malheureusement, certains inconvénients accompagnent également l'automatisation de techniques de laboratoire. En premier lieu, l'acquisition et la maintenance des tels systèmes créent généralement une augmentation du coût global d'un laboratoire. Ce coût est néanmoins amorti par l'augmentation du nombre de prélèvements traités et du volume d'activité rendant l'acquisition de ces systèmes généralement rentable financièrement. Deuxièmement, les personnels techniques de laboratoire voient souvent leur poste évoluer et doivent s'adapter à ce changement. En effet, ce changement implique des formations supplémentaires en informatique et gestion d'automates avec une diminution des facettes manuelles du travail en laboratoire. Enfin, l'intégration de systèmes automatisés ne peut être réalisée sans la mise en place de systèmes informatiques de laboratoires (SIL) nécessaires à la communication entre automates ainsi qu'à la transmission de résultats. Ces SIL rajoutent également une couche de complexité dans la gestion de laboratoire avec l'obligation de soit engager du personnel compétent en génie informatique soit d'assumer le coût d'une mise en sous-traitance. L'automatisation, bien qu'apparaissant de nos jours comme obligatoire pour les laboratoires, change drastiquement l'environnement et le flux de travail en laboratoire et demande une adaptation en conséquence des personnels et des budgets.



Figure 8 : Exemple de TLA, Chaine BD Kiestra™

Une autre facette de l'automatisation est la miniaturisation et la portabilité des techniques d'identification. La réduction de la taille des dispositifs ainsi que leur capacité à être déplacés permet de réaliser des analyses de façon délocalisée par rapport aux laboratoires centraux notamment en « Point-of-care », c'est-à-dire sur les lieux de consultation des patients. Ces évolutions se basent principalement sur des changements de support d'analyse avec notamment le développement de systèmes de micro fluidique (143) ou de téléphones portables pour l'acquisition et l'analyse d'images (144,145).

I.A.3.b. MALDI TOF MS sur échantillon natif

L'identification par MALDI TOF MS requiert une étape de culture dépendante du temps de croissance nécessaire pour isoler les bactéries et amplifier leur biomasse. Pour aller vers une technique d'identification plus rapide, beaucoup de développements ont été fait pour réduire ou même éviter l'étape de culture. De manière générale, l'identification rapide directement sur échantillons cliniques par MALDI TOF MS n'est pas envisageable du fait de la haute limite de détection de la technique et du manque du pureté des échantillons cliniques (souvent polymicrobien et/ou avec des charges bactériennes faibles) (60). Les scores de l'analyse MALDI TOF MS directement sur prélèvement sont diminués du fait d'un défaut de correspondance avec les bases de données créé par les protéines endogènes de l'échantillon (comme l' α -défensine dans les urines) (146). Enfin, comme c'est également le cas avec les protocoles d'identification MALDI TOF MS classiques, l'identification directe sur prélèvements est incapable d'identifier correctement des infections polymicrobiennes (147). L'identification rapide par MALDI TOF MS est principalement développée pour les hémocultures, car il est critique d'identifier aussi rapidement que possible les pathogènes responsables de septicémie, afin de définir et d'optimiser le traitement antibiotique (148).

Deux approches de développement se distinguent. La première vise à réaliser une analyse MALDI TOF MS après une courte incubation du prélèvement sur milieu solide. Ce protocole permet d'éliminer le sang et autres composants du prélèvement qui peuvent induire un bruit de fond dans l'analyse conduisant à des interférences et à de faibles corrélations avec les bibliothèques MALDI TOF MS (149,150). Des études ont montré que seulement 5h d'incubation permettent une identification précise de la plupart des bactéries, bien que les scores d'identification soient en deçà des performances de l'identification MALDI TOF MS conventionnelle (55). Cependant, moins de 50% des espèces anaérobies sont identifiées par cette approche (151).

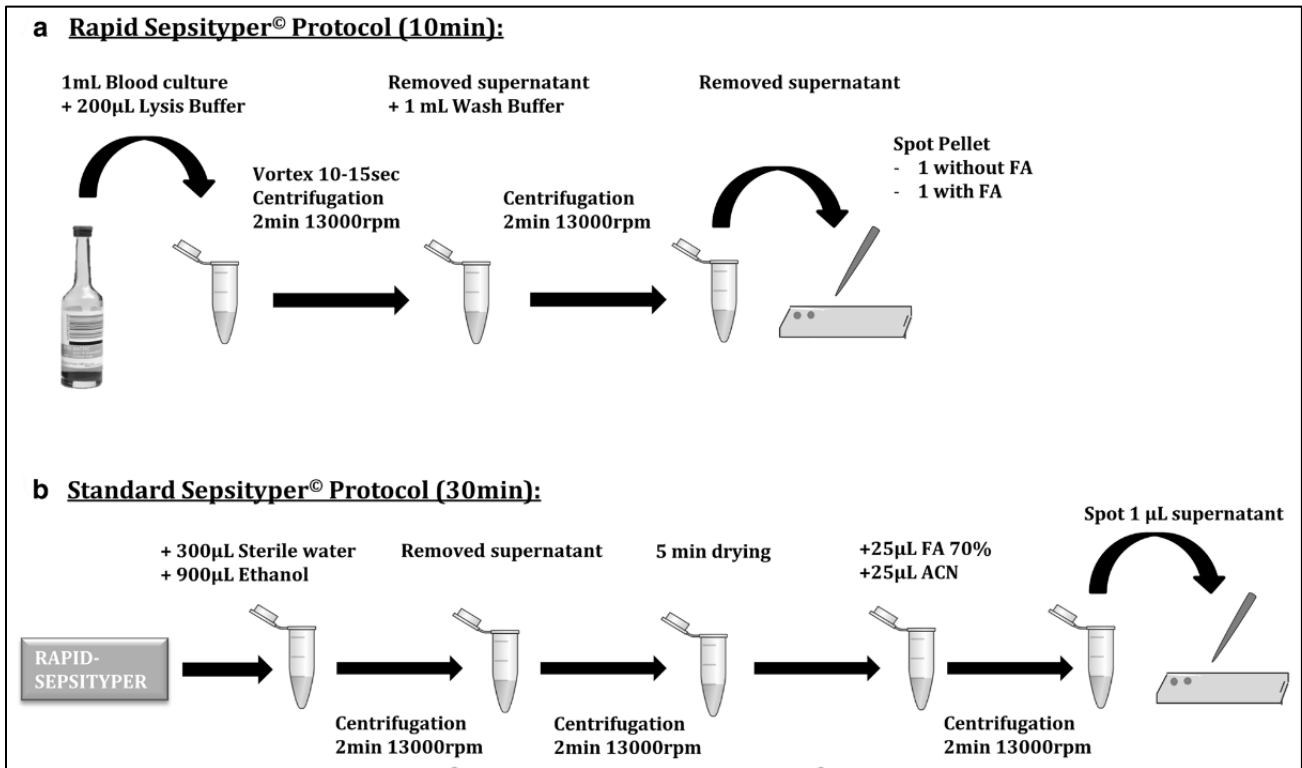


Figure 9: Protocole d'identification rapide MALDI-TOF Sepsityper d'après Ponderand et al, 2020 (38). FA : Acide formique. ACN : Acétonitrile

La seconde approche consiste à effectuer l'identification par MALDI TOF MS directement sur des culots d'hémocultures positives après des étapes d'extraction et de purification (152). Cette méthode peut être réalisée grâce à des kits commerciaux comme les kits MBT Sepsityper® ou Rapid Sepsityper® (Bruker) (Figure 9) ou des protocoles « maisons » (153,154). L'efficacité de cette approche est encourageante avec des concordances d'identification supérieures à 80% mais de grandes différences de résultats sont observées entre les bactéries Gram-positives et Gram-négatives (56,155). De plus, cette approche ne résout pas la problématique des échantillons polymicrobiens (156). Un module du logiciel d'analyse MBT Sepsityper® (Bruker) existe pour l'identification de deux microorganismes mélangés par MALDI TOF MS mais une étude conclue à de faibles résultats. Dans 34% des cas les deux bactéries étaient identifiées correctement, dans 55% des cas un seul microorganismes était identifié, dans 11% des cas aucune des deux bactéries n'était identifiée (157). L'application de l'identification directe par MALDI TOF MS a également été étudiée dans le cadre d'infections urinaires. Ces

dernières sont généralement mono-microbienne et avec une forte charge bactérienne, souvent supérieure à la limite de détection de la spectrométrie de masse (148). Cependant, il a été observé que les bactéries purifiées et extraites à partir d'urines ont un profil spectral légèrement différent des mêmes espèces après une culture sur milieu solide ce qui cause une baisse des scores d'identification lors de l'utilisation de bases de données conventionnelles (146,158). En conclusion, l'identification rapide et directe des prélèvements cliniques par MALDI TOF MS est une approche prometteuse et est actuellement en cours d'intégration dans de nombreux laboratoires de routine. Cependant, elle nécessite encore des améliorations pour pouvoir atteindre des sensibilités et spécificités similaire à celles des techniques de référence.

Comme expliqué précédemment, la spectrométrie de masse MALDI TOF échoue à différencier certaines espèces entre elles. Pour l'identification bactérienne, le MALDI TOF MS utilise les empreintes spectrales des protéines bactériennes (principalement les protéines ribosomales) dans une gamme m/Z (valeur de spectrométrie proportionnelle à la taille des molécules) entre 2000 et 12000 (51). Toute molécule avec un m/Z inférieur à 2000 est exclue de l'analyse pour éviter les interférences entre des ions générés par la matrice et les pics d'intérêt. Cette exclusion implique également l'éviction de macromolécules, telles que les lipides, des spectres analysés. Or, des études ont démontrées la possibilité d'identifier les espèces bactériennes en utilisant leur profil lipidique (159). Une étude récente a démontré le potentiel du profilage lipidique avec un MALDI TOF MS à pression atmosphérique liquide (160). Il est donc possible d'imaginer une future identification par MALDI TOF MS qui couple à la fois profilages protéique et lipidique pour améliorer le pouvoir et la profondeur d'identification des bactéries. Les systèmes de spectrométrie de masse actuels pour l'analyse des protéines bactériennes utilisent une ionisation positive. Cependant, les nouvelles approches émergentes d'analyse des lipides bactériens utilisent l'ionisation négative, complexifiant le couplage des deux approches. La société Bruker a récemment lancé la commercialisation d'un MALDI TOF MS à ions négatifs (MALDI Biotyper[®] Sirius). Cependant, aucune étude n'a à ce jour montré les bénéfices de ce système pour l'identification bactérienne. Pour l'instant, les MALDI à ions négatifs sont utilisés et commercialisés pour la détection rapide des résistances antibiotiques (161,162).

I.A.3.c. PCR digitale

Les méthodes d'identification basées sur la PCR évoluent également et plusieurs innovations ont été développées. Décrite pour la première fois dans les années 90, la PCR digitale (dPCR) est l'une d'entre-elles (163,164). Le principe de la dPCR consiste à isoler et répartir des molécules d'ADN dans des compartiments contenant au maximum un fragment d'ADN, créant ainsi une matrice de compartiments (Figure 10). De telles matrices peuvent être créées avec différentes méthodes de dispersion, comme les microgouttelettes, les micro puits ou la dispersion par impression (165). Une fois la matrice créée (avec des centaines de compartiments), le mix de PCR est ajouté à chaque compartiment et l'amplification en chaînes est réalisée. Les résultats sont analysés pour chaque compartiment. Ils sont binaires (0 ou 1) suivant si un fragment d'ADN est présent et amplifié ou non. Une fois les résultats compilés, une quantification absolue et précise de la concentration originelle en ADN cible est obtenue (166). Cette quantification absolue est le principal avantage de la dPCR par rapport à la qPCR dont la quantification est peu précise et requiert la création de courbes de calibration en amont de l'analyse (167). La dPCR est également moins sensible aux inhibiteurs de PCR en comparaison des autres méthodes moléculaires (168). Le multiplexage de cette technique est également plus facile et peut être poussé à un plus haut niveau en utilisant plusieurs combinaisons de fluorophores dans différentes zones d'une même matrice (169,170). Cependant, les systèmes de dPCR ont un débit d'analyse réduit avec des protocoles plus complexes que la qPCR. L'étape de dispersion augmente également le risque de contamination (171). Le très petit volume initial d'échantillon utilisé (par rapport à la qPCR) peut également provoquer une baisse de sensibilité pour certaines applications (172). Enfin, le besoin d'instruments et de consommables spécifiques se traduit en une augmentation du coût par analyse en comparaison des autres méthodes de PCR déjà coûteuses (171).

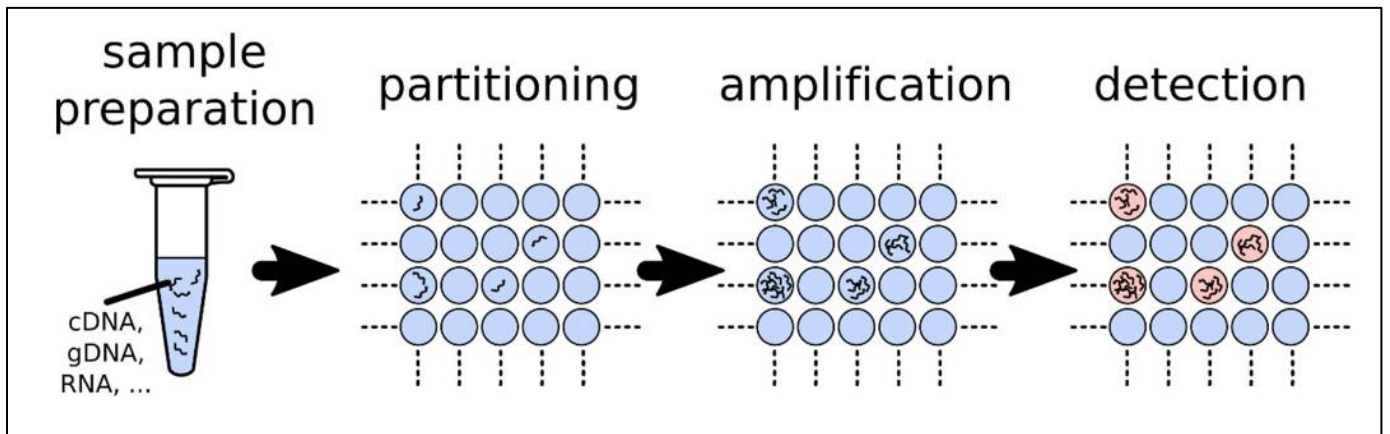


Figure 10 : Principe de PCR digitale d'après Quan et al, 2018 (153)

La dPCR a été utilisée pour identifier plusieurs espèces bactériennes (173–176) avec des performances similaires à l'identification par qPCR. Cependant, en bactériologie clinique, la quantification de la concentration initiale d'un pathogène est rarement un paramètre essentiel pour les cliniciens (à la différence de l'importance de la charge virale en virologie clinique) même si certaines publications établissent un lien entre la charge bactérienne et la gravité du pronostic du patient (177). Ainsi, dans un contexte de bactériologie clinique, la PCR digitale est, en l'état des choses, une méthode complexe et coûteuse, avec un débit d'analyse en deçà de la qPCR, et qui n'apporte pas assez d'avantages pour être utilisés largement dans les laboratoires de routine bactériologique. Elle reste cependant l'une des évolutions majeures des dernières décennies en termes de méthodes d'identification basées sur l'amplification de séquences nucléiques.

I.A.3.d. Techniques d'amplification moléculaire isothermale

Même si la PCR est largement utilisée pour l'identification bactérienne, elle demeure une technique onéreuse notamment car elle requiert l'utilisation de thermocycleurs. Le coût de cette technique fait qu'elle est difficilement utilisable dans des environnements aux ressources limitées. Pour outrepasser cette limitation financière, des méthodes d'amplification dites isothermales ont été développées. Ces techniques isothermales consistent à amplifier une séquence d'acides nucléiques cible en demeurant à une température constante, permettent de s'affranchir de thermocycleurs. Plusieurs approches isothermales ont été décrites comme l'amplification par polymérase-recombinase (RPA) (178), l'amplification hélicase-dépendante (179), l'amplification par cercle roulant (RCA) (180), l'amplification basée sur la séquence d'acides nucléiques (NASBA ou répllication de séquence auto-entretenu, 3SR) (181), ou encore l'amplification isothermale médiée par boucle (LAMP) (182). Une revue récente décrit en détails la plupart des techniques d'amplification isothermales (183). Parmi ces méthodes, la LAMP est la plus fréquemment utilisée. La LAMP est basée sur l'utilisation d'un ensemble de quatre à six amorces ce qui la rend très spécifique (Figure 11). Elle est réalisée à température constante entre 60 et 65 °C (184) grâce à une ADN polymérase Bst. Cette polymérase possède la capacité de se déplacer sur double brin et ne requiert donc pas d'étape de dissociation des doubles brins d'ADN. Une amplification par LAMP peut être réalisée en 30 à 60 minutes et produit environ cinquante fois plus d'amplicons qu'une PCR classique durant 90 minutes (185). Cette grande quantité d'amplicons participe à la très bonne sensibilité de cette technique (186).

Les tests basés sur la LAMP ont été utilisés pour détecter efficacement des bactéries comme *Mycobacterium tuberculosis* (187), *Salmonella typhi* et *paratyphi* (188), *Bordetella pertussis* (189) ainsi que des pathogènes respiratoires (190). Une étude rapporte une limite inférieure de détection d'un test LAMP (nommé RealAmp) de 1×10^3 CFU/mL contre 1×10^4 CFU/mL pour le test PCR lors de la détection d'*Acinetobacter baumannii* (191). De plus, la LAMP semble être moins sensible à certains inhibiteurs de PCR provenant de milieu de culture ou de prélèvements biologiques (192). Bien entendu, cette technique possède également plusieurs désavantages. Tout d'abord, le fait que la LAMP ait besoin de 6 jeux d'amorces pour être réalisée est un problème. Le design des amorces est sujet à beaucoup de restrictions

supplémentaires et peut être long et difficile (193). En effet, comme dans toute PCR, les amorces doivent avoir être spécifiques de la séquence ciblée. Cela pour pouvoir cibler le gène d'intérêt sans s'associer à d'autres amorces et ou à des séquences similaires dans le génome de l'organisme ciblé mais également au sein d'autres génomes pouvant potentiellement être dans les échantillons (humain, autres bactéries etc...). Des logiciels existent pour assister les utilisateurs dans cette tâche comme PrimerExplorer ou LAMPDesigner (Premier Biosoft). De plus, le multiplexage des méthodes LAMP est extrêmement complexe et restreint puisque chaque nouvelle cible rajoute un jeu de 6 amorces supplémentaires, augmentant encore la complexité du design des amorces (194). Une LAMP multiplexe a cependant déjà été développée pour l'identification des *Salmonella spp* et de *Vibrio parahaemolyticus* (195).

La présence de multiples amorces augmente également les risques d'hybridation amorce/amorce, ce qui augmente les risques de résultats faussement positifs (196). La technique LAMP est très sensible aux contaminations, de plus, ses amplicons finaux sont extrêmement stables ce qui augmente le risque de contamination entre analyses (197). Ce problème peut être partiellement résolu par l'utilisation de kits commerciaux qui ont des consommables à usage unique et des compartiments de tests scellés qui, une fois l'échantillon ajouté, ne seront plus ouverts même pour l'analyse et la visualisation des résultats. Enfin, le produit final de la LAMP est une longue chaîne de séquences amplifiées qui est inutilisable pour les analyses moléculaires supplémentaires comme le séquençage (198). La LAMP est donc une technique d'amplification prometteuse particulièrement dans des environnements à ressources limitées. Cependant, la PCR en temps réel reste généralement le premier choix pour le diagnostic par biologie moléculaire en laboratoire de routine à cause du manque de développement de solutions automatisées à haut débit pour les techniques LAMP.

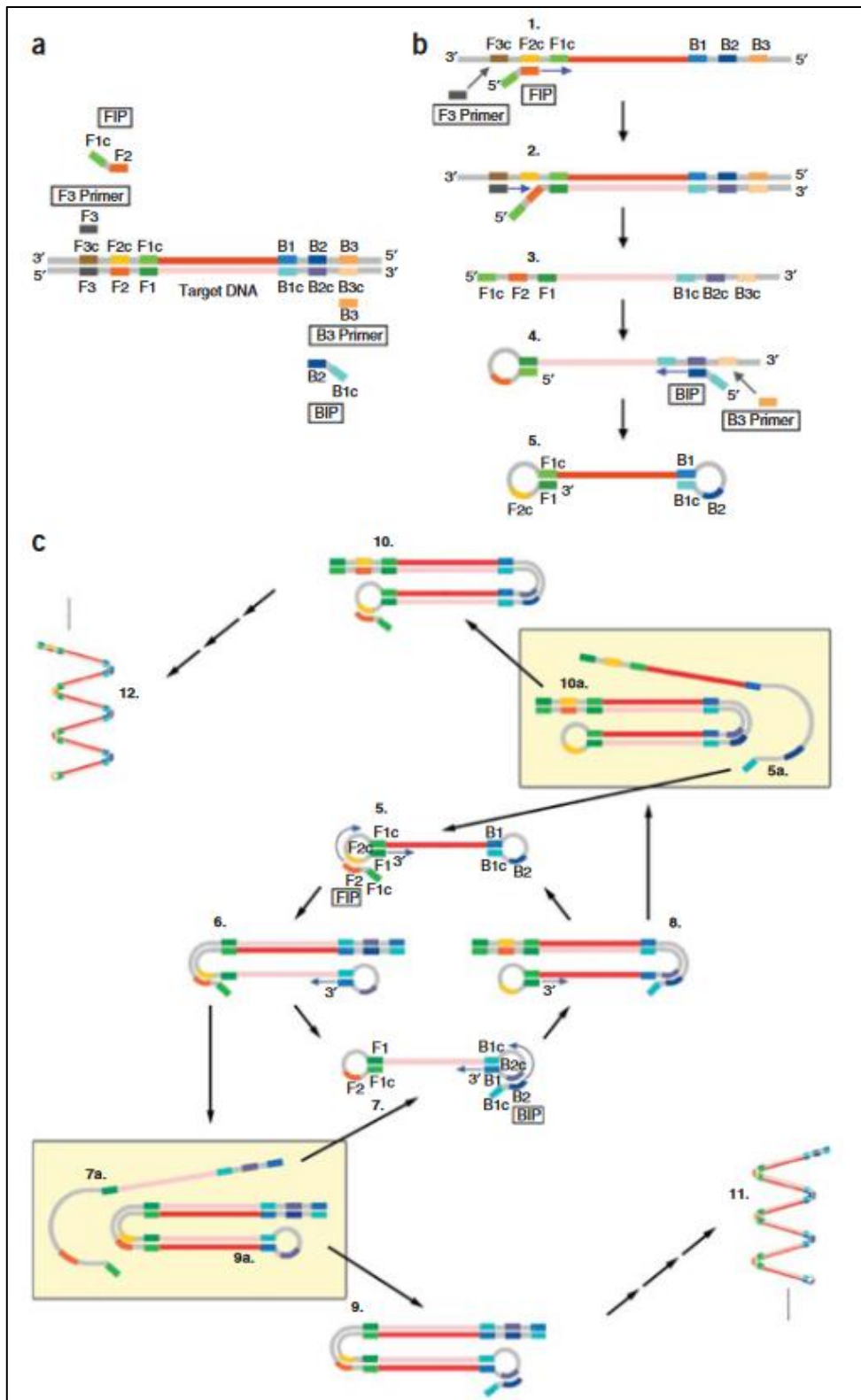


Figure 11 : Principe d'amplification isothermale m di e par boucle (LAMP) d'apr s Tomita et al, 2008 (199)

I.A.4. Technologies émergentes

En parallèle de l'évolution des techniques actuelles, de nouvelles technologies pour l'identification bactérienne apparaissent. Ces nouvelles approches sont en développement et non commercialisées, mais peuvent amener dans les prochaines années de nouveaux angles d'approches et outils pour la bactériologie.

I.A.4.a. Analyse des composés volatiles microbiens

Les composés volatiles organiques (VOCs) sont produits par les microorganismes et leur analyse peut avoir des intérêts diagnostics (200). Plusieurs études ont ainsi utilisé l'analyse de VOCs pour l'identification bactérienne. Des telles analyses sont généralement réalisées avec les systèmes de « nez électroniques ». Ces systèmes, inspirés du fonctionnement de l'odorat humain, permettent par divers moyens de capter et analyser les composés volatiles présents dans une atmosphère restreinte. Certains animaux peuvent également différencier des odeurs spécifiques à certaines infections bactériennes. En effet, des cricétomes de Gambie (rats géants d'Afrique) ont été utilisés pour identifier des infections à *Mycobacterium tuberculosis* directement à partir d'expectorations de patients (201). Cette méthode d'identification reste cependant imparfaite, avec une sensibilité de 80,3 % et une spécificité de 72,4 %. Bien que non applicable en routine et strictement limitée à une bactérie en particulier, cette approche avec des animaux reste peu chère et rapide en comparaison des techniques de détection de VOCs conventionnelles. Des analyses plus classiques des VOCs peuvent être réalisées par des méthodes vibrationnelles, de la chromatographie gazeuse ou par mesure de conductivité sur puces métalliques. Cette dernière méthode a été utilisée pour identifier neuf espèces bactériennes différentes avec un taux d'identification correcte de 87% (202). Cependant, cette identification a été réalisée pendant une étape de culture et d'amplification bactérienne et non directement sur échantillons cliniques.

La chromatographie gazeuse couplée à de la spectrométrie de masse est également utilisée pour l'identification de bactéries par analyse des VOCs comme par exemple pour *Mycobacterium tuberculosis* dans des expectorations (203) ou *Pseudomonas aeruginosa*, *Acinetobacter baumannii* et *Klebsiella pneumoniae* après une étape de culture et d'extraction (204). L'analyse des VOCs pour l'identification bactérienne peut être une approche viable mais reste pour l'instant à développer, inutilisable en routine en l'état et assez sensible au bruit de fond et aux contaminations (200).

I.A.4.b. Approches spectroscopiques et optiques

Les techniques de spectroscopie vibrationnelle sont basées sur l'analyse des variations des fréquences de vibrations d'ondes (lumière visible, infra-rouge etc...). Utilisées souvent pour l'identification de la composition d'un milieu, certaines de ces techniques comme la spectroscopie Raman, ou la spectroscopie infrarouge à transformée de Fourier (FTIR) ont été étudiées pour leur application à l'identification bactérienne. Ces techniques sont des approches non destructives ce qui est un avantage pour l'identification bactérienne. En effet, cela permet la réutilisation de la même colonie bactérienne pour des tests complémentaires, limitant ainsi les risques d'erreur lors de la sélection d'une colonie similaire.

Principalement utilisée pour des analyses chimiques, la spectroscopie Raman a également été appliquée à la caractérisation de matériaux biologiques comme l'analyse de cellule unique eucaryote, le tri cellulaire, l'analyse de tissus mais également à l'identification bactérienne (205). La capacité de cette technique à acquérir, sans aucun marquage, les spectres de cellules uniques en quelques secondes avec une forte spécificité rendent la spectroscopie Raman particulièrement intéressante pour l'identification bactérienne. Principalement car elle peut être utilisée sans étape préalable de culture des pathogènes. Cependant, les variations de spectres observées par spectroscopie Raman sont souvent très subtiles et donc facilement noyées dans le bruit de fond (206). Des méthodes d'amélioration du signal Raman existent avec l'utilisation de nanoparticules ou de supports nano structurés (207). La spectrométrie Raman nécessite également des stratégies

d'isolement des bactéries du milieu environnant. La composition complexe des échantillons cliniques cause des signaux non spécifiques aux bactéries et peut masquer des pics spécifiques aux bactéries dans des bruits de fond.

L'identification bactérienne rapide sans étape de culture (trois heures avec préparation des échantillons (208)) et la limite de détection basse (inférieure à 10^2 UFC/mL dans des échantillons de faible complexité comme l'urine (209)) sont des avantages forts de la spectroscopie Raman, néanmoins contrebalancés par plusieurs inconvénients. Tout d'abord, la limite de détection de la technique augmente rapidement avec la complexité des échantillons (210). Ensuite, les bases de données nécessaires pour l'identification par spectroscopie Raman sont, actuellement, dépendantes des systèmes d'acquisition et non transférables. De plus, ces bases de données doivent prendre en compte les conditions de pousse des bactéries (composition de la matrice de l'échantillon, temps d'incubation, température, hygrométrie) ce qui les rend très complexes à développer et à transférer (211). Pour le moment, la spectroscopie Raman a été utilisée pour discriminer plusieurs espèces bactériennes entre elles dans des échantillons cliniques comme les urines (208), le sang (212), les fluides d'ascites (208) ou les expectorations (213) mais également pour le typage de souches dans des études épidémiologiques (214).

L'application de la spectroscopie infrarouge à transformée de Fourier est une autre approche optique qui a été évaluée pour l'identification bactérienne (215). Cependant, cette technologie reste, pour le moment, uniquement utilisée pour le typage et l'épidémiologie de souches au sein d'une même espèce comme *Escherichia coli* ou *Salmonella enterica* (216) avec notamment le système IR Biotyper[®], une solution automatisée commercialisée par la société Bruker (217,218).

I.A.4.c. Autres approches

Beaucoup d'autres approches sont également à l'essai pour une application à l'identification bactérienne. Leur manque de développement, de standardisation et leur spectre d'identification étroit font qu'elles restent actuellement inutilisables en laboratoire de routine. Cependant, ces méthodes en devenir peuvent faire partie du futur de la bactériologie et compenser les inconvénients actuels des techniques de routine. Parmi ces techniques, on peut citer la résonance magnétique T2 pour la détection des produits de PCR (219,220), la détection photo-acoustique médiée par bactériophages (221), la résonance des plasmons de surface pour une identification médiée par des anticorps (222) ou l'utilisation d'aptamères d'ADN couplés avec de la fluorescence (223,224).

I.B. Intelligence artificielle et réseaux neuronaux convolutifs

I.B.1. Principe de fonctionnement d'un réseau neuronal

Basés sur le principe de régression logistique, les réseaux neuronaux permettent le développement de modèle de classification à partir de base de données. La régression logistique est une méthode d'analyse statistique permettant de prédire des valeurs à partir d'observations réelles. Elle consiste à déduire une formule de calcul à partir d'une base de données cumulant données et valeurs de résultats attendues (Figure 12). Cet apprentissage consiste en l'adaptation de deux coefficients, associés aux données initiales, nommés poids (w) et biais (b). Ces coefficients ainsi que la donnée initiale sont utilisés dans une fonction pour calculer la valeur finale estimée \hat{y} . Cette valeur \hat{y} est ensuite comparée à la valeur réelle attendue et l'erreur entre les deux est calculée. Cette erreur permet, par descente de gradient, une adaptation du poids et du biais associés à la donnée initiale permettant ainsi de recalculer une valeur estimée. L'enchaînement de ces cycles permet l'optimisation de la méthode de calcul et donc une estimation de plus en plus fiable de la valeur finale.

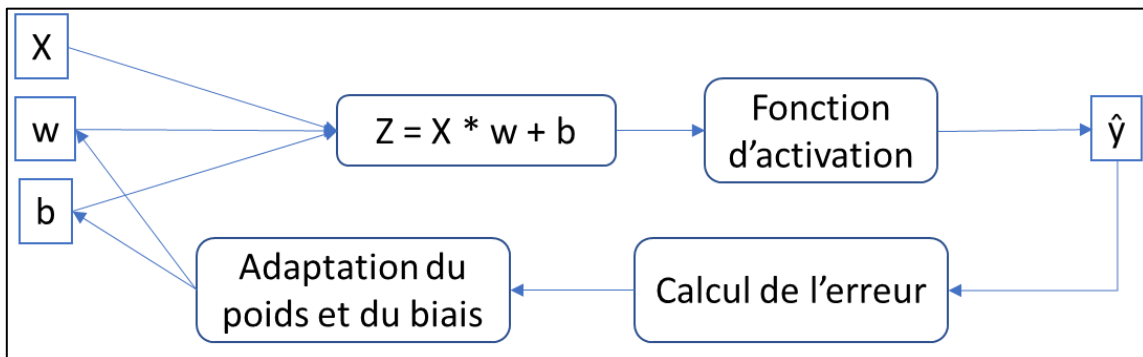


Figure 12 : Principe de régression logistique

Les réseaux neuronaux artificiels consistent en l'addition de plusieurs régressions logistiques pour pouvoir développer des modèles de classification plus complexes. Ces réseaux neuronaux, inspirés initialement de l'architecture du cerveau humain sont des successions de couches composées de neurones (Figure 13). Un neurone artificiel, au même titre qu'un neurone biologique, a pour fonction de recevoir une information (ou une donnée), de la traiter et de la transmettre. En pratique, un neurone artificiel effectue un calcul qui consiste à pondérer des données d'entrée par un poids synaptique (w) et fournir ainsi une valeur de sortie. L'accumulation de nombreux neurones permet ainsi la création d'un réseau dense de traitement de données. Pour commencer, les éléments de la base de données doivent être décomposés en différentes valeurs pour pouvoir être traités par le réseau neuronal. On peut, par exemple, imaginer définir une image par une liste des pixels qui la compose et la valeur des couleurs qui leur sont associés. Des étapes plus complexes de prétraitement des images peuvent être réalisées comme des déconvolutions ou des extractions de données par transformée de Fourier.

Une fois, ce traitement réalisé, la donnée initiale a donc été traduite en une liste de variables qui la définisse. Cette liste vient alors construire la première couche du réseau neuronal, la couche d'entrée. Cette couche permet de transférer ces données à la première couche cachée du réseau neuronal. Les couches cachées sont constituées de plusieurs neurones. Chaque neurone pondère toutes les données d'entrée par un poids synaptique (w) et somme les résultats. Ce résultat résume alors un traitement des données initiales et est transmis à la couche suivante du réseau. Chaque poids des différents neurones est généré aléatoirement ce qui permet des traitements différents pour chaque neurone. Les couches cachées peuvent être mise en série dans une architecture où chacune des couches traitent les données fournies par la couche précédente et calculent les données nécessaires à la couche suivante. L'enchaînement de couches cachées permet d'augmenter la complexité des données traitées par le réseau. Prenons l'exemple de l'utilisation d'un réseau neuronal pour de la reconnaissance faciale. L'image initiale peut être décomposée en pixels et les valeurs de chacun de ces pixels (pixel 1, pixel 2, ...) Ainsi la première couche cachée produit un traitement basé sur la variabilité de chaque pixel, la couche suivante n'a plus une liste de pixels, mais une liste de compilation de pixel qui est à son tour compilée. Ainsi, la première couche traite les valeurs propres des pixels, la seconde traite une compilation de ces pixels et est plutôt dans l'analyse d'un trait de visage. Au fur et à mesure que la complexité d'analyse augmente avec le nombre de couches cachées, les motifs traités par les

couches se complexifient et cela permet d'arriver au traitement d'une image par motif tels que la forme d'un nez, d'une oreille ou d'un œil. Enfin, une couche finale appelée couche de sortie compile tous les traitements réalisés par le réseau pour estimer un résultat de classification. Cette couche de sortie est généralement composée de neurones en même nombre que le nombre de classes à différencier, permettant ainsi de fournir une probabilité d'appartenance de la donnée initiale pour chaque classe. Cette étape de traitement de données est appelée « propagation avant » et permet d'estimer un résultat à partir de données d'entrée. C'est la première partie d'un cycle d'entraînement d'un réseau neuronal.

La seconde partie est appelée « rétropropagation ». Elle consiste à comparer la valeur du résultat estimé au résultat attendu. A partir de cette comparaison une fonction de perte est utilisée pour estimer l'erreur de calcul du réseau. Cela permet ensuite de modifier par une méthode de descente de gradient, en remontant les couches unes à unes, les différents poids associés aux neurones artificiels. L'enchaînement de cycles alternant propagation avant et rétropropagation permet ainsi l'évolution du modèle de calcul, un apprentissage à partir des erreurs calculées pour obtenir au final un modèle de calcul optimisé.

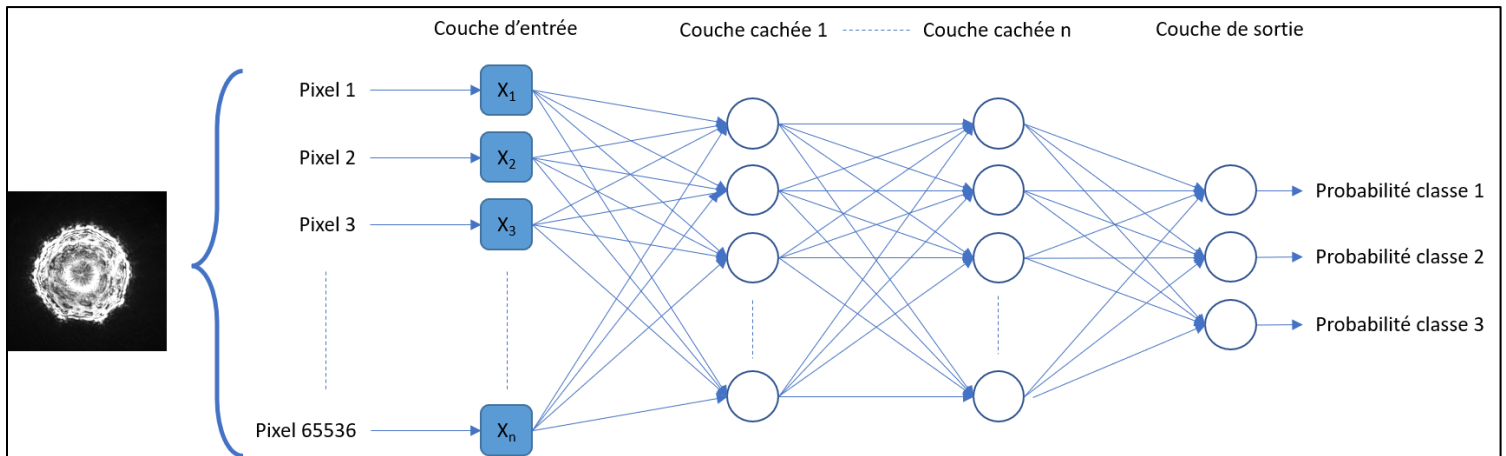


Figure 13 : Architecture d'un réseau neuronal

I.B.2. Apprentissage supervisé et non supervisé

Il existe principalement deux types d'apprentissages en intelligence artificielle : l'apprentissage supervisé et l'apprentissage non supervisé. Le fonctionnement du procédé en lui-même est sensiblement identique. La grande différence entre ces deux approches se situe au niveau de la base de données.

Lors d'un apprentissage supervisé, le réseau neuronal connaît les réponses qu'il doit obtenir, la base de données contient des données dites étiquetées. L'apprentissage consiste donc à développer le meilleur modèle pour classifier correctement les données initiales. Ce type d'apprentissage permet d'effectuer des tâches de régression qui consiste à attribuer une valeur mathématique à un jeu de données comme par exemple déduire le prix de vente d'un bien immobilier en fonction de sa superficie, sa localisation etc. L'apprentissage supervisé est également très utilisé pour des tâches de classification, comme c'est le cas dans les travaux ici présentés. Ces tâches consistent à attribuer une classe à des objets comme par exemple associer une espèce bactérienne à des patterns de diffraction LASER.

L'apprentissage non supervisé est un entraînement à partir de données non étiquetées. Le réseau neuronal crée alors ses propres classes pour y répartir les différentes données initiales. Ce type d'apprentissage est notamment utilisé pour des tâches de filtrage collaboratif. Les logiciels de préférence et de suggestions de vidéos des plateformes de vidéos en ligne utilise des systèmes d'apprentissages non supervisés pour étudier les préférences des utilisateurs et leur proposer du contenu personnalisé.

I.B.3. Réseau neuronal convolutif

I.B.3.a. Principe

Un réseau neuronal convolutif est similaire à une méthode d'apprentissage supervisé. A partir d'images ou de données en entrée, des caractéristiques (« features » en anglais) de ces dernières sont détectées et utilisées pour créer un modèle de classification. Cependant, dans le cas d'un réseau neuronal convolutif, l'extraction de ces caractéristiques est apprise et réalisée par l'algorithme lui-même. Les méthodes d'apprentissages supervisés classiques nécessitent elles une extraction en amont de l'algorithme selon des choix fournis par l'utilisateur de l'algorithme. L'extraction et la hiérarchisation automatique des caractéristiques permettent une minimisation des erreurs de classification ainsi qu'une optimisation des paramètres des modèles de classification. Le premier réseau neuronal convolutif a été ébauché dès 1989 (225) et proposé avec l'algorithme LeNet en 1998 par Yann LeCun (226). Mais c'est réellement en 2012 avec l'algorithme AlexNet que l'importance des réseaux neuronaux convolutifs pour la classification d'image a été démontrée (227).

L'apprentissage automatique des caractéristiques par un réseau neuronal convolutif consiste en l'ajout de couches de convolution et de sous-échantillonnage au début de l'algorithme. La convolution consiste à analyser par une approche locale une image. Un noyau de convolution est défini comme une zone d'une certaine taille, généralement de quelques pixels. En positionnant ce noyau sur une zone de l'image à traiter, il est possible d'extraire le motif ou la tendance générale de cette zone. Cela permet de résumer en une variable, les données de tous les pixels compris dans le noyau. En décalant pas à pas ce noyau sur l'image, il est ainsi possible d'extraire les caractéristiques d'une image zone par zone. Le calcul de la variable par un noyau de convolution est fait en appliquant une formule de calcul propre (multiplication, moyenne, addition, remplacement de valeurs...). En multipliant le nombre de neurones dans une couche de convolution, il est possible d'attribuer, par neurone, une méthode de calcul différente pour le noyau de convolution et ainsi obtenir plusieurs méthodes d'extraction de caractéristiques. C'est cet accès à un grand nombre de méthodes d'extractions qui permet l'apprentissage automatique par le réseau neuronal convolutifs des caractéristiques.

Plusieurs paramètres sont à prendre en compte dans la conception d'une couche de convolution (Figure 14). Tout d'abord, la taille du noyau de convolution, aussi appelé kernel, doit être définie. Il faut également définir le pas, ou stride en anglais, qui correspond à la distance entre chaque motif analysé. Enfin une fonction d'activation ou de correction est à définir. Cette fonction permet une homogénéisation des résultats de la convolution et évite des données aberrantes qui seront non analysables par la suite. Il est également possible de définir une marge, ou « padding » en anglais. Cela consiste à ajouter avant convolution une ou plusieurs rangées de pixels avec une valeur de 0 en périphérie de l'image. En effet, du fait du déplacement en chevauchement du noyau de convolution, les pixels en bordure de l'image se trouvent moins souvent inclus dans un noyau de convolution que ceux du centre. Déplacer artificiellement les bords d'une image permet ainsi de s'assurer une analyse homogène de tous les pixels d'une image.

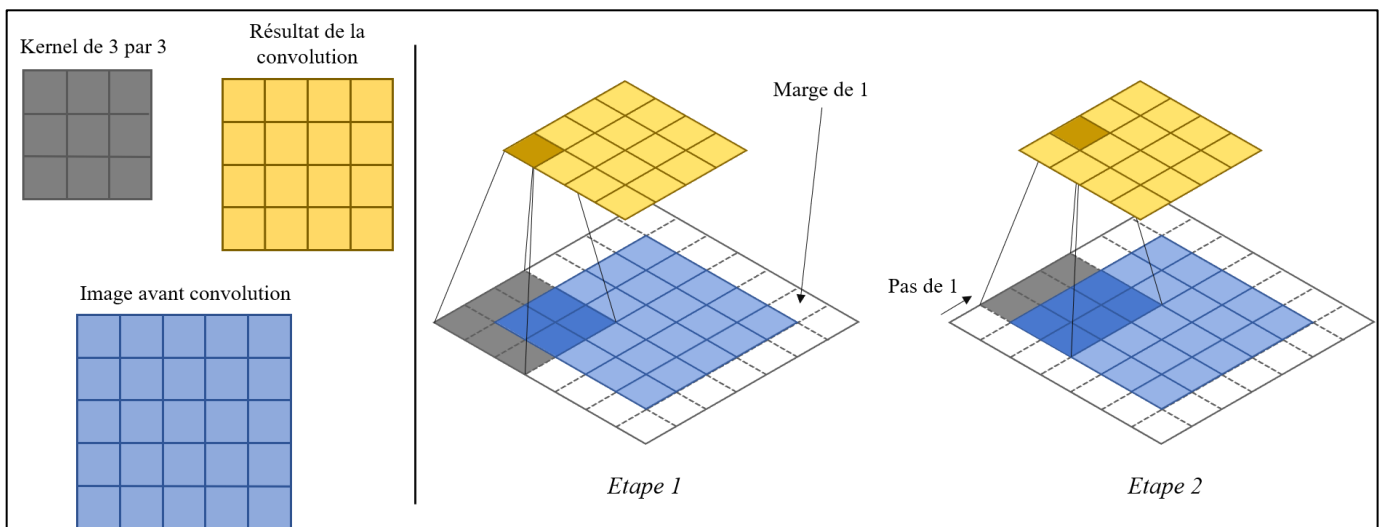


Figure 14 : Illustration des paramètres de convolution

Un réseau neuronal convolutif permet donc une analyse de l'image non plus dans son intégralité mais zone par zone. En prenant l'exemple de l'identification d'un animal sur une image, un modèle d'apprentissage supervisé classique analyse la forme générale de l'animal. Le réseau neuronal convolutif peut grâce à la convolution faire une analyse plus précise et baser son identification sur des zones spécifiques comme les oreilles, la queue, les pattes ou le nez.

I.B.3.b. Entraînement d'un réseau neuronal convolutif

Une fois un réseau neuronal constitué, l'entraînement de ce réseau sur une base de données a pour but de développer un modèle de classification. Pour cela, une partie de la base de données va être passée dans le réseau neuronal et permettre d'établir un premier modèle de classification lors de la « forward propagation ». Le reste de la base de données est alors utilisé en temps qu'échantillon de test sur le modèle tout juste établi. Les résultats de ce modèle de classification sont ensuite comparés aux résultats réels attendus pour définir le taux d'erreur du modèle lors de la « backward propagation ». Cette étape permet une adaptation des différents poids et biais des neurones artificiels du réseau pour le prochain cycle d'entraînement.

Il est à noter que les limitations de performances informatiques, ainsi que la taille des bases de données, ne permettent pas, dans la grande majorité des cas, de faire analyser l'intégralité d'une base de données par un réseau neuronal en une fois. C'est ici que les notions de taille de batch et d'itérations apparaissent. Pour pouvoir analyser une base de données entière, cette dernière est divisée en parties égales appelées « batch », un batch correspond donc au volume d'une base de données qui serait analysé d'un coup par le réseau neuronal. Il est possible de choisir la taille de ces batchs lors de la mise en place de l'entraînement en fonction des performances du matériel informatique disponible. Les itérations correspondent au nombre de batchs, ou de cycle d'entraînement, nécessaires pour faire analyser la totalité de la base de données par un réseau neuronal. Ce nombre dépend donc à la fois de la taille de la base de données et de celle des batchs lors de l'entraînement.

L'apprentissage d'un réseau neuronal convolutif est impacté par plusieurs paramètres. Tout d'abord, le ratio Train/Test permet de définir la proportion de la base de données utilisée pour établir un modèle de classification, l'autre partie de la base de données est quant à elle utilisée pour tester ce modèle et l'évaluer. Modifier ce ratio permet de dédier une plus grande partie de la base de données à l'entraînement de l'intelligence artificielle ou à l'évaluation du modèle de classification qui en découle.

Un autre paramètre modifiable lors de l'entraînement d'un algorithme basé sur l'intelligence artificielle est le nombre d'epochs réalisés durant l'entraînement. Un epoch correspond à un passage de la base de données dans le réseau neuronal soit à un cycle d'entraînement du réseau neuronal. Ainsi, augmenter le nombre d'epochs durant un apprentissage permet de diminuer le taux d'erreur d'un modèle de classification : plus le réseau a de cycles d'entraînements, plus il a d'occasion d'adapter les poids et biais de ces neurones pour établir un modèle de classification adapté. Un nombre insuffisant d'epoch mène à un problème de sous-apprentissage (underfitting en anglais). Ce terme désigne le fait que le réseau neuronal n'a pas eu suffisamment de cycles d'entraînements pour prendre en compte tous les paramètres des éléments de la base de données permettant de différencier les différentes classes de cette dernière. Cependant, un nombre d'epochs trop grand entraîne un phénomène de sur-apprentissage (overfitting en anglais). Après avoir pris en compte tous les paramètres pertinents pour la différenciation des classes de la base de données, le réseau neuronal, en essayant d'améliorer son modèle de classification, finit par prendre en compte d'autres paramètres et notamment les variations propres au bruit de fond des éléments de la base de données. Ce phénomène de sur-apprentissage base donc l'évolution du modèle de classification sur des variations non pertinentes et cause une perte de performance des modèles de classification. Le nombre idéal d'epochs pour l'apprentissage d'un réseau neuronal est dépendant des bases de données. Ainsi, il est impossible d'établir à l'avance le nombre d'epochs nécessaires pour un apprentissage optimal. Seul le test lors de divers entraînements du réseau neuronal en faisant varier le nombre d'epochs permet d'obtenir une réponse.

I.B.4. Evaluation de la performance d'un modèle de classification

Une fois l'entraînement d'un réseau neuronal effectué, il existe plusieurs métriques utilisées pour évaluer la performance des modèles de classification. Nous citerons ici : l'accuracy (terme gardé de l'anglais pour ne pas le confondre avec la précision), la précision, le rappel et le score F1. Ces quatre métriques sont représentées par des valeurs comprises entre 0 et 1. Un modèle de classification optimal étant un modèle où ces métriques tendent vers 1.

L'accuracy permet une mesure du taux de prédictions correctes par le modèle de classification. Elle prend en compte toutes les prédictions correctes qu'elles soient positives (vrais positifs) ou négatives (vrais négatifs). Elle se calcule grâce à la formule suivante :

$$\frac{\text{Vrais Positifs} + \text{Vrais Négatifs}}{\text{Nombre total de prédictions}}$$

La précision est l'équivalent à la valeur prédictive positive en statistique et épidémiologie. Elle correspond au taux de prédictions correctes parmi toutes les prédictions positives données par le modèle de classification. Elle évalue la capacité du modèle à ne pas faire d'erreur lors d'une prédiction positive. La précision est calculée à partir de la formule qui suit :

$$\frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Positifs}}$$

Le rappel est une métrique équivalente à la sensibilité en statistique et biologie. Il permet d'évaluer le taux de cas positifs détectés par le modèle de classification. Le rappel est calculé en utilisant la formule suivante :

$$\frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Négatifs}}$$

La précision et le rappel sont deux métriques qui perdent de leur pertinence dans le cas de base de données déséquilibrées. On entend par base de données déséquilibrée une base de données où le ratio entre cas positifs et négatifs s'éloigne de 50%. Dans le cas où très peu de cas positifs sont présents dans une base de données, la précision va avoir tendance à augmenter fortement sans pour autant être un bon indicateur des performances du modèle. Peu de positifs implique une tendance vers l'apprentissage au cas par cas du réseau neuronal et donc une incapacité à généraliser leur identification. Au contraire dans le cas d'une forte majorité de cas positifs dans la base de données, le rappel va augmenter pour des raisons symétriquement opposées. De même, l'accuracy est également peu pertinente dans le cas de données déséquilibrés puisqu'un manque de bonnes prédictions chez la catégorie minoritaire (positif ou négatif) sera masqué par l'omniprésence de la catégorie majoritaire. Dans le cas d'une base de données avec plusieurs classes, on la considère comme équilibrée lorsque le nombre de cas positifs par classe ainsi que les ratios entre cas positifs et négatifs pour chaque classe sont semblables.

Le score F1 permet de compenser les défauts de la précision et du rappel, principalement dans le cas de bases de données déséquilibrées. Il permet de synthétiser les valeurs de précision et de rappel en une seule métrique. Il est défini comme étant la moyenne harmonique de la précision et du rappel et se calcule grâce à la formule suivante :

$$2 \times \frac{\textit{Précision} \times \textit{Rappel}}{\textit{Précision} + \textit{Rappel}}$$

I.B.5. Application en bactériologie

L'arrivée de l'intelligence artificielle est en cours dans les domaines de santé. La capacité de traitement d'un grand volume de données et la possibilité d'extraire des tendances de ces données font de l'intelligence artificielle un outil redoutable pour plusieurs domaines de santé. Plusieurs études ont eu pour but d'appliquer une intelligence artificielle à un sujet de santé. Des études ont été menées sur l'utilisation d'intelligence artificielle pour l'évaluation de facteurs de risques en santé publique (228), la création de modèles de prédiction de risques en cas de maladies chroniques (229) ou encore la création d'algorithme de décision pour la prise d'antibiotiques chez les patients atteints d'infection urinaires (230). L'intégration de l'intelligence artificielle en microbiologie est reconnue par la communauté comme l'un des axes majeurs du futur de la discipline (231). En bactériologie, plusieurs études d'applications d'intelligence artificielle ont été réalisées. Des études ont démontré la capacité d'analyse et d'interprétation de géloses par intelligence artificielle sur prélèvements urinaires (232) ou pour la détection de *Staphylococcus aureus* résistants à la méticilline sur milieux gélosés chromogéniques (233) en utilisant une solution commercialisée : APAS Independence (Clever Culture Systems, Lbt Innovation, UK). Ces études démontrent de très bons résultats avec des concordances positives et négatives supérieures à 95 %. L'intelligence artificielle a également été appliquée à l'analyse de séquences génomiques pour la prédiction rapide de résistances antibiotiques de *Mycobacterium tuberculosis* avec des concordances de plus de 80% (234). Enfin, une étude récente a utilisé l'intelligence artificielle pour la détermination rapide des susceptibilités aux antibiotiques à partir de l'analyse de spectres de masse MALDI-TOF (235). Cette étude se conclue par une étude rétrospective démontrant que dans 89% des cas, cette nouvelle approche aurait permis une adaptation plus rapide et bénéfique de l'antibiothérapie des patients.

Certaines études se sont penchées sur l'application d'intelligence artificielle à l'analyse d'images. Des études rapportent l'efficacité de réseaux neuronaux convolutifs pour la lecture et l'analyse de coloration de Gram avec des sensibilités et spécificités supérieures à 90% (excepté pour la détection de Gram positive en chainettes où la sensibilité est de 75%) (236). Un programme similaire est commercialisé par la société Techcyte Inc. (Orem, USA).

L'arrivée de l'intelligence artificielle en laboratoire clinique est en cours. Cependant, elle se heurte à un problème majeur. Il existe de plus en plus de réglementation en termes de traçabilité pour les laboratoires cliniques, il doit être possible de s'avoir ce qui a été fait, comment et ce quel que soit le moment. Or, le développement d'un modèle de classification par une intelligence artificielle est fait sans supervision humaine. La plupart du temps, les utilisateurs ignorent ce qui constitue le modèle et comment il classifie. Ce constat pose donc des problèmes en termes de traçabilité. Les moyens de contrôle et de vérification des résultats fournis par une intelligence artificielle sont peu nombreux. De plus, la puissance d'un modèle de classification est dépendant de sa base de données. Une base de données qui ne représente pas suffisamment la réalité entraînera la création d'un modèle non pertinent. La validation de la qualité des bases de données est donc primordiale pour le développement de solutions basées sur l'intelligence artificielle. Un exemple flagrant vient de l'étude menée par Weis et al (235). L'un des premiers modèles de classification développé à partir de spectres de colonies bactériennes fournissait des prédictions de résistances antibiotiques que même les auteurs considéraient comme étonnamment parfaites. Après analyse de leurs données, il s'est avéré que 90 % des spectres de leur base de données avaient été fait à partir de colonies en croissance sur des milieux sélectifs avec pression antibiotique. Cependant, il s'avère que des composés de ces milieux gélosés étaient identifiables sur les spectres de masse. L'intelligence artificielle avait en réalité développé un modèle permettant d'identifier les milieux sélectifs sur lesquels poussaient les colonies analysées. Or sur ces milieux sélectifs, seules les bactéries avec des résistances antibiotiques pouvaient croître, la corrélation était donc parfaite. Malheureusement, le modèle de classification était incapable de trouver la résistance antibiotique pour une même souche bactérienne si celle-ci avait été cultivée sur un milieu gélosé non sélectif. Un biais de classification avait donc été involontairement introduit dans la base de données et avait rendu le modèle de classification non pertinent.

I.C. Diffraction de Fresnel

La diffraction est un phénomène apparaissant lorsqu'une onde traverse une ouverture dans une surface ou rencontre un objet. Ce phénomène peut se produire sur toute onde qu'elle soit mécanique (onde sonore, onde à la surface de l'eau, etc...) ou électromagnétique. La diffraction d'une source lumineuse est caractérisée par la déviation de rayons lumineux par un obstacle qui ne lui est pas totalement transparent ou par une ouverture de petite taille. Ce phénomène, bien que connu avant, est observé et décrit pour la première fois dans une étude scientifique par Francesco Grimaldi en 1665. Il sera ensuite interprété correctement par Christian Huygens et sa théorie ondulatoire de la lumière en 1690. Les travaux d'Augustin Fresnel aboutiront en 1818 à la description de la diffraction de la lumière selon le principe d'Huygens-Fresnel. Il énonce que tout point atteint par la lumière issue d'une source primaire, peut être considéré comme une source secondaire émettant une onde sphérique. Les déviations des ondes lumineuses sont dépendantes de la longueur d'onde de la lumière émise mais également en fonction de la forme de l'objet diffractant (Figure 15). Ainsi, la forme du pattern de diffraction change en fonction de l'objet diffractant et peut donner des informations sur cet objet.

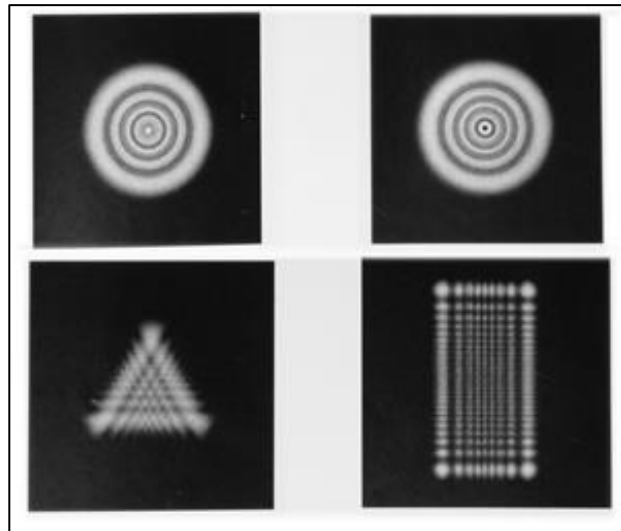


Figure 15 : Exemples de pattern de diffraction issus de la diffraction d'une onde lumineuse au travers d'ouverture de forme circulaire, triangulaire ou rectangulaire

Un LASER (acronyme pour Light Amplification by Stimulated Emission of Radiation) est un système optique permettant d'amplifier et de rassembler la lumière en un faisceau étroit. Les avantages de ce dispositif sont d'émettre une lumière directionnelle et pure spectralement (avec un spectre de longueur d'onde très étroit). Cela rend le LASER particulièrement intéressant pour toute expériences impliquant de la lumière diffractée.

L'idée d'utiliser la diffraction pour l'identification d'espèce bactérienne a été émise pour la première fois par P. Wyatt en 1968 (237). Ces expériences, réalisées sur des suspensions bactériennes en solution, permirent de différencier quelques espèces bactériennes entre elles (*Bacillus subtilis et anthracis, Staphylococcus aureus et epidermidis et Escherichia coli*). Entre 2004 et 2007, plusieurs études proposèrent le même principe d'expérience mais cette fois sur colonies bactériennes en culture sur milieu gélosé (238,239). En positionnant une colonie bactérienne entre un LASER et une caméra ou un écran, un pattern de diffraction peut être observé. La forme de ces patterns de diffraction semble être dépendante de l'espèces bactérienne de la colonie (240). Par la suite, la possibilité de déduire l'espèce bactérienne de la colonie à partir de son pattern de diffraction a été évoquée (241,242). Cependant, ces études restent très axées sur la partie optique du sujet (243) et se limitent à la comparaison de patterns de diffraction de quelques espèces bactériennes sans développement de bases de données (244). Elles fournissent cependant quelques informations sur les paramètres d'acquisition influant sur la forme de ces patterns. En effet, la distance séparant le LASER de la colonie influe sur le diamètre du pattern de diffraction. Le milieu gélosé utilisé pour la croissance des colonies a également une influence (245). Une étude fournit des informations sur la nécessité d'analyser des colonies isolées pour éviter la superposition de patterns de diffraction (244). Une autre étude indique des informations sur la taille des colonies à analyser suivant le diamètre du LASER utilisé. En effet, pour obtenir un pattern de diffraction d'une colonie entière, le LASER doit éclairer l'entièreté de cette dernière. Cependant, si la colonie est trop petite par rapport au LASER, les rayons lumineux non diffractés vont créer un bruit de fond trop conséquent pour une acquisition correcte des patterns de diffraction. Les colonies analysées doivent être d'un diamètre égal ou légèrement inférieur (au maximum de 0,3 mm) à celui du LASER (245).

D'un point de vue fondamental, aucune étude à notre connaissance n'a fourni, à ce jour, des informations sur les facteurs phénotypiques et moléculaires d'une bactérie qui causent la variation de patterns de diffraction.

II. Identification automatisée en non destructive par diffraction LASER
(BioKube)

II.A. Contexte et objectifs de la thèse

La microbiologie médicale est ainsi en train de vivre une véritable révolution technologique depuis le début du siècle. Le développement de nouvelles techniques d'identification plus fiables, plus rapides et avec des bases de données plus larges ont eu un impact bénéfique sur l'évolution de la morbi-mortalité des patients infectés. En parallèle, l'arrivée des chaînes automatisées permettant l'ensemencement des échantillons et l'incubation des géloses avec lecture digitale des cultures tend à automatiser le processus complet de l'échantillon, du patient au rendu final de l'identification bactérienne et de l'antibiogramme. Ces systèmes permettent d'améliorer la reproductibilité et la qualité des analyses, leur traçabilité ainsi que de réduire les tâches techniques manuelles fastidieuses et répétitives tout en permettant une gestion des échantillons 24h/24. Néanmoins, certaines parties de ce flux de travail nécessitent encore des étapes manuelles et ne peuvent être intégrées dans un flux totalement automatisé de par les technologies utilisées. De nouveaux procédés sont nécessaires afin de réduire encore les tâches manuelles en fournissant de nouvelles solutions implémentables directement sur les chaînes automatisées en bactériologie. Par ailleurs, certaines technologies utilisées à l'heure actuelle restent coûteuses ou présentes des limites pour l'identification de certains micro-organismes.

Bien que des techniques permettant une identification dans l'heure existent, telle que la PCR, elles ne permettent pas une identification sans *a priori* comme la spectrométrie de masse. Il est nécessaire de suspecter une espèce bactérienne en particulier pour démontrer ou non sa présence au sein du prélèvement par PCR. La spectrométrie de masse permet, quant à elle, une identification de toutes les espèces bactériennes présentes dans sa base de données. Par ailleurs, la spectrométrie de masse MALDI TOF est également une technique peu coûteuse (environ 1 euro par analyse) ce qui permet son utilisation sur un grand nombre de prélèvements. A contrario, les techniques de PCR sont plus coûteuses particulièrement pour les panels syndromiques (100 à 150 euros par test) et sont donc utilisés dans des contextes particuliers et d'urgence extrême où la rapidité d'analyse prévaut. Le travail réalisé dans cette thèse a vocation à présenter des travaux de développement d'une nouvelle technique potentielle d'identification bactérienne à partir de colonies sur milieux gélosés, non destructive et automatisable, et permettant d'identifier

rapidement la totalité des colonies isolées sur une gélose, en utilisant la diffraction LASER sur colonies bactériennes.

Lorsqu'un LASER est dirigé sur une colonie bactérienne, sa lumière se diffracte et l'on peut observer derrière la colonie un pattern de diffraction (cf. I.C. Diffraction de Fresnel). De précédents travaux ont montré des changements dans ces patterns de diffraction en fonction de l'espèce bactérienne des colonies (241,246). Il semble donc possible d'identifier à l'espèce des colonies bactériennes grâce à l'analyse de leur pattern de diffraction. Etudier cette possibilité est le sujet du projet de thèse présenté ici.

II.A.1. Contexte du projet de thèse

Ce projet a été initié par la compagnie BioKube Automation qui développe un automate d'identification bactérienne basé sur cette approche. Cette start-up a initié une collaboration avec le laboratoire de bactériologie du CHU Grenoble Alpes en 2018 ainsi qu'avec le Laboratoire d'Informatique de Grenoble (LIG) pour mettre en place ce projet dont découle la thèse ici décrite. Au commencement de ce projet, un prototype de l'automate BioKube avait déjà été développé. Ce prototype a été hébergé au laboratoire de bactériologie du CHU Grenoble Alpes. A son arrivée, il a été fourni avec un logiciel développé précédemment. Le but de cette collaboration avec BioKube automation était de développer et d'améliorer le prototype déjà existant puis d'étudier son application à l'identification bactérienne. Dans ce cadre, une étape d'analyse et de rétro-ingénierie de l'automate et principalement du logiciel était nécessaire pour la conception du cahier des charges du projet.

Pour permettre un développement optimal logiciel ainsi que la mise en place du réseau neuronal convolutif, un partenariat a été mis en place avec la société Revartis. Cette société basée à Zurich offrant des services de développement de solutions digitales sur mesure en fonction du besoin émis par ses clients. Dans le cadre du projet ici présenté, leur expertise a permis l'optimisation du logiciel de l'automate BioKube ainsi que la mise en place du logiciel de classification basé sur un réseau neuronal convolutif.

II.A.2. Objectifs de la thèse

Les objectifs du projet principal de la thèse ici présenté ont été les suivants :

- Poursuivre le développement d'un automate associé à des algorithmes d'intelligence artificielle permettant une identification bactérienne sur colonie par diffraction LASER
- Evaluer les performances de l'automate et de l'analyse de pattern de diffraction de colonie bactérienne en tant que nouvelle approche d'identification bactérienne.

Ainsi le projet de thèse réalisé se découpe en plusieurs axes. Tout d'abord, l'évaluation du prototype existant et de son logiciel a été effectuée pour permettre l'établissement du cahier des charges de développement. Dans un second temps, un développement informatique a été apporté au logiciel dans le but d'obtenir un prototype fonctionnel et robuste permettant une acquisition précise des patterns de diffraction. Les conditions d'acquisition de ces patterns ont ensuite été évaluées. La diffraction LASER sur colonie étant un sujet encore nouveau, il existe très peu d'information sur l'évaluation des critères d'acquisition d'un pattern diffraction ainsi que sur ses variations en fonction des conditions d'analyse. Il est donc nécessaire d'étudier ces points pour optimiser l'utilisation de l'automate. Enfin, le dernier point consiste à créer les bases de données nécessaires au fonctionnement de l'intelligence artificielle, d'entraîner cette dernière pour obtenir des modèles de classification puis d'évaluer les performances de ces modèles pour la discrimination des différentes espèces bactériennes présentes dans la base de données.

II.B. Matériel et Méthodes

II.B.1. Prototypé BioKube en développement

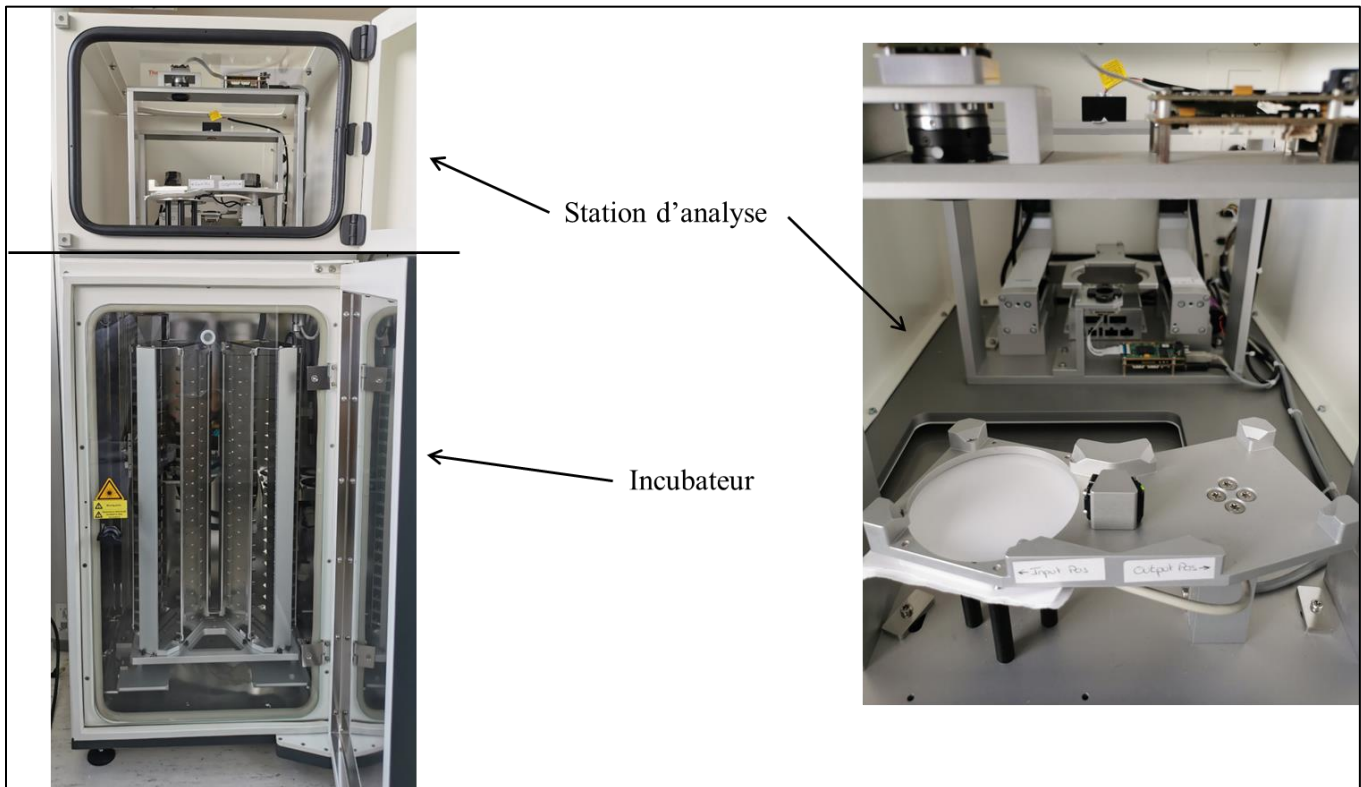


Figure 16 : Automate Biokube et station d'analyse

L'automate BioKube est composé de deux parties distinctes : un incubateur automatisé et une station d'analyse (Figure 16). La partie « incubateur » est commercialisée par l'entreprise Liconic (Mauren, Liechtenstein). L'incubateur permet une régulation de la température et possède 2 cassettes de stockage pouvant chacune héberger 22 géloses, portant ainsi la capacité maximale de l'automate Biokube à 44 géloses. L'incubateur possède également un bras robotisé qui peut déplacer les géloses entre l'incubateur et la station d'analyse grâce à une trappe les séparant. Les géloses sont donc stockées dans l'incubateur et transférées à la station d'analyse au moment voulu.

La station d'analyse est, quant à elle, un dispositif développé uniquement pour l'automate Biokube. Elle est disposée au-dessus de la partie « incubateur » de l'automate et y est reliée grâce à une trappe. Cette station d'analyse peut se décomposer en une partie frontale et une partie arrière. La station avant, ou station LED, est équipée d'un éclairage LED ainsi que d'une caméra (capteur CMOS, 1280x1024 pixels, PixeLink, Canada). Cette station permet de prendre en photo la gélose, de détecter, grâce à un programme informatique les colonies et de les sélectionner en fonction de critères de diamètre et de circularité. C'est à partir de cette image que les positions des colonies à scanner sont enregistrées. Cette station frontale sert également de zone d'entrée et de sortie des géloses dans l'automate. Enfin, la station arrière, ou station LASER, est équipée d'un LASER rouge (longueur d'onde de 635 nm, diamètre de faisceau de 1 mm), d'un capteur CMOS (capteur CMOS, 1280x1024 pixels, PixeLink, Canada) ainsi que d'un bras robotisé. Suite à l'analyse de la gélose et la récupération des coordonnées des colonies d'intérêt, ce bras robotisé permet de déplacer la gélose et ainsi positionner les colonies directement dans le trajet du LASER. Ainsi l'automate peut obtenir des patterns de diffraction des colonies analysées grâce à une caméra positionnée sous le bras robotisé. La distance entre la source du LASER et les colonies est de 14,5 cm, la caméra étant positionnée 1,5 cm sous les colonies (à 16 cm du LASER).

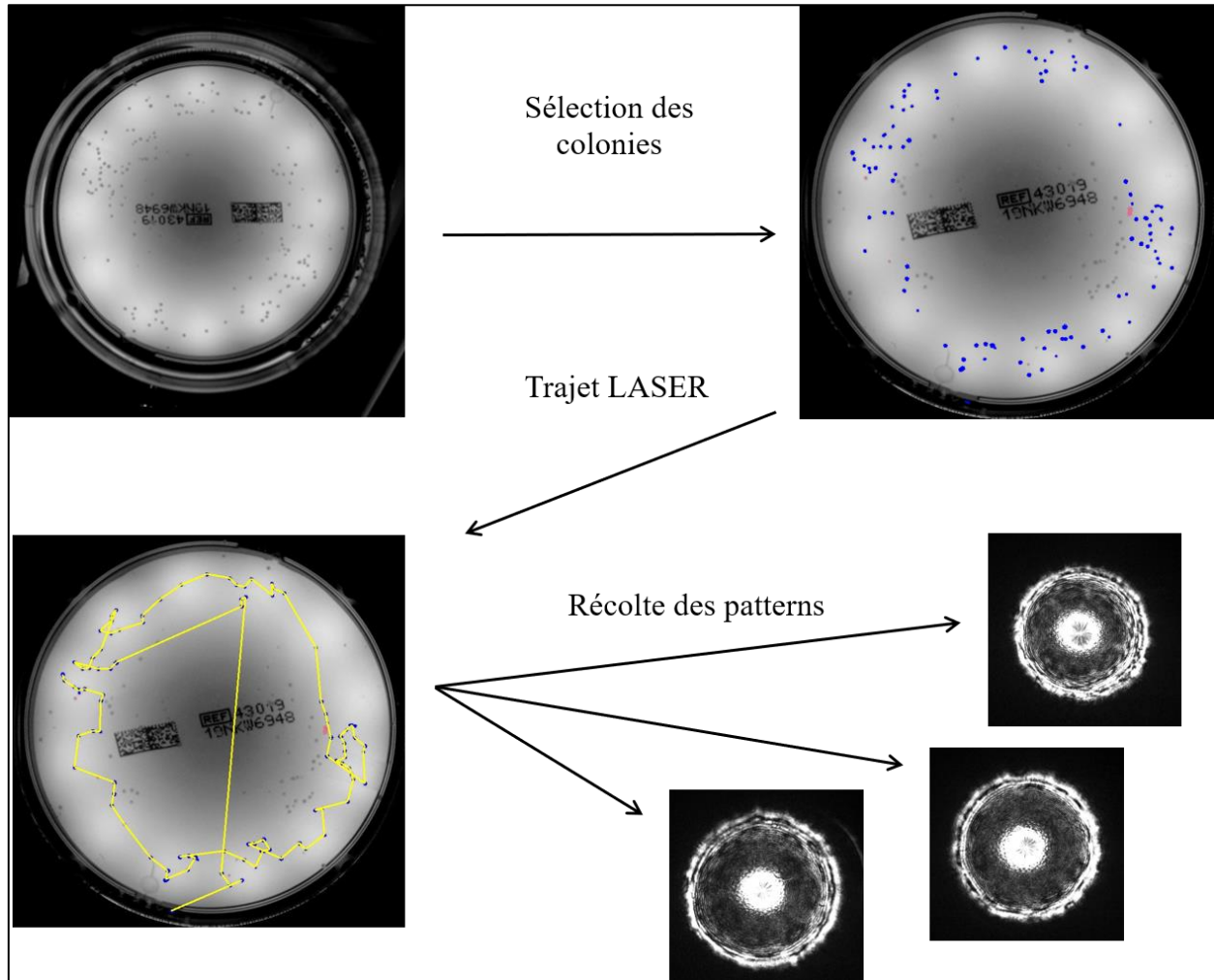


Figure 17 : Flux de fonctionnement d'acquisition de patterns

Une fois les patterns de diffraction acquis, il est bien entendu nécessaire de les analyser pour déduire l'espèce bactérienne des colonies observées. Pour ce faire, ces patterns sont comparés à une base de données contenant les patterns de diffraction de différentes espèces bactériennes. Pour cette analyse, l'utilisation d'une intelligence artificielle basée sur un réseau neuronal convolutif est toute indiquée. Cette intelligence artificielle peut permettre, à partir d'une base de données, de créer des modèles de classification. L'utilisation de ces modèles de classification permet une accélération de l'analyse des patterns de diffraction tout en ayant un grand nombre d'espèces bactériennes au sein des bases de données.

II.B.2. Architecture du réseau neuronal convolutif

L'intelligence artificielle ici utilisée fonctionne sur carte graphique (GPU, Graphics Processing Units) avec une architecture CUDA. L'architecture de l'intelligence artificielle est basée sur une architecture LeNet (225,226). Elle commence par une couche de convolution par un noyau de 5 par 5 pixels, 50 neurones, un pas (stride) de 1 et une fonction de correction ReLU (Unité de rectification linéaire). La fonction ReLU ($a = \max(0, z)$) permet de conserver toutes les valeurs positives et de convertir toutes les valeurs négatives en 0. Ensuite, les données passent par une couche de sous-échantillonnage, ou « pooling » en anglais, consistant en une fonction maximum (max-pooling) avec un filtre de 2 par 2 pixels et un pas de 2. Une seconde couche de convolution d'une taille de 100 neurones intervient ensuite. Cette fois-ci, la convolution est faite par un noyau de 5 par 5 pixels, un pas de 5 par 5, une marge (padding) de 1 par 1 et avec une fonction de correction ReLU. Pour finir le pré-traitement, une deuxième couche de sous-échantillonnage identique à la première est effectuée. Après le pré-traitement, une première couche cachée composée de 100 neurones cachés avec une fonction d'activation ReLU. La seconde couche cachée se compose de 50 neurones cachés et est également activée par une fonction ReLU. Enfin, la couche de sortie (output layer) est composée de 7 neurones et avec une fonction d'activation Softmax. La sortie d'une fonction Softmax permet d'établir des probabilités sur différents résultats possibles et est donc très utilisée pour la classification en classes multiples.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ pour tout } j \in \{1, \dots, K\}$$

Formule de la fonction d'activation Softmax

II.B.3. Souches bactériennes utilisées

90 souches cliniques (15 souches par espèce) ainsi que 6 souches de référence ont été utilisées tout au long des expériences. Le tableau suivant (Tableau 2) résume les différentes souches utilisées. Les souches cliniques sont issues de la collection du laboratoire de bactériologie du CHU Grenoble Alpes après leur isolement et leur identification depuis des prélèvements cliniques de patients du CHU Grenoble Alpes. Les différentes souches cliniques sont principalement issues de prélèvement urinaire et de flacons d'hémocultures reçus au laboratoire pour analyse microbiologique (Tableau 2). Six espèces bactériennes ont été retenues pour la constitution de la base de données : *Enterococcus faecalis*, *Enterococcus faecium*, *Enterobacter cloacae*, *Proteus vulgaris*, *Serratia marcescens* et *Staphylococcus saprophyticus*. Deux critères ont été pris en compte pour leur sélection. Tout d'abord, l'utilisation de phénomène de diffraction d'un LASER rouge nécessite l'utilisation de milieux gélosés transparents. Les espèces sélectionnées devaient donc pouvoir être mises en culture sur des géloses transparentes, dans le cas présent des géloses Tryptone Soy Agar (TSA). Enfin, pour avoir accès rapidement à un maximum de souches cliniques, les espèces bactériennes ont été sélectionnées parmi celles les plus souvent identifiées dans les prélèvements urinaires reçus au laboratoire. *Escherichia coli* est aussi une espèce remplissant ces critères, elle a cependant été exclue de la sélection pour des raisons détaillées plus loin (cf II.D.2 Diffraction LASER et espèces bactériennes).

Toutes les souches cliniques ont été identifiées en routine par spectrométrie de masse MALDI-TOF (Burker Biotyper MicroFlex LT/SH, base de données MBT Compass IVP V9) après mise en culture sur milieu gélosé pendant 18 à 24 heures. L'identification des souches de *Proteus vulgaris* a nécessité un test positif à l'indole en supplément. Après une identification de la souche bactérienne par l'analyse MALDI-TOF comme appartenant au groupe *Proteus vulgaris/penneri/hauseri*, un test à l'indole a été réalisé en utilisant le kit ID indole TDA (bioMérieux). C'est un test colorimétrique rapide qui s'effectue sur une colonie mise en suspension. Il permet, en quelques minutes, de différencier l'espèce *Proteus vulgaris* des espèces *Proteus penneri* et *hauseri*.

Toutes les souches utilisées ont été conservées dans des tubes à billes et stockées à -80°C.

Espèce bactérienne	Type de prélèvement	Nombre de souches
<i>Enterobacter cloacae</i> complexe	Ostéoarticulaire	2
	Ecouvillonnage cutané	1
	Hémoculture	3
	Urine	9
<i>Enterobacter cloacae</i>	Référence (ATCC 13047)	1
<i>Enterococcus faecalis</i>	Urine	15
	Référence (ATCC 29212)	1
<i>Enterococcus faecium</i>	Ostéoarticulaire	1
	Hémoculture	1
	Urine	13
	Référence (CIP 54.32)	1
<i>Proteus vulgaris</i>	Abcès	1
	Ostéoarticulaire	1
	Hémoculture	8
	Prélèvement digestif	2
	Urine	3
	Référence (ATCC 6380)	1
<i>Serratia marcescens</i>	Prélèvement respiratoire	1
	Ecouvillonnage cutané	1
	Hémoculture	9
	Urine	4
	Référence (CIP 103551)	1
<i>Staphylococcus saprophyticus</i>	Urine	15
	Référence (ATCC 35552)	1

Tableau 2 : Liste et origine des différentes souches bactériennes utilisées. Au total 90 souches cliniques et 6 souches de référence ont été utilisées.

Avant d'être utilisées pour expérimentation, toutes les souches ont été remises en culture 24 heures auparavant sur gélose Tryptone Soy Agar (TSA, bioMerieux) et incubées à 37°C en atmosphère supplémentée avec 5% de CO₂. Les souches ont ensuite été repiquées chaque jour sur gélose TSA. Pour éviter toute dérive génétique, les souches n'ont jamais été repiquées plus de 5 fois d'affilée. Passé ce nombre de repiquage, les souches ont à nouveau été remises en culture à partir du tube congelé primaire.

II.B.4. Identification des espèces du complexe *Enterobacter cloacae*

Pour identifier les espèces au sein du complexe *Enterobacter cloacae*, les souches cliniques ont été remises en culture sur gélose TSA et incubées 24 heures à 37°C en atmosphère supplémentée avec 5% de CO₂. L'identification a été réalisée par amplification puis séquençage du gène *hsp60* selon un protocole déjà utilisé au sein du laboratoire (61). A partir de quelques colonies suspendues dans 1 mL de PBS, l'ADN de ces souches a été extrait en utilisant le kit d'extraction NucleoSpin Blood Kit (Macherey-Nagel, Hoerd, France) suivant le protocole recommandé par le fournisseur. L'ADN extrait a été élué dans 50 µL de tampon d'éluion et dosé au Nanodrop N50 (ThermoFischer, USA). Un fragment de 840 paires de base du gène *hsp60* a été amplifié en utilisant le mix Ex Taq HS DNA polymérase (Takara Bio).

500 pmol des amorces suivantes ont été ajoutées à ce mix : HSP60 forward : 5'-GCTATTCAGGGCCGTGTTG-3' et HSP60 reverse : 5'ATGGGCGGCATGATGTAAT-3'. Après une première étape de 10 min à 95 °C, l'amplification des échantillons a été réalisée durant 36 cycles chacun consistant en 30 secondes de dénaturation à 95°C, 30 secondes d'hybridation à 55°C et 1 minute d'élongation à 72°C. Pour finir l'amplification, une étape finale d'élongation a été réalisée pendant 15 minutes à 72°C. Le séquençage des amplicons a été réalisé par Biofidal (Lyon, France). Les séquences forward et reverse ont été identifiées par comparaison avec les génomes des bactéries du complexe *Enterobacter cloacae* en utilisant l'outil d'analyse de séquence BLAST (NCBI). Les résultats avec la plus forte homologie et le plus haut pourcentage de couverture ont été retenus. Dans les cas où cela ne suffisait pas pour différencier deux identifications, l'identification avec le plus grand nombre de nucléotides concordants a été retenue.

II.B.5. Préparation des géloses et acquisition des patterns de diffraction LASER des colonies

Les souches bactériennes, repiquées à partir des tubes congelés, ont été mises en suspension dans du PBS à une densité optique à 600 nm de $0,5 \pm 0,03$ mesurée grâce à un spectrophotomètre (Ultrospec 10, Biochrom). Cette densité optique correspond à une concentration de 10^9 UFC/mL. Cet inoculum initial est ensuite dilué par un facteur de 10 000 pour obtenir une suspension bactérienne de 10^3 UFC/mL. Enfin, 100 μ L des suspensions à 10^3 UFC/mL sont déposés sur géloses TSA et étalés manuellement pour obtenir une centaine de colonies par gélose. Les géloses ont ensuite été incubées dans l'automate BioKube à une température de 37°C en atmosphère aérobie. L'acquisition des patterns de diffraction par le prototype BioKube a été réalisée après 12 et 16 heures d'incubation. Pour obtenir un pattern de diffraction optimal, le diamètre des colonies analysées doit être égal ou inférieur de moins de 2 dixième de millimètres à celui du LASER diffracté (241). Le LASER du prototype BioKube est d'un diamètre de 1 mm et ne peut pas être modifié. Ainsi, les colonies ont été sélectionnées avec un critère de diamètre entre 0,8 et 1 mm. Un critère de circularité inférieure à 0,2 a également été instauré.

II.B.6. Validation de la fonction de compensation

Après une analyse de chaque gélose au bout de 16 heures d'incubation, tous les patterns de diffraction avant et après compensation ont été enregistrés puis analysés grâce au logiciel ImageJ (247). Pour cette analyse, la sélection de tous les pixels de l'image ayant un niveau de gris supérieur ou égale à 40 a été faite. Ce seuil permet de sélectionner le pattern de diffraction tel qu'il est analysé par le programme de compensation du logiciel BioKube. Le pattern est ensuite analysé par le logiciel ImageJ pour en calculer la circularité. Il est important de noter ici que la circularité n'est pas calculée de la même façon par le logiciel ImageJ et le logiciel BioKube. ImageJ calcule la circularité grâce à l'équation suivante : $4 \times \frac{\text{Aire}}{\pi \times (\text{Diamètre max})^2}$. Cette équation donne un nombre entre 0 et 1. Plus la circularité d'une forme ainsi calculée se

rapproche de 1, plus cette forme se rapproche d'un cercle parfait. La distribution des circularités avant et après compensation a ensuite été étudiée. Les analyses statistiques ont été réalisées avec le logiciel Prism (GraphPad, version 9.0.0) en utilisant le test de Student. Les p-values inférieures à 0,05 ont été considérées comme significatives.

II.B.7. Construction des bases de données et développement des modèles de classification

Pour la constitution des bases de données, les différentes souches de référence et cliniques ont été analysées pour l'automate BioKube après 12 et 16 heures d'incubation. Pour permettre un entraînement optimal de l'intelligence artificielle, la base de données doit contenir 1000 images par classe. La base de données après 12 heures d'incubation ne comporte pas de classe pour *Staphylococcus saprophyticus*. En effet, la vitesse de croissance de cette espèce bactérienne est trop lente pour permettre l'obtention de colonies détectables et analysables après seulement 12 heures d'incubation. Après chaque analyse, les patterns remplissant les critères pour être utilisables pour la base de données sont sélectionnés. Ces critères de sélection ont été définis expérimentalement au cours de ce travail de thèse et sont détaillés dans le paragraphe II.C.4.b (Critères de sélection des patterns pour la constitution des bases de données, page 100). Dans le cas des bases de données basée sur les souches de référence, une souche par espèce a été analysée. Les analyses ont été répétées jusqu'à l'obtention de 1000 patterns utilisables par espèce et par temps d'incubation. Pour la base de données basée sur les souches cliniques, 15 souches par espèces ont été analysée. Les analyses de ces souches cliniques ont été répétées jusqu'à obtenir 66 ou 67 patterns de diffraction utilisables par souche soit 1000 patterns par espèce au total. En fonction de vitesse de croissance de chaque souche, les analyses ont été répétées entre une et 15 fois avec à chaque analyse 4 à 16 gélosesensemencées pour chaque souche.

II.B.8. Classification des patterns de diffraction

Une fois acquis, les patterns de diffraction sont analysés par le réseau neuronal convolutif. Pour chaque pattern analysé, le modèle de classification correspondant est sélectionné, en fonction du temps d'incubation et de la volonté d'analyser le pattern selon la base de données issue de souches de référence ou de souches cliniques. Le résultat de cette analyse est donné sous la forme d'une appartenance à l'une des classes de modèle avec un score de pertinence. Les résultats de cette classification sont ensuite comparés à l'identification de l'espèce bactérienne de la souche réalisée en routine dans le cadre du diagnostic de prélèvements de patients. Ces identifications sont réalisées par identification par spectrométrie de masse MALDI-TOF après 18 à 24 heures de culture sur milieu gélosé. Un test à l'indole est effectué pour finaliser l'identification d'une souche en tant que *Proteus vulgaris*. Ces analyses sont réalisées par le personnel technique du laboratoire et validées par un biologiste en tenant compte de tous les critères disponibles (coloration de Gram, morphologie des colonies sur les différents milieux de culture, contexte clinique du patient, score MALDI-TOF > 2,0, etc...). De plus, ces analyses sont réalisées selon les recommandations du Référentiel en Microbiologie Clinique (REMIC, Société Française de Microbiologie (9)) et du Comité Français d'Accréditation (COFRAC). Il a été décidé, en cas de résultat non concordant rendu par l'automate BioKube, de considérer l'identification de routine comme méthode de référence et l'identification par analyse de pattern de diffraction comme fausse.

II.B.9. Représentation graphique des résultats

Les résultats présentés en boxplot ont été réalisés avec le logiciel Prism (GraphPad). Pour tous les boxplots, les extrémités de chaque boîte correspondent aux premiers et troisièmes quartiles, la ligne traversant chaque boîte représente la médiane de la série de données. Enfin, les extrémités supérieures et inférieures correspondent respectivement au maximum et minimum de chaque série.

II.C. Résultats

II.C.1. Analyse du prototype initial

L'étape de rétro-ingénierie de l'automate initial a permis de comprendre son fonctionnement ainsi que les améliorations nécessaires. Son logiciel est codé en langage C sharp. La partie principale fonctionne sur plateforme Windows. En revanche, un second logiciel, en charge de la classification des patterns, est codé pour MATLAB sur Linux. Cette diversité de plateformes de fonctionnement empêche un traitement automatique des patterns recueillis et demande à l'utilisateur une manipulation informatique pour faire la classification. Le flux de fonctionnement physique de l'automate est détaillé dans le paragraphe suivant.

Au niveau du matériel physique, le panneau d'éclairage LED ne permet pas un éclairage homogène de la gélose lors de l'acquisition d'images. Le manque d'homogénéité de l'éclairage créé une zone sombre au centre de la gélose empêchant la détection des colonies à cet endroit. Cette zone de non détection est observable sur la figure 18.

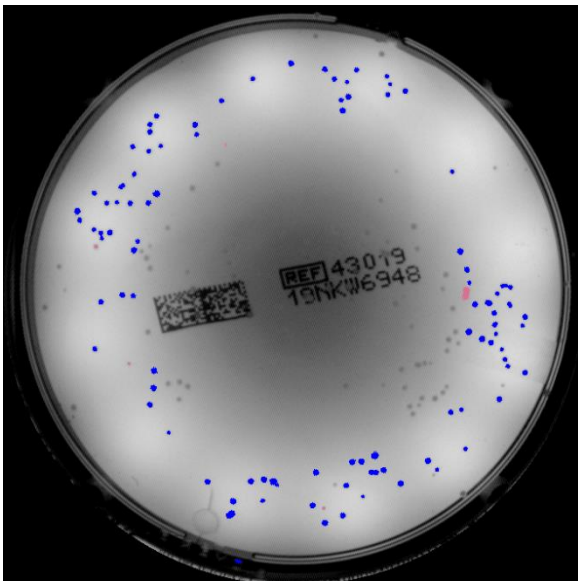


Figure 18 : Illustration de la zone de non détection des colonies due au manque d'homogénéité de l'éclairage LED. Les colonies surlignées correspondent aux colonies détectées, sélectionnées (bleu) ou non (rose) pour analyse par critère de diamètre pour analyse ou non

L'analyse du programme principal en charge de la gestion de l'automate révèle plusieurs points à corriger. Tout d'abord, ce logiciel est adapté d'un ancien automate n'ayant pas la même disposition que le prototype actuel au niveau de la station d'analyse. Cet ancien dispositif avait vraisemblablement une trajectoire rectiligne lors du transfert de gélises entre la zone d'acquisition LED et la station LASER. Or, dans le cas du prototype actuel, le transfert entre ces deux stations est opéré en faisant faire une rotation à la gélise. Si cet angle de rotation n'est pas pris en compte dans le calcul des coordonnées des colonies détectées, le positionnement de ces colonies sous le LASER ne peut pas être correct. La figure 19 illustre la rotation d'une gélise analysée lors de son transfert de la station LED (A) à la station LASER (B).

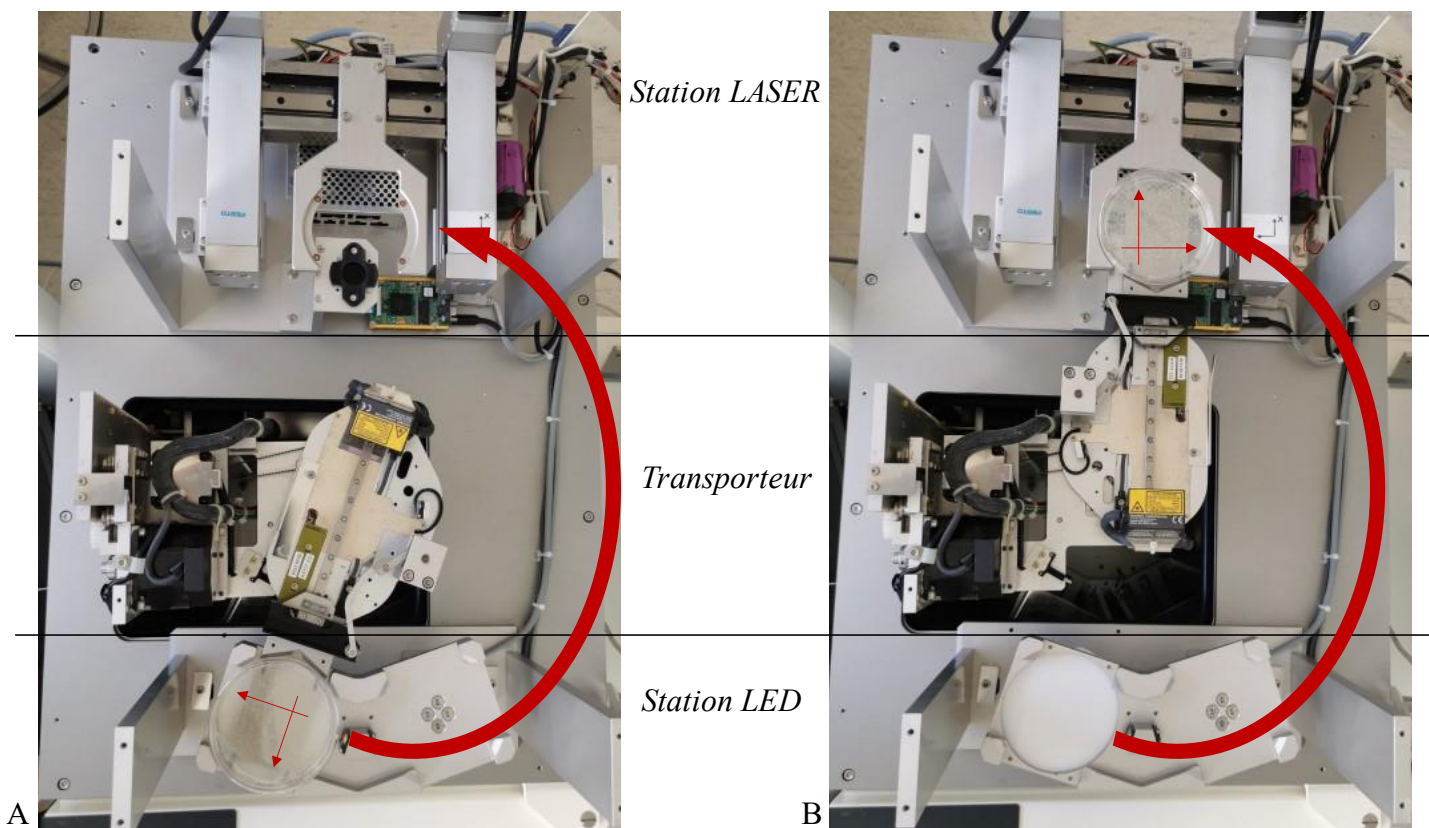


Figure 19 : Illustration de la rotation des gélises durant leur transfert entre la station LED (A) et la station LASER (B). Vue supérieure de la zone d'analyse du prototype BioKube

Par ailleurs, le programme de calibration de l'automate, servant à calculer la corrélation entre pixels sur l'image des géloses analysées et les distances réelles, est basé sur une plaque de calibration percée de trous servant de repères. Cependant, le schéma de répartition de ces trous n'était pas disponible avec l'automate et aucune plaque de calibration n'était à disposition pour le projet. Il est donc nécessaire d'adapter ce programme avec un nouveau schéma de calibration. En l'état, la calibration se fait par une mesure de l'utilisateur du nombre de pixel correspondants à un millimètre. Cette mesure est faite manuellement à partir d'une longueur connue prise en photo sur la station LED de l'automate. La figure 20 illustre la méthode de calcul du coefficient entre distance en pixels et distance en millimètres. A partir d'une image de papier gradué de 0,5 en 0,5 mm, plusieurs segments de 1 cm sont déterminés. Les coordonnées en pixels des points de ces segments sur l'image sont relevées. La distance correspondante en pixel est calculée à partir de la formule suivante en prenant les points $A(x_A, y_A)$ et $B(x_B, y_B)$ extrémités du segment : $d = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$ Ce calcul permet d'obtenir en moyenne une distance de 79,95 pixels arrondie à 80 pixels pour un segment de 10 millimètres, soit un ratio de 8 pixels par millimètres.

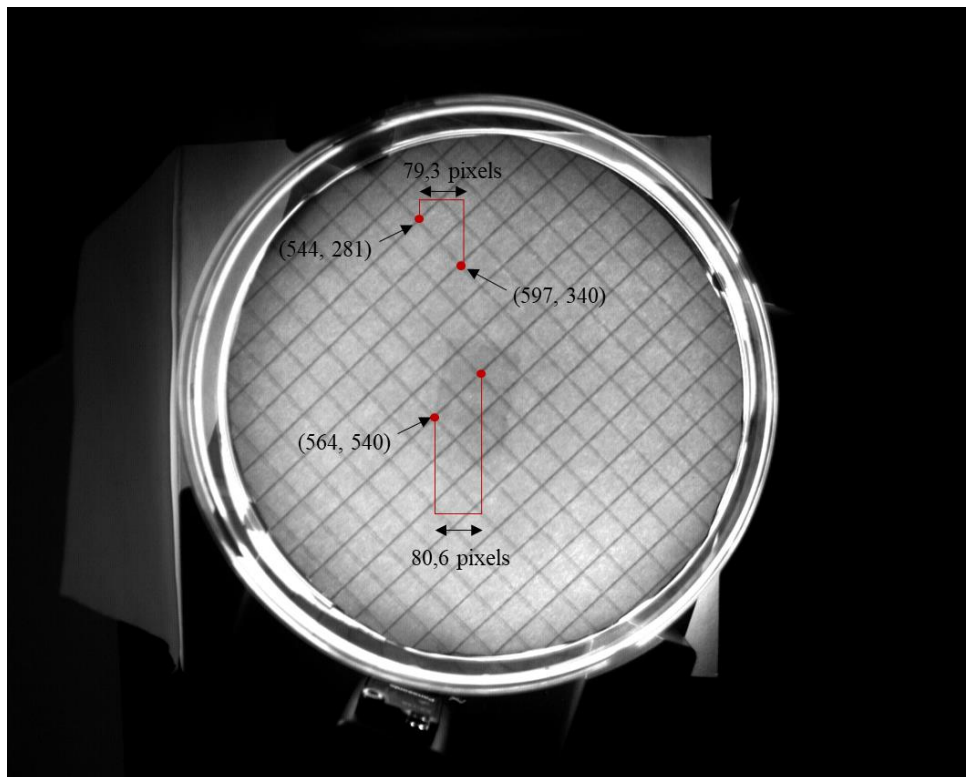


Figure 20 : Estimation du ratio pixels par millimètres à partir d'une image de papier gradué. Chaque graduation correspond à 0,5 mm de distance

En plus, d'une calibration peu précise de l'automate, aucun moyen de vérification ni même de compensation du centrage du LASER sur une colonie après visée n'est disponible avant l'acquisition des patterns de diffraction. Cela cause un positionnement approximatif du LASER par rapport aux colonies analysées et donc une mauvaise acquisition des patterns de diffraction.

Au niveau de la gestion des géloses en cours d'analyse dans l'automate, le programme initial ne dispose pas d'outils permettant la gestion du flux d'analyse. Lorsque l'analyse d'une gélose est prévue alors qu'une autre analyse est déjà en cours, le programme ne peut la mettre en attente le temps que l'analyse en cours se termine. Ainsi la seconde analyse n'est simplement pas effectuée. Il appartient à l'utilisateur de prendre en compte les analyses déjà prévues en termes d'horaire et de durée d'analyse lorsqu'il ajoute une nouvelle analyse pour éviter un chevauchement temporel des acquisitions.

Enfin, le logiciel de classification effectue une série d'extraction de caractéristiques d'images prédéfinie. Il utilise ensuite ces extractions pour comparer l'image à analyser à tous les patterns de la base de données un par un. Le résultat de cette analyse est donné sous la forme de distances d'homologie pour chacun des patterns constituant la base de données. Cette méthode de classification n'est pas basée sur l'utilisation d'une intelligence artificielle. L'extraction et la classification des patterns est lente avec un délai d'analyse d'environ 2 minutes par pattern classifié lors de l'utilisation d'une base de données de 500 patterns de diffraction au total. De plus, l'analyse par l'utilisateur des résultats est complexe car elle nécessite la compilation manuelle de toutes les distances d'homologie.

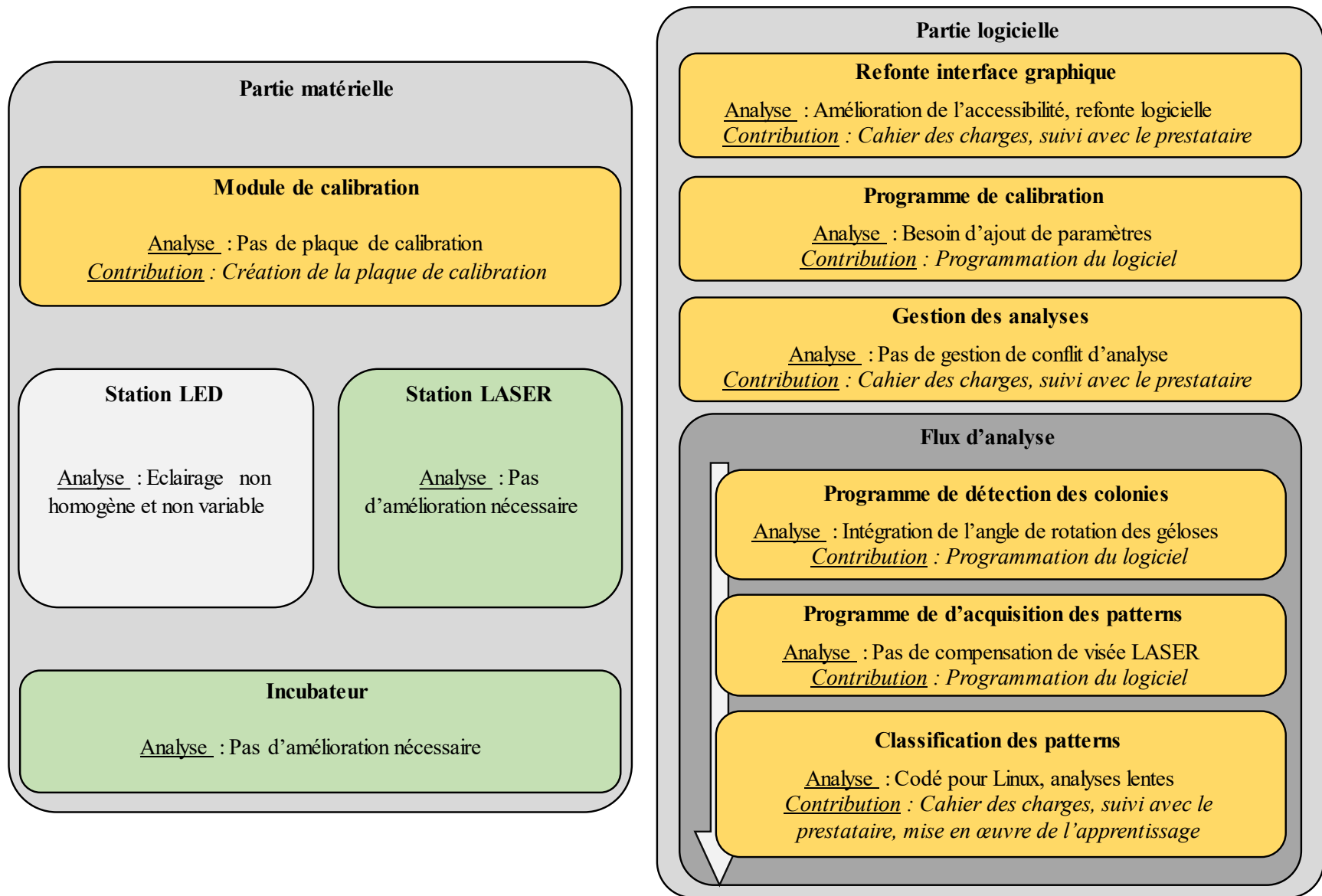


Figure 21 : Synthèse d'analyse du prototype initial et répartition des contributions à son amélioration.
Vert : pas de contribution, Orange : Contributions apportées

II.C.2. Flux de fonctionnement de l'automate BioKube

Le développement d'un automate, d'un logiciel affilié et d'un logiciel hébergeant les modèles de classification pour l'identification bactérienne par diffraction LASER a permis d'aboutir à un prototype stabilisé dont voici le flux de fonctionnement optimisé à ce jour (Figure 22). Lorsqu'une gélose est déposée dans l'automate, elle est transférée et stockée dans l'incubateur, après avoir renseigné manuellement son nom ainsi que le ou les temps d'incubations à attendre avant de réaliser l'acquisition et l'identification des patterns de diffraction des colonies. Une fois la période d'incubation effectuée, la gélose est transportée de l'incubateur à la station LED. A ce niveau, l'éclairage LED est allumé et une image de la gélose est prise. C'est à ce niveau que doit être intégrée une translation des coordonnées en fonction de l'angle de rotation subi par la gélose lors de son transfert vers la station LASER. L'image avant analyses est tournée pour simuler la rotation résultant du transfert de la gélose de la station LED à la station LASER. Le programme utilise alors une analyse des niveaux de gris pour détecter tous les pixels de l'image pouvant correspondre à des colonies. En effet, l'éclairage LED étant sous la gélose tout élément autre que la gélose elle-même apparaissent plus foncés que l'agar. Des niveaux de gris minimum et maximum sont fixés à l'avance par l'utilisateur. Ces bornes de niveaux de gris sont utilisées pour la détection des colonies, leur modification par l'utilisateur peut permettre d'adapter cette détection en fonction de l'opacité de la gélose et donc de la couleur de « fond » de l'image acquise. Cette image est définie informatiquement par une matrice de pixels, chacun associé à une valeur d'intensité. Cette intensité est représentée par des nombres allant de 0 à 255, 0 correspondant au noir et 255 au blanc. Pour les acquisitions présentées tout au long des résultats, la gamme de gris considérée comme pouvant correspondre à une colonie à détecter est de 25 à 195. Le logiciel analyse donc un à un tous les pixels de l'image pour définir lesquels ont un niveau de gris étant dans la gamme fixée. Par la suite, le logiciel utilise cette liste de pixels pour définir toutes les zones de l'image qui peuvent correspondre à des objets étant potentiellement des colonies. A partir d'un pixel dans la gamme de gris désirée, le logiciel associe tout pixel étant directement à son contact et qui possède également une intensité dans la gamme de gris demandée. Ainsi toutes les zones de l'image ayant des intensités de gris recherchées sont définies et étiquetées.

Le logiciel informatique définit ensuite les contours de ces zones et calcule leur aire, la position de leur centroïde et leur circularité. Un centroïde, ou centre de masse, est le point de centre géométrique de toutes figures quel que soit leur forme. Ce point correspond au centre pour un cercle. Dans le cas d'une figure géométrique en deux dimensions, les coordonnées X et Y du centroïde correspondent à la moyenne des coordonnées x et y de tous les points de la figure géométrique. Le calcul de la circularité d'une forme permet de définir si cette forme est proche ou non d'un cercle. Dans le cas ici présenté, la circularité est calculée de la manière suivante. A partir de l'aire des zones, un rayon théorique est calculé en supposant que ces zones sont parfaitement circulaires. Par la suite, le logiciel calcule pour chaque point du contour de la forme la distance entre ce point et le centroïde de la forme, son « rayon » réel. La variance entre chaque « rayon » réel et le rayon théorique de la forme est calculée. Le rapport entre la moyenne de ces variances et le rayon théorique de la forme correspond à la circularité de la forme. Plus ce rapport sera proche de 0, plus les différents rayons réels auront une taille homogène proche de celle du rayon théorique et plus la forme étudiée sera similaire à un cercle. A ce stade, le logiciel a détecté toutes les zones sombres de l'acquisition LED de la gélose. Leur aire, diamètre et circularité sont aussi connus. Dès lors, il faut utiliser ces paramètres pour pouvoir différencier des colonies d'autres objets ou artefacts présents sur la gélose tels que des écritures, des codes-barres ou des défauts de gélose. Le logiciel peut filtrer la liste d'éléments détectés grâce à des critères de diamètre et de circularité fixés par l'utilisateur pour ne retenir que des colonies ayant une taille permettant l'acquisition de leur pattern de diffraction. Les colonies bactériennes ayant en général des formes ovoïdes, le critère de circularité permet ainsi d'éliminer toute forme pouvant être notamment des chiffres ou des codes-barres imprimés au fond de la gélose. Le critère de diamètre permet quant à lui de limiter l'analyse à des colonies ayant des tailles proches et légèrement inférieures au diamètre du LASER (ici un diamètre entre 0,8 et 1 mm) pour pouvoir acquérir un pattern de diffraction.

Une fois cette sélection faite, le logiciel conserve les coordonnées des centroïdes de toutes les colonies détectées pour permettre au bras LASER de les cibler lors de l'acquisition des patterns de diffraction. Une fois la gélose transférée sur la station LASER, le bras automatisé positionne chaque colonie dans le trajet LASER grâce aux centroïdes communiqués. La caméra située sous la gélose acquiert les patterns de diffraction de chaque colonie pour une identification de l'espèce bactérienne par le modèle de classification.

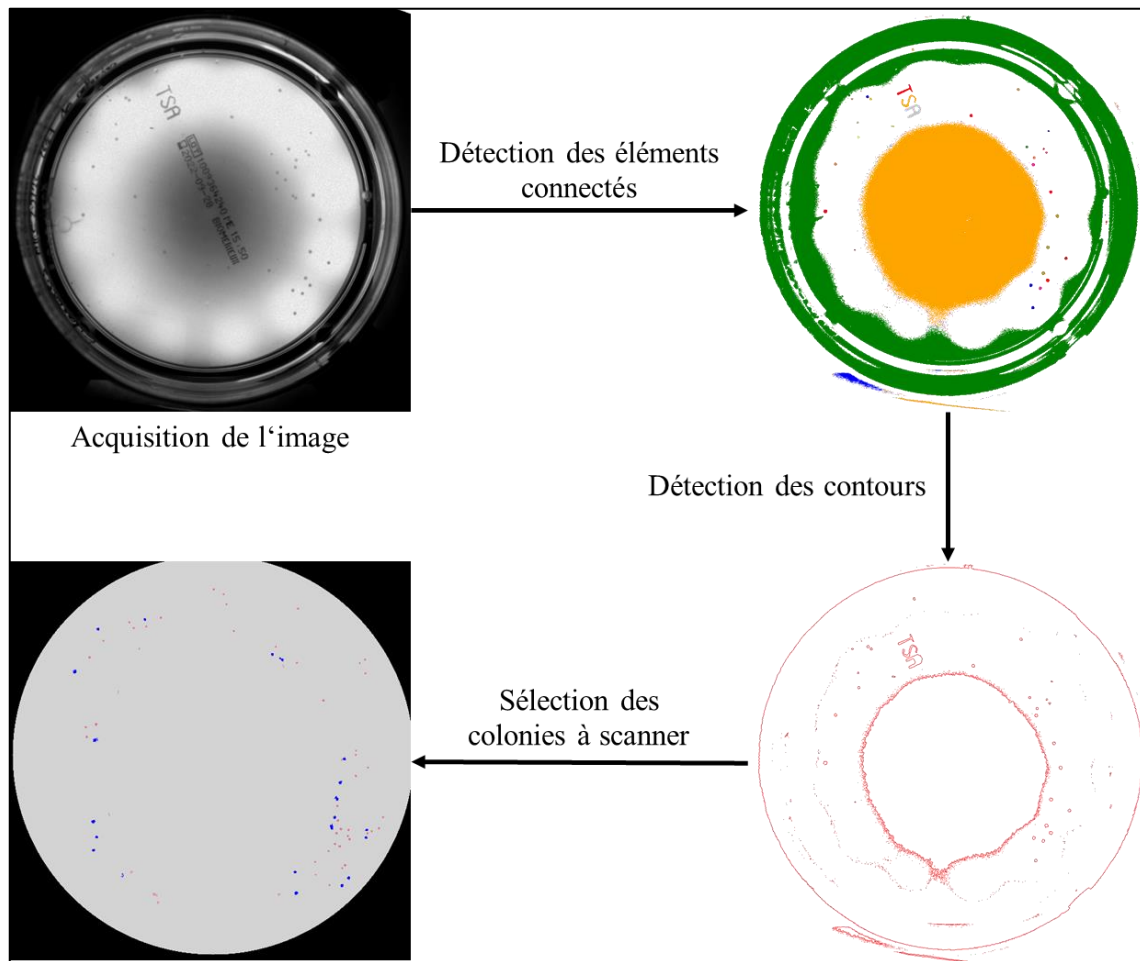


Figure 22 : Processus de détection des colonies

II.C.3. Développement logiciel

II.C.3.a. Programme de calibration

Afin de pouvoir faire fonctionner correctement les programmes de détection des colonies et d'acquisition des patterns, le logiciel doit connaître un certain nombre de paramètres physiques de la station d'analyse de l'automate. Pour ce faire, un programme de calibration est nécessaire. Suite aux conclusions de l'analyse du prototype initial, le programme de calibration déjà existant a été modifié pour permettre une acquisition semi-automatique et précise de ces paramètres.

Le programme originel de calibration ne comprenait que le calcul des coefficients de la fonction permettant de calculer des distances réelles à partir de distances en pixels. De plus, ce calcul se base sur de coordonnées de points précis sur une plaque de calibration. Or, le schéma de la plaque de calibration initialement développée était inconnu au démarrage du projet. L'adaptation du programme de calibration a donc consisté en le développement d'une nouvelle plaque de calibration ainsi qu'en l'ajout de calcul de paramètres supplémentaires. Ces paramètres sont l'angle de rotation appliqué aux géloses lors de leurs transfert, ainsi que la position du centre du LASER sur les images issus la caméra de la station LASER. Cette adaptation a été réalisée par le doctorant sur Visual Studio 2019 (Microsoft)

Tout d'abord, les coordonnées des centroïdes des colonies détectées sont donnés initialement en pixels par rapport à l'image LED de la gélose. Cependant, le bras robotisé situé au niveau de la station LASER requiert des coordonnées en millimètres pour pouvoir déplacer en conséquence la gélose. Le programme doit donc connaître la corrélation entre la distance en pixels entre deux points sur l'image LED et la distance réelle qui sépare ces points sur la gélose. Pour ce faire, un programme de calibration ainsi qu'une plaque de calibration ont été mis au point. La plaque de calibration est un cylindre plein aux dimensions d'une gélose. Elle est percée de sept trous de 1 mm de diamètre. Ces trous sont positionnés suivant un schéma précis qui permet de connaître exactement les distances qui les séparent. Ici, le schéma est une forme en croix centrée sur le centre de la plaque de calibration. Sur deux branches de cette croix, deux

points supplémentaires sont rajoutés en périphérie. Le schéma exact de la plaque de calibration a été mis au point au cours de la thèse et est représenté sur la figure 23.

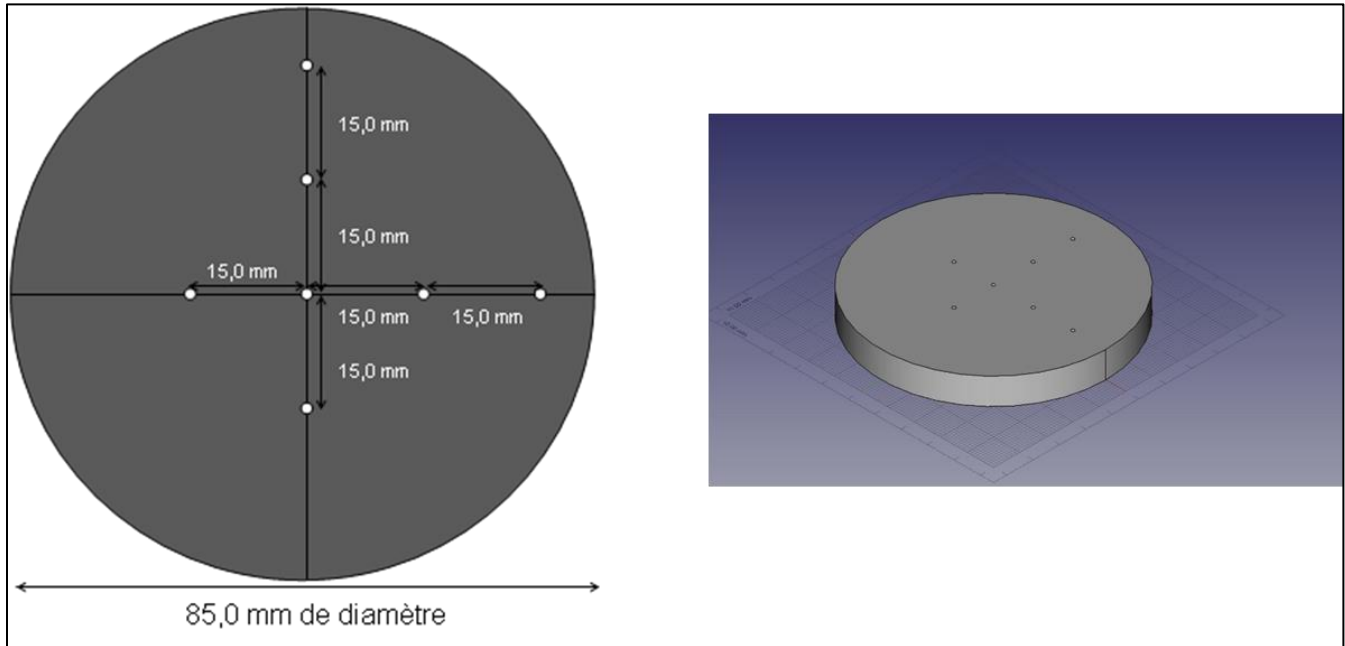


Figure 23 : Plan et simulation 3D de la plaque de calibration de l'automate BioKube

Cette plaque de calibration est ajoutée dans l'automate comme une gélose à analyser. Après le lancement du programme de calibration, la plaque est transférée sur la station LED et une image est prise. A partir de cette image, il est demandé à l'utilisateur de sélectionner dans un ordre précis les différents points de cette plaque. La distance réelle entre les différents points étant connu par le logiciel, ce dernier peut à partir des coordonnées de ces mêmes points sur l'image déterminer les facteurs de l'équation permettant de calculer une distance réelle à partir d'une distance en pixel. Cette équation est, dans le cas présent, une équation du second degré : $y = ax^2 + bx + c$. Bien qu'une équation du premier degré suffise pour corréliser une distance sur une image à une distance réelle, l'ajout du terme du second degré permet de prendre en compte également un potentiel manque de parallélisme entre la caméra et la gélose lors de l'acquisition. En effet, si le plan d'acquisition de la caméra et la gélose ne sont pas parfaitement parallèles, certaines zones de la gélose seront plus proches de la caméra que d'autres créant ainsi un

phénomène de zoom à certains endroits. Dans ce cas, une même distance réelle peut ne pas avoir la même distance sur l'image suivant où elle est située sur la gélose.

L'application d'une équation du second degré permet de prendre en compte ce phénomène potentiel. Après l'analyse des coordonnées de la plaque de calibration, le logiciel déduit donc les facteurs a , b et c de l'équation de second degré et peut ainsi utiliser cette équation dès que nécessaire pour transformer des distances et coordonnées en pixels en distances et coordonnées réelles.

Un deuxième point nécessitant un paramètre à calibrer est l'angle de rotation de la gélose entre la station LED et la station LASER. En effet, lors du transfert de la gélose entre ces deux stations, le bras automatisé, après avoir pris la gélose, effectue une rotation pour atteindre la seconde station. La gélose est donc tournée d'un certain angle pendant ce trajet. Comme expliqué précédemment, le logiciel transfère les coordonnées de centroïdes des colonies détectées à partir de l'image LED au bras LASER. Si ces coordonnées ne sont pas translatées par une rotation similaire à celle réalisée en réalité par le bras automatisé, ils ne correspondront plus aux positions réelles des colonies au niveau de la station LASER.

Le calcul de cet angle est donc la deuxième étape du programme de calibration. La plaque de calibration est transférée à la station LASER et il est demandé à l'utilisateur de déplacer le bras automatisé pour positionner, tour à tour et dans un ordre précis, les différents points de la plaque de calibration dans le trajet LASER. Les coordonnées de ces points sont alors enregistrées. Le logiciel ayant en mémoire la position des points de calibration au départ (au niveau de la station LED) et à la fin de la rotation (au niveau de la station LASER), il est ainsi capable de calculer l'angle avec lequel la gélose a été tournée. Cet angle peut ainsi être utilisé à chaque analyse de gélose.

Enfin, pour pouvoir faire fonctionner le programme de compensation présenté dans le paragraphe suivant, le logiciel doit connaître la position du centre du faisceau LASER sur les images prises par la caméra LASER. Pour cela, une image est prise par la caméra LASER du faisceau lumineux sans présence de gélose au milieu. Le faisceau LASER non diffracté est alors observé. Cette image est ensuite analysée par le même procédé que celui de détection de colonies pour obtenir le centroïde de la tâche LASER et donc la position du centre du faisceau lumineux sur la caméra LASER.

II.C.3.b. Programme de compensation de la visée du LASER

Même si la calibration de l'automate paraît correcte, une certaine marge d'erreur est toujours présente entre la position visée par le LASER et le centre réel des colonies analysées. Cependant, pour pouvoir acquérir les patterns de diffraction correctement, il est important que le LASER soit parfaitement centré avec la colonie. En effet, si le centre du faisceau lumineux n'est pas sur le centre de la colonie, des déformations du pattern de diffractions sont observées. Après plusieurs acquisitions sur une même colonie en décalant légèrement le LASER de 0,2 mm dans toutes les directions, il a été observé que lorsque le pattern se déforme, le vecteur entre le centre du pattern et son point le plus éloigné semble donner une information sur la direction où se situe le centre de la colonie (Figure 24). De plus, plus le LASER est centré sur le centre de la colonie, plus le pattern de diffraction de cette dernière a une forme circulaire. Il est donc possible en analysant le pattern de diffraction, de déterminer dans quelle direction déplacer le LASER pour le positionner sur le centre de la colonie. Ces deux constats ont donc été utilisés pour définir un procédé automatique de repositionnement du LASER pour qu'il soit aligné au maximum avec le centre des colonies analysées. Ainsi un programme de compensation LASER a été écrit et ajouté au logiciel de l'automate par le doctorant.

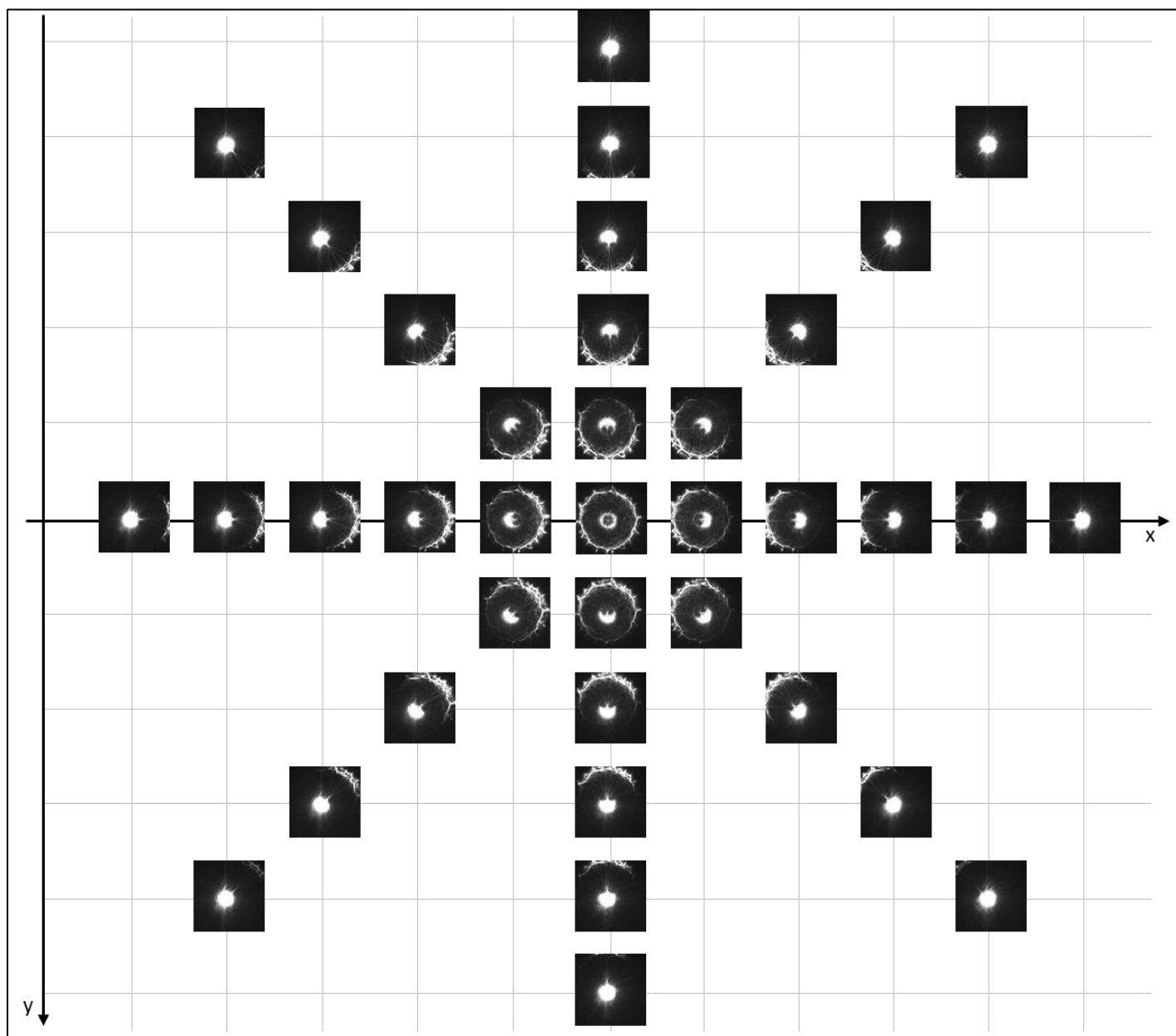


Figure 24 : Evolution du pattern de diffraction d'une colonie bactérienne en fonction de la position de LASER par rapport au centre de la colonie observée. Dans le cas présenté, les patterns de diffraction sont ceux d'une colonie de *Serratia marcescens* CIP 103551 après 16 heures d'incubation. Chaque graduation correspond à une distance de 0,2 mm.

Ce programme analyse le premier pattern de diffraction recueilli par l'automate. Dans un premier temps, la circularité du pattern est calculée. Si cette dernière est considérée comme inférieure au seuil fixé, le programme considère le LASER comme suffisamment centré et utilise le pattern pour la phase d'identification. Dans le cas contraire, le pattern est analysé pour définir dans quelle direction déplacer le LASER. Pour ce faire, le point du pattern le plus distant de son centre est déterminé. Les coordonnées de ce point permettent ainsi de calculer le vecteur entre le centre du pattern et son point le plus éloigné. Ce vecteur permet de définir la direction du mouvement de compensation du LASER. Pour définir la distance de mouvement du LASER, la longueur du vecteur est calculée puis multipliée par un facteur défini par l'utilisateur. La distance de mouvement va influencer sur la capacité de compensation du logiciel. En effet, si cette distance est trop faible, la compensation sera trop faible pour atteindre le centre de la colonie. Au contraire, si la distance de mouvement est trop grande, le mouvement de compensation dépassera le centre de la colonie. Cette distance a donc besoin de pouvoir être modulée par l'utilisateur pour affiner la compensation.

Une fois cette étape de compensation effectuée par le logiciel, un nouveau pattern de diffraction est acquis et analysé et le programme de compensation se répète. Soit le pattern répond au critère de circularité (dans les travaux présentés ici une circularité inférieure à 0,3) et est utilisé pour analyse soit la position du LASER est de nouveau corrigée. Pour éviter des temps d'analyses trop long et une potentielle boucle infinie de compensation dans le cas où le logiciel ne réussisse pas à positionner correctement le LASER, le nombre maximal d'essai de compensation est fixé à 8 dans le logiciel. Une fois ce nombre d'essai de compensation atteint, le dernier pattern de diffraction est conservé, peu importe sa qualité, pour identification.

Afin d'étudier l'impact de la fonction de compensation précédemment décrite, la circularité des patterns de diffraction d'une même colonie avant et après compensation ont été calculées et comparées. Pour cela chacune des 6 souches de référence des 6 espèces bactériennes choisies initialement pour la création de la première base de données du projet (*Enterococcus faecalis*, *Enterococcus faecium*, *Enterobacter cloacae*, *Proteus vulgaris*, *Serratia marcescens* et *Staphylococcus saprophyticus*) ont été ensemencées sur gélose TSA et analysée par l'automate BioKube après 16h d'incubation. Pour 30 colonies de chaque gélose, les patterns de diffraction avant et après ont été récupérés et analysés en utilisant le logiciel ImageJ pour calculer leur circularité. Les distributions avant et après compensation des différents patterns sont représentées pour chaque espèce dans la figure 25.

La circularité moyenne des patterns avant compensation est de 0,69. Après compensation, les patterns de diffraction ont une circularité moyenne de 0,90. Le logiciel de compensation permet donc une amélioration significative de la visée des colonies pour l'acquisition de leur pattern de diffraction. Il permet ainsi un meilleur alignement du LASER avec le centre de la colonie observée et donc une homogénéisation de la qualité des patterns.

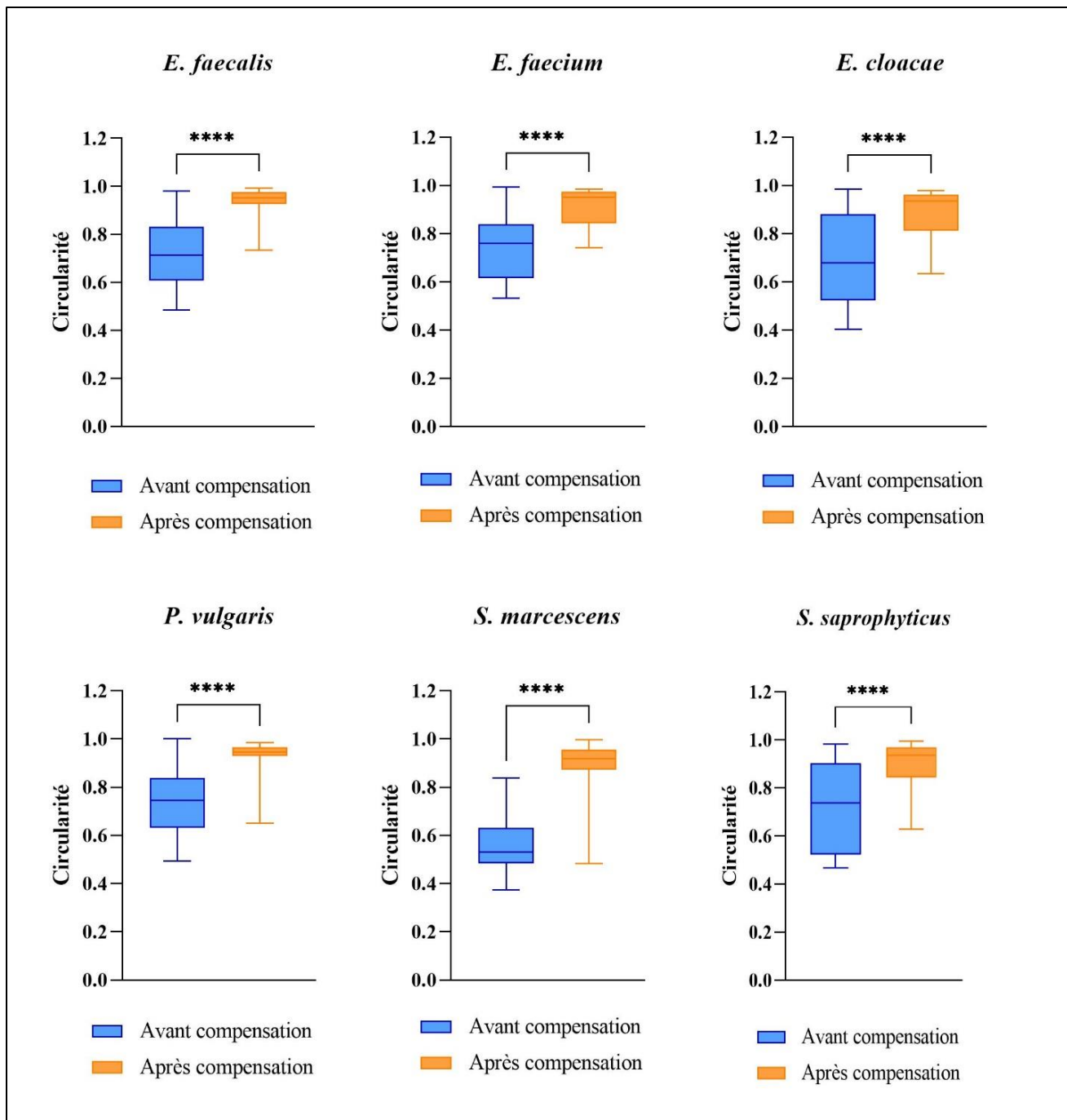


Figure 25 : Répartition des circularités des patterns de diffraction avant et après compensation de la position du LASER par rapport au centre des colonies observées. Les patterns acquis pour ces mesures sont ceux de colonies après 16 heures d'incubation pour les souches suivantes : *Enterococcus faecalis* ATCC 29212, *Enterobacter cloacae* ATCC 13047, *Enterococcus faecium* CIP 5432, *Proteus vulgaris* ATCC 6380, *Serratia marcescens* CIP 103551 et *Staphylococcus saprophyticus* ATCC 35552. **** : $p < 0,0001$. Caractéristiques des boxplots décrites en II.B.9, page 79

II.C.3.c. Développement d'une nouvelle interface graphique

Une refonte complète de l'interface graphique du logiciel a été opérée en collaboration avec la société Revertis. En effet, pour pouvoir utiliser au mieux l'automate BioKube, certaines fonctionnalités devaient être ajoutées au programme initial. La société Revertis a réalisé, sous la supervision de doctorant, le développement de l'interface graphique ainsi que l'intégration du nouveau module de classification

Tout d'abord, l'interface de gestion des analyses en cours a été modifiée (Figure 26 et 27). Ces modifications permettent de gérer dès l'entrée d'une gélose à analyser son positionnement dans les cassettes de l'incubateur en fonction des places disponibles mais également les temps d'incubation avant analyse ainsi que les horaires d'analyse. Pour compléter la gestion des analyses, une base de données relationnelle codée en SQL (Structure Query Language) a été implémentée au programme. Une base de données relationnelles est une base de données contenant des ensembles de tableaux à deux entrées. Elle permet la gestion des relations entre données. Dans le cas présent, elle est utilisée pour la gestion des géloses en cours d'analyse dans l'automate, particulièrement pour la hiérarchisation des horaires d'analyses. Cela permet de gérer un conflit quand deux analyses se chevauchent et de mettre l'une des deux en attente le temps que la première se termine.

Pour pouvoir gérer les réglages de l'automate de façon plus simple, une interface de réglages a été ajoutée. Elle permet la gestion de tous les paramètres modifiables du logiciel de manière facile et intuitive. Les versions précédentes du logiciel ne permettaient la modification des réglages que par modification du code informatique lui-même. Cette interface permet donc de définir les paramètres de détection de colonies (niveaux de gris, taille et seuil de circularité acceptée), les paramètres des caméras (gain, temps d'exposition). Les paramètres de calibrations restent, quant à eux, modifiables uniquement par le programme de calibration et non par l'utilisateur.

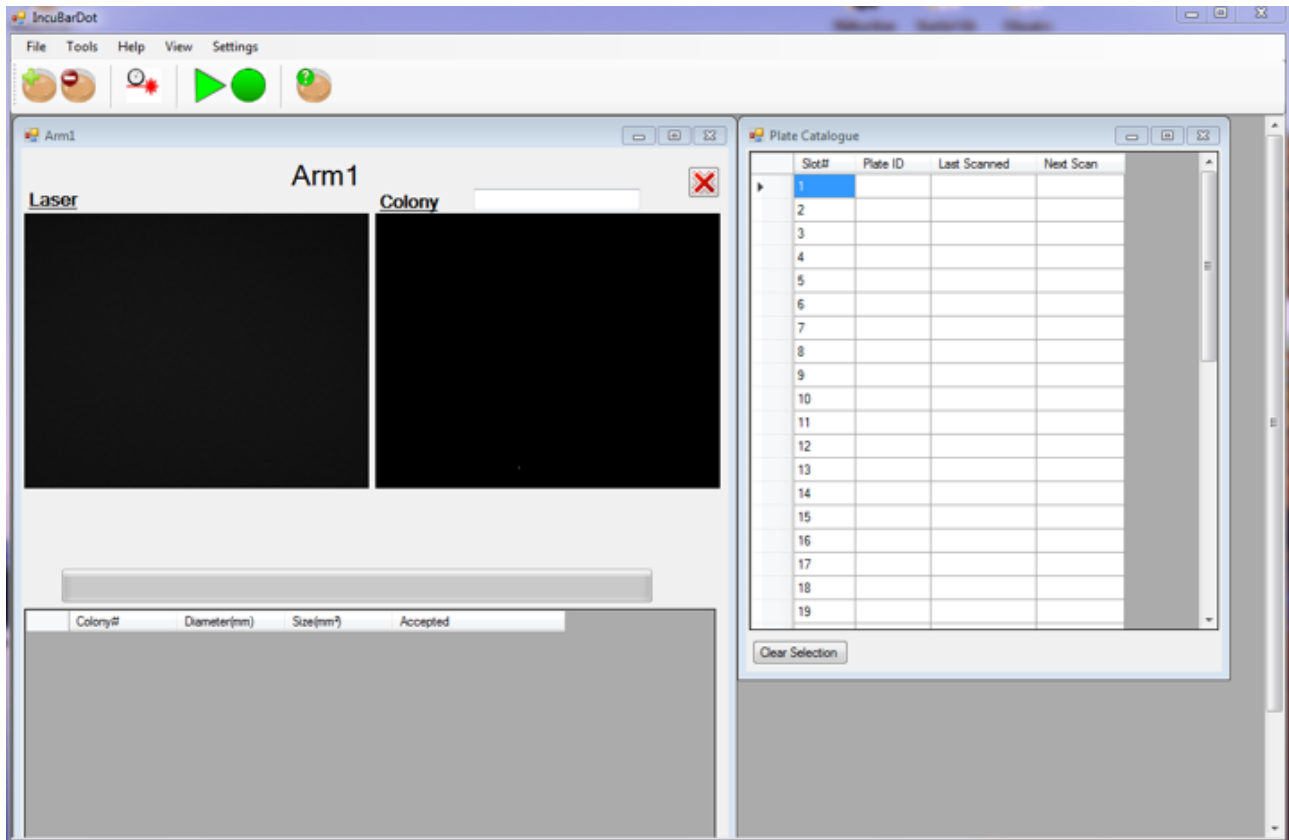


Figure 26 : Ancienne interface utilisateur du logiciel BioKube

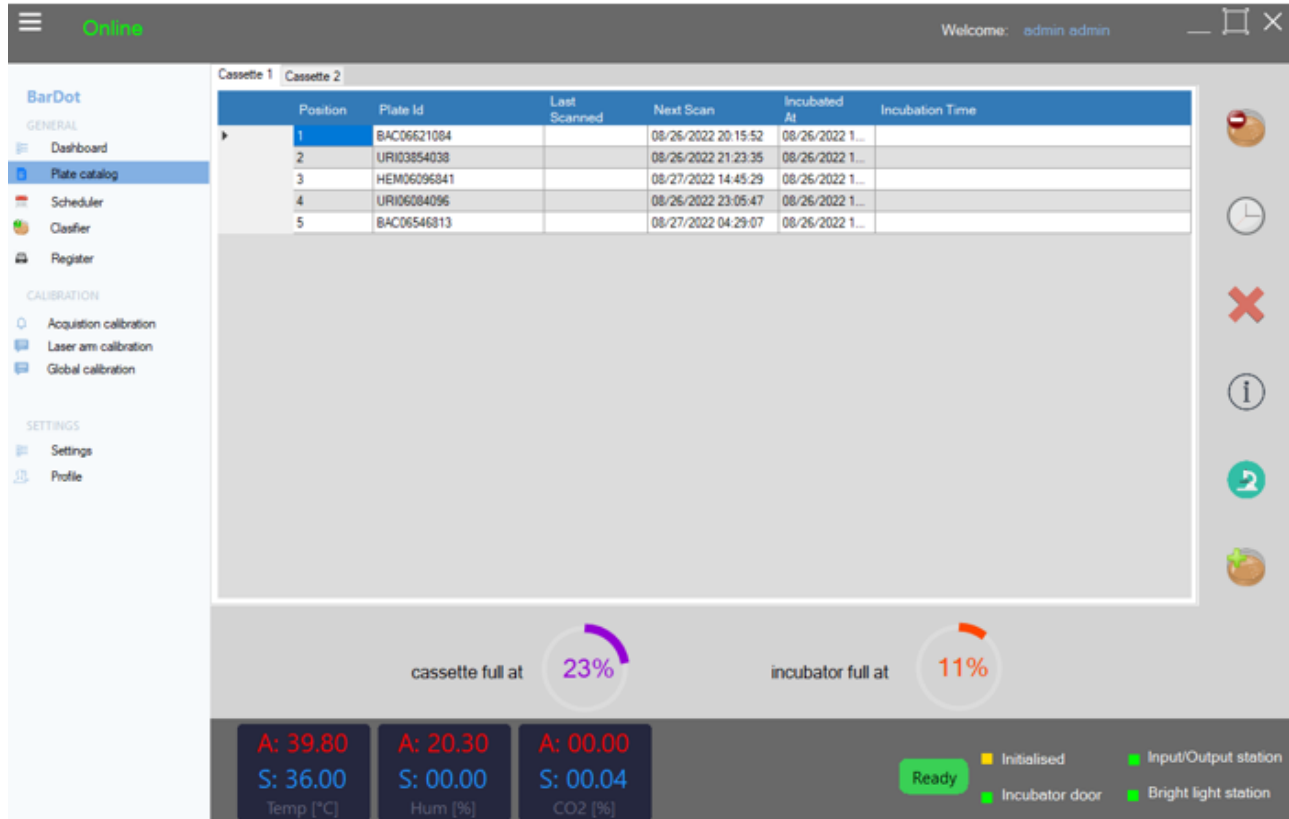


Figure 27 : Fenêtre de gestion des analyses en cours de la nouvelle interface utilisateur

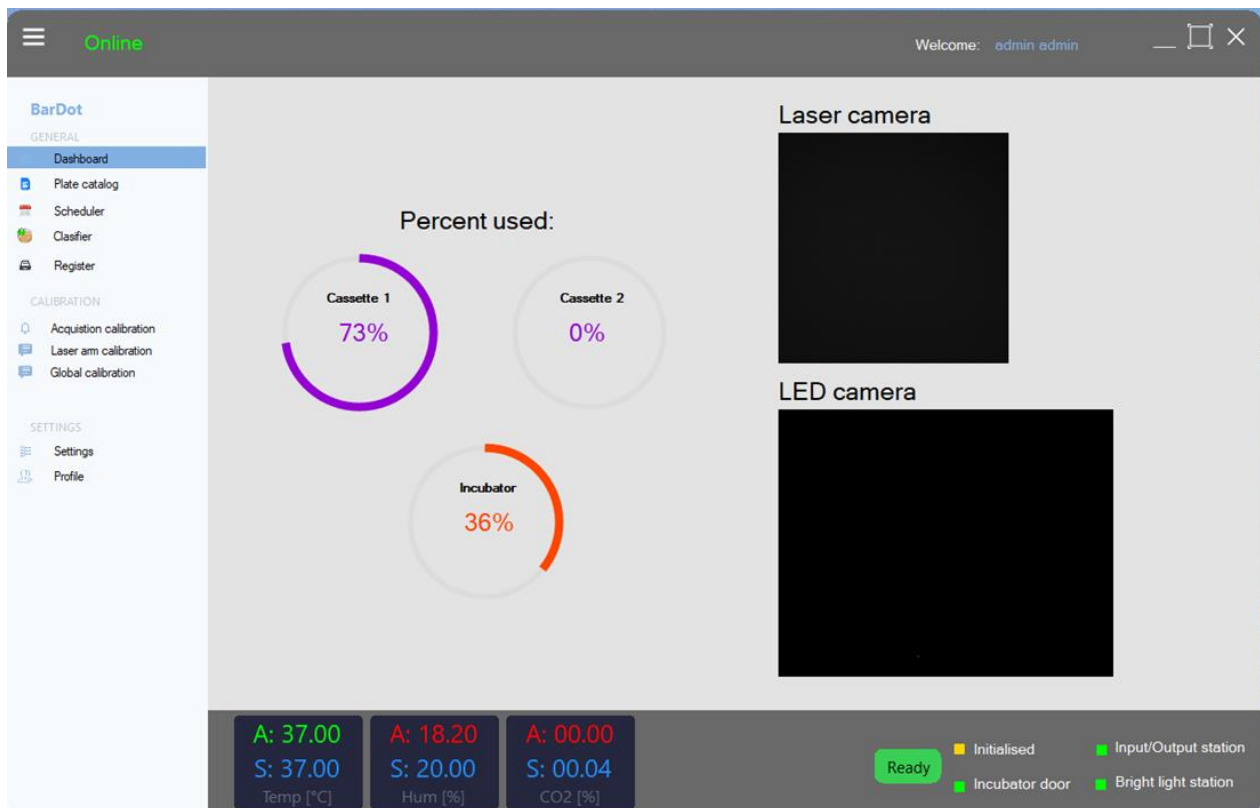


Figure 28 : Nouvelle fenêtre d'accueil du logiciel BioKube

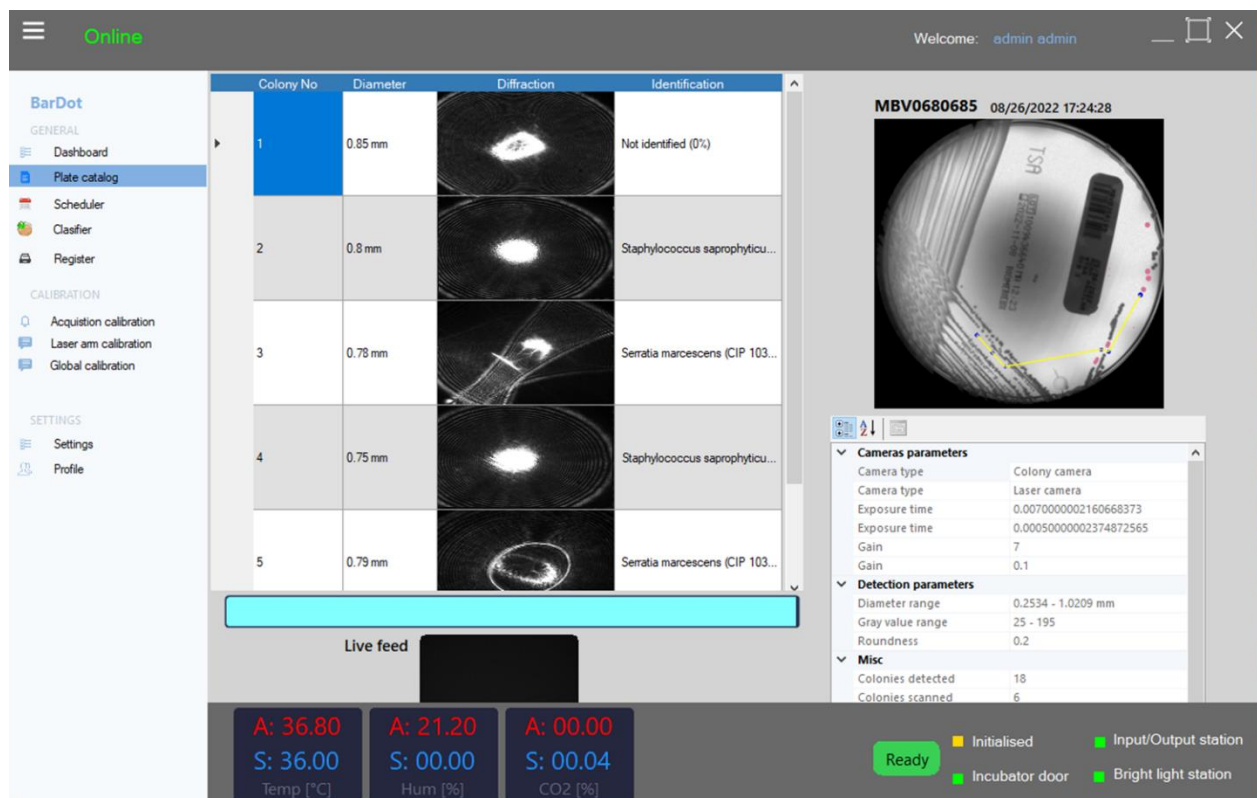


Figure 29 : Fenêtre de résultats d'analyse de la nouvelle interface utilisateur

Les développements apportés au logiciel permettent donc une utilisation plus fiable du prototype BioKube pour l'acquisition de patterns de diffraction de colonies bactériennes. Les travaux effectués sur le module de classification ne sont pas détaillés ici car son développement et son intégration ont été réalisés par la société Revartis. Seule la validation de son fonctionnement sur les bases de données a été réalisée par le doctorant et est exposée dans les paragraphes suivants. Une fois l'obtention d'un prototype fonctionnel, il est nécessaire de déterminer la méthodologie d'acquisition des patterns de diffraction avant leur utilisation pour la constitution des bases de données.

II.C.4. Acquisition des patterns

II.C.4.a. Choix des temps d'incubation pour l'établissement des bases de données de spectre de diffraction

Pour établir les temps d'incubation après lesquelles les bases de données ont été constituées, des acquisitions de patterns de diffraction ont été réalisées toutes les heures sur les six espèces de références. La figure 30 montre des exemples de patterns de diffraction pour ces espèces. Ces patterns montrent l'évolution heure par heure des patterns de diffraction pour une même espèce.

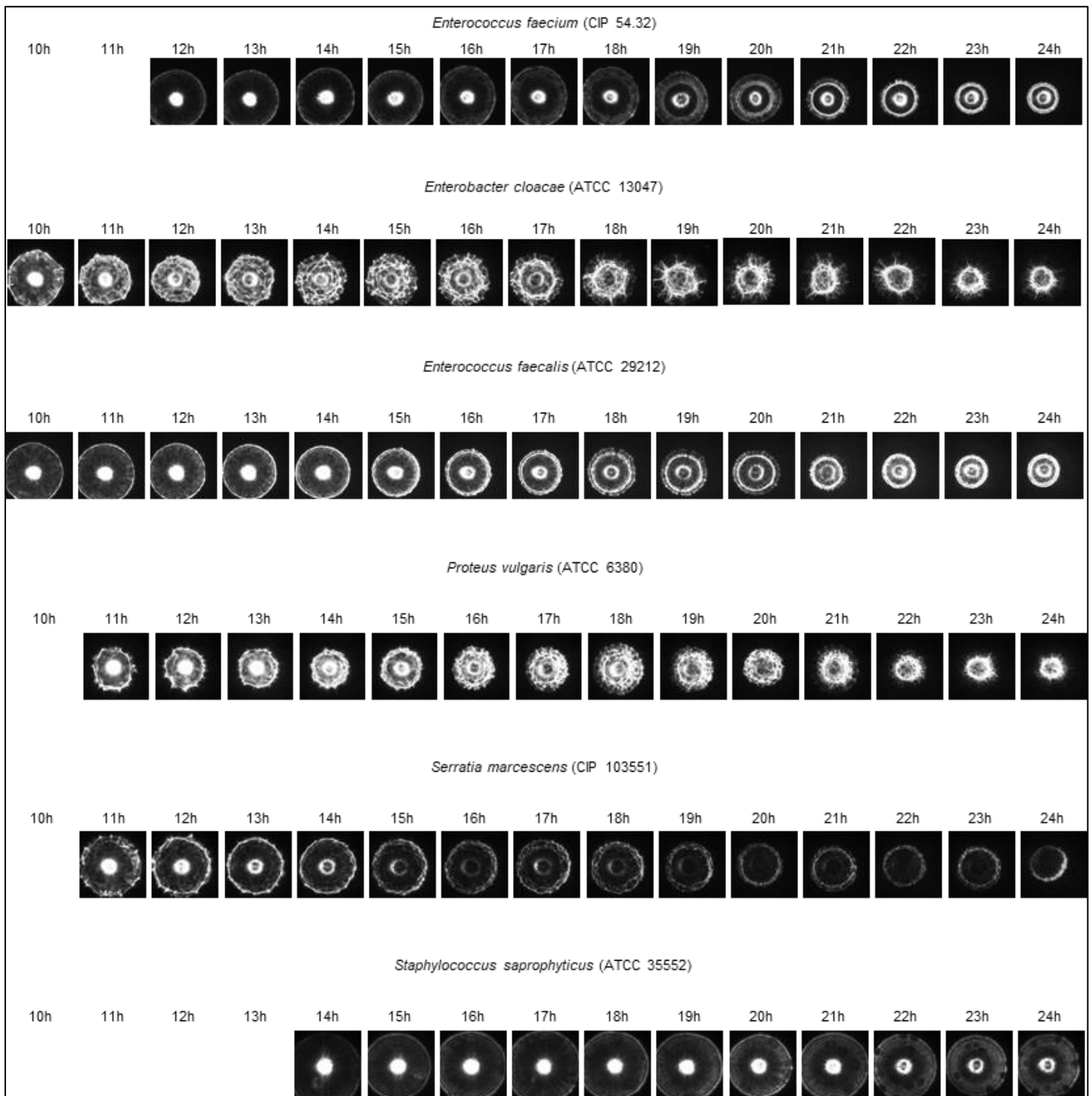


Figure 30 : Evolution de pattern de diffraction au cours de 24 heures d'incubation

Ces observations ont permis d'identifier le temps d'incubation minimal requis pour l'obtention de colonies de taille suffisante pour leur analyse. Ce temps est variable en fonction de la vitesse de croissance des bactéries. Il est de 10h pour *Enterobacter cloacae*, *Enterococcus faecalis*, de 11 h *Proteus vulgaris* et *Serratia marcescens* et de 12h pour *Enterococcus faecium*. Pour *Staphylococcus saprophyticus*, ce temps est sensiblement plus long et se situe entre 14 et 15 heures. De plus, on constate une diminution de la taille des patterns de diffraction au cours du temps. Aux vues de l'évolution des patterns de diffraction au cours du temps pour une même espèce, il est apparu nécessaire de créer des bases de données pour un temps d'incubation particulier. Voulant démontrer la capacité d'identification précoce de l'analyse de patterns de diffraction, il a donc été choisi de constituer des bases de données pour deux temps d'incubation : 12 heures, temps pour lequel seules 5 des 6 espèces étudiées feront partie de la base de données (toutes excepté *Staphylococcus saprophyticus*) et 16 heures où les six espèces seront incluses dans la base de données.

II.C.4.b. Critères de sélection des patterns pour la constitution des bases de données

Afin de pouvoir constituer des bases de données pertinentes, il est important d'établir des critères de sélection pour les patterns qui constitueront ces banques d'images. L'utilisation de patterns de diffraction de colonies bactériennes étant une approche nouvelle, peu d'informations existent dans la littérature sur l'évaluation de pattern de diffraction bactérien. Par ailleurs, la corrélation entre des facteurs biologiques et phénotypiques des bactéries et la forme de leur pattern de diffraction est encore inconnue, il est donc impossible de cibler quelle partie ou quelle composante du pattern est impliquée dans la discrimination d'espèces bactériennes. De plus, il est important de conserver pour une espèce donnée (considérée comme une classe dans la base de données) une hétérogénéité maximale des patterns pour que l'intelligence artificielle dispose d'images englobant le maximum des variations de patterns pouvant être présentes au sein d'une même espèce. La présence de cette hétérogénéité au sein d'une même classe de la base de données permet à l'intelligence artificielle de ne pas prendre en compte des éléments non pertinents pour la classification d'espèces bactérienne, éléments qui seraient simplement dus à cette variabilité des patterns obtenus pour une même souche ou espèce. En revanche, éliminer

des patterns selon un critère défini par le concepteur de la base de données, enlève fatalement ce critère de la liste des critères potentiellement utilisables par l'intelligence artificielle pour l'établissement d'un modèle de classification. Ainsi un seul critère de sélection a été défini et est basé sur le fait que l'un des critères premiers utilisé par l'intelligence artificielle pour son entraînement est le contour des formes présentes dans l'image. Or, dans le cas où deux colonies bactériennes sont physiquement proches sur la gélose analysée, il est possible d'observer le pattern de diffraction de l'une se superposer partiellement à celui de la seconde lors de l'acquisition de son pattern de diffraction (Figure 31). L'apparition de ces phénomènes sur une acquisition de patterns a donc été considérée comme un facteur d'élimination du pattern pour la constitution des bases de données.

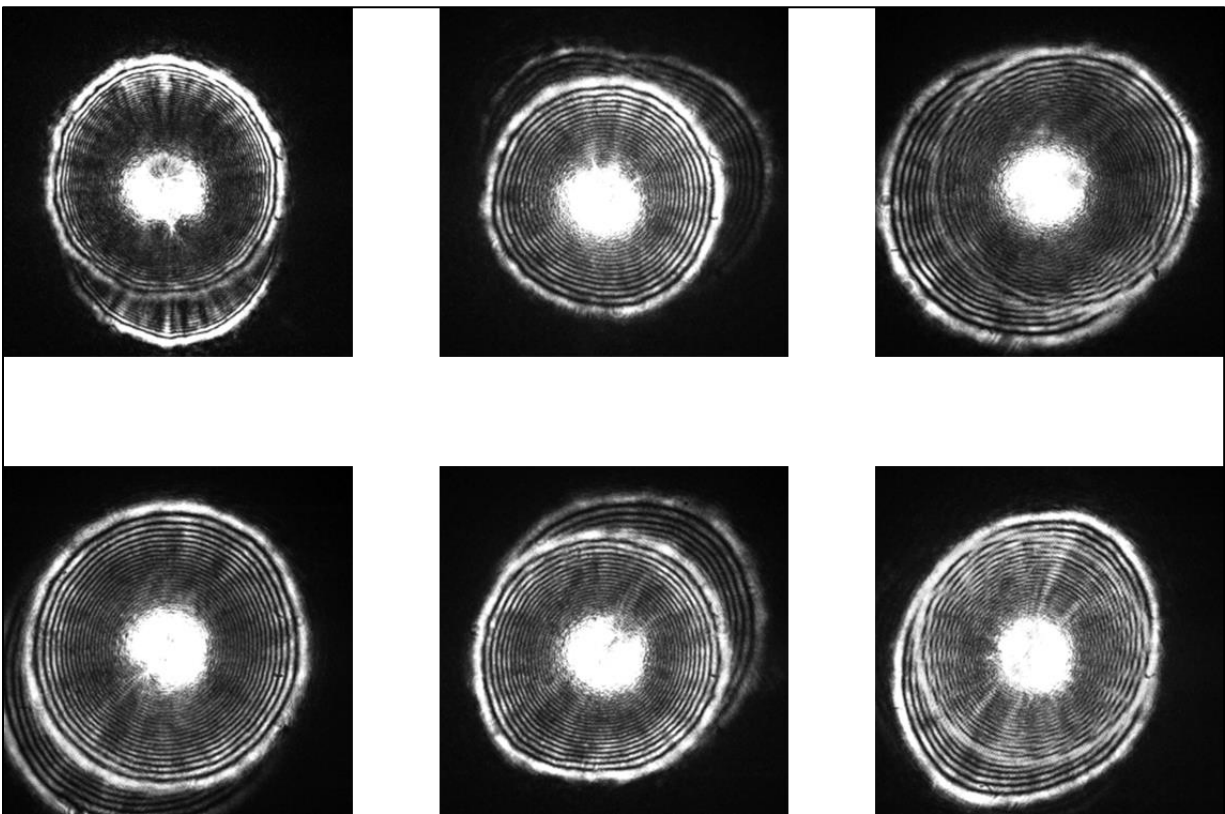


Figure 31 : Images de patterns avec un artefact causé par la présence proche d'une seconde colonie. Ces patterns sont issus de l'analyse de colonies d'Enterococcus faecalis après 16h d'incubation

Des formes de patterns aberrantes ont également été observées sans que les causes de ces formes soient identifiées (Figure 32). Ces patterns n'ont également pas été sélectionnés pour la constitution de la base de données.

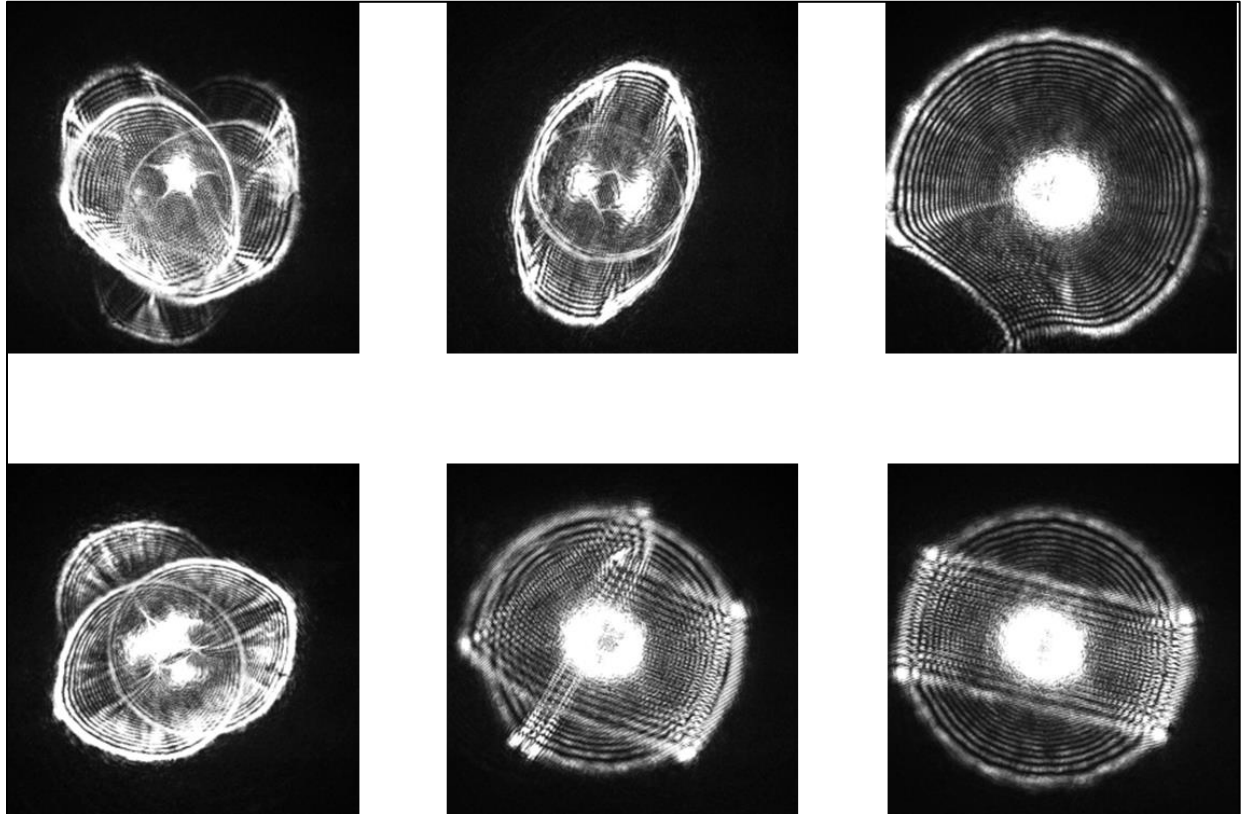


Figure 32 : Images de patterns ayant des formes aberrantes. Ces exemples de patterns de diffraction sont issus de l'analyse de colonies d'Enterococcus faecalis après 16h d'incubation

Les modalités d'acquisition et de sélection des patterns pour la constitution des bases de données étant déterminées, la constitution de ces dernières peut commencer. Dans un premier temps, la construction des bases de données sur les 6 espèces bactériennes sélectionnées a été faite à partir de souches de références, une par espèce.

II.C.5. Classification à partir de souches de référence

II.C.5.a. Création des bases de données à 12h et 16h d'incubation à partir de souches de référence

Il a été décidé d'établir deux modèles de classification à deux temps d'incubation différents : 12 heures et 16 heures. Il est donc nécessaire de créer une base de données de patterns par temps d'incubation. Pour optimiser l'entraînement du logiciel de classification, ces bases doivent être composées de 1000 images par classe. Ainsi leur construction nécessite l'acquisition d'un millier de patterns de diffraction pour chaque espèce bactérienne à chaque temps d'incubation. Par ailleurs, chaque base de données comporte une classe par espèce et également une classe appelée « aucun pattern détecté ». Cette classe regroupe des images de LASER non diffracté, c'est à dire sans colonies dans le trajet lumineux (Figure 33). Le fonctionnement d'un réseau neuronal convolutif fait que le résultat rendu ne peut être qu'une appartenance à l'une de classe de la base de données avec un score de pertinence. La possibilité de refuser une identification car l'image analysée est trop éloignée des différentes classes de la base de données n'existe pas. Ajouter une classe de LASER non diffracté permet de donner la possibilité au logiciel de classification de classer une acquisition de LASER non diffracté comme telle dans le cas où la colonie n'ait pas été correctement ciblée. Cela dans le but de pouvoir analyser les erreurs de ciblage de l'automate comme des acquisitions où aucun pattern de diffraction n'a été détecté au lieu de rendre une identification erronée.

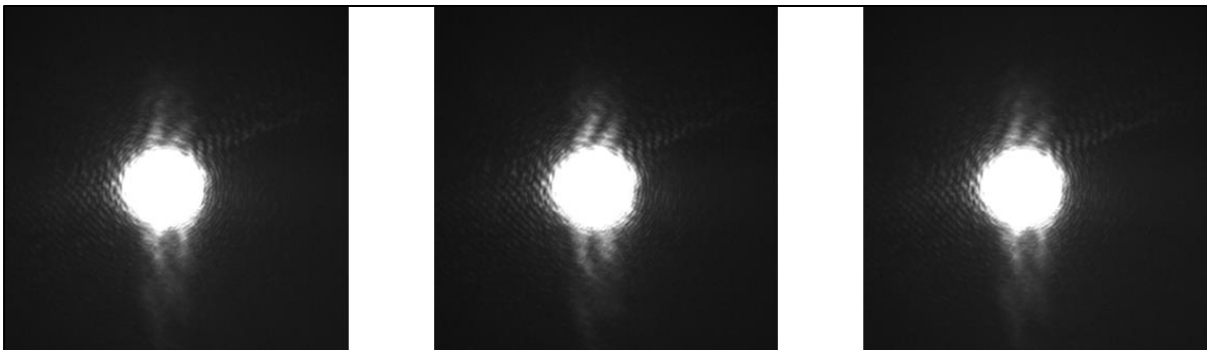


Figure 33 : Images de la classes "aucun pattern détecté"

II.C.5.b. Détermination du nombre d'epochs pour l'apprentissage du réseau neuronal convolutif

Une fois la base de données à partir des souches de références terminée, l'apprentissage par l'intelligence artificielle peut être lancé. Cet apprentissage est réalisé avec un ratio Train/Test de 80/20 suivant les recommandations de la société Revartis. C'est-à-dire que l'intelligence artificielle utilise 80 % de la base de données, (répartis équitablement parmi les différentes classes, soit 800 patterns par classe) pour s'entraîner et créer un modèle de classification et calculera les performances de ce modèle en le testant sur les 20 % restant de la base de données (soit 200 patterns par classe). Un des paramètres d'apprentissage influant principalement sur la qualité du modèle de classification est le nombre d'epochs que fera l'intelligence artificielle durant le développement du modèle de classification. Pour rappel, le nombre d'epochs représente le nombre de fois que l'entièreté de la partition d'entraînement (subset TRAIN) de la base de données est passée dans le réseau neuronal. Après chaque passage de la base de données (donc après chaque epochs), les différents poids dans le réseau neuronal sont adaptés. Plus le nombre d'epochs est grand, plus les poids du réseau neuronal seront changés. Cependant, au-delà d'un certain nombre d'epochs, un phénomène de surapprentissage peut apparaître entraînant une baisse des performances du modèle de classification. Le nombre d'epochs idéal pour un algorithme varie en fonction des bases de données et il faut donc tester des nombre d'epochs différents pour évaluer le nombre optimal de répétition. Pour cela, plusieurs entraînements de l'intelligence artificielle sur les deux bases de données (après 12 et 16 heures d'incubation) ont été réalisés et comparés. Les résultats de cette analyse sont présentés dans les tableaux et graphiques suivants. Le temps requis pour un entraînement du réseau neuronal dépend du nombre d'epochs, de la taille de la base de données mais également des performances de l'ordinateur sur lequel il est effectué. Les différents entraînements réalisés durant les travaux ici présentés ont été réalisés sur une carte graphique GeForce RTX 2080 MaxQ Design (Nvidia) et ont duré entre 40 minutes et 3 heures.

Nombre d'epochs	Accuracy	Precision	Recall	F1 Score
20	0.9450	0.9475	0.9450	0.9450
30	0.9500	0.9519	0.9500	0.9500
40	0.9483	0.9511	0.9483	0.9489
50	0.9517	0.9571	0.9517	0.9514
100	0.9683	0.9697	0.9683	0.9684
120	0.9625	0.9625	0.9625	0.9624

Tableau 3 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches de références après 12h d'incubation avec un ratio Train/Test de 80/20. La ligne grisée représente le modèle de classification avec les meilleurs résultats

Nombre d'epochs	Accuracy	Precision	Recall	F1 Score
20	0.9400	0.9412	0.9400	0.9399
30	0.9250	0.9306	0.9250	0.9243
40	0.9314	0.9337	0.9314	0.9314
50	0.9486	0.9490	0.9486	0.9485
100	0.9357	0.9380	0.9357	0.9359

Tableau 4 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches de références après 16h d'incubation. La ligne grisée représente le modèle de classification avec les meilleurs résultats

A partir de ces résultats, un modèle par temps d'incubation a été choisi. Après 12h d'incubation, le modèle de classification créé après 100 epochs d'entraînement est sélectionné comme étant le plus performant (Tableau 3). Il permet d'atteindre des métriques de performances entre 0,9683 et 0,9697 alors que l'entraînement avec 20 epochs ne permet pas d'avoir des métriques de performances supérieures à 0,9412. Pour les modèles après 16h d'incubation, un entraînement de 50 epochs suffit à fournir le modèle de classification le plus performant (Tableau 4) avec des métriques de performances entre 0,9485 et 0,9490. A ce temps d'incubation, l'entraînement avec 100 epochs résulte en une accuracy de 0,9357 et une précision de 0,9380.

II.C.5.c. Performances des modèles de classification établis à partir des bases de données de patterns de souches de référence

Maintenant que le nombre d'epochs optimal pour l'entraînement sur les bases de données est établi pour chaque temps d'incubation, il convient d'évaluer plus précisément les performances des deux modèles de classification en termes de sensibilité et de spécificité d'identification des différentes espèces bactériennes. Le réseau neuronal convolutif ayant utilisé 20 % de la base de données (soit 200 patterns par classe) comme test pour évaluer le modèle, il est possible d'accéder à la classification des ces patterns et donc de déterminer classe par classe les sensibilités et spécificités des deux modèles de classification retenus. Les tableaux suivants présentes les résultats de la phase de test des deux modèles de classification ainsi que les sensibilités et spécificités d'identification classe par classe. Ces résultats sont présentés sous forme de matrice de confusion. Ces matrices consistent à comparer les résultats d'identification obtenus par les modèles avec les identifications réelles des patterns classifiés. Dans toutes les matrices présentées dans ce document, les identifications du modèle sont en colonnes et les identifications réelles en lignes.

	<i>E. cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	Aucun pattern détecté
<i>E cloacae</i>	199	0	0	1	0	0
<i>E faecalis</i>	0	194	4	1	1	0
<i>E faecium</i>	0	19	180	1	0	0
<i>P vulgaris</i>	5	2	0	193	0	0
<i>S marcescens</i>	2	2	0	0	196	0
Aucun pattern détecté	0	0	0	0	0	200

Tableau 5 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de 5 souches de référence après 12 heures d'incubation. 200 patterns par classe ont été testés.

	Sensibilité	Spécificité
<i>E cloacae</i>	99,5 %	99,3 %
<i>E faecalis</i>	97,0 %	97,7 %
<i>E faecium</i>	90,0 %	99,6 %
<i>P vulgaris</i>	96,5 %	99,7 %
<i>S marcescens</i>	98,0 %	99,9 %
Aucun pattern détecté	100,0 %	100,0 %

Tableau 6 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns de 5 souches de référence après 12 heures d'incubation. 200 patterns par classe ont été testés.

De très bonnes sensibilités et spécificités ont été observées avec le modèle de classification développé à partir de la base de données après 12 heures d'incubation (Tableau 6). Les spécificités par classe sont toutes supérieures à 99 % excepté pour *Enterococcus faecalis* avec 97,7 % de spécificité. Les sensibilités, quant à elles, sont toutes supérieures à 96,5 % sauf pour la classe *Enterococcus faecium* qui a une sensibilité de 90,0%. On peut également noter la sensibilité et spécificité parfaite pour la classe « aucun pattern détecté » (100%). Aucune image sans pattern de diffraction n'a été identifiée comme appartenant à une classe d'espèces bactériennes (Tableau 5). A l'inverse, aucun pattern de diffraction de colonies n'a été faussement identifié comme une image de LASER non diffracté. L'analyse des résultats de ce test montre que la plus grande confusion d'identification a lieu entre les deux espèces d'entérocoques (*E. faecium* et *E. faecalis*). 19 des 200 patterns d'*Enterococcus faecium*, soit 10 %, ont été classifiés comme *Enterococcus faecalis* par le modèle de classification. Cela peut refléter une forte similitude entre les patterns de ces deux espèces, ce qui peut sembler attendu pour deux espèces du même genre.

	<i>E cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	<i>S saprophyticus</i>	Aucun pattern détecté
<i>E cloacae</i>	183	0	0	17	0	0	0
<i>E faecalis</i>	0	190	10	0	0	0	0
<i>E faecium</i>	0	15	185	0	0	0	0
<i>P vulgaris</i>	26	3	0	171	0	0	0
<i>S marcescens</i>	0	1	0	0	199	0	0
<i>S saprophyticus</i>	0	0	0	0	0	200	0
Aucun pattern détecté	0	0	0	0	0	0	200

Tableau 7 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de 6 souches de référence après 16 heures d'incubation. 200 patterns par classe ont été testés.

Espèces bactérienne	Sensibilité	Spécificité
<i>E cloacae</i>	91,5 %	97,8 %
<i>E faecalis</i>	95,0 %	98,4 %
<i>E faecium</i>	92,5 %	99,2 %
<i>P vulgaris</i>	85,5 %	98,6 %
<i>S marcescens</i>	99,5 %	100,0 %
<i>S saprophyticus</i>	100,0 %	100,0 %
Aucun pattern détecté	100,0 %	100,0 %

Tableau 8 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns de 6 souches de référence après 16 heures d'incubation. 200 patterns par classe ont été testés.

Les résultats du modèle de classification basé sur la base de données après 16 heures d'incubation sont légèrement moins bons qu'après 12 heures d'incubation mais restent quand même encourageants. Les classes *Serratia marcescens* et *Staphylococcus saprophyticus* sont particulièrement bien identifiées (respectivement 99,5 % et 100 % de sensibilité et 100 % de spécificité pour les deux classes). Les spécificités sont très bonnes, la plus basse étant de 97,8 % pour *Enterobacter cloacae*. On observe des sensibilités un peu plus faibles, particulièrement pour la classe *Proteus vulgaris* avec une sensibilité de 82,5 %. Mis à part cette classe, les sensibilités restent cependant supérieures à 91 % (Tableau 8). Comme cela est le cas pour le modèle à 12 heures d'incubation, la classe « aucun pattern détecté » est identifiée avec une sensibilité et une spécificité de 100 %. A partir de la matrice de confusion (Tableau 7), on observe les plus grands taux d'erreurs d'identification entre 2 couples d'espèces : *Enterobacter cloacae* et *Proteus vulgaris* ainsi qu'entre les deux espèces d'*Enterococcus* (*E. faecalis* et *E. faecium*). Ces résultats bien que fortement encourageant ne restent que des évaluations des modèles de classification sur leur propre base de données et ne suffisent donc pas à fournir une évaluation réelle des capacités de ces modèles pour l'identification de souches cliniques. Pour cela, il est important de tester les capacités d'identification des modèles sur des souches cliniques n'ayant pas servies à développer les bases de données.

II.C.5.d. Identification de patterns de souches cliniques par les modèles de classification basés sur des souches de référence

Les souches de référence pouvant présenter des variations génétiques et phénotypiques par rapport aux souches cliniques d'une même espèce (notamment à cause des repiquages nombreux sur des milieux gélosés qui ont pu diriger de manière artificielle l'évolution de la souche, ou de la variabilité des souches au sein d'une même espèce), il est nécessaire d'évaluer les performances de catégorisation de patterns de souches cliniques appartenant aux 6 espèces ayant servi à construire les bases de données sur souches de référence. Pour cela, 15 souches cliniques de chaque espèce ont été testées pour mettre à l'épreuve les modèles de classification dans des conditions plus proches de la réalité. Chaque souche a été analysée sur gélose TSA par l'automate BioKube et les patterns de diffraction acquis après 12 et 16 heures d'incubation ont été identifiés grâce aux modèles développés précédemment. 4 à 8 géloses de chaque souche ont été analysées après 12 et 16 heures d'incubation. Entre 50 et 250 patterns par souche et par temps d'incubation ont été identifiés. Au total, 7538 patterns acquis après 12 heures d'incubation et 11037 patterns après 16 heures d'incubation ont été classifiés durant ces tests pour une moyenne de 112 patterns par souches et 1688 patterns par classe. Les résultats de ces classifications sont présentés dans les tableaux ci-dessous.

	<i>E cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	Aucun pattern détecté	Patterns analysés totaux
<i>Complexe E cloacae</i>	290	30	4	1029	146	471	1970
<i>E faecalis</i>	8	1071	92	42	8	1	1222
<i>E faecium</i>	0	206	870	4	0	177	1257
<i>P vulgaris</i>	166	214	9	914	3	1	1307
<i>S marcescens</i>	234	1034	120	75	311	8	1782

Tableau 9 : Matrice de confusion des résultats de classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 5 souches de références après 12 heures d'incubation

	Sensibilité	Spécificité
<i>E cloacae</i>	20,0 %	93,7 %
<i>E faecalis</i>	87,6 %	76,5 %
<i>E faecium</i>	69,2 %	96,4 %
<i>P vulgaris</i>	69,9 %	81,5 %
<i>S marcescens</i>	17,5 %	97,3 %

Tableau 10 : Sensibilités et spécificités d'identification par classe de la classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 5 souches de références après 12 heures d'incubation

	<i>E cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	<i>S saprophyticus</i>	Aucun pattern détecté	Patterns analysés totaux
<i>Complexe E cloacae</i>	419	15	1	238	172	534	197	1576
<i>E faecalis</i>	19	1466	260	42	6	0	0	1793
<i>E faecium</i>	0	306	1032	0	0	414	0	1752
<i>P vulgaris</i>	281	605	21	1185	4	112	30	2238
<i>S marcescens</i>	81	185	1	69	1620	99	0	2055
<i>S saprophyticus</i>	0	0	1	0	0	1471	1	1473

Tableau 11 : Matrice de confusion des résultats de classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 16 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 6 souches de références après 16 heures d'incubation.

	Sensibilité	Spécificité
<i>E cloacae</i>	26,6 %	95,9 %
<i>E faecalis</i>	81,8 %	87,8 %
<i>E faecium</i>	58,9 %	96,9 %
<i>P vulgaris</i>	52,9 %	96,0 %
<i>S marcescens</i>	78,9 %	97,9 %
<i>S saprophyticus</i>	99,9 %	87,9 %

Tableau 12 : Sensibilités et spécificités d'identification par classe de la classification de patterns de diffraction provenant de 15 souches cliniques de chaque espèce bactérienne après 16 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données des 6 souches de références après 16 heures d'incubation

L'analyse de patterns issus de souches cliniques montrent une nette perte de performances des modèles de classification développés à partir de souches de référence. 7538 patterns acquis après 12 heures d'incubation et 10887 patterns acquis après 16 heures d'incubation ont été classifiés par les deux modèles de classification basés sur souches de référence (Tableaux 9 et 11). Sans considération pour les différentes espèces bactériennes, le nombre d'identifications correctes est de respectivement 3456 et 7193 après 12 et 16 heures d'incubation soit 45,8 % et 66,1 % d'identification correcte. Cette baisse de performance est notamment très forte au niveau des sensibilités d'identification avec des sensibilités par classe descendant jusqu'à 17,5 % pour *Serratia marcescens* après 12 heures d'incubation. Mise à part une très bonne sensibilité pour la classe *Staphylococcus saprophyticus* à 16 heures d'incubation (99,9 %), toutes les classes voient leur sensibilité chuter (entre 17,5 et 87,6 % à 12 heures d'incubation et 26,6 et 81,8 % à 16h d'incubation). Les spécificités par classe sont également diminuées mais moins fortement que les sensibilités (76,5 – 96,4 % pour 12 heures et 81,8 – 97,9 % pour 16 heures) (Tableaux 10 et 12).

Cette perte de performances lors de la classification de patterns issus de souches cliniques à partir d'un modèle basé sur des patterns obtenus à partir de souches de références peut notamment s'expliquer par une hétérogénéité parmi les différentes souches d'une même espèce bactérienne. En effet, lorsque que l'on observe les taux d'identification correcte des patterns par souche clinique, une grande disparité existe entre différentes souches d'une même espèce (Figures 34 et 35).

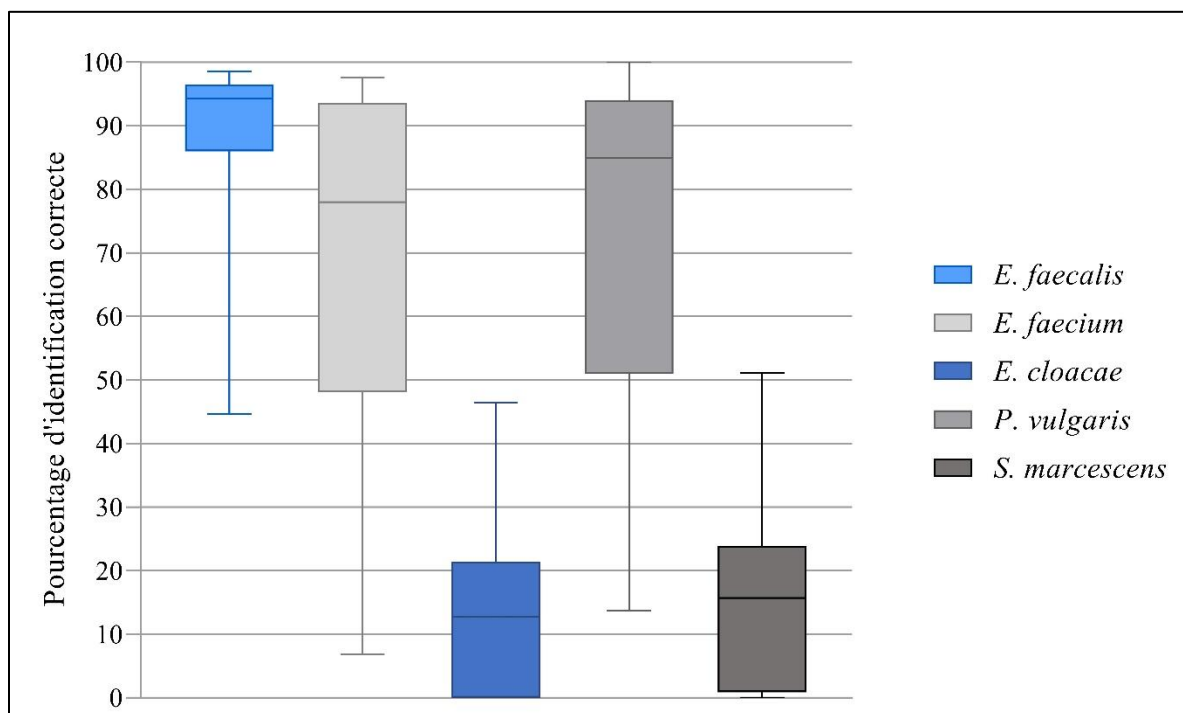


Figure 34: Répartition des pourcentages d'identifications correctes des patterns par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de 5 souches de références après 12 heures d'incubation. Caractéristiques des boxplots décrites en II.B.9, page 79

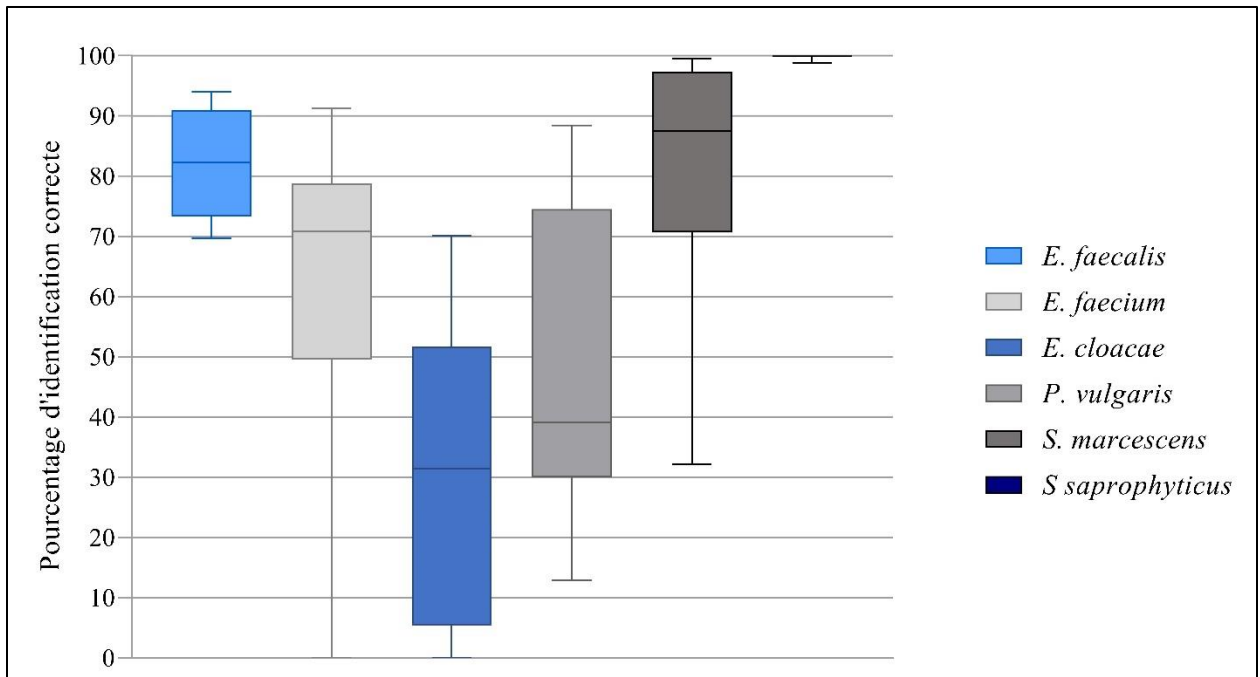


Figure 35 : Répartition des pourcentages d'identifications correctes des patterns par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 16h d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches de référence après 16 heures d'incubation. Caractéristiques des boxplots décrites en II.B.9, page 79

Par exemple, les différentes souches cliniques de *Proteus vulgaris* ont des pourcentages d'identification correcte allant de 18,75 % à 88,41 % après 16 heures d'incubation. Ce phénomène d'hétérogénéité s'observe par des différences de patterns de diffraction observables à l'œil nu (Figures 36 et 37).

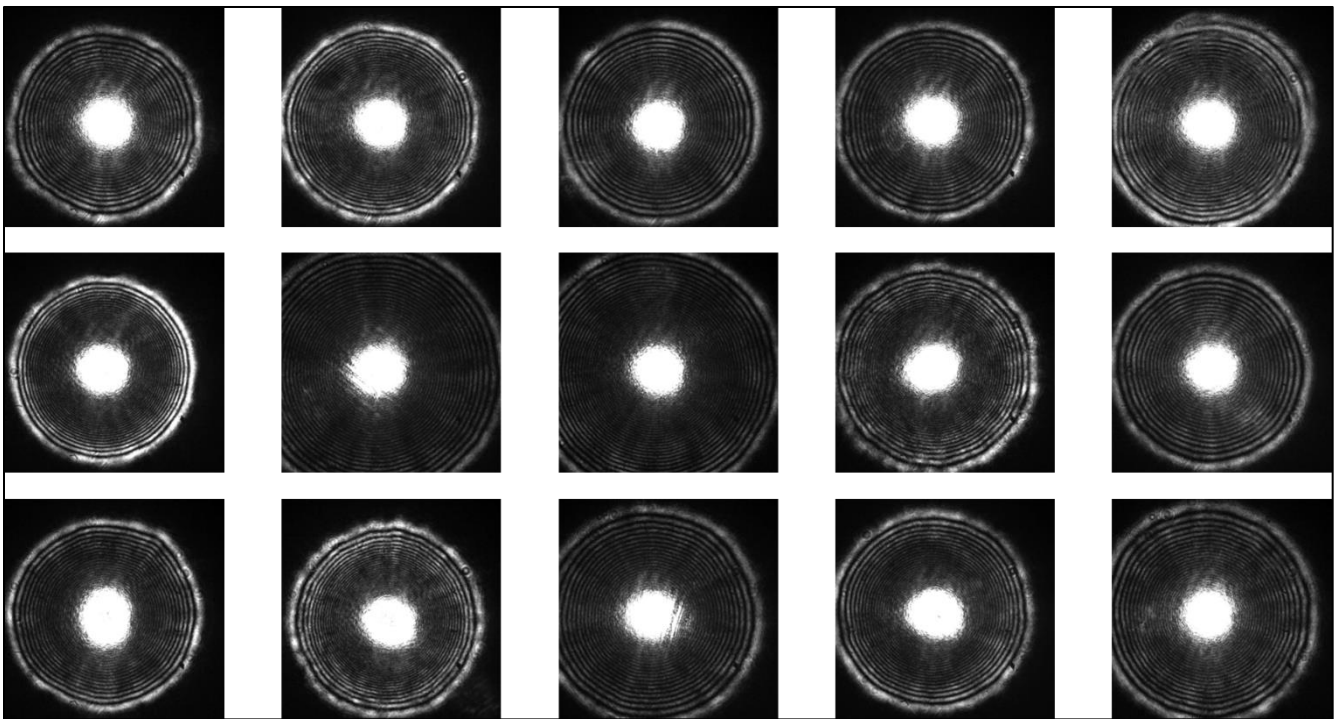


Figure 36 : Exemple d'hétérogénéité de patterns de diffraction pour Enterococcus faecium après 12 heures d'incubation. Chaque pattern est issu d'une souche clinique différente

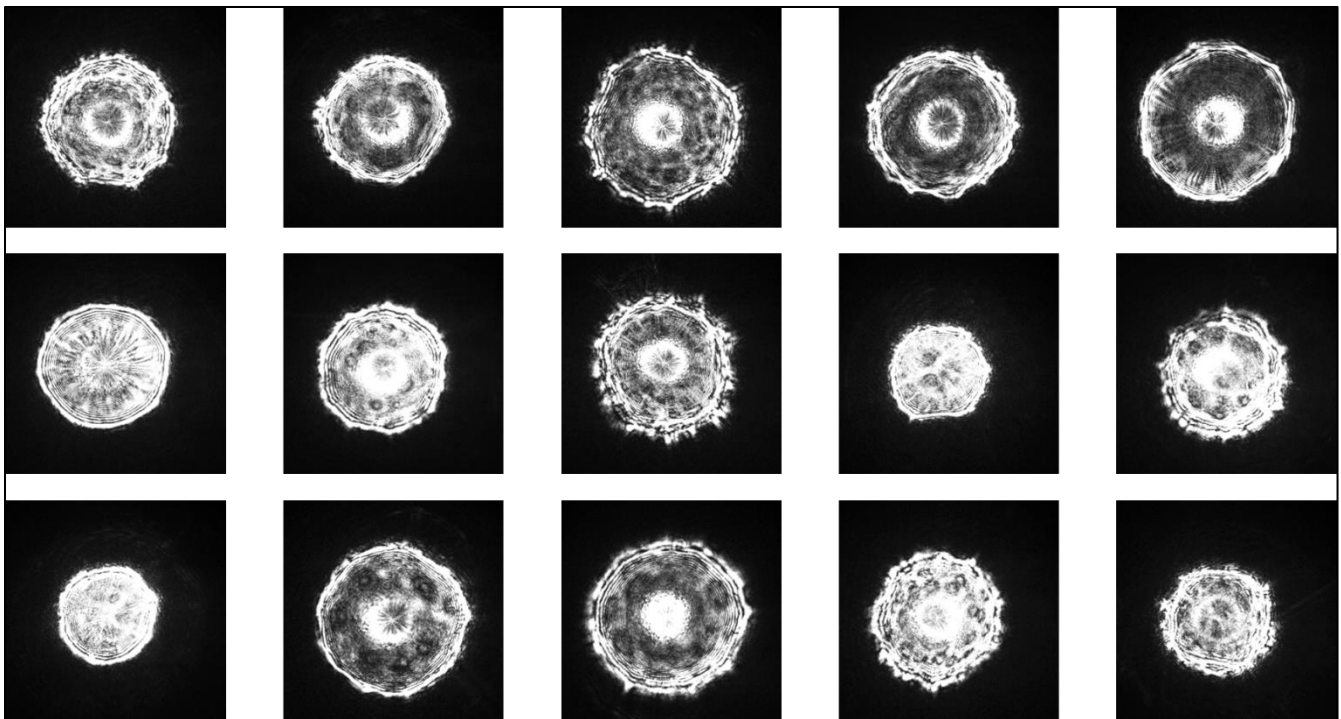


Figure 37: Exemple d'hétérogénéité de patterns de diffraction pour Proteus vulgaris après 16 heures d'incubation. Chaque pattern est issu d'une souche clinique différente

Cette hétérogénéité parmi les différentes souches d'une même espèce bactérienne ne peut être prise en compte dans une base de données faite à partir d'une seule souche référence. Ainsi malgré les premiers résultats encourageants des modèles de classification, l'utilisation, pour l'identification de souches cliniques, d'une base de données basée uniquement sur une souche de référence par espèce ne permet pas d'obtenir des résultats fiables d'identification sur des souches cliniques. Cela a donc conduit à la réalisation d'une base de données de patterns issus des plusieurs souches cliniques pour chaque espèce bactérienne représentée.

II.C.6. Bases de données basées sur souches cliniques

- Complexe *Enterobacter cloacae*

Par ailleurs, la faible sensibilité de classification des patterns issus des souches d'*Enterobacter cloacae* peut également s'expliquer par le fait que la spectrométrie de masse MALDI-TOF ne permet pas la différenciation de plusieurs espèces du genre *Enterobacter*. Cette analyse se limite donc à l'identification d'un complexe *Enterobacter cloacae* qui regroupe au total 22 espèces bactériennes dont *Enterobacter cloacae*, *Enterobacter absuriae*, *Enterobacter bugandensis*, *Enterobacter cancerogenus*, *Enterobacter hormaechei*, *Enterobacter kobei*, *Enterobacter ludwigii*, *Enterobacter mori*, *Enterobacter nimipressuralis*, *Enterobacter rogenkampii* et *Enterobacter soli* (248,249). Cependant, d'un point de vue clinique, l'identification d'une appartenance au complexe *Enterobacter cloacae* est suffisante pour pouvoir rendre un diagnostic. En effet, bien que les différentes espèces de ce complexe puissent avoir des pathogénicités différentes pour les humains (*E. nimipressuralis* est considéré comme une bactérie environnementale non pathogène), le traitement à mettre en place en cas d'infections est relativement similaire. Même s'il existe un intérêt à pouvoir différencier les espèces du complexe *Enterobacter cloacae* entre elles, les souches cliniques identifiées en laboratoire le sont comme appartenant au complexe sans que des précisions sur l'espèce exacte soit considérées comme indispensables. Les quinze souches cliniques récupérées de routine pour les expériences qui sont présentées par la suite ont donc seulement une identification comme appartenant à ce complexe. Pour connaître l'espèce bactérienne exacte de ses souches, une amplification et un séquençage du gène *hsp60* ont été réalisés.

Espèce bactérienne	Nombre de souches
<i>Enterobacter cloacae</i>	10
<i>Enterobacter hormaechei</i>	3
<i>Enterobacter kobei</i>	1
<i>Enterobacter ludwigii</i>	1

Tableau 13 : Identification des souches cliniques appartenant au complexe *Enterobacter cloacae*

Ainsi 10 des 15 des souches cliniques du complexe *Enterobacter cloacae* utilisées appartiennent bien à l'espèce *Enterobacter cloacae*. Trois souches ont été identifiées en tant qu'*Enterobacter hormaechei*, une souche comme appartenant à l'espèce *Enterobacter kobei* et une autre à l'espèce *Enterobacter ludwigii* (Tableau 13). Il serait possible de créer différentes classes pour chaque espèce bactérienne du complexe au sein de la base de données. Cependant, la récupération d'un nombre suffisant de souches cliniques pour chaque espèce serait longue. Les différentes souches ont donc été réunies pour les analyses qui suivent dans la classe Complexe *Enterobacter cloacae* pour évaluer la capacité d'identifier ce complexe et non les espèces qui le constituent par analyse des patterns de diffraction LASER.

II.C.6.a. Développement de modèles de classification basés sur 15 souches cliniques par espèce

Le procédé de constitution de base de données précédemment présenté sur souches de référence a été répété mais en utilisant des patterns acquis à partir de souches cliniques. Pour cela, les 15 souches cliniques de chacune des six espèces bactériennes d'intérêt ont été à nouveau analysées par l'automate BioKube. Leurs patterns de diffraction ont été récupérés et sélectionnés selon les mêmes critères que pour la base de données basée sur les souches de référence. Les bases de données devant toujours être constituées de 1000 patterns par classe, il est important que les différentes souches de chaque espèce soient représentées équitablement au sein d'une même classe. Ainsi, 67 (ou 66) patterns de chaque souche ont été sélectionnés pour pouvoir constituer des classes des 1000 patterns de diffraction.

Une fois la base de données constituée pour les deux temps d'incubation (12 et 16 heures), il était nécessaire de réévaluer le nombre optimal d'epochs pour la création des modèles de classification. En effet, ce nombre est dépendant de la base de données utilisée. Cette dernière ayant changée, le nombre d'epochs choisi pour les modèles basés sur les souches de référence peut ne pas être adéquat pour celles basées sur souches cliniques.

Nombre d'epochs	Accuracy	Precision	Recall	F1 Score
20	0.9500	0.9508	0.9500	0.9500
30	0.9592	0.9595	0.9592	0.9590
40	0.9500	0.9519	0.9500	0.9503
50	0.9608	0.9619	0.9608	0.9610
70	0.9508	0.9531	0.9508	0.9511
100	0.9608	0.9617	0.9608	0.9609

Tableau 14 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches cliniques après 12 heures d'incubation. La ligne grisée représente le modèle de classification avec les meilleurs résultats

Nombre d'epochs	Accuracy	Precision	Recall	F1 Score
20	0.9007	0.9058	0.9077	0.8995
30	0.9521	0.9533	0.9521	0.9522
40	0.9593	0.9608	0.9593	0.9593
50	0.9529	0.9532	0.9529	0.9527
100	0.9214	0.9334	0.9214	0.9198

Tableau 15 : Evaluation des différents modèles de classification en fonction du nombre d'epochs. Les modèles de classification sont issus de l'entraînement de l'intelligence artificielle sur la base de données de souches de cliniques après 16h d'incubation. La ligne grisée représente le modèle de classification avec les meilleurs résultats

Cette évaluation a permis de décider du nombre d'epochs qui a été utilisé pour l'entraînement de l'intelligence artificielle à partir des bases de données créées à partir de patterns de souches cliniques. Ce nombre est différent en fonction du temps d'incubation. Il est de 50 epochs pour le modèle de classification après 12 heures d'incubation, avec des métriques de performances entre 0,9608 et 0,9619 (Tableau 14) et de 40 epochs pour celui après 16 heures d'incubation avec des métriques de performances de 0,9593 et 0,9608 (Tableau 15).

II.C.6.b. Performances des modèles de classification établis à partir des bases de données de patterns de souches cliniques

Ainsi deux modèles de classification basés sur des souches cliniques ont été développés. L'évaluation de ces modèles sur leur propre base de données est présentée dans les tableaux suivants.

	<i>Complexe E cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	Aucun pattern détecté
<i>Complexe E cloacae</i>	189	0	0	11	0	0
<i>E faecalis</i>	0	192	8	0	0	0
<i>E faecium</i>	0	3	194	3	0	0
<i>P vulgaris</i>	2	2	4	192	0	0
<i>S marcescens</i>	6	2	1	5	186	0
Aucun pattern détecté	0	0	0	0	0	200

Tableau 16 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de souches cliniques après 12h d'incubation. 200 patterns par classe ont été testés.

	Sensibilité	Spécificité
<i>Complexe E cloacae</i>	94,5 %	99,2 %
<i>E faecalis</i>	96,0 %	99,3 %
<i>E faecium</i>	98,5 %	98,7 %
<i>P vulgaris</i>	96,0 %	98,1 %
<i>S marcescens</i>	93,0 %	100,0 %
Aucun pattern détecté	100,0 %	100,0 %

Tableau 17 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns de souches cliniques après 12 heures d'incubation

Le modèle de classification développé après 12 heures d'incubation à partir d'une base de données de souches cliniques donne des résultats satisfaisants lorsqu'il est évalué avec les 20 % de sa base de données ayant servis de sous-ensemble de test pour le modèle (Tableau 16). Les spécificités par classe sont excellentes (entre 98,1 et 100,0 %) et similaires à celles obtenues avec la base de données de souches de référence (comprises entre 90 et 100 % (Tableau 6)). Les sensibilités par classe, quant à elles, sont légèrement inférieures à celles obtenues avec une base de données de souches de référence (> 99% (Tableau 6)) mais tout en restant très bonnes (93 % pour *Serratia marcescens*, 98,5 % pour *Enterococcus faecium*) (Tableau 17).

	<i>Complexe E cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	<i>S saprophyticus</i>	Aucun pattern détecté
<i>Complexe E cloacae</i>	178	1	0	13	8	0	0
<i>E faecalis</i>	0	196	2	2	0	0	0
<i>E faecium</i>	0	5	192	1	0	2	0
<i>P vulgaris</i>	1	4	0	195	0	0	0
<i>S marcescens</i>	2	8	1	6	183	0	0
<i>S saprophyticus</i>	0	0	1	0	0	199	0
Aucun pattern détecté	0	0	0	0	0	0	200

Tableau 18 : Matrice de confusion des résultats du test du modèle de classification établi à partir de la base de données de patterns de souches cliniques après 16h d'incubation. 200 patterns par classe ont été testés.

	Sensibilité	Spécificité
<i>Complexe E cloacae</i>	89,0 %	99,8 %
<i>E faecalis</i>	98,0 %	98,5 %
<i>E faecium</i>	96 %	99,7 %
<i>P vulgaris</i>	97,5 %	98,2 %
<i>S marcescens</i>	91,5 %	99,3 %
<i>S saprophyticus</i>	99,5 %	99,8 %
Aucun pattern détecté	100,0 %	100,0 %

Tableau 19 : Sensibilités et spécificités d'identification par classe du modèle de classification établi à partir de la base de données de patterns après 16h d'incubation

Le modèle de classification après 16 heures d'incubation est également satisfaisant, même si avec des résultats légèrement inférieurs au modèle de classification après 12 heures d'incubation basé sur les souches de référence (sensibilités : 85,5 % - 100 %, spécificités : 97,8 % - 100 %). Les sensibilités par classe sont bonnes (89 % pour la classe Complexe *Enterobacter cloacae*, 99,5 % pour *Staphylococcus saprophyticus*). Les spécificités par classe sont excellentes, étant toutes supérieures à 98 % (98,2 % pour *Proteus vulgaris*) (Tableau 19).

Il est à noter que, pour les deux modèles de classification, la classe « aucun pattern détecté » a une sensibilité et spécificité de 100%. La confusion préalablement observée entre *Proteus vulgaris* et *Enterobacter cloacae* semble persister. Respectivement 100 % et 59 % des faux négatifs pour la classe Complexe *Enterobacter cloacae* après 12 et 16 heures d'incubation sont identifiés à tort comme *Proteus vulgaris*. Le nombre de résultats faussement négatifs pour la classe Complexe *Enterobacter cloacae* est parmi les plus élevés des classes avec 11 résultats faux négatifs pour 189 vrais positifs à 12 heures d'incubation (Tableau 16) et 22 faux négatifs pour 178 vrais positifs à 16 heures d'incubation (Tableau 18). Les erreurs d'identification entre les deux espèces d'*Enterococcus* semblent, quant à elles, persister mais à un moindre niveau : 100 % (8 sur 8) à 12 heures et 50 % (2 sur 4) à 16 heures des faux négatifs d'*Enterococcus faecalis* ont été identifiés en *Enterococcus faecium* et 50 % (3 sur 6) à 12 heures et 62 % (5 sur 8) à 16 heures des faux négatifs d'*Enterococcus faecium* ont été identifiés en *Enterococcus faecalis*. Cependant, les résultats faux positifs restent peu fréquents : 4% et 2% pour *Enterococcus faecalis* après 12 et 16 heures d'incubation respectivement et 3% pour *Enterococcus faecium* après les deux temps d'incubation. De fait, l'évaluation de confusions particulières entre deux classes reste du domaine de l'observation et ne permet pas de fournir une conclusion fiable.

II.C.6.c. Identification de patterns des souches cliniques par les modèles de classification basés sur des souches cliniques

Pour évaluer plus précisément les modèles de classifications basés sur les souches cliniques, il était nécessaire de les tester sur des patterns ne faisant pas partie des bases de données qui ont servi à les développer. Pour cela, la banque de patterns déjà utilisée précédemment pour tester les modèles de classification basés sur souches de référence, mais, n'ayant pas servi à la construction de la base de données sur souches cliniques, a été réanalysé à l'aide des nouveaux modèles de classification basés sur les souches cliniques.

	<i>Complexe E cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	Aucun pattern détecté	Patterns analysés totaux
<i>Complexe E cloacae</i>	1860	0	0	53	57	0	1970
<i>E faecalis</i>	2	1197	16	6	1	0	1222
<i>E faecium</i>	0	14	1236	3	4	0	1257
<i>P vulgaris</i>	7	22	23	1243	12	0	1307
<i>S marcescens</i>	71	25	23	21	1642	0	1782

Tableau 20 : Matrice de confusion des résultats de classification des patterns de souches cliniques après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 12 heures d'incubation

	Sensibilité	Spécificité
<i>Complexe E cloacae</i>	90,4 %	99,2 %
<i>E faecalis</i>	98,0 %	99,0 %
<i>E faecium</i>	98,3 %	99,0 %
<i>P vulgaris</i>	95,1 %	98,7 %
<i>S marcescens</i>	92,1 %	98,7 %

Tableau 21 : Sensibilités et spécificités d'identification par classe de la classification des patterns de souches cliniques après 12 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 12 heures d'incubation

	<i>Complexe E cloacae</i>	<i>E faecalis</i>	<i>E faecium</i>	<i>P vulgaris</i>	<i>S marcescens</i>	<i>S saprophyticus</i>	Aucun pattern détecté	Patterns analysés totaux
<i>Complexe E cloacae</i>	1378	8	1	110	79	0	0	1576
<i>E faecalis</i>	1	1703	25	63	1	0	0	1793
<i>E faecium</i>	0	49	1675	7	0	21	0	1752
<i>P vulgaris</i>	9	36	25	2168	0	0	0	2238
<i>S marcescens</i>	38	64	19	87	1847	0	0	2055
<i>S saprophyticus</i>	0	0	18	0	0	1455	0	1473

Tableau 22 : Matrice de confusion des résultats de classification des patterns de souches cliniques après 16 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 16 heures d'incubation

	Sensibilité	Spécificité
<i>Complexe E cloacae</i>	87,4 %	99,5 %
<i>E faecalis</i>	95,0 %	98,3 %
<i>E faecium</i>	95,6 %	99,0 %
<i>P vulgaris</i>	96,9 %	96,9 %
<i>S marcescens</i>	89,9 %	99,1 %
<i>S saprophyticus</i>	98,8 %	99,8 %

Tableau 23 : Sensibilités et spécificités d'identification par classe de la classification des patterns de souches cliniques après 16 heures d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 16 heures d'incubation

7538 patterns acquis après 12 heures d'incubation et 10887 patterns acquis après 16 heures d'incubation ont été classifiés par les deux modèles de classification basés sur souches cliniques (Tableaux 20 et 22). Sans considération pour les différentes espèces bactériennes, le nombre d'identifications correctes est de respectivement 7178 et 10226 après 12 et 16 heures d'incubation soit 95,2 et 93,9 % d'identification correcte. Seules 4,8% (360 patterns) et 6,1% (661 patterns) des patterns de diffraction acquis après respectivement 12 et 16 heures d'incubation, ont été incorrectement classifiés.

Une nette amélioration des performances de classification est observée en comparaison de l'utilisation de modèles de classification basés sur des souches de références. Après 12 heures d'incubation, les spécificités d'identification par classe sont homogènes et comprises entre 98,7 et 99,2 % (Tableau 21). Les sensibilités par classe sont quant à elles supérieures à 90 % (90,4 % pour la classe Complexe *Enterobacter cloacae*) avec un maximum de 98,3 % pour *Enterococcus faecium*. Les résultats de l'évaluation du modèle de classification après 16 heures d'incubation sont également très encourageants bien que légèrement en deçà de ceux après 12 heures d'incubation (Tableau 23) avec notamment des sensibilités inférieures à 90 % pour les classes Complexe *Enterobacter cloacae* (87,4 % contre 90,4 % à 12 heures) et *Serratia marcescens* (89,9 % contre 92,1 % à 12 heures). Les spécificités de classe pour ce temps d'incubation sont

comprises entre 96,9 % (*Proteus vulgaris*) et 99,8 % (*Staphylococcus saprophyticus*). Les sensibilités, quant à elles, sont plus basses avec deux classes ayant des sensibilités inférieures à 90% : Complexe *Enterobacter cloacae* (87,4 %) et *Serratia marcescens* (89,9 %). Pour les quatre autres classes, les sensibilités sont toutes supérieures à 95 % avec un maximum de 98,8 % pour *Staphylococcus saprophyticus*.

La problématique de l'hétérogénéité intra-espèce semble être prise en compte par ces nouveaux modèles de classification. En effet, l'étude des pourcentages d'identification correcte par souches montre une augmentation générale des taux d'identification correcte ainsi qu'une baisse des disparités pour les souches au sein d'une même espèce (Figures 38 et 39). Par exemple, la classe *Proteus vulgaris* après 16 heures d'incubation, qui avec le modèle de classification basé sur souches de références avait des scores d'identification correcte par souches allant de 18,75 % à 88,41 %, a des pourcentages d'identification correcte entre 83,8 et 100 % avec le modèle de classification basé sur souches cliniques.

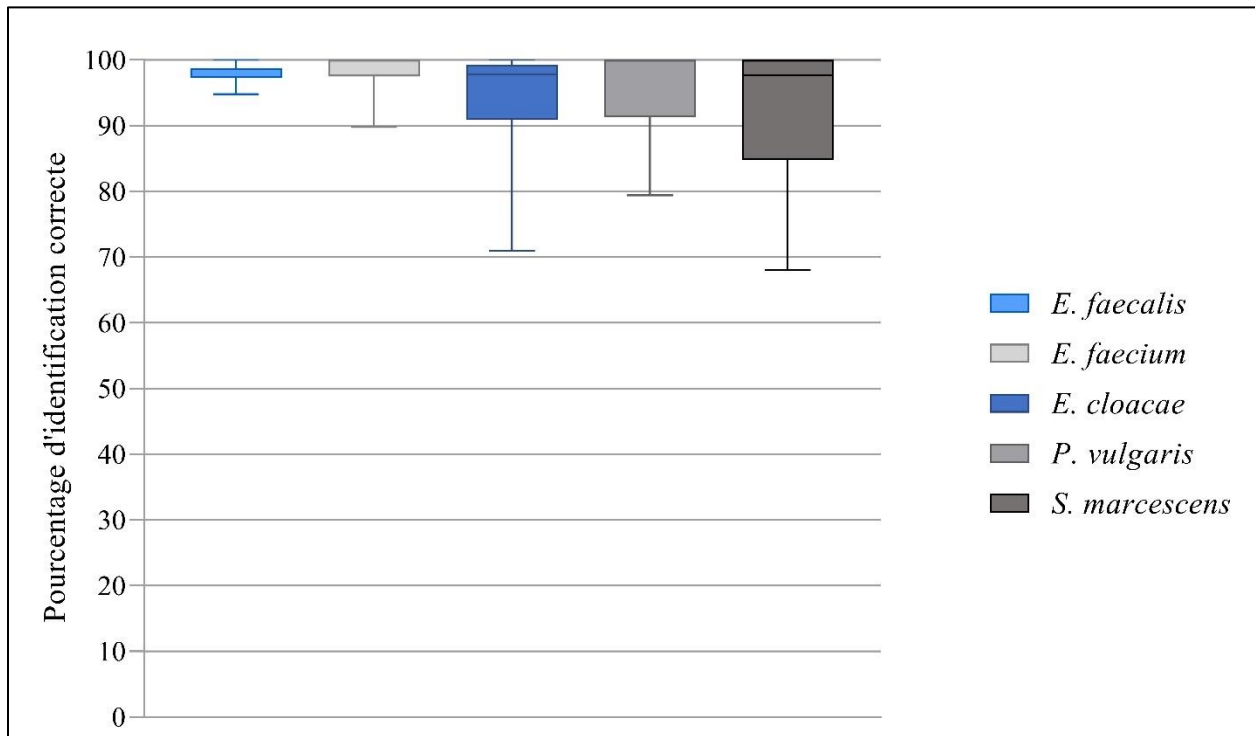


Figure 38 : Pourcentages d'identifications correctes par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 12h d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 12 heures d'incubation. Caractéristiques des boxplots décrites en

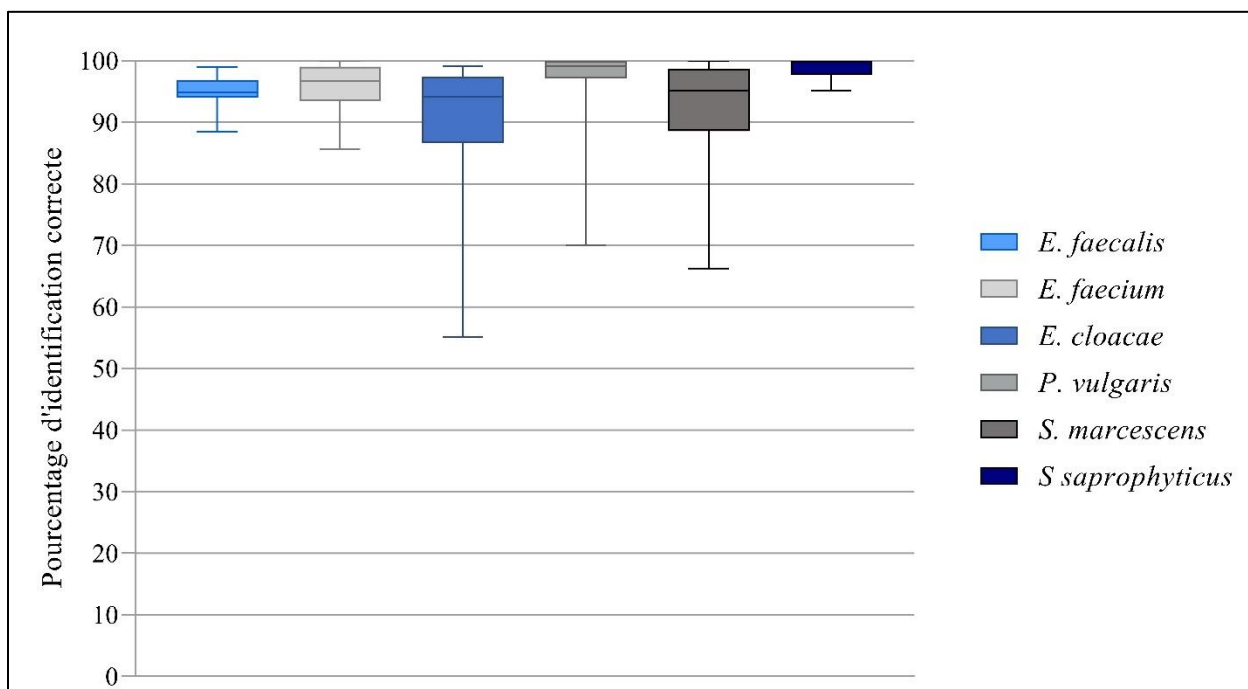


Figure 39 : Répartition des pourcentages d'identifications correctes par souche clinique pour chaque espèce bactérienne après classification de patterns acquis après 16h d'incubation. La classification a été réalisée en utilisant le modèle de classification basé sur la base de données de souches cliniques après 16 heures d'incubation. Caractéristiques des boxplots décrites en II.B.9, page 79

Ces résultats confirment l'importance de l'utilisation de souches cliniques lors de la construction des bases de données ainsi que l'utilisation possible de la diffraction LASER pour l'identification d'espèces bactériennes. Il reste cependant plusieurs points à étudier pour comprendre le champ d'application de ces modèles ainsi que la pertinence des modèles de classification. Il est notamment question de savoir si ces bases de données sont transférables à l'identification de colonies en culture sur des milieux gélosés différents. De plus, les résultats des modèles de classification ont été jusqu'ici évalués uniquement sur l'appartenance à une classe sans prise en compte des scores de pertinences associés à ces classifications. Ces scores peuvent permettre une meilleure compréhension des performances d'identification des modèles développés.

II.C.7. Evaluation des modalités d'acquisition et de classification des patterns

II.C.7.a. Impact du type de gélose sur la classification

Jusqu'ici toutes les expériences ont été réalisées sur gélose TSA. Cependant différents milieux gélosés existent et sont utilisés en routine. Deux autres milieux gélosés incolores ont été testés : les milieux gélosés Mueller-Hinton (MHE) et Brain-Heart Infusion (BHI). Les milieux gélosés ayant des compositions différentes, il est connu que le comportement des colonies bactériennes varie en fonction du milieu gélosé que ce soit en termes de forme ou de vitesse de croissance. Des études ont même démontré des changements dans les patterns de diffraction de colonies bactériennes en fonction des milieux de culture utilisés (245). Le but de cette expérience est d'avoir un début de réponse sur la transférabilité des bases de données entre différents milieux gélosés. 3 espèces bactériennes ont été utilisés ici : *Enterococcus faecalis*, *Proteus vulgaris* et *Serratia marcescens*. Pour chaque espèce, 1 souche clinique a été ensemencé sur géloses TSA, BHI et MHE puis les géloses ont été analysées par le prototype BioKube après 16 heures d'incubation.

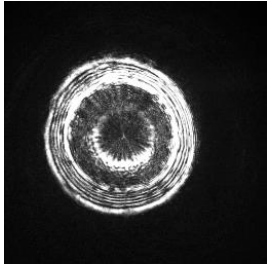
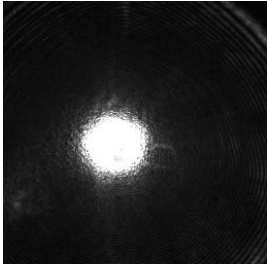
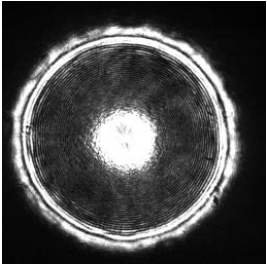
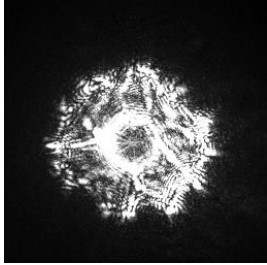
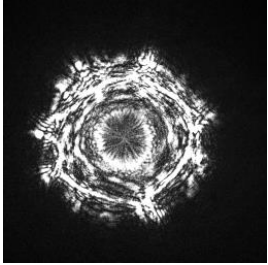
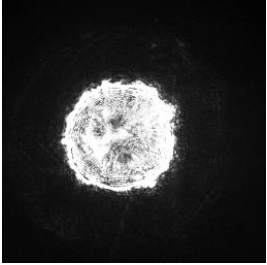
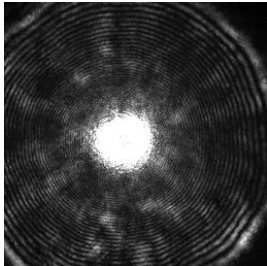
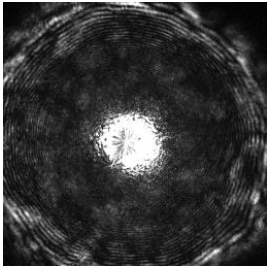
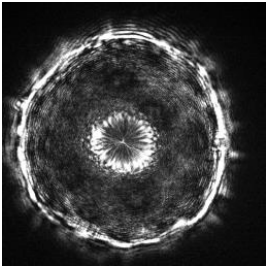
	BHI	MHE	TSA
<i>Enterococcus faecalis</i>	 <p>Complexe <i>E. cloacae</i> (95,8%)</p>	 <p><i>S. saprophyticus</i> (54,7%)</p>	 <p><i>E. faecalis</i> (99,9%)</p>
<i>Proteus vulgaris</i>	 <p><i>P. vulgaris</i> (100%)</p>	 <p><i>P. vulgaris</i> (90,7%)</p>	 <p><i>P. vulgaris</i> (99,8%)</p>
<i>Serratia marcescens</i>	 <p><i>S. saprophyticus</i> (81,1%)</p>	 <p><i>S. saprophyticus</i> (49,8%)</p>	 <p><i>S. marcescens</i> (99,9%)</p>

Figure 40 : Patterns de diffraction de colonies sur 3 milieux gélosés différents : BHI, TSA et MHE. Les patterns ont été acquis après 16 heures d'incubation et sont issus de colonies de la même souche clinique d'*Enterococcus faecalis*, de *Proteus vulgaris* et *Serratia marcescens*. Pour chaque pattern, l'identification rendue par le modèle de classification basés sur les souches cliniques (géluses TSA) après 16h d'incubation est donnée avec son score de pertinence

Les différences sur les patterns de diffraction causées par le changement de milieu gélosé sont observables à l'œil nu (Figure 40). Des patterns complètement différents sont observables, particulièrement pour les souches cliniques d'*Enterococcus faecalis* et *Serratia marcescens*. A part pour la souche clinique de *Proteus vulgaris*, les résultats de la classification par le modèle de classification basé sur les patterns de souches cliniques sont complètement aberrants. Ces tests montrent que le type de gélose est également un facteur important pour la classification d'espèce bactérienne par diffraction LASER.

II.C.7.b. Seuil de classification

Lorsque le modèle de classification rend un résultat, un score de pertinence est également fourni. Ce score, fourni par le modèle sous forme de pourcentage, permet d'évaluer la certitude de la classification donnée par le réseau neuronal. Pour pouvoir utiliser ce score de pertinence, il est important de déterminer des seuils qui serviront à l'évaluer. Ces seuils, qui seront nommés seuils de confiance ne peuvent être déterminés que par l'expérience. Pour cela, les données de classification de la banque de patterns déjà utilisées précédemment ont été regroupées en deux groupes. Sans prendre en compte l'espèce identifiée, les scores de pertinence ont été séparés entre ceux représentant une identification correcte fournie par le modèle de classification et ceux représentant une identification erronée. Un seuil de confiance a par la suite été fixé et les pourcentages de score supérieurs à ce seuil pour chaque groupe ont été calculés. En faisant varier ce seuil, il est ainsi possible de déterminer pour quelle valeur les meilleures performances de classification sont obtenues. Les résultats de ces analyses sont présentés dans le tableau 24.

Valeur seuil	Identification correcte 12h	Identification erronée 12h	Identification correcte 16h	Identification erronée 16h
50	99.86 %	96.11 %	99.85 %	96.36 %
55	99.54 %	89.17 %	99.29 %	88.33 %
60	99.25 %	82.50 %	98.77 %	81.97 %
65	98.98 %	78.33 %	98.12 %	75.15 %
70	98.51 %	69.72 %	97.43 %	68.94 %
75	98.05 %	63.06 %	96.60 %	62.58 %
80	97.35 %	56.94 %	95.47 %	55.45 %
85	96.35 %	49.72 %	94.23 %	49.24 %
90	95.14 %	41.39 %	91.83 %	40.15 %
95	93.03 %	32.50 %	88.27 %	31.52 %
96	92.06 %	29.72 %	86.65 %	28.33 %
97	90.87 %	25.83 %	84.61 %	24.85 %
98	88.80 %	22.50 %	81.38 %	21.52 %
99	84.98 %	16.94 %	75.08 %	15.76 %

Tableau 24 : Proportion d'identifications correctes et erronées ayant un score de pertinence supérieur au seuil fixé. La classification a été réalisée en utilisant les modèles de classification basés sur les bases de données de souches cliniques après 12 et 16 heures d'incubation

Après analyse des résultats, un seuil fixé à 90 % pour les scores de pertinence permettrait de retenir 95% des identifications correctes après 12 heures d'incubation et 91 % après 16 heures d'incubation. Cependant, encore 41 % et 40 % d'identification erronées respectivement après 12 et 16 heures d'incubation demeurent au-dessus du seuil de confiance. Il est cependant important de noter que le nombre d'identifications correctes est de 7178 après 12 heures d'incubation et de 10226 après 16 heures d'incubation alors que le nombre d'identification erronées est de 360 et 660 respectivement après 12 et 16 heures d'incubation. Ainsi même sans seuil de confiance fixé, les taux d'identification erronée sont de 4,8 % et 6% après respectivement 12 et 16 heures d'incubation. Bien que ces fréquences d'erreurs de classification soient faibles, l'utilisation des scores de pertinence peut aider à diminuer ces taux d'erreurs. Il est à préciser ici ces seuils de confiance n'auront aucun impact sur les taux d'identifications correctes des modèles. En effet, les identifications erronées seront simplement considérées comme non utilisables.

II.D. Discussion et perspectives

II.D.1. Evaluation et adaptation du prototype initial

Au niveau de la partie matérielle de l'automate, deux points ont été relevés comme problématiques. Le premier était l'absence de plaque de calibration pour pouvoir paramétrer correctement l'automate. Cette plaque a été développée et usinée pour permettre l'utilisation du programme de calibration modifié. Le second point matériel concernait l'éclairage LED de l'automate. En effet, cet éclairage n'est pas homogène et d'intensité non réglable. Le manque d'homogénéité de cet éclairage crée une zone sombre au centre de la gélose. La détection de colonies étant basée sur la détection de niveau de gris, cette zone centrale est détectée comme un composant à part entière. Sa taille demeure cependant trop grande pour qu'elle soit considérée comme une colonie à analyser. Cependant, la détection de cette zone comme un composant par le logiciel empêche la détection de colonies d'intérêt se trouvant dans cette zone. Ainsi, en l'état actuel, les colonies présentes au centre des géloses analysées ne peuvent pas être détectées. Le fait de ne pas pouvoir régler l'intensité de cet éclairage pose également un problème et crée une potentielle surexposition des images avec une perte d'information sur les colonies détectées. Actuellement la forte intensité de l'éclairage est compensée par les réglages d'acquisition de la caméra.

D'un point de vue logiciel, l'analyse a révélé le besoin d'adapter le programme de calibration de l'automate ainsi que la nécessité d'intégrer un programme de compensation de visée LASER. Une restructuration de l'interface graphique du logiciel a été jugée nécessaire pour faciliter l'utilisation et l'accès aux différentes fonctions du logiciel. Le module de classification initial s'est révélé lent et peu ergonomique, notamment avec la nécessité d'un changement de système d'exploitation pour son utilisation. Son remplacement par un module de classification basé sur un réseau neuronal convolutif et ayant une communication automatique avec le logiciel principal s'est avéré nécessaire.

L'adaptation du programme de calibration a permis de prendre en compte les nouveaux paramètres requis par la disposition actuelle du poste d'analyse, notamment le calcul de la rotation subit par les géloses lors de leur analyse. La création de la plaque de calibration a permis l'utilisation du programme de manière semi-automatique. La sélection manuelle des différents points de calibration est toujours requise pour mener à bien la calibration. L'ajout en complément d'un programme de compensation de visée LASER permet au logiciel d'améliorer sa précision de détection et de ciblage des colonies à analyser.

La restructuration de l'interface graphique a permis une navigation plus facile dans le logiciel de l'automate avec un accès simplifié aux différentes parties du programme (analyses en cours, réglages des paramètres). L'intégration de la base de données SQL permet une gestion facilitée des analyses en cours avec notamment la gestion automatique d'analyses devant être réalisée au même moment.

Toutes ces modifications ont permis l'obtention d'un prototype fonctionnel et suffisamment fiable pour pouvoir être acquérir et classifier des patterns de diffraction issus de colonies bactériennes.

II.D.2. Identification des souches bactériennes

Les résultats obtenus pendant ce projet de thèse permettent d'établir une preuve de concept démontrant que l'analyse de pattern de diffraction LASER à partir de colonies peut être utilisée pour distinguer des espèces bactériennes entre elles et pourrait permettre d'identifier des bactéries. Après 12 et 16 heures d'incubation, le système a permis l'identification de toute colonie isolée correspondant aux critères de taille fixés. Cette approche présente plusieurs avantages. Tout d'abord, elle permet l'analyse d'une multitude de colonies sur une gélose, au contraire de techniques comme la spectrométrie de masse MALDI-TOF qui sont utilisées pour l'analyse d'une ou quelques colonies sélectionnées. La sélection *a priori* d'une colonie à identifier se fait par un personnel formé en se basant notamment sur la forme, la taille et la couleur des colonies. Si deux colonies ont un aspect différent, elles sont supposées comme appartenant à deux espèces différentes et toutes les deux sélectionnées pour analyse. Cependant

cela peut mener à passer outre la présence d'une espèce bactérienne dans les cas de colonies semblables ou d'une espèce minoritaire dans l'échantillon. Dans le cas de l'automate BioKube, l'identification d'une colonie est complètement indépendante des autres colonies présentes sur la gélose. Cela permet une identification de plusieurs espèces bactériennes sur une gélose et donc l'analyse d'échantillons pluri-bactériens. Le temps d'analyse d'une gélose est principalement dépendant du nombre de colonies détectées et dont les patterns de diffraction seront analysés. Ainsi, une gélose avec une centaine de colonies analysables sera analysée en environ 8 minutes. Pour une dizaine de colonies, le temps d'analyse sera plus court aux alentours de 4 minutes. Ce temps d'analyse, bien que court, est à prendre en compte dans l'optique d'une technique automatisée à haut débit. Il est envisageable de faire évoluer la détection des colonies avec un logiciel capable, par analyse d'image, de réaliser une sélection des 10 ou 20 colonies les plus représentatives de l'ensemble de la population de la gélose pour limiter au maximum les temps d'analyses sur des colonies redondantes.

Un autre avantage de l'approche par diffraction LASER est qu'elle est non destructive. Ainsi, à l'opposé de la plupart des autres techniques d'identification, les colonies analysées peuvent être réutilisées pour d'autres analyses que ce soit pour des analyses complémentaires d'identification ou pour l'étude des résistances antibiotiques des bactéries. Enfin un dernier point fort de cette approche est sa simplicité d'utilisation. En effet, la seule étape de manipulation de l'échantillon consiste à son ensemencement sur une gélose, tâche qui, de surcroît, est actuellement facilement automatisable. La manipulation de l'automate et de son logiciel est simple et ne requiert que très peu de temps de formation. De plus, les consommables requis pour les analyses se limitent aux géloses, ce qui fait de cette technique une approche peu coûteuse.

II.D.2.a. Performance de classification des modèles basés sur souches cliniques

Après la constitution de bases de données de patterns de diffraction de souches cliniques, avec l'inclusion de 15 souches cliniques par espèces bactériennes, les résultats de classification se trouvent nettement améliorés. De manière générale, les modèles de classification après 12 et 16 heures d'incubation fournissent des taux d'identifications correctes de respectivement 95,2% et 93,9 %. Ces résultats sont équivalents avec les performances de techniques d'identification actuellement utilisées comme la spectrométrie de masse MALDI-TOF décrite comme ayant des taux d'identification correcte de 96% sur prélèvements cliniques après isolation et amplification de la biomasse par culture (250). Bien entendu cette comparaison est purement illustrative, les conditions d'analyses ayant fourni les résultats ne permettent pas de les comparer avec les capacités d'une technique validée et éprouvée en analyse de routine. Tout d'abord, les bases de données sur lesquelles sont basées ces identifications n'incluent que 6 espèces bactériennes. Avant toute comparaison sérieuse avec des techniques commercialisées, il est nécessaire d'augmenter drastiquement le nombre d'espèces dans les bases de données. En effet, l'ajout d'espèces bactériennes supplémentaires peut tout à fait venir diminuer les performances d'identification.

De plus, les résultats obtenus l'ont été à partir de la classification de patterns de diffraction certes différents de ceux utilisés dans les bases de données mais issus des mêmes souches cliniques. Il est nécessaire donc de continuer cette étude avec l'identification de souches cliniques n'ayant pas été incluses dans les bases de données pour valider les performances des modèles de classification développés.

Enfin, des études complémentaires sont à mener sur la validation des résultats de classification, plus précisément sur les scores de confiance donnés par les modèles de classification. Ces scores peuvent permettre d'établir des seuils de confiance en dessous desquels l'identification sera refusée. L'établissement de ces seuils de confiance pourrait permettre d'éviter des résultats erronés (4,8% et 6% respectivement après 12 et 16 heures d'incubation dans les derniers modèles de classification basés sur souches cliniques). L'étude de la répartition

de ces scores de confiance n'a pas permis d'évaluer le positionnement de ce seuil. Des études plus poussées sont nécessaires pour pouvoir utiliser les scores de confiance comme critères d'évaluation du résultat de classification. Ces études devraient inclure de plus grand nombre de patterns analysés et surtout incorrectement identifiés mais également des patterns jugés de mauvaise qualité (mal centrés, avec des artefacts) pour évaluer également l'impact de la qualité des patterns sur la pertinence de la classification.

II.D.2.b. Souches de référence et souches cliniques

L'utilisation de souches de référence pour la constitution de bases de données et de modèles de classification dont le but final est l'identification de souches cliniques s'est montrée peu concluante. Les très faibles sensibilités d'identification lors de la classification de souches cliniques à partir des modèles de classification basés sur souches de référence (17,5% de sensibilité pour l'identification de *Serratia marcescens* après 12 heures d'incubation) ont démontré que les souches de référence constituent des modèles trop éloignés de la réalité clinique pour être représentatives des espèces bactériennes concernées. De plus, la variabilité intra-espèce des souches cliniques ne peut être prise en compte en utilisant une seule souche par espèce bactérienne pour la constitution des bases de données. Il est donc fondamental de constituer les bases de données à partir de plusieurs souches cliniques d'une même espèce bactérienne pour permettre la prise en compte de leur hétérogénéité par le modèle de classification. L'utilisation de souches d'espèces différentes au sein du complexe *Enterobacter cloacae* peut également avoir un impact sur la capacité d'identification des modèles. La classe de souches cliniques correspondant à ce complexe est composée de patterns issus de quatre espèces différentes : *Enterobacter cloacae* (10 souches), *Enterobacter hormaechei* (3 souches), *Enterobacter kobei* (1 souche) et *Enterobacter ludwigii* (1 souche). La composition de cette classe est critiquable. Elle ne possède pas suffisamment d'espèces du complexe et en nombre homogène pour en être pleinement représentative de ce dernier. Par ailleurs, pas assez de souches d'*Enterobacter cloacae* n'ont pu être collectées durant le projet de thèse pour établir une classe pour cette seule espèce. Le complexe *Enterobacter cloacae* comportant à ce jour 22 espèces

bactériennes, il apparaît difficile de composer une classe unique de 1000 patterns de diffraction comprenant toutes ces espèces avec un nombre équivalent de patterns par espèce particulièrement si la contrainte d'avoir plusieurs souches par espèce bactérienne est ajoutée. Il est donc nécessaire de collecter plus de souches cliniques pour établir des bases de données plus complètes avec des classes pour chaque espèce appartenant à ce complexe. Par ailleurs, certaines souches de *Proteus hauseri* peuvent également être positives au test à l'indole et donc avoir été identifiées à tort comme *Proteus vulgaris*. Il est tout à fait envisageable que ces souches, si elles sont présentes dans la classe *Proteus vulgaris* des bases de données, viennent influencer sur la capacité d'identification des modèles. Par ailleurs ces exemples démontrent les limites de la spectrométrie de masse en tant que technique de référence, particulièrement dans le cas de complexe d'espèces bactériennes. Dans le cadre de l'agrandissement des bases de données, il sera important de considérer d'autres techniques d'identification telles que la PCR 16S pour pallier ponctuellement aux manques de performances du MALDI-TOF.

II.D.3. Diffraction LASER et espèces bactériennes

II.D.3.a. Causes biologiques des changements de formes de patterns de diffraction

Bien que certaines des variations de patterns de diffraction soient visuellement observables entre différentes espèces bactériennes, les causes biologiques qui sont responsables de ces changements sont actuellement inconnues. Plusieurs éléments peuvent rentrer en jeu dans la formation de ces patterns et ce à plusieurs niveaux. Tout d'abord, à l'échelle d'une bactérie, il est possible de supposer que toute différence phénotypique peut entraîner une modification des patterns de diffraction. Comme le montre l'utilisation de la spectrométrie de masse en identification bactérienne, les variations d'expression protéique entre les espèces bactériennes sont nombreuses et suffisamment fortes pour permettre leur discrimination. Ces différences dans la composition protéique d'une bactérie pourraient être à l'origine de variations des patterns de diffraction. Les lipides pourraient également jouer un rôle. Leur présence au sein des membranes

bactériennes est variable en fonction des espèces bactériennes. La composition lipidique des bactéries est d'ailleurs un sujet d'étude pour être appliquée à l'identification bactérienne par spectrométrie de masse (160).

Vraisemblablement, l'architecture et l'organisation des colonies bactériennes pourraient être à l'origine de variations de patterns de diffraction. En effet, les débuts historiques de la classification bactérienne se basaient sur la forme, couleur (et odeur) des colonies bactériennes. De plus, les colonies au fur et à mesure de leur croissance vont développer des zones de lyses et d'absence de croissance (251) qui pourraient influencer la formation des patterns de diffraction et leur évolution au fur et à mesure du temps d'incubation. Il est important de noter que les colonies d'une même espèce peuvent avoir des formes différentes suivant les milieux sur lesquelles elles sont (252). Ceci peut être une explication principale des changements de patterns observés entre plusieurs types de géloses. Cependant, l'architecture d'une colonie est conditionnée par les connexions établies entre bactéries et donc par leurs compositions protéiques et lipidiques principalement. Certaines molécules sont connues pour jouer un rôle dans l'agrégation et la morphogénèse des colonies bactériennes comme les acides aminés, les oligopeptides et les polypeptides (253). Il est donc possible que les protéines et molécules membranaires aient un rôle important dans la forme de patterns de diffraction. Bien entendu, cela reste à ce stade une hypothèse et des études approfondies sur ce sujet seraient à mener. Il peut être notamment envisager d'utiliser une banque de mutants créée à partir d'une même souche bactérienne et de déterminer les mutations qui provoquent des modifications dans le pattern de diffraction. Cette étude permettrait d'évaluer quel type d'éléments ou structures bactériens ont des impacts sur la forme des patterns de diffraction. Connaître les causes des variations des patterns de diffraction permettrait de savoir quelles bactéries sont différenciables par diffraction LASER et jusqu'à quel niveau phylogénétique cette approche permet d'aller.

Un questionnement existe également sur l'hétérogénéité des patterns suivant le diamètre des colonies au sein de la gamme de tailles sélectionnées dans les travaux ici présentés (0,8 à 1 mm). En effet, il est important de pouvoir évaluer si des variations de diamètre de moins de deux dixièmes de millimètres suffisent à faire varier suffisamment les formes des patterns de diffraction pour influencer sur leur classification. Si cela est le cas, deux possibilités peuvent être envisagées. La première est de réduire la gamme de diamètres utilisée pour la sélection des

colonies, de manière à éliminer l'hétérogénéité de patterns créée par les variations de tailles de colonies. La seconde consiste en la prise en compte, au sein des classes constituant les bases de données, de l'hétérogénéité due aux variations de tailles de colonies en s'assurant de la présence, dans les patterns sélectionnés, de l'entièreté de la gamme de diamètres de colonies sélectionnées. Cependant, la faible résolution spatiale des images de géloses prises avec la station LED de l'automate (1 pixel correspond à 0,125 mm) ne permet pas en l'état une mesure fine du diamètre des colonies. Ce point d'évolution de l'automate est abordé dans le paragraphe II.D.4.

II.D.3.b. Sérotypes et sérovars

L'une des espèces bactériennes les plus utilisées en recherche ne fait pas partie des bases de données qui ont été constituées dans ce projet : *Escherichia coli*. Les premières acquisitions des patterns sur *Escherichia coli* ont montré de fortes différences de patterns de diffraction entre plusieurs souches de référence de cette espèce (Figure 41). Des variations entre différentes souches d'*Escherichia coli* sont donc responsables de changement de forme de pattern de diffraction. Il est possible que les différents sérotypes au sein de cette espèce aient chacun des patterns de diffraction spécifiques. L'espèce bactérienne *Escherichia coli* regroupe environ 180 sérotypes différents. Cependant, l'identification de ces sérotypes n'est jamais réalisée en routine de bactériologie car, excepté pour les quelques sérotypes producteurs de toxine Shiga-like (O157 H7 étant le plus courant) le sérotypage d'*Escherichia coli* n'est pas médicalement pertinent. Pour pouvoir étudier l'identification d'*Escherichia coli* par diffraction LASER, il aurait été nécessaire de créer une banque de données par sérotype pour cette espèce. Cependant, peu de sérotypes sont disponibles en souches de référence et le sérotypage des souches cliniques nécessite l'utilisation de techniques peu répandues et coûteuses en matière de temps et de finances. Les travaux menés pendant la thèse ici décrite étant dans un but de preuve de concept, il a donc été décidé de ne pas inclure *Escherichia coli* dans les espèces bactériennes présentes dans les bases de données.

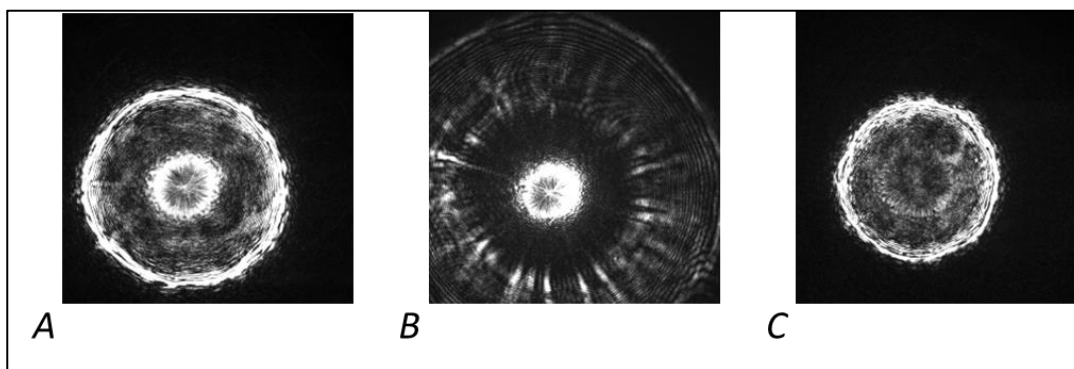


Figure 41 : Patterns de diffraction pour les souches de références d'*Escherichia coli* ATCC 10536 (A), ATCC 25922 (B) et ATCC 35218 (C)

La perspective d'une différenciation possible de sérotypes par diffraction LASER pourrait poser un problème pour certaines espèces bactériennes comme *Escherichia coli*. En effet, cela obligerait la création de base de données pour chaque sérotype d'une espèce sans que la caractérisation du sérotype soit pertinente et intéressante d'un point de vue biologique et médicale. Cependant, pour d'autres espèces, être capable d'identifier les différents sérotypes ou sérovars peut être d'une importance capitale. C'est notamment le cas pour les Salmonelles.

L'identification des différentes sérovars de Salmonelles est nécessaire d'un point de vue épidémiologique et de pathogénicité et est actuellement réalisée par agglutination de bille de latex. Cette technique demande beaucoup de temps de manipulation car elle nécessite de tester les différents sérovars. C'est dans ce type de contexte que la capacité d'identifier des sérotypes ou des sérovars pourrait être un avantage pour l'analyse de pattern de diffraction. Bien que des différences soient observables à l'œil nu entre des patterns de diffractions issus de plusieurs sérovars de *Salmonella sp* (Figure 42), cette capacité de différenciation est encore hypothétique et des travaux sur la constitution de base de données incluant les différents sérovars de *Salmonella* doivent être réalisés avant que toute conclusion ne puisse être tirée.

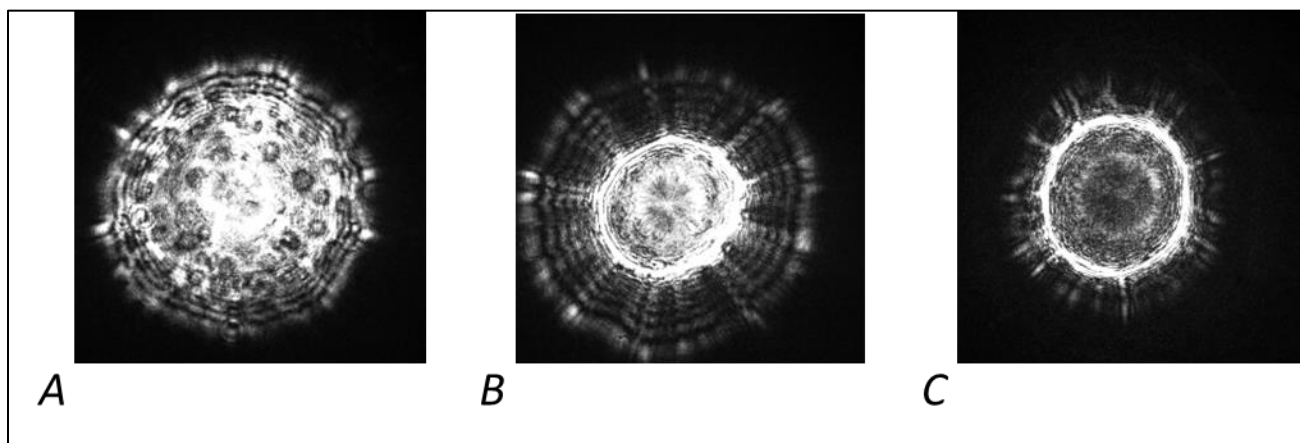


Figure 42 : Patterns de diffraction de sérovars de *Salmonella* sp. A : Ajiobo. B : Enteritidis. C : Typhimurium

II.D.3.c. Acquisition des patterns

Le principal délai d'analyse de l'identification par analyse de patterns de diffraction est le temps d'incubation requis avant l'acquisition des patterns. Ce problème est commun à toutes les techniques d'identification basées sur la culture des bactéries. Les temps d'incubation ici étudiés sont de 12 et 16 heures. Ces temps permettent de positionner l'identification par pattern de diffraction dans des délais de rendus similaires à ceux de la spectrométrie de masse MALDI-TOF. Toutefois, l'identification par MALDI-TOF MS n'étant pas totalement automatisée nécessite le dépôt, après croissance sur milieu gélosé, de la colonie à analyser sur la plaque MALDI par le personnel de laboratoire. Dans le cas de laboratoires n'ayant pas de personnel présent en continu, cette action manuelle peut créer des allongements de temps de rendu. En effet, les analyses pouvant être réalisées durant la nuit ne seront faites que le lendemain. Cela est un des avantages de l'automatisation complète du procédé d'analyse de l'automate BioKube qui, une fois la gélose mise en incubation, ne nécessite plus d'intervention humaine pour réaliser son analyse. Cependant, ces délais de rendus sont dépendant de la vitesse de croissance des bactéries. Le fait que *Staphylococcus saprophyticus* n'ait pas pu être inclus dans la base de données après 12 heures d'incubation en est un exemple parfait. Ce constat impacte notamment l'expansion des bases de données à d'autres espèces bactériennes. A chaque temps d'incubation, la base de

données correspondante ne pourra inclure que les espèces bactériennes avec une croissance suffisamment rapide pour permettre la détection et l'analyse des colonies.

Il est également important d'évaluer la fenêtre d'acquisition d'un pattern en fonction du temps d'incubation sur lequel les modèles de classification sont basés. Ce point rejoint un questionnement sur l'évolution du pattern de diffraction d'une colonie au cours du temps et en fonction de la taille de cette dernière. Il est important de connaître les intervalles de taille et de temps durant lesquels l'utilisation d'un modèle de classification permet une classification fiable. Est-il possible de classer correctement un pattern de diffraction acquis après 13 heures d'incubation avec un modèle de classification basé sur des patterns acquis après 12 heures d'incubation ? A partir de quel diamètre de colonie un pattern ne peut plus être correctement classifié par un modèle de classification basé sur des patterns de colonies ayant des diamètres compris entre 0,8 et 1 mm ? Répondre à ces questions permettrait notamment d'améliorer la constitution des bases de données ainsi que la qualité de classification en termes de critères de sélection du diamètre des colonies. Cela permettrait également l'amélioration de la gestion du flux d'analyse lorsque plusieurs géloses sont à analyser en même temps, en connaissant les bornes de temps dans lesquelles l'analyse peut être effectuée sans perte de qualité du résultat.

Un second point à développer lors de l'acquisition des patterns de diffraction est l'évaluation de leur qualité avant leur classification. En effet, dans le cas d'une identification non concluante, il est important de pouvoir déterminer si le manque de précision vient de la classification du pattern ou de son acquisition. Ce procédé de qualification existe notamment en identification par spectrométrie de masse où la qualité du spectre est évaluée suivant de nombreux critères (254). Il est possible ici d'imaginer quelques critères pour évaluer la qualité d'un spectre comme la présence ou non d'artefact causée par une colonie adjacente ou encore la précision du centrage du LASER par rapport à la colonie. Cependant, il est nécessaire de savoir dans quelle mesure ces variations d'acquisition influencent le résultat de la classification mais également quelles causes peuvent créer des variations de patterns délétères pour l'identification des espèces. Ces informations ne peuvent être connues sans mener des études plus fondamentales sur les causes de changement de forme des patterns de diffraction.

Certains patterns de diffraction (notamment pour *Staphylococcus saprophyticus*) étaient plus grands la zone d'acquisition de la caméra LASER. L'impact de ce phénomène sur les performances de classification des modèles n'a cependant pas été étudié durant le projet. Cependant, il a été démontré que modifier la distance entre la source lumineuse et la colonie bactérienne permettait d'influer sur la taille des patterns de diffraction (245). Une modification du système d'acquisition pour permettre de modifier cette distance pourrait permettre de n'utiliser uniquement des patterns de diffraction capturés en entier par le capteur CMOS.

II.D.4. Utilisation d'un réseau neuronal convolutif pour la classification de pattern de diffraction

L'utilisation d'un réseau neuronal convolutif pour la classification de pattern de diffraction apporte plusieurs avantages pour une technique d'identification bactérienne. Tout d'abord, la rapidité d'analyse des patterns permet une utilisation en haut-débit de cette technique. La classification du pattern de diffraction peut ainsi être rendu moins d'une seconde après son acquisition par l'automate. Ensuite, les capacités de création de modèles de classification en apprentissage supervisé permettent la création de modèle de classification fiables à partir d'une base de données de grande taille. Ce développement peut de surcroit être fait sans connaissance particulière des moyens de différenciation entre les différentes classes de la base de données puisque les critères de classification en sont pas fournis au réseau neuronal mais bien déduis par ce dernier. Ce dernier point est particulièrement intéressant dans le cas ici présenté puisque les connaissances sur les raisons de variations de formes des patterns sont peu connues. Cependant, la puissance de classification d'un modèle dépend entièrement de la base de données utilisée pour son développement. En effet, la base de données doit être représentative de la population cible de la classification. L'utilisation des souches de références pour la classification de souches cliniques dans le cas présent en est l'exemple parfait. Si la base de données n'est pas une bonne illustration de la population cible ou n'est représentative que d'une portion de cette population, les modèles de classification qui seront développés ne pourront pas être utilisables. Dans le cas présent, la constitution d'une base de données basée sur une souche de référence par espèce bactérienne n'a pas permis au réseau neuronal de prendre en compte les variations existantes entre différentes souches cliniques d'une même espèce bactérienne.

Une attention particulière doit être également portée aux méthodes par lesquelles les souches incluses dans les bases de données ont été identifiées. En effet, l'identification par MALDI-TOF MS n'est pas fiable dans tous les cas. L'exemple du complexe *Enterobacter cloacae* ou l'impossibilité de différencier certaines espèces du genre *Proteus* montrent le besoin fort d'analyses plus poussées de l'espèce bactérienne des souches avant leur ajout dans les bases de données. Ainsi les futures bases de données développées pour la classification des patterns de diffraction doivent se baser sur des identifications fiables des espèces bactériennes. La

confirmation en cas de doute par des techniques comme la PCR 16S ou le séquençage de génome entier est nécessaire pour une qualité optimale des bases de données et donc des modèles de classification.

Bien que le terme intelligence artificielle soit employé, un réseau neuronal convolutif ne peut qu'extrapoler des informations à partir d'une base de données existante. Les modèles de classification développés ont pour but de trouver des méthodes de classification pour différencier au mieux les différentes classes de la base de données. Ainsi une attention particulière doit être portée sur la constitution des bases de données et tout particulièrement sur la sélection des patterns de diffraction. Une sélection humaine des patterns basée sur des critères subjectifs peut induire un biais de classification dans les modèles. Ce biais peut être d'autant plus compliqué à éviter quand les connaissances sur les formes des patterns ainsi que les causes de ces formes sont inconnues. N'ayant pas de critères spécifiques définis de manière scientifique sur la sélection de patterns, il est d'autant plus facile de se baser sur des critères subjectifs et peu pertinents. Il est donc d'autant plus important pour la continuation des travaux présentés ici d'approfondir les connaissances fondamentales sur la forme de patterns de diffraction et leur corrélation avec les différentes espèces bactériennes. Ces connaissances permettront d'améliorer la qualité des bases de données et peut être le développement de protocoles d'évaluation objectifs de la qualité des patterns acquis.

II.D.5. Perspectives d'optimisation de l'automate et de son logiciel

L'automate Biokube, dans la configuration présentée dans cette thèse, semble confirmer la possibilité de réaliser une identification bactérienne par analyse des patterns de diffraction LASER à partir de colonies ayant poussées sur milieu gélosé, de façon non destructive et sans intervention manuelle. Le développement de cet automate, de plusieurs bases de données et de modèles de classification des patterns acquis, a permis d'aboutir à un système capable de réaliser la détection de colonies sur une gélose incolore, l'acquisition de leur pattern de diffraction ainsi que l'identification des bactéries associées à ces patterns par comparaison avec une base de données.

Bien que l'automate Biokube soit fonctionnel et permette l'acquisition de patterns de diffraction, certaines améliorations peuvent être apportées et seront nécessaires dans l'optique d'une utilisation de routine. La principale facette à améliorer est le panneau LED éclairant la gélose lors de la détection des colonies. Actuellement, cet éclairage n'est pas homogène et d'une intensité non réglable. Le manque d'homogénéité de l'éclairage crée une zone sombre au centre de la gélose empêchant la détection des colonies à cet endroit. Changer le panneau LED pour un éclairage homogène est nécessaire pour permettre la détection de la totalité des colonies présentes sur une gélose. Il est également important de pouvoir agir sur l'illumination de la gélose pour permettre d'optimiser la qualité d'image lors de la détection des colonies. De plus, dans l'optique d'une utilisation de plusieurs types de géloses, l'opacité de ces dernières peut être amenée à varier. Il serait donc intéressant de pouvoir adapter la puissance d'éclairage en fonction de l'opacité des géloses analysées.

Un second point d'amélioration est la calibration de l'automate. Actuellement, cette calibration bien que fonctionnelle manque encore de précision. Ce manque de précision est en partie compensé par le programme de compensation du logiciel lors de l'acquisition des patterns. Plusieurs améliorations peuvent permettre d'agir sur la précision de calibration. Tout d'abord, l'éclairage LED peut jouer un rôle. Le manque d'homogénéité de cet éclairage, ainsi que l'épaisseur de la plaque de calibration, crée un éclairage non parallèle aux trous de la plaque de calibration ce qui résulte en une image de cette plaque avec des trous déformés (Figure 43). Cela impacte le positionnement de ces trous de calibration par l'utilisateur et donc fatalement, la précision de la calibration qui en résulte. La fréquence à laquelle l'automate doit être recalibré est actuellement inconnue. Seule une utilisation intensive et prolongée pourra permettre d'acquérir des informations sur ce point.

Un autre élément pouvant être amélioré est la caméra d'acquisition des images de géloses. Actuellement, la caméra prend des images en 1280 par 1024 pixels. Cette résolution donne actuellement une correspondance pour un pixel de 0,125 millimètre. Ainsi, 0,125 millimètre est la limite inférieure de mesure de distance de l'automate. Augmenter la résolution de la caméra peut ainsi permettre de diminuer la distance réelle correspondant à un pixel et ainsi améliorer la précision générale de mesure de l'automate que ce soit au niveau de la détection des positions des colonies ou lors de la mesure de leur diamètre.

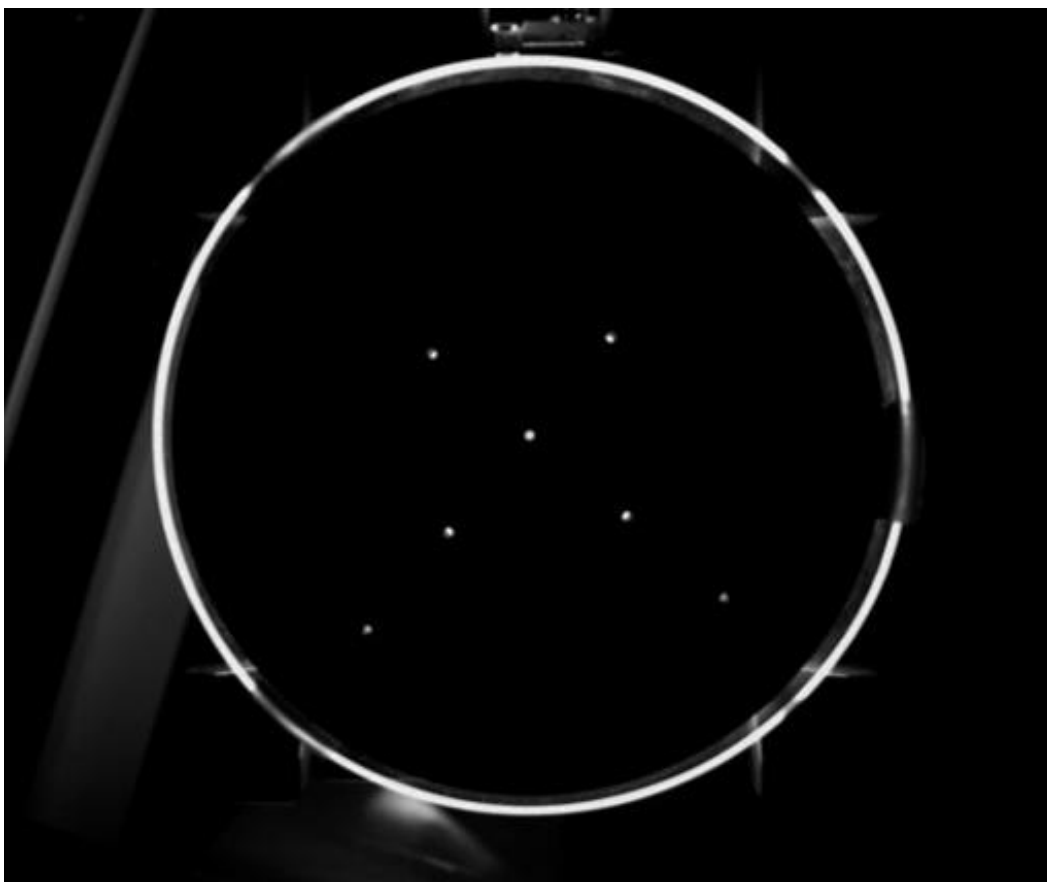


Figure 43 : Acquisition de la plaque de calibration. On y observe une déformation des trous de calibration créée par l'éclairage LED non homogène

Enfin, un dernier élément pouvant être amené à évoluer est le LASER utilisé par l'automate. Actuellement, ce LASER est de couleur rouge (longueur d'onde de 635 nm). Bien qu'étant tout à fait fonctionnel avec les géloses incolores de type TSA, ce LASER pose problème dans l'optique de l'utilisation de géloses au sang. En effet, ces géloses étant de couleur rouge, le LASER est naturellement diffracté par la gélose elle-même et donc ne permet pas l'acquisition de patterns de diffraction. Cependant, les géloses au sang restent parmi les plus utilisées en routine de bactériologie. Ainsi ne pas pouvoir faire d'identification sur ce type de gélose peut constituer un handicap pour une future technique d'identification bactérienne.

Ce problème pourrait être résolu en utilisant un LASER de couleur verte (532 nm de longueur d'onde). Cette longueur d'onde, ne diffractant pas naturellement sur des milieux gélosés au sang, permettrait l'application de la diffraction LASER à de nouveaux types de géloses.

III. Conclusion et perspectives

Les techniques d'identification bactérienne sont en constante évolution. Le besoin permanent d'accélération du temps de rendu de résultats fait de la bactériologie, un domaine encore très manuel, un terrain de choix pour la mise en place de solutions automatisées et équipées d'intelligence artificielle. Aux critères de qualité et de rapidité de résultats s'ajoutent également des critères financiers. Les techniques d'identification se doivent également d'être facilement utilisables et avec peu de temps de manipulation pour limiter les risques de contamination à la fois des prélèvements et du personnel de laboratoire.

L'analyse de patterns de diffraction issus de colonies bactériennes semble être tout à fait en adéquation avec ces critères. Tout d'abord, cette approche implique une technologie simple et un automate peu complexe. Les coûts d'acquisition et d'analyse de cette technique sont donc peu élevés, le seul consommable requis pour cette analyse étant les milieux gélosés, qui sont dans tous les cas utilisés pour la mise en culture des prélèvements. Les temps de manipulations des prélèvements sont également minimales puisqu'ils se résument à l'ensemencement du prélèvement sur gélose, effectué soit manuellement ou soit de manière automatisée. Cette approche, consistant uniquement à éclairer une colonie avec un LASER, est également non destructive fournissant ainsi la possibilité de réutilisation des colonies pour des analyses complémentaires. De plus, elle permet une classification colonie par colonie et est donc tout à fait applicable à des prélèvements polymicrobiens.

Au démarrage de ce projet, un prototype a été mis à disposition par la société BioKube. Une refonte de l'interface graphique ainsi que l'ajout de plusieurs fonctions telles que des programmes de calibration et de compensation de visée LASER ont été réalisés. Une adaptation de ce logiciel pour pouvoir communiquer avec le réseau neuronal en charge de la classification des patterns a également été réalisée. Ces développements ont permis l'obtention d'un prototype, certes encore perfectible, mais fonctionnel et permettant la détection de colonies bactériennes ainsi que l'acquisition et la classification de leur pattern de diffraction.

La construction des bases de données a été réalisée avec des patterns de diffraction issus de colonies bactériennes après 12 et 16 heures d'incubation et ce pour 6 espèces bactériennes : *Enterococcus faecalis*, *Enterococcus faecium*, *Enterobacter cloacae*, *Proteus vulgaris*, *Serratia marcescens* et *Staphylococcus saprophyticus*. Dans un premier temps, une souche de référence par espèce bactérienne a été utilisée pour récolter les patterns de diffraction et construire les premières bases de données. Les modèles de classification issus de l'apprentissage du réseau neuronal sur ces bases de données ont d'abord donné des résultats intéressants lors de leur auto-évaluation mais se sont révélées très peu efficace lors de l'analyse de patterns de diffraction issus de souches cliniques avec des taux globaux d'identification correcte de 45,8% après 12 heures d'incubation et de 67,2% après 16 heures d'incubation. L'utilisation de souches de référence ainsi que d'une seule souche par espèce ne permet donc pas de représenter fidèlement la population clinique d'une espèce bactérienne. En effet, une forte hétérogénéité entre souches cliniques d'une même espèce bactérienne existe et doit être prise en compte dans la constitution des bases de données.

Des bases de données ont donc été réalisés à partir de souches cliniques en incluant 15 souches différentes par espèce bactérienne. Les modèles de classification créés à partir de ces dernières bases de données ont permis une amélioration des performances de classification avec des taux d'identification correcte de 95,4% et 94% après respectivement 12 et 16 heures d'incubation. Ces performances sont très intéressantes mais demeurent des résultats initiaux qui doivent être à confirmer. Tout d'abord, les patterns utilisés pour évaluer les performances de classification sont issus des souches cliniques utilisées dans les bases de données. Il est nécessaire d'évaluer les capacités d'identification des modèles sur d'autres souches cliniques pour s'assurer d'une prise en compte totale de la diversité intra-espèce. De plus, le nombre d'espèces bactériennes incluses dans les bases de données n'est pas suffisant pour pleinement évaluer les capacités d'identification de la technique. Un agrandissement des bases de données avec l'inclusion de nouvelles espèces bactériennes est également nécessaire. Néanmoins, l'analyse de pattern de diffraction de colonies bactériennes semble être applicable à des fins d'identification bactérienne. Avec des temps de rendu de résultats de l'ordre de 12 à 16 heures, cette approche apparait comme légèrement plus rapide que les techniques basées sur la culture

bactérienne actuellement utilisées et notamment l'identification par spectrométrie de masse MALDI-TOF avec un temps de rendu de résultats de 18 à 24 heures.

Un travail d'étude plus fondamental sur la formation des patterns de diffraction et la corrélation entre leur forme et les caractéristiques moléculaires des bactéries serait également intéressant à entreprendre. Actuellement, la relation entre forme de pattern et espèces bactériennes est uniquement observationnelle et factuelle. Comprendre la mécanistique de cette corrélation pourrait permettre une meilleure analyse et classification des patterns de diffraction mais également l'application de cette approche à l'identification de sérotypes bactériens ou de résistances antibiotiques.

Cette approche innovatrice peut donc être utilisée en identification bactérienne mais nécessite encore des études et développement. L'élargissement de la base de données peut s'avérer fastidieux dans une optique d'atteindre une capacité d'identification englobant les 2000 et quelques espèces bactériennes pathogènes pour l'humain. De plus, aucune certitude n'existe sur la capacité à différencier l'intégralité de ces espèces seulement par l'analyse des patterns de diffraction issus de colonies. Cependant, de par la détection des colonies avant récupération de leur pattern de diffraction, certaines informations sur l'espèce bactérienne peuvent être disponible pour faciliter l'identification. Comme l'a démontré l'étude annexe réalisée sur l'automate ScanStation (Interscience), il est possible d'imager la croissance des colonies sur milieux gélosés et ainsi obtenir des informations sur la forme, couleur et vitesse de croissance des colonies bactériennes. Ces informations, bien qu'insuffisante en elles-mêmes pour identifier une espèce bactérienne, peuvent cependant être très utiles pour affiner le panel d'espèces bactériennes compatibles avec un comportement spécifique des colonies sur gélose. D'autres approches comme l'analyse de composés organiques volatiles peuvent également être envisagées pour venir épauler la classification par analyse de pattern de diffraction. Il est donc tout à fait possible d'envisager un système composite incluant différentes approches d'identification. Plusieurs algorithmes d'intelligence artificielle déduiraient ainsi une identification de l'espèce bactérienne multimodale en se basant sur des patterns de diffraction, les caractéristiques morphologiques des colonies et l'analyse de composés volatiles qu'elles sécrètent par exemple.

Un grand nombre d'approches innovantes sont à l'étude pour l'identification d'espèces bactériennes. L'arrivée des intelligences artificielles en bactériologie ouvre la possibilité d'un traitement de très grands volumes de données permettant ainsi l'analyse de données multifactorielles. Ainsi, il est fort probable que les techniques d'identification de demain soient des regroupements de plusieurs approches se complétant les unes les autres pour obtenir la meilleure capacité d'identification possible et que l'analyse de patterns de diffraction y ait toute sa place.

REFERENCES

1. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol.* sept 2020;38(9):1079-86.
2. Lanyi B. Classical and rapid identification methods for medically important bacteria. In: *Methods in microbiology.*
3. Bascomb S. Enzymes tests in bacterial identification. In: *Methods in microbiology.* 1987. p. 105-60.
4. McGregor A, Schio F, Beaton S, Boulton V, Perman M, Gilbert G. The MicroScan WalkAway diagnostic microbiology system — an evaluation. *Pathology (Phila).* avr 1995;27(2):172-6.
5. Funke G, Monnet D, Debernardis C, Graevenitz AV, Freney J. Evaluation of the VITEK 2 System for Rapid Identification of Medically Relevant Gram-Negative Rods. *J Clin Microbiol.* juill 1998;36(7):1948-52.
6. Eigner U, Schmid A, Wild U, Bertsch D, Fahr AM. Analysis of the Comparative Workflow and Performance Characteristics of the VITEK 2 and Phoenix Systems. *J Clin Microbiol.* août 2005;43(8):3829-34.
7. Liu H, Du Z, Wang J, Yang R. Universal Sample Preparation Method for Characterization of Bacteria by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *Appl Environ Microbiol.* mars 2007;73(6):1899-907.
8. Bizzini A, Greub G. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. *Clin Microbiol Infect.* nov 2010;16(11):1614-9.
9. SFM. REMIC, Référentiel en Microbiologie Médicale. 6^e éd. 2018.
10. Bryan A, Youngster I, McAdam AJ. Shiga Toxin Producing Escherichia coli. *Clin Lab Med.* juin 2015;35(2):247-72.
11. Chandrasekaran R, Lacy DB. The role of toxins in Clostridium difficile infection. *FEMS Microbiol Rev.* nov 2017;41(6):723-50.
12. Xu F, Ge C, Luo H, Li S, Wiedmann M, Deng X, et al. Evaluation of real-time nanopore sequencing for Salmonella serotype prediction. *Food Microbiol.* août 2020;89:103452.
13. Kumar A. Optimizing antimicrobial therapy in sepsis and septic shock. *Crit Care Nurs Clin North Am.* mars 2011;23(1):79-97.

14. Whiley H, Taylor M. Legionella detection by culture and qPCR: Comparing apples and oranges. *Crit Rev Microbiol.* janv 2016;42(1):65-74.
15. Sorlozano A. Comparative Evaluation of Three Culture Methods for the Isolation of Mycobacteria from Clinical Samples. *J Microbiol Biotechnol* [Internet]. mai 2009 [cité 20 sept 2022];19. Disponible sur: <http://jmb.inforang.com/>
16. Monsen T, Rydén P. Flow Cytometry Analysis Using Sysmex UF-1000i Classifies Uropathogens Based on Bacterial, Leukocyte, and Erythrocyte Counts in Urine Specimens among Patients with Urinary Tract Infections. Onderdonk AB, éditeur. *J Clin Microbiol.* janv 2015;53(2):539-45.
17. Fritzenwanker M, Imirzalioglu C, Chakraborty T, Wagenlehner FM. Modern diagnostic methods for urinary tract infections. *Expert Rev Anti Infect Ther.* sept 2016;14(11):1047-63.
18. Falbo R, Sala MR, Bussetti M, Cappellini F, Giacobone C, Fania C, et al. Performance evaluation of a new and improved cuvette-based automated urinalysis analyzer with phase contrast microscopy. *Clin Chim Acta.* avr 2019;491:126-31.
19. Íñigo M, Coello A, Fernández-Rivas G, Carrasco M, Marcó C, Fernández A, et al. Evaluation of the SediMax automated microscopy sediment analyzer and the Sysmex UF-1000 i flow cytometer as screening tools to rule out negative urinary tract infections. *Clin Chim Acta.* mai 2016;456:31-5.
20. Coopman R, Speeckaert MM, Aps JK, Delanghe JR. Flow cytometry-based analysis by Sysmex-UF1000i® is an alternative method in the assessment of periodontal inflammation. *Clin Chim Acta.* sept 2014;436:176-80.
21. Maleb A, Bouayadi O, El Malki J, Rifai S, Lamrabat S, Benaissa E, et al. Cytological examination of cerebrospinal fluid: Sysmex UF-1000i versus optical microscopy. *Anal Biochem.* janv 2020;589:113501.
22. Wu Y, Yang X, Wang H, Wang T. Evaluation of the Sysmex UF-1000i system as an alternative for the screening of genital tract inflammation in male infertility patients. *Andrologia.* mai 2018;50(7):e13054.
23. Bartholomew JW, Mittwer T. The gram stain. *Bacteriol Rev.* mars 1952;16(1):1-29.
24. La Scolea LJ, Dryja D. Quantitation of bacteria in cerebrospinal fluid and blood of children with meningitis and its diagnostic significance. *J Clin Microbiol.* févr 1984;19(2):187-90.
25. Canis F, Chatelin N, Guérin F, Ranque S. Infections urinaires. In: REMIC. 6^e éd. 2018. p. 181-97.
26. Deun AV, Hossain MA, Gumusboga M, Rieder HL. Ziehl-Neelsen staining: theory and practice. *Int J Tuberc Lung Dis.* janv 2008;12(1):108-10.

27. Vilchèze C, Kremer L. Acid-Fast Positive and Acid-Fast Negative Mycobacterium tuberculosis : the Koch Paradox. *Microbiol Spectr.* mars 2017;5(2).
28. Kim SY, Kim YJ, Hwang SH, Kim HH, Lee EY, Chang CL. Evaluation of an acid-fast bacilli autostainer for concentrated sputum smears. *Int J Tuberc Lung Dis.* janv 2008;12(1):39-43.
29. Qiang L, Lin ZY, Qing W, Feng DY, Di X, Guo YH, et al. Evaluation of the New Automatic Mycob.T Stainer and Scanner for Detecting Acid-fast Bacilli in China. *Biomed Environnemental Sci.* août 2018;31(8):572-8.
30. Doern GV. Detection of Selected Fastidious Bacteria. *Clin Infect Dis.* janv 2000;30(1):166-73.
31. Emery CL, Appleman MD, Siders JA, Davis TE. Rapid devices and instruments for the identification of anaerobic bacteria. In: *Manual of Commercial Methods in Clinical Microbiology.* 2nd éd. Willey Blackwell; 2016. p. 56-86.
32. Desjardins M, Sant N, Miron-Celis M, Gosal J, Jémus MF, Jémus-Gonzalez E. Impact of reduced incubation times on culture and susceptibility testing of urine cultures incubated in the BD Kiestra ReadA Compact incubators. *Diagn Microbiol Infect Dis.* janv 2020;96(1):114899.
33. Morel P, Naegelen C, Deschaseaux M, Bardiaux L. Prévention du risque bactérien : inactivation des pathogènes/détection des bactéries. *Transfus Clin Biol.* mai 2013;20(2):109-14.
34. Li G, Sun J, Pan S, Li W, Zhang S, Wang Y, et al. Comparison of the Performance of Three Blood Culture Systems in a Chinese Tertiary-Care Hospital. *Front Cell Infect Microbiol.* août 2019;9:285.
35. Perry JD, Freydière AM. The application of chromogenic media in clinical microbiology: Application of chromogenic media. *J Appl Microbiol.* juill 2007;103(6):2046-55.
36. Darkoh C, DuPont HL, Kaplan HB. Novel One-Step Method for Detection and Isolation of Active-Toxin-Producing Clostridium difficile Strains Directly from Stool Samples. *J Clin Microbiol.* déc 2011;49(12):4219-24.
37. Perry JD. A Decade of Development of Chromogenic Culture Media for Clinical Microbiology in an Era of Molecular Diagnostics. *Clin Microbiol Rev.* janv 2017;30(2):449-79.
38. Siller-Ruiz M, Hernández-Egido S, Sánchez-Juanes F, González-Buitrago JM, Muñoz-Bellido JL. Fast methods of fungal and bacterial identification. MALDI-TOF mass spectrometry, chromogenic media. *Enfermedades Infecc Microbiol Clin Engl Ed.* mai 2017;35(5):303-13.

39. Foschi C, Gaibani P, Lombardo D, Re MC, Ambretti S. Rectal screening for carbapenemase-producing Enterobacteriaceae: a proposed workflow. *J Glob Antimicrob Resist.* juin 2020;21:86-90.
40. Carroll KC, Weinstein MP. Manual and automated systems for detection and identification of microorganisms. In: *Manual of clinical microbiology*. 9th Ed. 2007. p. 192-217.
41. Biomérieux. Fiche technique API 20E. mai 2010;
42. Jorgensen JH, Pfaller MA, éditeurs. *Manual of Clinical Microbiology*. 11. ed. Washington, DC: ASM Press; 2015.
43. Jin WY, Jang SJ, Lee MJ, Park G, Kim MJ, Kook JK, et al. Evaluation of VITEK 2, MicroScan, and Phoenix for identification of clinical isolates and reference strains. *Diagn Microbiol Infect Dis.* août 2011;70(4):442-7.
44. Anhalt JP, Fenselau Catherine. Identification of bacteria using mass spectrometry. *Anal Chem.* févr 1975;47(2):219-25.
45. Tanaka K. The Origin of Macromolecule Ionization by Laser Irradiation (Nobel Lecture). *Angew Chem Int Ed.* août 2003;42(33):3860-70.
46. Krishnamurthy T, Ross PL. Rapid Identification of Bacteria by Direct Matrix-assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Whole Cells. *Rapid Commun Mass Spectrom.* oct 1996;10(15):1992-6.
47. Claydon MA, Davey SN, Edwards-Jones V, Gordon DB. The rapid identification of intact microorganisms using mass spectrometry. *Nat Biotechnol.* nov 1996;14(11):1584-6.
48. Barenfanger J, Drake C, Kacich G. Clinical and Financial Benefits of Rapid Bacterial Identification and Antimicrobial Susceptibility Testing. *J Clin Microbiol.* mai 1999;37(5):1415-8.
49. Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom Rev.* déc 2001;20(4):157-71.
50. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, Francois P, et al. Comparison of Two Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Methods with Conventional Phenotypic Identification for Routine Identification of Bacteria to the Species Level. *J Clin Microbiol.* avr 2010;48(4):1169-75.
51. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P, Rolain JM, et al. Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *Clin Infect Dis.* août 2009;49(4):543-51.

52. Bizzini A, Durussel C, Bille J, Greub G, Prod'hom G. Performance of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Identification of Bacterial Strains Routinely Isolated in a Clinical Microbiology Laboratory. *J Clin Microbiol.* mai 2010;48(5):1549-54.
53. Enroth H, Retz K, Andersson S, Andersson C, Svensson K, Ljungström L, et al. Evaluation of QuickFISH and maldi Sepsityper for identification of bacteria in bloodstream infection. *Infect Dis.* févr 2019;51(4):249-58.
54. Eigner U, Holfelder M, Oberdorfer K, Betz-Wild U, Bertsch D, Fahr AM. Performance of a Matrix-Assisted Laser Desorption Ionization-Time-of-Flight Mass Spectrometry System for the Identification of Bacterial Isolates in the Clinical Routine Laboratory. *Clin Lab.* août 2009;55:289-96.
55. Verroken A, Defourny L, Lechgar L, Magnette A, Delmée M, Glupczynski Y. Reducing time to identification of positive blood cultures with MALDI-TOF MS analysis after a 5-h subculture. *Eur J Clin Microbiol Infect Dis.* janv 2015;34(2):405-13.
56. Ponderand L, Pavese P, Maubon D, Giraudon E, Girard T, Landelle C, et al. Evaluation of Rapid Sepsityper® protocol and specific MBT-Sepsityper module (Bruker Daltonics) for the rapid diagnosis of bacteremia and fungemia by MALDI-TOF-MS. *Ann Clin Microbiol Antimicrob.* déc 2020;19(1):60.
57. van Belkum A, Welker M, Erhard M, Chatellier S. Biomedical Mass Spectrometry in Today's and Tomorrow's Clinical Microbiology Laboratories. *J Clin Microbiol.* avr 2012;50(5):1513-7.
58. Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: a Fundamental Shift in the Routine Practice of Clinical Microbiology. *Clin Microbiol Rev.* juill 2013;26(3):547-603.
59. Pavlovic M, Konrad R, Iwobi AN, Sing A, Busch U, Huber I. A dual approach employing MALDI-TOF MS and real-time PCR for fast species identification within the *Enterobacter cloacae* complex. *FEMS Microbiol Lett.* mars 2012;328(1):46-53.
60. Patel R. MALDI-TOF MS for the Diagnosis of Infectious Diseases. *Clin Chem.* janv 2015;61(1):100-11.
61. Souna D, Drissi M, Almahmoud I, Maurin M. *Enterobacter cloacae* Complex and CTX-M Extended-Spectrum β -Lactamases in Algeria. *Microb Drug Resist.* déc 2021;mdr.2020.0535.
62. Fang X yu, Li W bo, Zhang C fan, Huang Z da, Zeng H yi, Dong Z, et al. Detecting the presence of bacterial DNA and RNA by Polymerase Chain Reaction to diagnose suspected periprosthetic joint infection after antibiotic therapy. *Orthop Surg.* janv 2018;10(1):40-6.

63. Cirino NM, Tavakoli NP, Madison-Antenucci S, Egan C. Multiplex rtPCR in Microbiology. In: Real-Time PCR in Microbiology: From Diagnosis to Characterization. 2009.
64. Branger S, Casalta JP, Habib G, Collard F, Raoult D. Streptococcus pneumoniae Endocarditis: Persistence of DNA on Heart Valve Material 7 Years after Infectious Episode. J Clin Microbiol. sept 2003;41(9):4435-7.
65. Rovey C, Greub G, Lepidi H, Casalta JP, Habib G, Collart F, et al. PCR Detection of Bacteria on Cardiac Valves of Patients with Treated Bacterial Endocarditis. J Clin Microbiol. janv 2005;43(1):163-7.
66. Bauman JG. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. Exp Cell Res. août 1980;128(2):485-90.
67. Richter E, Schlüter C, Duchrow M, Hahn M, Rüsç-Gerdes S, Galle J, et al. An improved method for the species-specific assessment of mycobacteria in routinely formalin-fixed and paraffin-embedded tissues. J Pathol. janv 1995;175(1):85-92.
68. Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, et al. PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. Nature. oct 1993;365(6446):566-8.
69. Lefmann M, Schweickert B, Buchholz P, Gobel UB, Ulrichs T, Seiler P, et al. Evaluation of Peptide Nucleic Acid-Fluorescence In Situ Hybridization for Identification of Clinically Relevant Mycobacteria in Clinical Specimens and Tissue Sections. J Clin Microbiol. oct 2006;44(10):3760-7.
70. Gescher DM, Kovacevic D, Schmiedel D, Siemoneit S, Mallmann C, Halle E, et al. Fluorescence in situ hybridisation (FISH) accelerates identification of Gram-positive cocci in positive blood cultures. Int J Antimicrob Agents. nov 2008;32:S51-9.
71. Carretto E, Bardaro M, Russello G, Mirra M, Zuelli C, Barbarini D. Comparison of the Staphylococcus QuickFISH BC Test with the Tube Coagulase Test Performed on Positive Blood Cultures for Evaluation and Application in a Clinical Routine Setting. J Clin Microbiol. déc 2013;51(1):131-5.
72. Frickmann H, Alnamar Y, Essig A, Clos J, Racz P, Barth TF, et al. Rapid identification of Leishmania spp. in formalin-fixed, paraffin-embedded tissue samples by fluorescence in situ hybridization: Rapid identification of Leishmania spp. Trop Med Int Health. juill 2012;17(9):1117-26.
73. Schillinger C, Petrich A, Lux R, Riep B, Kikhney J, Friedmann A, et al. Co-Localized or Randomly Distributed? Pair Cross Correlation of In Vivo Grown Subgingival Biofilm Bacteria Quantified by Digital Image Analysis. Semple MG, éditeur. PLoS ONE. mai 2012;7(5):e37583.

74. Frickmann H, Hänle A, Essig A, Dekker D, Boahen K, Acquah S, et al. Fluorescence in situ hybridization (FISH) for rapid identification of *Salmonella* spp. from agar and blood culture broth - An option for the tropics? *Int J Med Microbiol.* juill 2013;303(5):277-84.
75. Rodriguez-Mateos P, Azevedo NF, Almeida C, Pamme N. FISH and chips: a review of microfluidic platforms for FISH analysis. *Med Microbiol Immunol (Berl).* janv 2020;209:373-91.
76. Frickmann H, Zautner AE, Moter A, Kikhney J, Hagen RM, Stender H, et al. Fluorescence in situ hybridization (FISH) in the microbiological diagnostic routine laboratory: a review. *Crit Rev Microbiol.* janv 2017;43(3):263-93.
77. Poppert S, Nickel D, Berger A, Yildiz T, Kaestner N, Mauerer S, et al. Rapid identification of beta-hemolytic streptococci by fluorescence in situ hybridization (FISH). *Int J Med Microbiol.* août 2009;299(6):421-6.
78. Drobniewski FA, More PG, Harris GS. Differentiation of *Mycobacterium tuberculosis* Complex and Nontuberculous Mycobacterial Liquid Cultures by Using Peptide Nucleic Acid-Fluorescence In Situ Hybridization Probes. *J Clin Microbiol.* janv 2000;38(1):444-7.
79. Charnot-Katsikas A, Tesic V, Love N, Hill B, Bethel C, Boonlayangoor S, et al. Use of the Accelerate Pheno System for Identification and Antimicrobial Susceptibility Testing of Pathogens in Positive Blood Cultures and Impact on Time to Results and Workflow. Bourbeau P, éditeur. *J Clin Microbiol.* déc 2017;56(1):e01166-17.
80. Lutgring JD, Bittencourt C, McElvania TeKippe E, Cavuoti D, Hollaway R, Burd EM. Evaluation of the Accelerate Pheno System: Results from Two Academic Medical Centers. Bourbeau P, éditeur. *J Clin Microbiol.* mars 2018;56(4):e01672-17.
81. Mackay IM. Real-Time PCR : History and Fluorogenic Chemistries. In: *Real-Time PCR in Microbiology : From Diagnosis to Characterization.* 2009.
82. Cockerill FR. Application of Rapid-Cycle Real-Time Polymerase Chain Reaction for Diagnostic Testing in the Clinical Microbiology Laboratory. *Arch Pathol Lab Med.* sept 2003;127(9):1112-20.
83. Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, et al. Real-Time PCR in Clinical Microbiology: Applications for Routine Laboratory Testing. *Clin Microbiol Rev.* janv 2006;19(1):165-256.
84. Lehmann LE, Hunfeld KP, Emrich T, Haberhausen G, Wissing H, Hoefl A, et al. A multiplex real-time PCR assay for rapid detection and differentiation of 25 bacterial and fungal pathogens from whole blood samples. *Med Microbiol Immunol (Berl).* sept 2008;197(3):313-24.
85. Korber F, Zeller I, Grünstäudl M, Willinger B, Apfalter P, Hirschl AM, et al. SeptiFast versus blood culture in clinical routine – A report on 3 years experience. *Wien Klin Wochenschr.* févr 2017;129(11-12):427-34.

86. Deboode F, Marien A, Janssen E, Bragard C, Berben G. The influence of amplicon length on real-time PCR results. *Biotechnol Agron Soc Env.* janv 2017;21(1):3-11.
87. Barken KB, Haagensen JAJ, Tolker-Nielsen T. Advances in nucleic acid-based diagnostics of bacterial infections. *Clin Chim Acta.* sept 2007;384(1-2):1-11.
88. Mothershed EA, Whitney AM. Nucleic acid-based methods for the detection of bacterial pathogens: Present and future considerations for the clinical laboratory. *Clin Chim Acta.* janv 2006;363(1-2):206-20.
89. Kemp M, Jensen KH, Dargis R, Christensen JJ. Routine ribosomal PCR and DNA sequencing for detection and identification of bacteria. *Future Microbiol.* juill 2010;5(7):1101-7.
90. van Pelt-Verkuil E, van Belkum A, Hays JP. Taq and other thermostable DNA polymerases. In: *Principles and Technical aspects of PCR amplification.* Springer Netherlands; 2008. p. 103-18.
91. van Pelt-Verkuil E, van Belkum A, Hays JP. PCR primers. In: *Principles and Technical aspects of PCR amplification.* Springer Netherlands; 2008. p. 63-90.
92. van Pelt-Verkuil E, van Belkum A, Hays JP. Analysis of PCR amplification products. In: *Principles and Technical aspects of PCR amplification.* Springer Netherlands; 2008. p. 141-82.
93. Claustres M, Kjellberg P. Detection of deletions by the amplification of exons (multiplex PCR) in Duchenne muscular dystrophy. *J Genet Hum.* sept 1989;37(3):251-7.
94. Mackay IM, Gardam T, Arden KE, McHardy S, Whiley DM, Crisante E, et al. Co-detection and discrimination of six human herpesviruses by multiplex PCR-ELAHA. *J Clin Virol.* déc 2003;28(3):291-302.
95. Melo AC, Almeida AMP, Leal NC. Retrospective study of a plague outbreak by multiplex-PCR. *Lett Appl Microbiol.* oct 2003;37(5):361-4.
96. Nolte FS. Molecular Microbiology. In: *Manual of Clinical Microbiology.* 11th Ed. 2015. p. 54-90.
97. Dunbar SA, Vander Zee CA, Oliver KG, Karem KL, Jacobson JW. Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP™ system. *J Microbiol Methods.* mai 2003;53(2):245-52.
98. Zink S, Cirino NM, Egan C. Multiplex PCR product: Detection and Discrimination. In: *Molecular microbiology: Diagnostic principles and practice.* 2nd Ed. ASM Press; 2011. p. 325-41.
99. Smith K. Automation of a fluorescence-based multiplex PCR for the laboratory confirmation of common bacterial pathogens. *J Med Microbiol.* févr 2004;53(2):115-7.

100. Templeton KE, Scheltinga SA, Beersma MFC, Kroes ACM, Claas ECJ. Rapid and Sensitive Method Using Multiplex Real-Time PCR for Diagnosis of Infections by Influenza A and Influenza B Viruses, Respiratory Syncytial Virus, and Parainfluenza Viruses 1, 2, 3, and 4. *J Clin Microbiol.* avr 2004;42(4):1564-9.
101. Poritz MA, Blaschke AJ, Byington CL, Meyers L, Nilsson K, Jones DE, et al. FilmArray, an Automated Nested Multiplex PCR System for Multi-Pathogen Detection: Development and Application to Respiratory Tract Infection. Costa C, éditeur. *PLoS ONE.* oct 2011;6(10):e26047.
102. de Boer RF, Ott A, Kesztyus B, Kooistra-Smid AMD. Improved Detection of Five Major Gastrointestinal Pathogens by Use of a Molecular Screening Approach. *J Clin Microbiol.* nov 2010;48(11):4140-6.
103. Coste JF, Vuiblet V, Moustapha B, Bouin A, Lavaud S, Toupance O, et al. Microbiological Diagnosis of Severe Diarrhea in Kidney Transplant Recipients by Use of Multiplex PCR Assays. *J Clin Microbiol.* mai 2013;51(6):1841-9.
104. Westh H, Lisby G, Breysse F, Böddinghaus B, Chomarat M, Gant V, et al. Multiplex real-time PCR and blood culture for identification of bloodstream pathogens in patients with suspected sepsis. *Clin Microbiol Infect.* juin 2009;15(6):544-51.
105. Blaschke AJ, Heyrend C, Byington CL, Fisher MA, Barker E, Garrone NF, et al. Rapid identification of pathogens from positive blood cultures by multiplex polymerase chain reaction using the FilmArray system. *Diagn Microbiol Infect Dis.* déc 2012;74(4):349-55.
106. Wojewoda CM, Sercia L, Navas M, Tuohy M, Wilson D, Hall GS, et al. Evaluation of the Verigene Gram-Positive Blood Culture Nucleic Acid Test for Rapid Detection of Bacteria and Resistance Determinants. *J Clin Microbiol.* juin 2013;51(7):2072-6.
107. Halse TA, Musser KA, Limberger RJ. A multiplexed real-time PCR assay for rapid detection of *Chlamydia trachomatis* and identification of serovar L-2, the major cause of Lymphogranuloma venereum in New York. *Mol Cell Probes.* oct 2006;20(5):290-7.
108. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Fox AJ, Kaczmarek EB. Simultaneous Detection of *Neisseria meningitidis*, *Haemophilus influenzae*, and *Streptococcus pneumoniae* in Suspected Cases of Meningitis and Septicemia Using Real-Time PCR. *J Clin Microbiol.* avr 2001;39(4):1553-8.
109. Tzanakaki G, Tsopanomichalou M, Kesanopoulos K, Matzourani R, Sioumala M, Tabaki A, et al. Simultaneous single-tube PCR assay for the detection of *Neisseria meningitidis*, *Haemophilus influenzae* type b and *Streptococcus pneumoniae*. *Clin Microbiol Infect.* mai 2005;11(5):386-90.
110. Banada PP, Deshpande S, Banik S, Shah D, Koshy R, Patel B, et al. Multiplex Detection of Three Select Agents Directly from Blood by Use of the GeneXpert System. Carroll KC, éditeur. *J Clin Microbiol.* avr 2019;57(5):e00036-19.

111. Ramanan P, Bryson AL, Binnicker MJ, Pritt BS, Patel R. Syndromic Panel-Based Testing in Clinical Microbiology. *Clin Microbiol Rev.* nov 2017;31(1):e00024-17.
112. Patel J. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mol Diagn.* déc 2001;6(4):313-21.
113. Petti CA. Detection and Identification of Microorganisms by Gene Amplification and Sequencing. *Clin Infect Dis.* avr 2007;44(8):1108-14.
114. Maiwald M. Broad-range PCR for Detection and Identification of Bacteria. In: *Molecular Microbiology: Diagnostic Principles and Practices*. 2nd Ed. 2011. p. 491-505.
115. Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect.* oct 2008;14(10):908-34.
116. Woese C. The universal ancestor. *Proc Natl Acad Sci.* juin 1998;95(12):6854-9.
117. Carroll K, Patel R. Systems for Identification of Bacteria and Fungi. In: *Manual of Clinical Microbiology*. 11^e éd. ASM Press; 2015. p. 29-43.
118. Mollet C, Drancourt M, Raoult D. rpoB sequence analysis as a novel basis for bacterial identification. *Mol Microbiol.* déc 1997;26(5):1005-11.
119. Renvoisé A, Brossier F, Sougakoff W, Jarlier V, Aubry A. Broad-range PCR: Past, present, or future of bacteriology? *Médecine Mal Infect.* août 2013;43(8):322-30.
120. Rantakokko-Jalava K, Nikkari S, Jalava J, Eerola E, Skurnik M, Meurman O, et al. Direct Amplification of rRNA Genes in Diagnosis of Bacterial Infections. *J Clin Microbiol.* janv 2000;38(1):32-9.
121. Niemz A, Ferguson TM, Boyle DS. Point-of-care nucleic acid testing for infectious diseases. *Trends Biotechnol.* mai 2011;29(5):240-50.
122. Schuurman T, de Boer RF, Kooistra-Smid AMD, van Zwet AA. Prospective Study of Use of PCR Amplification and Sequencing of 16S Ribosomal DNA from Cerebrospinal Fluid for Diagnosis of Bacterial Meningitis in a Clinical Setting. *J Clin Microbiol.* févr 2004;42(2):734-40.
123. Heining A, Binder M, Ellinger A, Botzenhart K, Unertl K, Doring G. DNase Pretreatment of Master Mix Reagents Improves the Validity of Universal 16S rRNA Gene PCR Results. *J Clin Microbiol.* avr 2003;41(4):1763-5.
124. Hilali F, Saulnier P, Chachaty E, Andreumont A. Decontamination of polymerase chain reaction reagents for detection of low concentrations of 16S rRNA genes. *Mol Biotechnol.* juin 1997;7(3):207-16.

125. Wilson KH, Blitchington RB. Human colonic biota studied by ribosomal DNA sequence analysis. *Appl Environ Microbiol.* juill 1996;62(7):2273-8.
126. Choi BK, Paster BJ, Dewhirst FE, Göbel UB. Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. *Infect Immun.* mai 1994;62(5):1889-95.
127. Tanner MA, Shoskes D, Shahed A, Pace NR. Prevalence of Corynebacterial 16S rRNA Sequences in Patients with Bacterial and “Nonbacterial” Prostatitis. *J Clin Microbiol.* juin 1999;37(6):1863-70.
128. Rolph HJ, Lennon A, Riggio MP, Saunders WP, MacKenzie D, Coldero L, et al. Molecular Identification of Microorganisms from Endodontic Infections. *J Clin Microbiol.* sept 2001;39(9):3282-9.
129. Schabereiter-Gurtner C, Maca S, Kaminsky S, Rölleke S, Lubitz W, Barisani-Asenbauer T. Investigation of an anaerobic microbial community associated with a corneal ulcer by denaturing gradient gel electrophoresis and 16S rDNA sequence analysis. *Diagn Microbiol Infect Dis.* juill 2002;43(3):193-9.
130. Trees E, Rota PA, MacCannell D, Gerner-Smidt P. Molecular Epidemiology. In: *Manual of Clinical Microbiology.* 11th Ed. 2015. p. 131-61.
131. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. *Pathology (Phila).* avr 2015;47(3):199-210.
132. Levy SE, Myers RM. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet.* juin 2016;17(1):95-115.
133. Bertelli C, Greub G. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. *Clin Microbiol Infect.* sept 2013;19(9):803-13.
134. Oniciuc E, Likotrafiti E, Alvarez-Molina A, Prieto M, Santos J, Alvarez-Ordóñez A. The Present and Future of Whole Genome Sequencing (WGS) and Whole Metagenome Sequencing (WMS) for Surveillance of Antimicrobial Resistant Microorganisms and Antimicrobial Resistance Genes across the Food Chain. *Genes.* mai 2018;9(5):268.
135. Yarbrough ML, Lainhart W, McMullen AR, Anderson NW, Burnham CAD. Impact of total laboratory automation on workflow and specimen processing time for culture of urine specimens. *Eur J Clin Microbiol Infect Dis.* sept 2018;37(12):2405-11.
136. Burckhardt I. Laboratory Automation in Clinical Microbiology. *Bioengineering.* nov 2018;5(4):102.
137. Phillips LE, Verma S, Surapaneni BK, Dutta SK. Potential clinical application of an automated fluorescent microbial cell counter in the detection of urinary tract infection. *J Clin Lab Anal [Internet].* avr 2020 [cité 17 sept 2020];34(8). Disponible sur: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcla.23334>

138. Qi H, Wang Y, Wu P, Ma Y, Wang J. Rapid and fully-automated detection of *Clostridium difficile* Toxin B via magnetic-particle-based chemiluminescent immunoassay. *Am J Transl Res.* août 2020;12(8):4228-36.
139. Bourbeau PP, Ledebor NA. Automation in Clinical Microbiology. *J Clin Microbiol.* mai 2013;51(6):1658-65.
140. Bailey AL, Ledebor N, Burnham CAD. Clinical Microbiology Is Growing Up: The Total Laboratory Automation Revolution. *Clin Chem.* mai 2019;65(5):634-43.
141. Snyder JW, Thomson GK, Heckman S, Jamros K, AbdelGhani S, Thomson KS. Automated preparation for identification and antimicrobial susceptibility testing: evaluation of a research use only prototype, the BD Kiestra IdentifA/SusceptA system. *Clin Microbiol Infect.* mai 2021;27(5):783.e1-783.e5.
142. Cherkaoui A, Renzi G, Martischang R, Harbarth S, Vuilleumier N, Schrenzel J. Impact of Total Laboratory Automation on Turnaround Times for Urine Cultures and Screening Specimens for MRSA, ESBL, and VRE Carriage: Retrospective Comparison With Manual Workflow. *Front Cell Infect Microbiol.* oct 2020;10:552122.
143. Nasser B, Soleimani N, Rabiee N, Kalbasi A, Karimi M, Hamblin MR. Point-of-care microfluidic devices for pathogen detection. *Biosens Bioelectron.* oct 2018;117:112-28.
144. Müller V, Sousa JM, Ceylan Koydemir H, Veli M, Tseng D, Cerqueira L, et al. Identification of pathogenic bacteria in complex samples using a smartphone based fluorescence microscope. *RSC Adv.* oct 2018;8(64):36493-502.
145. Kim S, Lee MH, Wiwasuku T, Day AS, Youngme S, Hwang DS, et al. Human sensor-inspired supervised machine learning of smartphone-based paper microfluidic analysis for bacterial species classification. *Biosens Bioelectron.* sept 2021;188:113335.
146. Kohling HL, Bittner A, Muller KD, Buer J, Becker M, Rubben H, et al. Direct identification of bacteria in urine samples by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and relevance of defensins as interfering factors. *J Med Microbiol.* mars 2012;61(3):339-44.
147. Wang XH, Zhang G, Fan YY, Yang X, Sui WJ, Lu XX. Direct identification of bacteria causing urinary tract infections by combining matrix-assisted laser desorption ionization-time of flight mass spectrometry with UF-1000i urine flow cytometry. *J Microbiol Methods.* mars 2013;92(3):231-5.
148. Bryson AL, Hill EM, Doern CD. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight: the Revolution in Progress. *Clin Lab Med.* sept 2019;39(3):391-404.
149. Klein S, Zimmermann S, Kohler C, Mischnik A, Alle W, Bode KA. Integration of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in blood culture diagnostics: a fast and effective approach. *J Med Microbiol.* mars 2012;61(3):323-31.

150. Drancourt M. Detection of microorganisms in blood specimens using matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a review. *Clin Microbiol Infect.* nov 2010;16(11):1620-5.
151. Shannon S, Kronemann D, Patel R, Schuetz AN. Routine use of MALDI-TOF MS for anaerobic bacterial identification in clinical microbiology. *Anaerobe.* déc 2018;54:191-6.
152. Dubourg G, Lamy B, Ruimy R. Rapid phenotypic methods to improve the diagnosis of bacterial bloodstream infections: meeting the challenge to reduce the time to result. *Clin Microbiol Infect.* sept 2018;24(9):935-43.
153. Yonetani S, Ohnishi H, Ohkusu K, Matsumoto T, Watanabe T. Direct identification of microorganisms from positive blood cultures by MALDI-TOF MS using an in-house saponin method. *Int J Infect Dis.* nov 2016;52:37-42.
154. Yuan Y, Wang J, Zhang J, Ma B, Gao S, Li Y, et al. Evaluation of an optimized method to directly identify bacteria from positive blood cultures using MALDI-TOF mass spectrometry. *J Clin Lab Anal.* avr 2020;34(4):e23119.
155. Nonnemann B, Tvede M, Bjarnsholt T. Identification of pathogenic microorganisms directly from positive blood vials by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *APMIS.* janv 2013;121(9):871-7.
156. Fiori B, D'Inzeo T, Giaquinto A, Menchinelli G, Liotti FM, de Maio F, et al. Optimized Use of the MALDI BioTyper System and the FilmArray BCID Panel for Direct Identification of Microbial Pathogens from Positive Blood Cultures. Richter SS, éditeur. *J Clin Microbiol.* févr 2016;54(3):576-84.
157. Scohy A, Noël A, Boeras A, Brassinne L, Laurent T, Rodriguez-Villalobos H, et al. Evaluation of the Bruker® MBT Sepsityper IVD module for the identification of polymicrobial blood cultures with MALDI-TOF MS. *Eur J Clin Microbiol Infect Dis.* août 2018;37(11):2145-52.
158. Pinault L, Chabrière E, Raoult D, Fenollar F. Direct Identification of Pathogens in Urine by Use of a Specific Matrix-Assisted Laser Desorption Ionization–Time of Flight Spectrum Database. Carroll KC, éditeur. *J Clin Microbiol.* mars 2019;57(4):e01678-18.
159. Hamid AM, Jarmusch AK, Pirro V, Pincus DH, Clay BG, Gervasi G, et al. Rapid Discrimination of Bacteria by Paper Spray Mass Spectrometry. *Anal Chem.* juill 2014;86(15):7500-7.
160. Lellman SE, Cramer R. Bacterial identification by lipid profiling using liquid atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Clin Chem Lab Med CCLM.* déc 2019;58(6):930-8.
161. Dortet L, Potron A, Bonnin RA, Plesiat P, Naas T, Filloux A, et al. Rapid detection of colistin resistance in *Acinetobacter baumannii* using MALDI-TOF-based lipidomics on intact bacteria. *Sci Rep.* nov 2018;8(1):16910.

162. Furniss RCD, Dortet L, Bolland W, Drews O, Sparbier K, Bonnin RA, et al. Detection of Colistin Resistance in *Escherichia coli* by Use of the MALDI Biotyper Sirius Mass Spectrometry System. Ledebøer NA, éditeur. *J Clin Microbiol.* nov 2019;57(12):e01427-19.
163. Sykes PJ, Neoh SH, Brisco MJ, Condon J, Morley AA. Quantitation of targets for PCR by use of limiting dilution. *BioTechniques.* sept 1992;13(3):444-9.
164. Vogelstein B, Kinzler KW. Digital PCR. *Proc Natl Acad Sci.* août 1999;96(16):9236-41.
165. Sreejith KR, Ooi CH, Jin J, Dao DV, Nguyen NT. Digital polymerase chain reaction technology – recent advances and future perspectives. *Lab Chip.* oct 2018;18(24):3717-32.
166. Salipante SJ, Jerome KR. Digital PCR—An Emerging Technology with Broad Applications in Microbiology. *Clin Chem.* janv 2020;66(1):117-23.
167. Huggett JF, Foy CA, Benes V, Emslie K, Garson JA, Haynes R, et al. The Digital MIQE Guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. *Clin Chem.* juin 2013;59(6):892-902.
168. Dingle TC, Sedlak RH, Cook L, Jerome KR. Tolerance of Droplet-Digital PCR vs Real-Time Quantitative PCR to Inhibitory Substances. *Clin Chem.* nov 2013;59(11):1670-2.
169. Whale AS, Huggett JF, Tzonev S. Fundamentals of multiplexing with digital PCR. *Biomol Detect Quantif.* déc 2016;10:15-23.
170. Zhong Q, Bhattacharya S, Kotsopoulos S, Olson J, Taly V, Griffiths AD, et al. Multiplex digital PCR: breaking the one target per color barrier of quantitative PCR. *Lab Chip.* mai 2011;11(13):2167.
171. Kuypers J, Jerome KR. Applications of Digital PCR for Clinical Microbiology. Kraft CS, éditeur. *J Clin Microbiol.* mai 2017;55(6):1621-8.
172. Quan PL, Sauzade M, Brouzes E. dPCR: A Technology Review. *Sensors.* avr 2018;18(4):1271.
173. Roberts C h., Last A, Molina-Gonzalez S, Cassama E, Butcher R, Nabicassa M, et al. Development and Evaluation of a Next-Generation Digital PCR Diagnostic Assay for Ocular *Chlamydia trachomatis* Infections. *J Clin Microbiol.* juin 2013;51(7):2195-203.
174. Kelley K, Cosman A, Belgrader P, Chapman B, Sullivan DC. Detection of Methicillin-Resistant *Staphylococcus aureus* by a Duplex Droplet Digital PCR Assay. *J Clin Microbiol.* juin 2013;51(7):2033-9.
175. Talarico S, Safaeian M, Gonzalez P, Hildesheim A, Herrero R, Porras C, et al. Quantitative Detection and Genotyping of *Helicobacter pylori* from Stool using Droplet Digital PCR Reveals Variation in Bacterial Loads that Correlates with *cagA* Virulence Gene Carriage. *Helicobacter.* août 2016;21(4):325-33.

176. Devonshire AS, Honeyborne I, Gutteridge A, Whale AS, Nixon G, Wilson P, et al. Highly Reproducible Absolute Quantification of Mycobacterium tuberculosis Complex by Digital PCR. *Anal Chem.* févr 2015;87(7):3706-13.
177. Waterer G, Rello J. Why should we measure bacterial load when treating community-acquired pneumonia?: *Curr Opin Infect Dis.* avr 2011;24(2):137-41.
178. Piepenburg O, Williams CH, Stemple DL, Armes NA. DNA Detection Using Recombination Proteins. *Haber J, éditeur. PLoS Biol.* juin 2006;4(7):e204.
179. Vincent M, Xu Y, Kong H. Helicase-dependent isothermal DNA amplification. *EMBO Rep.* août 2004;5(8):795-800.
180. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet.* juill 1998;19(3):225-32.
181. Compton J. Nucleic acid sequence-based amplification. *Nature.* mars 1991;350:91-2.
182. Notomi T, Mori Y, Tomita N, Kanda H. Loop-mediated isothermal amplification (LAMP): principle, features, and future prospects. *J Microbiol.* janv 2015;53(1):1-5.
183. Zhao Y, Chen F, Li Q, Wang L, Fan C. Isothermal Amplification of Nucleic Acids. *Chem Rev.* nov 2015;115(22):12491-545.
184. Chen HW, Weissenberger G, Atkins E, Chao CC, Suputtamongkol Y, Ching WM. Highly Sensitive Loop-Mediated Isothermal Amplification for the Detection of *Leptospira*. *Int J Bacteriol.* janv 2015;2015:147173.
185. Seyrig G, Ahmad F, Stedtfeld RD, Turlousse DM, Hashsham SA. Simple, powerful, and smart: Using LAMP for low-cost screening of multiple waterborne pathogens. *Environ Microbiol Curr Technol Water Appl.* janv 2011;(4):103-25.
186. Noren T, Alriksson I, Andersson J, Akerlund T, Unemo M. Rapid and Sensitive Loop-Mediated Isothermal Amplification Test for *Clostridium difficile* Detection Challenges Cytotoxin B Cell Test and Culture as Gold Standard. *J Clin Microbiol.* févr 2011;49(2):710-1.
187. Kaewphinit T, Ckumdee J, Chansiri K, Santiwatanakul S. Development and Evaluation of a Loop-mediated Isothermal Amplification Combined with Au-nanoprobe Assay for Rapid Detection of *Mycobacterium tuberculosis*. *Indian J Med Microbiol.* avr 2017;35(2):302-4.
188. Kaur A, Ruhela A, Sharma P, Khariwal H, Seth S, Kumar A, et al. Simultaneous and high sensitive detection of *Salmonella typhi* and *Salmonella paratyphi a* in human clinical blood samples using an affordable and portable device. *Biomed Microdevices.* nov 2019;21(4):95.

189. Kamachi K, Moriuchi T, Hiramatsu Y, Otsuka N, Shibayama K. Evaluation of a commercial loop-mediated isothermal amplification assay for diagnosis of *Bordetella pertussis* infection. *J Microbiol Methods*. févr 2017;133:20-2.
190. Vergara A, Boutal H, Ceccato A, López M, Cruells A, Bueno-Freire L, et al. Assessment of a Loop-Mediated Isothermal Amplification (LAMP) Assay for the Rapid Detection of Pathogenic Bacteria from Respiratory Samples in Patients with Hospital-Acquired Pneumonia. *Microorganisms*. janv 2020;8(1):103.
191. Wang Q, Zhou Y, Li S, Zhuo C, Xu S, Huang L, et al. Real-Time Fluorescence Loop Mediated Isothermal Amplification for the Detection of *Acinetobacter baumannii*. Xu S yong, éditeur. *PLoS ONE*. juill 2013;8(7):e66406.
192. Kaneko H, Kawana T, Fukushima E, Suzutani T. Tolerance of loop-mediated isothermal amplification to a culture medium and biological substances. *J Biochem Biophys Methods*. avr 2007;70(3):499-501.
193. Wong YP, Othman S, Lau YL, Radu S, Chee HY. Loop-mediated isothermal amplification (LAMP): a versatile technique for detection of micro-organisms. *J Appl Microbiol*. mars 2018;124(3):626-43.
194. Dhama K, Karthik K, Chakraborty S, Tiwari R, Kapoor S, Kumar A, et al. Loop-mediated Isothermal Amplification of DNA (LAMP): A New Diagnostic Tool Lights the World of Diagnosis of Animal and Human Pathogens: A review. *Pak J Biol Sci*. janv 2014;17(2):151-66.
195. Liu N, Zou D, Dong D, Yang Z, Ao D, Liu W, et al. Development of a multiplex loop-mediated isothermal amplification method for the simultaneous detection of *Salmonella* spp. and *Vibrio parahaemolyticus*. *Sci Rep*. mars 2017;7(1):45601.
196. Watts MR, James G, Sultana Y, Ginn AN, Outhred AC, Kong F, et al. A Loop-Mediated Isothermal Amplification (LAMP) Assay for *Strongyloides stercoralis* in Stool That Uses a Visual Detection Method with SYTO-82 Fluorescent Dye. *Am J Trop Med Hyg*. févr 2014;90(2):306-11.
197. Fischbach J, Xander NC, Frohme M, Glökler JF. Shining a light on LAMP assays— A comparison of LAMP visualization methods including the novel use of berberine. *BioTechniques* [Internet]. avr 2015 [cité 5 févr 2020];58(4). Disponible sur: <https://www.future-science.com/doi/10.2144/000114275>
198. Sahoo PR, Sethy K, Mohapatra S, Panda D. Loop mediated isothermal amplification: An innovative gene amplification technique for animal diseases. *Vet World*. mai 2016;9(5):465-9.
199. Tomita N, Mori Y, Kanda H, Notomi T. Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nat Protoc*. mai 2008;3(5):877-82.

200. Korpi A, Järnberg J, Pasanen AL. Microbial Volatile Organic Compounds. *Crit Rev Toxicol.* févr 2009;39(2):139-93.
201. Mgode GF, Weetjens BJ, Nawrath T, Cox C, Jubitana M, Machang'u RS, et al. Diagnosis of Tuberculosis by Trained African Giant Pouched Rats and Confounding Impact of Pathogens and Microflora of the Respiratory Tract. *J Clin Microbiol.* janv 2012;50(2):274-80.
202. Bruins M, Bos A, Petit PLC, Eadie K, Rog A, Bos R, et al. Device-independent, real-time identification of bacterial pathogens with a metal oxide-based olfactory sensor. *Eur J Clin Microbiol Infect Dis.* févr 2009;28(7):775-80.
203. O'Sullivan DM, Nicoara SC, Mutetwa R, Mungofa S, Lee OYC, Minnikin DE, et al. Detection of *Mycobacterium tuberculosis* in Sputum by Gas Chromatography-Mass Spectrometry of Methyl Mycocerosates Released by Thermochemolysis. *Nigou J, éditeur. PLoS ONE.* mars 2012;7(3):e32836.
204. Karami N, Karimi A, Aliahmadi A, Mirzajan F, Rezadoost H, Ghassempour A, et al. Identification of bacteria using volatile organic compounds. *Cell Mol Biol.* févr 2017;63(2):112-21.
205. Krafft C, Popp J. The many facets of Raman spectroscopy for biomedical analysis. *Anal Bioanal Chem.* janv 2015;407(3):699-717.
206. Butler HJ, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, et al. Using Raman spectroscopy to characterize biological materials. *Nat Protoc.* avr 2016;11(4):664-87.
207. Le Ru E, Etchegoin P. A quick overview of surface-enhanced Raman Spectroscopy. In: *Principles of Surface-Enhanced Raman Spectroscopy.* 1st Edition. Oxford: Elsevier; 2009. p. 27.
208. Kloß S, Kampe B, Sachse S, Rösch P, Straube E, Pfister W, et al. Culture Independent Raman Spectroscopic Identification of Urinary Tract Infection Pathogens: A Proof of Principle Study. *Anal Chem.* oct 2013;85(20):9610-6.
209. Pahlow S, Weber K, Popp J, Wood BR, Kochan K, Rüther A, et al. Application of Vibrational Spectroscopy and Imaging to Point-of-Care Medicine: A Review. *Appl Spectrosc.* sept 2018;72(1 (suppl)):52-84.
210. Pahlow S, Meisel S, Cialla-May D, Weber K, Rösch P, Popp J. Isolation and identification of bacteria by means of Raman spectroscopy. *Adv Drug Deliv Rev.* juill 2015;89:105-20.
211. Walter A, Kuhri S, Reinicke M, Bocklitz T, Schumacher W, Rösch P, et al. Raman spectroscopic detection of Nickel impact on single *Streptomyces* cells - possible bioindicators for heavy metal contamination: Raman spectroscopic detection of Nickel impact on single *Streptomyces* cells. *J Raman Spectrosc.* août 2012;43(8):1058-64.

212. Boardman AK, Wong WS, Premasiri WR, Ziegler LD, Lee JC, Miljkovic M, et al. Rapid Detection of Bacteria from Blood with Surface-Enhanced Raman Spectroscopy. *Anal Chem.* juill 2016;88(16):8026-35.
213. Wu X, Chen J, Li X, Zhao Y, Zughair SM. Culture-free diagnostics of *Pseudomonas aeruginosa* infection by silver nanorod array based SERS from clinical sputum samples. *Nanomedicine Nanotechnol Biol Med.* nov 2014;10(8):1863-70.
214. Ghebremedhin M, Heitkamp R, Yesupriya S, Clay B, Crane NJ. Accurate and Rapid Differentiation of *Acinetobacter baumannii* Strains by Raman Spectroscopy: a Comparative Study. Munson E, éditeur. *J Clin Microbiol.* août 2017;55(8):2480-90.
215. Zarnowiec P, Lechowicz L, Czerwonka G, Kaca W. Fourier Transform Infrared Spectroscopy (FTIR) as a Tool for the Identification and Differentiation of Pathogenic Bacteria. *Curr Med Chem.* avr 2015;22(14):1710-8.
216. Novais Â, Freitas AR, Rodrigues C, Peixe L. Fourier transform infrared spectroscopy: unlocking fundamentals and prospects for bacterial strain typing. *Eur J Clin Microbiol Infect Dis.* mars 2019;38(3):427-48.
217. Hong JS, Kim D, Jeong SH. Performance Evaluation of the IR Biotyper® System for Clinical Microbiology: Application for Detection of *Staphylococcus aureus* Sequence Type 8 Strains. *Antibiotics.* juill 2022;11(7):909.
218. Passaris I, Mauder N, Kostrzewa M, Burckhardt I, Zimmermann S. Validation of Fourier Transform Infrared Spectroscopy for Serotyping of *Streptococcus pneumoniae*. *J Clin Microbiol.* juill 2022;60(7):16.
219. Clancy CJ, Nguyen MH. T2 magnetic resonance for the diagnosis of bloodstream infections: charting a path forward. *J Antimicrob Chemother.* mars 2018;73(Suppl 4):iv2-5.
220. De Angelis G, Posteraro B, De Carolis E, Menchinelli G, Franceschi F, Tumbarello M, et al. T2 Bacteria magnetic resonance assay for the rapid detection of ESKAPEc pathogens directly in whole blood. *J Antimicrob Chemother.* mars 2018;73(suppl_4):iv20-6.
221. Edgar RH, Cook J, Noel C, Minard A, Sajewski A, Fitzpatrick M, et al. Bacteriophage-mediated identification of bacteria using photoacoustic flow cytometry. *J Biomed Opt.* nov 2019;24(11):1.
222. Dudak FC, Boyacı İH. Rapid and label-free bacteria detection by surface plasmon resonance (SPR) biosensors. *Biotechnol J.* juill 2009;4(7):1003-11.
223. Hong KL, Sooter LJ. Single-Stranded DNA Aptamers against Pathogens and Toxins: Identification and Biosensing Applications. *BioMed Res Int.* juin 2015;2015:1-31.
224. Shin HS, Gedi V, Kim JK, Lee D ki. Detection of Gram-negative bacterial outer membrane vesicles using DNA aptamers. *Sci Rep.* sept 2019;9(1):13167.

225. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* 1989;1:541-51.
226. LeCun Y, Bottou L, Bengio Y, Ha P. Gradient-Based Learning Applied to Document Recognition. *Proc IEEE.* 1998;46.
227. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* 2012;25(6):84-90.
228. Dolley S. Big Data's Role in Precision Public Health. *Front Public Health.* mars 2018;6:68.
229. Williamson J. Using Machine Learning Algorithms to Predict Risk for Development of Calciphylaxis in Patients with Chronic Kidney Disease. *AMIA Jt Summits Transl Sci Proc.* mai 2018;8.
230. Kanjilal S, Oberst M, Boominathan S, Zhou H, Hooper DC, Sontag D. A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection. *Sci Transl Med.* nov 2020;12(568):eaay5067.
231. Doern CD, Miller MB, Alby K, Bachman MA, Brecher SM, Casiano-Colon A, et al. Proceedings of the Clinical Microbiology Open 2018 and 2019 - a Discussion about Emerging Trends, Challenges, and the Future of Clinical Microbiology. Humphries RM, éditeur. *J Clin Microbiol.* juill 2022;60(7):e00092-22.
232. Brenton L, Waters MJ, Stanford T, Giglio S. Clinical evaluation of the APAS® Independence: Automated imaging and interpretation of urine cultures using artificial intelligence with composite reference standard discrepant resolution. *J Microbiol Methods.* oct 2020;177:106047.
233. Gammel N, Ross TL, Lewis S, Olson M, Henciak S, Harris R, et al. Comparison of an Automated Plate Assessment System (APAS Independence) and Artificial Intelligence (AI) to Manual Plate Reading of Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus aureus* CHROMagar Surveillance Cultures. Richter SS, éditeur. *J Clin Microbiol.* oct 2021;59(11):e00971-21.
234. Kuang X, Wang F, Hernandez KM, Zhang Z, Grossman RL. Accurate and rapid prediction of tuberculosis drug resistance from genome sequence data using traditional machine learning algorithms and CNN. *Sci Rep.* déc 2022;12(1):2427.
235. Weis C, Cuénod A, Rieck B, Dubuis O, Graf S, Lang C, et al. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. *Nat Med.* janv 2022;28(1):164-74.
236. Smith KP, Kang AD, Kirby JE. Automated Interpretation of Blood Culture Gram Stains by Use of a Deep Convolutional Neural Network. Bourbeau P, éditeur. *J Clin Microbiol.* mars 2018;56(3):e01521-17.

237. Wyatt PJ. Differential Light Scattering: a Physical Method for Identifying Living Bacterial Cells. *Appl Opt.* oct 1968;7(10):1879.
238. Bae E, Banada PP, Huff K, Bhunia AK, Robinson JP, Hirleman ED. Biophysical modeling of forward scattering from bacterial colonies using scalar diffraction theory. *Appl Opt.* juin 2007;46(17):3639.
239. Guo S. Optical scattering for bacterial colony detection and characterization. [Lafayette, Indiana, US]: Purdue University; 2004.
240. Bae E, Aroonual A, Bhunia AK, Hirleman ED. On the sensitivity of forward scattering patterns from bacterial colonies to media composition. *J Biophotonics.* avr 2011;4(4):236-43.
241. Suchwalko A, Buzalewicz I, Podbielska H. Computer-based classification of bacteria species by analysis of their colonies Fresnel diffraction patterns. In: Miller BL, Fauchet PM, éditeurs. San Francisco, California, USA; 2012 [cité 2 oct 2020]. p. 82120R. Disponible sur: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.907420>
242. Bae E, Ying D, Kramer D, Patsekina V, Rajwa B, Holdman C, et al. Portable bacterial identification system based on elastic light scatter patterns. *J Biol Eng.* août 2012;6(1):1-11.
243. Bae E, Bai N, Aroonual A, Robinson JP, Bhunia AK, Hirleman ED. Modeling light propagation through bacterial colonies and its correlation with forward scattering patterns. *J Biomed Opt.* juill 2010;15(4):045001.
244. Bae E, Patsekina V, Rajwa B, Bhunia AK, Holdman C, Davisson VJ, et al. Development of a microbial high-throughput screening instrument based on elastic light scatter patterns. *Rev Sci Instrum.* avr 2012;83(4):044304.
245. Buzalewicz I, Wieliczko A, Podbielska H. Influence of various growth conditions on Fresnel diffraction patterns of bacteria colonies examined in the optical system with converging spherical wave illumination. *Opt Express.* oct 2011;19(22):21768.
246. Suchwalko A, Buzalewicz I, Podbielska H. Bacteria identification in an optical system with optimized diffraction pattern registration condition supported by enhanced statistical analysis. *Opt Express.* oct 2014;22(21):26312-27.
247. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* juill 2012;9(7):671-5.
248. Girlich D, Ouzani S, Emeraud C, Gauthier L, Bonnin RA, Le Sache N, et al. Uncovering the novel *Enterobacter cloacae* complex species responsible for septic shock deaths in newborns: a cohort study. *Lancet Microbe.* oct 2021;2(10):e536-44.

249. Wu W, Feng Y, Zong Z. Precise Species Identification for Enterobacter : a Genome Sequence-Based Study with Reporting of Two Novel Species, *Enterobacter quasiroggenkampii* sp. nov. and *Enterobacter quasimori* sp. nov. Lal R, éditeur. *mSystems*. août 2020;5(4):e00527-20.
250. Ge MC, Kuo AJ, Liu KL, Wen YH, Chia JH, Chang PY, et al. Routine identification of microorganisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Success rate, economic analysis, and clinical outcome. *J Microbiol Immunol Infect*. oct 2017;50(5):662-8.
251. Wimpenny JWT. The Bacterial Colony. In: *CRC Handbbok of the Laboratory model systems for microbial ecosystems*. CRC Press. 1988. p. 109-36.
252. Pipe LZ, Grimson MJ. Spatial-temporal modelling of bacterial colony growth on solid media. *Mol Biosyst*. janv 2008;4(3):192.
253. Shapiro JA. The significances of bacterial colony patterns. *BioEssays*. juill 1995;17(7):597-607.
254. Bittremieux W, Tabb DL, Impens F, Staes A, Timmerman E, Martens L, et al. Quality control in mass spectrometry-based proteomics. *Mass Spectrom Rev*. sept 2018;37(5):697-711.
255. Lucena Baeza L, Pfennigwerth N, Hamprecht A. Rapid and Easy Detection of Carbapenemases in Enterobacterales in the Routine Laboratory Using the New GenePOC Carba/Revogene Carba C Assay. Carroll KC, éditeur. *J Clin Microbiol*. sept 2019;57(9):e00597-19.
256. Lawrence C, Lucet JC. Prélèvements de dépistage-portage, recherche de bactéries multirésistantes. In: *Référentiel de Microbiologie Clinique*. 2018. p. 5.
257. Zhu Y, Huang WE, Yang Q. Clinical Perspective of Antimicrobial Resistance in Bacteria. *Infect Drug Resist*. mars 2022;Volume 15:735-46.
258. European Center for Disease Prevention and Control (ECDC). Antimicrobial resistance surveillance in Europe 2022. *Annu Rep Eur Antimicrob Resist Surveill Netw*. mars 2022;
259. World Health Organization. Global guidelines for the prevention of surgical site infection. 2^e éd. 2018. 184 p.
260. Matsumura Y, Pitout JD. Recent advances in the laboratory detection of carbapenemase-producing Enterobacteriaceae. *Expert Rev Mol Diagn*. juill 2016;16(7):783-94.
261. El-Jade MR, Parcina M, Schmithausen RM, Stein C, Meilaender A, Hoerauf A, et al. ESBL Detection: Comparison of a Commercially Available Chromogenic Test for Third Generation Cephalosporine Resistance and Automated Susceptibility Testing in Enterobacteriaceae. Chang YF, éditeur. *PLOS ONE*. août 2016;11(8):e0160203.

262. Croxatto A, Marcelpoil R, Orny C, Morel D, Prod'hom G, Greub G. Towards automated detection, semi-quantification and identification of microbial growth in clinical bacteriology: A proof of concept. *Biomed J.* déc 2017;40(6):317-28.
263. Faron ML, Buchan BW, Samra H, Ledeboer NA. Evaluation of WASPLab Software To Automatically Read chromID CPS Elite Agar for Reporting of Urine Cultures. Burnham CAD, éditeur. *J Clin Microbiol.* déc 2019;58(1):e00540-19.
264. Dauwalder O, Michel A, Eymard C, Santos K, Chanel L, Luzzati A, et al. Use of artificial intelligence for tailored routine urine analyses. *Clin Microbiol Infect.* août 2021;27(8):1168.e1-1168.e6.
265. Uwamino Y, Nagata M, Aoki W, Kato A, Daigo M, Ishihara O, et al. Efficient automated semi-quantitative urine culture analysis via BD Urine Culture App. *Diagn Microbiol Infect Dis.* janv 2022;102(1):115567.
266. Faron ML, Buchan BW, Vismara C, Lacchini C, Bielli A, Gesu G, et al. Automated Scoring of Chromogenic Media for Detection of Methicillin-Resistant *Staphylococcus aureus* by Use of WASPLab Image Analysis Software. Richter SS, éditeur. *J Clin Microbiol.* mars 2016;54(3):620-4.
267. Girlich D, Poirel L, Nordmann P. Comparison of the SUPERCARBA, CHROMagar KPC, and Brilliance CRE screening media for detection of Enterobacteriaceae with reduced susceptibility to carbapenems. *Diagn Microbiol Infect Dis.* févr 2013;75(2):214-7.
268. Girlich D, Groperrin V, Naas T, Dortet L. CHROMagarTM ESBL/mSuperCARBA bi-plate medium for detection of ESBL- and carbapenemase-producing Enterobacteriaceae from spiked stools. *Diagn Microbiol Infect Dis.* oct 2019;95(2):107-12.

PUBLICATIONS ET COMMUNICATIONS

Publications en cours de préparation :

- Revue bibliographique

J. Peyroux, Y. Caspar, A. Lbath, M. Maurin, Methods for bacterial infection identification : overview of available techniques and innovations for the future, article en préparation, soumission prévue dans la revue *Frontiers in Microbiology*

- Etude sur l'application de l'analyse de patterns de diffraction pour l'identification bactérienne

J. Peyroux, Y. Caspar, L. Lbath, M. Maurin, Rapid non-destructive and automated bacterial identification by image-based diffraction pattern analysis, article en préparation, publication dans une revue à déterminer

J. Peyroux, S. Oualibouch, Y. Caspar, A. Lbath, Using CNN algorithm for image analysis-based bacterial identification, article en préparation pour une conférence internationale avec acte et comité de lecture

- Etude sur le dépistage rapide et automatisé des BMR aux antibiotiques sur milieux chromogènes à l'aide de l'automate Scanstation

J. Peyroux, I. Almahmoudh, E. Prebe-Coquerel, T. Girard, M. Maurin, Y. Caspar, Rapid and automated screening of MDR bacteria using chromogenic agar media and ScanStation device, article en préparation, soumission prévue dans *Journal of Clinical Microbiology*

Communications :

J. Peyroux, Y. Caspar, L. Lbath, M. Maurin, Application of diffraction pattern analysis and deep learning for fast non destructive and automated bacterial identification, congrès ASM Microbe 2021 (17 juin 2021), congrès à distance, présentation d'un poster

J. Peyroux, Y. Caspar, L. Lbath, M. Maurin, Rapid non destructive and automated bacterial identification by image based diffraction pattern analysis, congrès ECCMID 2021 (9 juillet 2021), congrès à distance, présentation orale

J. Peyroux, Y. Caspar, M. Maurin, Real-time automated screening of MDR bacteria using Scanstation device and chromogenic media, congrès ECCMID 2022 (26 avril 2022), Lisbonne, présentation orale

J. Peyroux, I. Almahmoudh, E. Prebe-Coquerel, T. Girard, M. Maurin, Y. Caspar, Rapid and automated screening of MDR bacteria using chromogenic agar media and ScanStation device, Poster au congrès SFM Microbe 2022 (3 octobre 2022), Montpellier, présentation d'un poster

ANNEXE

Dépistage automatisé de bactéries multi-
résistantes aux antibiotiques sur géloses
chromogènes : évaluation des performances
de l'automate ScanStation (Interscience)

Dans le cadre de l'intérêt porté à l'étude et l'application de nouvelles techniques d'imagerie pour l'identification bactérienne, l'occasion s'est présentée d'étudier l'automate ScanStation de la société Interscience qui utilise une autre approche d'identification rapide de pathogènes bactériens. Ce automate est commercialisé pour la surveillance sanitaire en agro-alimentaire. Il permet l'identification et le compte automatique de colonies sur milieux sélectifs pour la détection de bactéries pathogènes d'intérêt sanitaire (*Listeria*, *Salmonella spp...*). En effet, le ScanStation permet d'aborder une approche de pré-identification en fonction des caractères physiques d'une colonie (taille, couleur, vitesse de croissance). La méthode de détection des colonies et d'analyse de leur couleur peut être également appliquée pour la détection de résistances antibiotiques. Nous avons étudié l'application de l'automate ScanStation combiné avec des milieux chromogéniques pour la détection de bactéries multi-résistantes.

I. Introduction et problématique

Durant la dernière décennie, de nombreux développements ont été réalisés sur l'automatisation en laboratoire de bactériologie, permettant un raccourcissement des temps d'analyse et une augmentation du volume d'échantillons traités pour l'identification bactérienne et l'évaluation des résistances antimicrobiennes (142). Bien que les méthodes moléculaires puissent remplacer les méthodes de culture dans certains cas, leur coût élevé reste un facteur limitant pour leur application à haut-débit (255). En conséquence, la culture bactérienne reste actuellement la méthode de référence dans la plupart des laboratoires, en particulier pour le dépistage systématique de bactérie multi-résistantes (256).

Les *Staphylococcus aureus* résistants à la méticilline (SARM), les bactéries productrices de β -lactamases à spectre étendu (BLSE) ou productrices de carbapénémases (BPC) sont des pathogènes massivement dépistés chaque jour en laboratoire dans les prélèvements nasaux (SARM) et rectaux (BLSE et BPC) (257). Ils font partie de la liste des pathogènes prioritaires à surveiller, établie par l'Organisation Mondiale de la Santé (OMS), due à la forte augmentation de la prévalence des résistances antibiotiques sur les dernières décennies (258). Le dépistage automatique des patients, particulièrement ceux avec des facteurs à risques comme un retour de l'étranger, une chirurgie cardiaque ou orthopédique ou une admission en unité de soins intensifs, est nécessaire pour prévenir des contaminations nosocomiales et une épidémie hospitalière, ceci grâce à des mesures de contrôle de l'infection rapides (259). Ce dépistage est généralement effectué par culture sur milieux gélosés sélectifs suivie par une identification de l'espèce bactérienne par spectrométrie de masse MALDI-TOF, puis par une confirmation du phénotype de résistance antibiotique par des tests comme l'immunochromatographie rapide, les tests colorimétriques ou moléculaires ou d'autres méthodes de tests de susceptibilité antimicrobienne (39,260,261).

Lors de l'étape de culture, le principal facteur responsable des temps de rendu d'analyse est le temps de croissance sur milieu sélectif qui est dépendant de la bactérie. Un autre facteur est

la fréquence de lecture des géloses par le personnel de laboratoire. En effet, même si les colonies ont une croissance rapide sur les milieux gélosés, le rendu de résultats ne peut être fait qu'après la lecture de la gélose. Actuellement, les géloses sont généralement lues une fois par jour en se basant sur les recommandations de temps de lecture données par les fournisseurs des milieux sélectifs. La détection automatisée en temps réel des colonies peut permettre de réduire drastiquement le temps de rendu de résultats. L'automate ScanStation 100 (Interscience, Saint Nom, France) est un incubateur capable de suivre au cours du temps la croissance de colonies bactériennes sur milieux gélosés en imageant les géloses fréquemment. Le logiciel de cet automate permet la détection et la numération des colonies bactériennes mais également la détection de leurs couleurs suivant des filtres prédéfinis par l'utilisateur (jusqu'à 4 couleurs différentes). En combinant cet automate avec l'utilisation de milieux gélosés chromogéniques, il est possible de détecter et catégoriser, en fonction de leur couleur, des colonies bactériennes. Cette approche peut avoir un intérêt pour le dépistage rapide et automatisé des bactéries multirésistantes sur milieux sélectifs chromogéniques permettant ainsi de conduire plus rapidement des tests de confirmation et accélérer la mise en place de mesure de contrôle des infections. Les travaux présentés ici visent donc à évaluer les performances de l'automate ScanStation en combinaison avec l'utilisation de milieux sélectifs chromogéniques CHROMagar™ mSuperCARBA, ESBL et MRSA pour la détection précoce de bactéries productrices de carbapénémases, ou de β -lactamases à spectre étendu dans des prélèvements rectaux ainsi que de *Staphylococcus aureus* résistants à la méticilline (SARM) dans des prélèvements nasaux.



Figure 44 : Automate ScanStation 100 (Interscience)

II. Matériel et méthode

II.A. Automate ScanStation

L'automate ScanStation est un incubateur automatisé. Un bras robotisé permet de transférer les géloses sous une caméra pouvant prendre des images des géloses toutes les 30 minutes. L'analyse de ces images permet la détection des colonies se développant sur les géloses ainsi que la détection de leurs couleurs. Le logiciel de l'automate permet la configuration de filtres couleurs jusqu'au nombre de 4. La configuration de ces filtres est faite en amont pour permettre de classifier les différentes colonies suivant les gammes de couleurs possibles lors de l'utilisation de milieux chromogènes. L'automate peut ainsi quantifier les colonies présentes sur la gélose mais également donner une suspicion d'identification en fonction des géloses chromogènes utilisées. La version ScanStation 100 de l'automate qui a été utilisée dans cette étude peut contenir en même temps un maximum de 100 géloses à analyser.

II.B. Prélèvements cliniques

L'étude comparative présentée ici a été menée au laboratoire de bactériologie du CHU Grenoble Alpes. Des prélèvements rectaux et nasaux de patients de plus de 18 ans, conditionnés en EswabTM (Copan, Italie) et envoyés au laboratoire de bactériologie entre Janvier et Juin 2021 ont été inclus. Cette étude non interventionnelle comprenant des prélèvements et informations de patients a été menée en accord avec la déclaration d'Helsinki ainsi qu'avec la réglementation française en vigueur. L'investigateur principal (Dr. Yvan Caspar) a signé un engagement à respecter la méthodologie de référence n°004 émise par les autorités françaises de la Commission Nationale de l'Informatique et des Libertés (CNIL). Les participants de l'étude ont été informés et n'ont pas émis d'objection à leur inclusion. Le consentement écrit pour participation n'était pas requis pour cette étude, en accord avec la réglementation nationale et institutionnelle.

II.C. Ensemencement des milieux chromogènes

10 µL des prélèvements ont été ensemencés sur géloses chromogènes CHROMATMAgarTM MRSA dans le cas de prélèvements nasaux et sur géloses chromogènes CRHOMATMAgarTM mSuperCARBA et ESBL dans le cas de prélèvements rectaux. L'ensemencement a été réalisé de manière automatisée par une chaîne d'ensemencement automatisée BD KiestraTM (Inocula specimen processor, BD Diagnostics, France) en utilisant un schéma d'épuisement spécifique (4 quadrants S200 pattern, BD Kiestra). Les géloses ont ensuite été insérées dans l'automate ScanStation, incubées à 37 °C pendant 24 heures et avec une acquisition d'image par l'automate toutes les 30 minutes. Chaque gélose a été automatiquement analysée par le logiciel de l'automate pour la détection de colonies, leur compte et leur classification basée sur leur couleur grâce à des filtres de couleurs spécifiques fournis par Interscience pour chaque milieu gélosé chromogénique utilisé dans cette étude en accord avec leur spécification.

II.D. Analyses de résultats

Sur les milieux CHROMATMAgarTM MRSA, les colonies de SARM sont colorées en rose. Sur milieux CHROMATMAgarTM mSuperCARBA, seules les colonies d'espèces résistantes aux carbapénémases peuvent se développer. Les colonies d'*Escherichia coli* sont colorées en rose foncé, les colonies de coliformes (*Klebsiella* sp, *Enterobacter* sp, *Citrobacter* sp et *Serratia* sp) sont colorées en bleu métallique tandis que celles d'autres microorganismes producteurs de carbapénémases se développent en colonies incolores. Dans le cas des géloses CHROMATMAgarTM ESBL servant à la détection de bactéries productrices de BLSE, les colonies d'*Escherichia coli* sont également colorées en rose foncé, les colonies de *Klebsiella*, d'*Enterobacter* et de *Citrobacter* sont colorées en bleu métallique. Les colonies de *Proteus spp* se développent sur ces géloses avec une couleur marron, celles d'*Acinetobacter spp* avec une couleur crème tandis que les colonies de *Pseudomonas spp* sont translucides et celles de *Stenotrophomonas spp* restent incolores. Cependant, l'automate ScanStation ne permet la distinction que de 4 couleurs au maximum alors que les géloses chromogènes CHROMATMAgarTM ESBL peuvent contenir des

colonies de 6 couleurs différentes. Les couleurs crème, translucide et incolore ont donc été regroupées dans un seul filtre couleur correspondant à une catégorie nommée « incolore ».

Les temps de détection de la première colonie, de 50 % et de 95 % des colonies par gélose ainsi que les temps d'identification des colonies ont été extraits des rapports d'analyses de l'automate. Les temps dits d'identification des colonies correspondent au moment où l'automate caractérise correctement les colonies selon leur couleur.

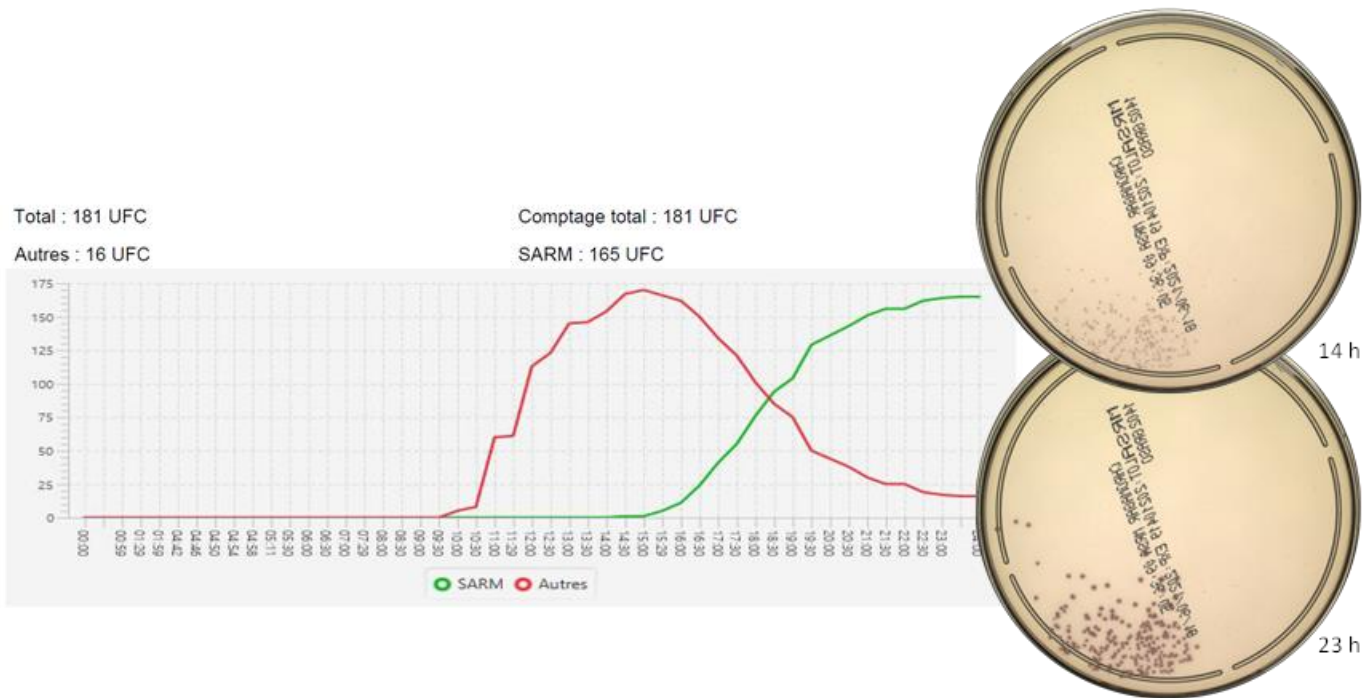


Figure 45 : Exemple du délai entre la détection des colonies et leur identification dans la catégorie de couleur correcte pour un dépistage de SARM dans un prélèvement nasal. Le graphique montre le nombre de colonies détectées toutes les 30 minutes pour deux catégories de couleur : rose, correspondant à une suspicion de SARM (ligne verte) et incolore correspondant à d'autres microorganismes pouvant se développer sur ces géloses (ligne rouge). Les images à droite montrent des images de la gélose analysée prises par l'automate ScanStation après 14 et 24 heures d'incubation

II.E. Observations visuelles des géloses

Les résultats du ScanStation ont été comparés avec une analyse visuelle des géloses réalisé à la fin des analyses du ScanStation (soit après 24 heures d'incubation) et donc en concordance avec les standards de diagnostics réalisés en laboratoire. La couleur de colonies, leur nombre (classé en catégories : 0 à 10, 11 à 100, 101 à 300 et plus de 300 colonies) ainsi que la confluence ou non des colonies ont été relevés pour chaque prélèvement analysé. Certaines géloses contenaient des colonies de très petites tailles visibles en fin d'incubation par observation visuelle (Figure 46C). Ces petites colonies à pousse tardive ne sont pas compatibles avec des espèces bactériennes multirésistantes et ont donc été considérés lors de l'analyse visuelle comme des résultats négatifs pour le dépistage de bactéries multirésistantes. Les concordances positives (Positive Percentage Agreement ou PPA) et négatives (Negative Percentage Agreement ou NPA) ainsi que les valeurs prédictives positives (Positive Predictive Value ou PPV) et négatives (Negative Predictive Value ou NPV) du dépistage de bactéries multirésistantes par l'automate ScanStation en comparaison de résultats obtenus par observation visuelle des géloses ont été déterminés.

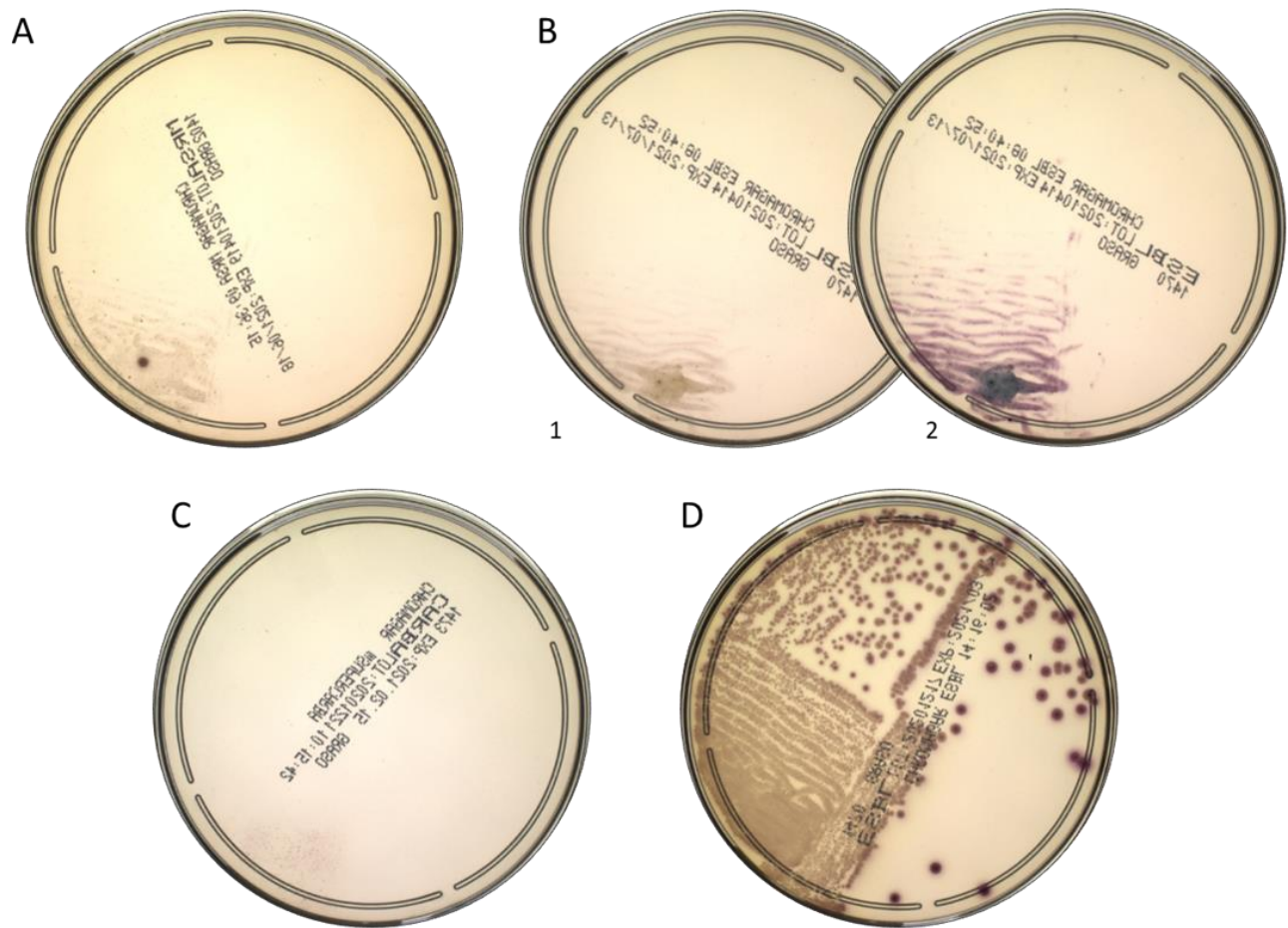


Figure 46 : Exemple de résultats faux positifs et faux négatifs observés durant l'analyse ScanStation. A : Colonie rose identifiée à tort comme étant incolore pendant un dépistage de SARM. B : Artefacts et coloration de la gélose causés par le dépôt d'un prélèvement rectal juste après son ensemencement (1) et après 24 heures d'incubation (2). Ces artefacts résultent en une coloration de la gélose sans présence de colonies bactériennes. C : Petites colonies à pousse tardive apparaissant après 20 heures d'incubation. D : Epuisement du composé chromogène créant un palissement des colonies roses menant à leur fausse identification comme colonies incolores

III. Résultats

Au total, 419 prélèvements nasaux et 501 prélèvements rectaux ont été analysés au cours de cette étude. Les concordances positives et négatives ainsi que les valeurs prédictives positives et négatives pour la détection et la classification en temps réel de colonies par l'automate ScanStation en comparaison de l'analyse visuelle des géloses chromogènes sont présentées dans le tableau 25.

	PPA	NPA	PPV	NPV	Vrais Positifs	Vrais Négatifs	Faux Positifs	Faux Négatifs
SARM	78.57	98.27	61.11	99.25	11	398	7	3
BPC								
Global	100.00	88.27	70.32	100.00	109	346	46	0
E. coli	100.00	78.23	11.67	100.00	14	381	106	0
Coliformes	100.00	94.06	30.95	100.00	13	459	29	0
Autres	77.32	95.30	79.79	94.59	75	385	19	22
BP-BLSE								
Global	100.00	85.94	81.03	100.00	188	269	44	0
E. coli	100.00	70.00	35.50	100.00	71	301	129	0
KEC	95.08	98.41	89.23	99.31	58	433	7	3
Proteus	100.00	80.04	17.24	100.00	20	385	96	0
APS	45.45	90.33	46.05	90.12	35	393	41	42

Tableau 25 : Concordances positives et négatives et valeurs prédictives du dépistage de bactéries multirésistantes par l'automate ScanStation sur milieux gélosés chromogéniques

CHROMagar™

PPA: Concordance positive, NPA : Concordance négative, PPV : Valeur prédictive positive, NPV : Valeur prédictive négative, KEC : *Klebsiella*, *Enterobacter*, *Citrobacter* ; Coliformes : *Klebsiella*, *Enterobacter*, *Citrobacter* and *Serratia*, SARM: *Staphylococcus Aureus* résistants à la métiline, BPC : Bactéries Productrices de Carbapénèmes, BP-BLSE : Bactéries productrices de β -lactamases à spectre étendu

III.A. Dépistage de *Staphylococcus aureus* résistants à la méticilline

Après 24 heures d'incubation, l'observation visuelle des géloses CHROMAgar™ MRSA ensemencées avec les prélèvements nasaux a permis d'identifier 14 géloses sur 419 (3,3%) avec des colonies de couleurs roses et 405 (96,7%) géloses négatives dont 179 avec uniquement des colonies incolores ou de petites colonies à croissance tardive et 226 géloses sans aucune croissance bactérienne (Tableau 26). L'analyse de ces géloses par le ScanStation a identifié 18 géloses sur 419 comme positives (4,3%). Cependant, 7 de ces 18 résultats (39%) étaient de faux positifs. Ces faux positifs ont été causés par la mauvaise classification comme rose de colonies à croissance tardive dans 57% des cas (4/7), par la fausse détection d'une ou de quelques colonies roses au sein d'un groupe de colonies incolores dans 29% des cas (2/7, Figure 46A) et dans un seul cas par la coloration de la gélose par l'échantillon (14%). Parmi les 401 / 419 (95,7%) géloses analysées comme négatives pour la présence de colonies roses par l'automate, 258 (soit 64,3%) ont été classées avec une absence de croissance bactérienne et 143 (35,7%) avec la présence de colonies incolores. Trois de ces géloses étaient des résultats faux négatifs avec des colonies roses vues par l'automate comme étant incolores (Tableau 26).

SARM			Analyse ScanStation				
			Rose	Incolore	Mix	Négatif	Total
Observation visuelle	Positif	Rose	1	0	1	0	2
		Incolore	0	126	2	0	128
		Deux couleurs	0	3	9	0	12
	Négatif	Croissance tardive	2	9	2	38	51
		Vide	0	5	1	220	226
		Total	3	143	15	258	419

Tableau 26 : Corrélation entre l'observation visuelle et les résultats du ScanStation pour l'analyse de Staphylococcus aureus résistants à la méticilline. Les résultats positifs sont différenciés par la couleur des colonies présentes. Les résultats négatifs par observation visuelle regroupent les géloses sans colonie et celles avec seulement de petites colonies à croissance tardive non compatible avec la présence de colonies de SARM

III.B. Dépistage de bactéries productrices de carbapénémases

L'observation visuelle des géloses CHROMAgar™ mSuperCARBA après 24 heures d'incubation a révélé 109 géloses positives sur 501 (24,8%) : 6 géloses avec des colonies roses évocatrices de la présence d'*Escherichia coli*, 4 avec des colonies bleues métalliques compatibles avec la présence de coliformes, 85 géloses avec des colonies incolores évocatrices d'autres bactéries à Gram négatif et 14 géloses avec un mélange de colonies de couleurs différentes (Tableau 27). Parmi les 392 géloses négatives en observation visuelle, 368 étaient vides de toutes colonies et 24 présentaient de petites colonies à croissance tardive. Indépendamment de la classification par couleur, le ScanStation a identifié 155 géloses sur 501 (30,9%) comme ayant des colonies d'intérêt et 346 (69,1%) comme étant vides de toutes colonies. Parmi les 155 résultats considérés comme positifs par l'automate, 46 (29,7%) étaient des résultats faux positifs. Ils étaient dus à une coloration de la gélose ou à un artefact causés par le dépôt du prélèvement dans 17 cas (37%) (Figure 46B), par des colonies à croissance tardive détectées par l'automate et considérées comme négatives en observation visuelle dans 17 cas (37%), par l'apparition de condensation sur le couvercle de la gélose pour 6 cas (13%), des défauts de gélose dans 4 cas (9%) et par la détection des bords de la gélose dans 2 cas (4%).

Après analyse des couleurs de colonies, aucun résultat faux négatif n'a été identifié pour les colonies roses et bleues métallique. Cependant, l'analyse par le ScanStation a identifié des colonies roses dans 120 prélèvements parmi lesquels seuls 14 (12%) étaient de vrais positifs. Les résultats faux positifs incluaient 34 géloses (32%) négatives (21 géloses vides et 13 avec des colonies à croissance tardive), des colonies incolores identifiées comme roses dans 69 cas (65%) et des colonies bleues identifiées comme roses dans les 3 cas restants (3%). De manière similaire, sur les 42 géloses analysées comme ayant des colonies bleues métalliques, seulement 13 étaient positives en observation visuelle (31%). Les 29 résultats faux positifs étaient dus à la mauvaise classification comme bleues de colonies incolores dans 21 cas, et de colonies roses dans 1 cas. Les 7 autres résultats faux positifs pour la classification de colonies bleues étaient dus à la fausse détection de 7 géloses négatives (6 géloses vides et 1 avec des colonies à croissance tardive). Enfin, la détection de colonies incolores a résulté en 94 analyses positives donc 75 étaient de vrais positives (80%). L'identification erronée de colonies incolores sur des géloses

négatives a été observée dans 15 cas (8 géloses vides et 7 avec des colonies à croissance tardive). Quatre autres cas étaient dus à l'identification de colonies roses comme étant incolores.

BPC		Résultats ScanStation						
		Rose	Bleu	Incolore	Mix	Négatif	Total	
Observation visuelle	Rose	4	0	0	2	0	6	
	Positif	Bleu	0	1	0	3	0	4
		Incolore	18	3	16	48	0	85
		Mix	0	0	0	14	0	14
	Négatif	Croissance tardive	9	1	3	4	7	24
		Vides	18	2	5	4	339	368
Total		49	7	24	75	346	501	

Tableau 27 : Corrélation entre l'observation visuelle et les résultats du ScanStation pour l'analyse de bactéries productrices de carbapénèmes. Les résultats positifs sont différenciés par la couleur des colonies présentes. Les résultats négatifs par observation visuelle regroupent les géloses sans colonie et celles avec seulement de petites colonies à croissance tardive non compatible avec la présence de colonies de BPC. Mix: au moins 2 catégories de couleurs différentes présentes sur la gélose

III.C. Dépistage de bactéries productrices de β -lactamases à spectre étendu

Les 501 prélèvements rectaux ont été également ensemencés sur géloses CHROMAgar™ ESBLE et analysés pour le dépistage de Bactéries productrices de β -lactamases à spectre étendu (BP-BLSE). Comme seulement 4 catégories de couleurs sont autorisées par l'automate ScanStation, seules sont distinguées les colonies roses compatibles avec la présence d'*E. coli*, celles de couleur bleue métallique indiquant la présence d'espèces du groupe KEC (*Klebsiella spp*, *Enterobacter spp*, *Citrobacter spp*), les colonies marrons pour les espèces de *Proteus* et les colonies incolores regroupant les espèces du groupe appelé APS (*Acinetobacter spp*, *Pseudomonas spp* et *Stenotrophomonas spp*). Après 24 heures d'incubation, l'observation visuelle des géloses indique 188 géloses positives (37,5%) sur 501 : 48 avec des colonies roses, 36 avec des colonies bleues métalliques, 8 avec des colonies marrons, 57 présentant des colonies incolores et 39 avec des colonies de plusieurs catégories de couleurs (Tableau 28). Les 313 géloses (62,5%) restantes étaient négatives. Indépendamment de la classification par catégorie de couleur, 232 géloses (46,3%) ont été identifiées par le ScanStation comme positives. Aucun résultat faux négatif n'a été observé mais 44 géloses parmi les 232 rendues positives par l'automate étaient de faux positifs (19%). Ces faux positifs étaient dus à une coloration ou à un artefact causés par le dépôt de l'échantillon dans 23 cas (52%), à la détection de petites colonies à croissance tardive dans 9 cas (21%), à la détection des bords de la gélose dans 6 cas (14%), à l'apparition de condensation dans 5 cas (11%) et à des défauts dans la gélose dans 1 cas (2%).

Après la catégorisation par couleur de l'automate, 200 prélèvements ont été analysés par le ScanStation comme présentant des colonies roses. Parmi ces 200 analyses, 129 (64,5%) étaient de faux positifs (32 avec des colonies bleues, 6 avec des colonies marrons, 51 avec des colonies incolores et 40 géloses négatives). La détection par le ScanStation de colonies bleues métalliques a résulté en 58 géloses sur 65 (89,2%) vraies positives et 3 géloses sur 436 (0,7%) faussement négatives (dans les 3 cas du fait d'une identification des colonies bleues comme étant roses). Les 7 résultats faux positifs résultent de la mauvaise identification de colonies roses comme étant bleues dans 2 cas, de colonies incolores classées bleues dans 3 cas et à la fausse détection de colonies bleues sur des géloses négatives dans 2 cas. La détection de colonies marrons se conclue par 20 résultats sur 116 (17%) vrais positifs et 96 (83%) résultats faux positifs dus à la mauvaise

classification comme marron de colonies roses dans 27 échantillons, de colonies bleues dans 18 échantillons, de colonies incolores dans 44 échantillons ainsi que la détection erronée de colonies sur 7 géloses négatives. Enfin, la détection de colonies incolores résulte en 35 géloses sur 76 (46%) vraies positives et 41 faussement positives. Ces résultats faux positifs sont dus à la mauvaise catégorisation comme incolore des colonies roses dans 10 cas, de colonies bleues dans 23 cas, de colonies marrons dans 2 cas et à la fausse détection de colonies sur 6 géloses négatives. 42 résultats faux négatifs ont été également observés. Ils sont dus à des colonies incolores identifiées à tort comme étant roses dans 26 cas, comme bleues dans 1 cas et comme étant marron dans 15 cas.

BP-BLSE		Résultats ScanStation							
		Rose	Bleu	Marron	Incolore	Mix	Négatif	Total	
Observation visuelle	Rose	29	0	0	0	19	0	48	
	Bleu	2	5	0	0	29	0	36	
	Marron	0	0	2	0	6	0	8	
	Incolore	17	0	7	2	31	0	57	
	Mix	0	0	0	0	39	0	39	
	Négatif	Croissance tardive	6	0	0	0	3	6	15
		Vide	27	0	1	2	5	263	298
Total		81	5	10	4	132	269	501	

Tableau 28 : Corrélation entre l'observation visuelle et les résultats du ScanStation pour l'analyse de bactéries productrices de β -lactamases à spectre étendu. Les résultats positifs sont différenciés par la couleur des colonies présentes. Les résultats négatifs par observation visuelle regroupent les géloses sans colonie et celles avec seulement de petites colonies à croissance tardive non compatible avec la présence de colonies de BP-BLSE. Mix: au moins 2 catégories de couleurs différentes présentes sur la gélose

III.D. Délai entre détection et catégorisation correcte des couleurs de colonies

Les temps de détection de la première colonie, de 50% et de 95% des colonies ainsi que les temps auxquels les colonies ont été correctement classées par couleur sont présentés dans le tableau 29. Les premières colonies roses sur milieux chromogènes SARM ont été détectées en moyenne en 13 heures et 19 minutes avec un délai avant une classification de couleur correcte de 17 minutes en moyenne. Sur milieu CHROMAgar™ mSuperCARBA, la détection des premières colonies a eu lieu en moyenne au bout de 13 heures et 19 minutes également mais avec un délai moyen avant identification correcte de 1 heure et 59 minutes. Enfin, la détection des premières colonies présument de la présence de BP-BSLE a été réalisée en moyenne en 10 heures et 23 minutes, le délai moyen avant l'identification correcte de leur couleur étant de 2 heures et 15 minutes.

		Temps moyen (min - max)		
		SARM	BPC	BP-BLSE
Première colonie	Détection	13h19 (5h – 24h)	13h19 (5h – 24h)	10h23 (4h30 – 24h)
	Couleur	13h36 (5h – 24h)	15h18 (5h – 24h)	12h38 (5h – 24h)
	Délai	0h17 (0h – 16h)	1h59 (0h – 18h)	2h15 (0h - 15h30)
50% des colonies	Détection	17h31 (7h30 – 24h)	16h34 (5h – 24h)	13h22 (5h – 24h)
	Couleur	17h42 (7h30 – 24h)	17h32 (6h – 24h)	15h22 (5h – 24h)
	Délai	0h11 (0h – 13h30)	0h58 (0h – 9h)	2h00 (0h – 12h)
95% des colonies	Détection	20h24 (8h – 24h)	18h41 (5h – 24h)	16h16 (5h – 24h)
	Couleur	20h35 (8h – 24h)	19h29 (5h – 24h)	18h12 (5h – 24h)
	Délai	0h11 (0h – 11h30)	0h48 (0h – 8h)	1h56 (0h – 13h)

Tableau 29 : Temps moyen de détection et d'identification correcte de la couleur lors des analyses ScanStation pour le dépistage de SARM, de BPC et de BP-BLSE

IV. Discussion et perspectives

La réduction du temps de détection de bactéries multirésistantes pouvant avoir un impact majeur sur la limitation de leur transmission au sein des établissements de santé. Toute amélioration technologique permettant la diminution des temps d'analyses peut avoir un intérêt tant que son coût d'application reste compatible avec une utilisation pour des dépistages massifs. L'automatisation des laboratoires avec des analyses automatisées utilisant l'analyse d'images et une intelligence artificielle s'est largement développée ces dernières années (262,263). Des solutions telles que PhenoMATRIX™ ou BD Kiestra™ Urine Culture App ont été développées pour le compte et l'analyse des couleurs automatisés de colonies sur milieux gélosés chromogéniques, principalement pour les prélèvements urinaires. La concordance globale de ces approches en comparaison des techniques traditionnelles est aux alentours de 90 % (142,264,265). Avec des sensibilités proches des 100%, au détriment de spécificités qui se situent entre 80 et 90 %, ces solutions présentent une réduction des temps de rendu de résultats principalement pour les résultats négatifs (39,266,267). Cependant, les systèmes automatisés actuels font des acquisitions d'images de géloses seulement une fois par jour. Dans cette étude, les performances d'une stratégie combinant les milieux gélosés chromogéniques et un automate permettant la détection précoce de colonies par suivi en temps réel des géloses ont été évaluées.

Comme la plupart des milieux chromogéniques, les géloses CHROMAgar™ ont une forte sensibilité avec une spécificité moyenne (par exemple, les géloses CRHOMAgar™ mSuperCARBA ont une sensibilité entre 96,5 et 97,8 % pour des spécificités se situant entre 60,7 et 66,1%) (267,268). Il a été décidé de comparer les résultats des analyses de l'automate ScanStation non pas au diagnostic clinique effectué sur les prélèvements, mais à une observation visuelle des géloses après incubation dans l'automate pour ne pas prendre en compte, dans l'évaluation du ScanStation, les sensibilités et spécificités des géloses CHROMAgar™. La détection de colonies compatibles avec la présence de bactéries multirésistantes par l'automate ScanStation sur de tels milieux gélosés apparaît comme applicable au dépistage de bactéries productrices de carbapénémases ou de β -lactamases à spectre étendu. En effet, sans prendre en compte la catégorisation par couleur, une concordance positive de 100 % a été observée en comparaison des résultats par observation visuelle des géloses. Ce procédé a une efficacité

réduite cependant pour le dépistage de SARM car il nécessite une analyse de couleur (les colonies incolores étant dans ce cas considérées comme des résultats négatifs). Le ScanStation a classé comme négatifs 3 des 14 cas positifs (79% de concordance positive) à cause de la présence de quelques colonies roses non détectées au sein de groupes confluent de colonies incolores. En revanche, les résultats sont très concordants dans le cas d'échantillons négatifs (concordance négative de 98,3%). Il est important également de noter la faible proportion de prélèvements positifs lors du dépistage de SARM au sein de la cohorte d'étude (1 prélèvement positif lors du diagnostic clinique, 14 sur 419 prélèvements positifs sur milieu chromogénique), cette faible proportion peut avoir un impact sur les concordances positives et valeurs prédictives positives pour le dépistage des SARM.

La concordance négative des résultats est moins bonne dans le cas des dépistages de bactéries productrices de carbapénémases ou de β -lactamases à spectre étendu (88% pour les BPC et 85% pour les BP-BLSE) du fait de la détection de nombreux artefacts (coloration de la gélose, dépôt de l'échantillon, colonies à croissance tardive, détection des bords de gélose, condensation ou défaut de gélose). En prenant en compte l'entièreté des échantillons n'ayant pas montrés de croissance bactérienne sur milieu chromogénique, 97 des 1110 géloses vides (8,7%) ont été faussement identifiées par l'automate comme positives (1,7% pour les SARM, 11,7% pour les BPC et 14,1% pour les BP-BLSE). Une des causes principales de ce problème est la coloration de la gélose ou la présence d'artefacts dus au dépôt du prélèvement (45% des cas de faux positifs). La composition de certains prélèvements apparaît comme pouvant causer une réaction chimique avec les milieux chromogéniques créant ainsi une coloration de la gélose sans présence de colonie bactérienne (figure 46B). Les autres cas de faux positifs pour des géloses négatives sont dus à la détection de colonies à croissance tardive dans 28% des cas, d'artefacts causés par des défauts de géloses dans 5% des cas, à la détection des bords de géloses dans 8% des cas et à une mauvaise identification de la couleur des colonies dans 2% des cas (uniquement pour la détection de SARM où des colonies incolores sont catégorisées comme négatives).

Les valeurs prédictives négatives sont très hautes pour tout type de dépistage (100% pour les BPC et BP-BLSE, 99,3% pour les SARM) confirmant la possibilité d'utiliser cette approche pour l'élimination automatique des résultats négatifs lors de campagnes de dépistage. Cette stratégie peut donc permettre une gestion accélérée des résultats négatifs, avec un gain de temps technique et de rendu des résultats négatifs pour l'arrêt rapide des mesures de contrôles des infections. Dans l'étude ici présentée, c'est le cas de respectivement 54%, 69% et 95% des prélèvements ensemencés respectivement sur géloses chromogéniques CHROMAgar™ ESB, CHROMAgar™ mSuperCARBA et CHORMAgar™ MRSA. Comme attendu, les valeurs prédictives positives étaient assez basses (61,1% pour les SARM, 70,3% pour les BPC et 81% pour les BP-BLSE) à cause de la présence nombreuses d'artefacts. Cela montre la nécessité d'une étape de validation des résultats par contrôle visuel des géloses pour les prélèvements considérés comme positifs par le ScanStation. Cela permettrait d'éliminer rapidement les résultats faussement positifs de l'automate à cause d'artefacts ou de colonies à croissance tardive qui représentent dans notre étude 30 à 39% de la totalité des analyses rendues positives. 31 % des faux positifs totaux de l'étude sont dus à la détection de colonies à croissance tardive.

Pour la détection de couleurs sur les milieux chromogéniques mSuperCARBA et ESB, l'automate ScanStation est fiable pour la détection des colonies bleues mais montre des difficultés à différencier les colonies roses de celles marrons ou incolores. La détection des colonies bleues montre de très bonnes concordances positives et négatives et valeurs prédictives négatives (respectivement 100%, 64,1% et 100% sur mSuperCARBA et 95,1%, 98,4% et 100% sur milieu ESB). La détection de colonies roses a de fortes concordances positives et valeurs prédictives négatives (100% pour les 2 milieux chromogéniques) mais de faibles concordances négatives et valeurs prédictives positives (respectivement 78,2% et 11,7% sur mSuperCARBA et 70% et 35,5% sur milieu ESB). Une proportion significative des classifications de couleurs faussement positives par le ScanStation est due à la mauvaise catégorisation de colonies incolores comme étant roses (86,7% pour mSuperCARBA et 60% pour ESB) et comme étant marron (35,3% sur milieu ESB). Les colonies initialement incolores, se foncent au fur à mesure de l'analyse et mènent à des résultats erronés. L'amélioration des systèmes d'acquisition d'images de l'automate pourrait améliorer la discrimination des différentes couleurs.

Les catégorisations par couleur incorrectes peuvent également être corrélées à la confluence des colonies. Parmi les 227 cas de mauvaises identifications des couleurs de colonies, 98 (43%) géloses ont un nombre de colonies supérieur à 100, parmi lesquelles 95% montrent une confluence des colonies bactériennes. La proximité des colonies entre elles peut ainsi empêcher l'automate de détecter et de catégoriser correctement toutes les colonies présentes sur une gélose. De plus, la présence d'un grand nombre de colonies sur une gélose semble créer une variation dans la coloration des colonies de par l'épuisement des composés chromogènes des milieux gélosés. Cela cause une hétérogénéité dans la couleur des colonies qui apparaissent plus pâle (figure 46D). Toutes ces difficultés permettent d'expliquer pourquoi l'automate ScanStation a détecté plus de catégories de couleurs que réellement présents sur les géloses dans 25,6% des prélèvements rectaux (17,1% avec une couleur supplémentaire, 6,7% avec 2 et 1,8 avec 3) (Tableaux 30 et 31).

Plusieurs développements futurs pour l'acquisition d'images et le logiciel du ScanStation peuvent être proposés pour réduire les cas faux positifs et améliorer la catégorisation des couleurs. Par exemple, la détection comme positifs des artefacts peut être partiellement résolue en implémentant une élimination automatique des artefacts détectés très tôt (en moins de 3 heures d'incubation) mais également en éliminant les colonies à croissance tardive détectées très tard dans la phase d'incubation (les petites colonies à croissance tardive non compatibles avec une suspicion de bactérie multirésistante apparaissent généralement après 20 heures d'incubation et ont des tailles significativement inférieures après 24 heures d'incubation). L'amélioration des systèmes d'acquisition d'images pourrait également permettre une meilleure distinction des couleurs de colonies. Enfin, la gélose elle-même des milieux a une couleur crème, les colonies ayant une couleur semblable sont ainsi plus difficiles à catégoriser, changer les composés chromogènes roses par une autre couleur pourrait ainsi aider à l'amélioration de ce point.

Dans la majorité des laboratoires de bactériologies, les bactéries ayant une croissance sur des milieux sélectifs pour le dépistage de SARM, de BPC et de BP-BLSE sont détectées après 18 à 48 heures d'incubation, en se basant sur une observation quotidienne des géloses, soit manuelle ou en utilisant des systèmes automatisés d'acquisition inclus dans les systèmes comme les BD Kiestra™ ou WASP-LAB®. L'analyse avec images toutes les 30 minutes par l'automate

ScanStation montre la possibilité de détection des premières colonies précoce (après seulement 10 à 14 heures de croissance) dans tous les prélèvements et de 95 % des colonies 5 à 8 heures plus tard. L'identification rapide de ces colonies peut par la suite être réalisée par spectrométrie de masse MALDI-TOF. Des tests de confirmations de résistances antibiotique pourraient ainsi être réalisés plus tôt avec un test rapide colorimétrique donnant un premier résultat en quelques minutes et une analyse complète des susceptibilités antibiotiques rendue le jour d'après.

En conclusion, l'automate ScanStation combiné avec l'utilisation de milieux chromogéniques est une approche rapide et automatisée pour le dépistage de *Staphylococcus aureus* résistants à la méticilline (SARM) dans des prélèvements nasaux et les bactéries productrices de β -lactamases à spectre étendu (BLSE) ou productrices de carbapénémases (BPC) dans des prélèvements rectaux. Cette approche est applicable en laboratoire de microbiologie clinique. Avec son acquisition fréquente d'image, le ScanStation accélère la détection des cultures positives et permet le tri des résultats négatifs et positifs. Ses fortes valeurs prédictives négatives rendent possible l'élimination automatique des géloses négatives où aucune croissance bactérienne n'est détectée. Bien que des améliorations soient à apporter notamment sur la détection des couleurs de colonies, cette technologie peut être très intéressante pour accélérer le dépistage de bactéries multirésistantes et la mise en place des mesures de contrôle des infections qui en découlent. Cette accélération peut être d'un grand avantage dans la prévention des épidémies nosocomiales au sein d'établissements de santé.

ScansStation Observation Visuelle	Rose	Bleu	Incolore	Rose + Bleu	Rose + Incolore	Bleu + Incolore	Rose + Bleu + Incolore
Rose	4				1		
Bleu		1		3			
Incolore	18	3	16	1	31	2	15
Rose + Bleu							2
Rose + Incolore					4		1
Bleu + Incolore						1	5
Rose + Bleu + Incolore							1

Tableau 30 : Corrélation entre l'observation visuelle et les résultats ScanStation pour les prélèvements rectaux avec des populations mixtes (plusieurs couleurs) sur milieu mSuperCARBA

ScanStation Observation Visuelle	Rose	Bleu	Marro n	Incolor e	Rose + Bleu	Rose + Marro n	Bleu + Marro n	Bleu + Incolor e	Marron + Incolore	Rose + Bleu + Marro n	Rose + Bleu + Incolor e	Rose + Marron + Incolore	Rose + Bleu + Marron + Incolore
Rose	29				1	13				1		4	
Bleu	2	5			6			2		1	8		12
Marron			2			4			2				
Incolore	17		7	2	1	5			6	1	17		1
Rose + Bleu					3					3			6
Rose + Marron						5						1	
Rose + Incolore						3							
Bleu + Marron							1						
Bleu + Incolore					2	1				3	1		3
Marron + Incolore						2			1			2	
Rose + Bleu + Incolore													2

Tableau 31 : Corrélation entre l'observation visuelle et les résultats ScanStation pour les prélèvements rectaux avec des populations mixtes (plusieurs couleurs) sur milieu ESBL