

**Exploring ambient mass spectrometry capacities
for rapid detection and phenotyping of
Pseudomonas aeruginosa infection
in cystic fibrosis patients**

A thesis submitted by
Elodie Emmanuelle BARDIN

For the degree of Doctor of Philosophy (PhD)

Imperial College London
Faculty of Medicine

Declaration of originality

I certify that this thesis and the research it rests upon are the result of my own work. Other people collaborating to this project, as well as external sources are fully acknowledged or referenced according to the standard practices of the discipline.

E. Emmanuelle BARDIN

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

I would like to dedicate this thesis to my dearest friend Robbi,
who inspired me to start the PhD and supported me to achieve it.

Acknowledgements

Research is a team effort and I am infinitely grateful to the many great people and friends who have contributed to this project.

Foremost, I would like to thank my supervisors, Professors Zoltan Takats, Jane Davies, Andy Bush and Eric Alton, for entrusting me with this huge and meaningful project and for giving me the opportunity to work in such a stimulating environment as Imperial College London. I am particularly grateful to Prof Jane Davies for her constant support throughout those four years, and to Prof Andy Bush for his faith and wise advice, especially during the writing-up.

I am also very thankful to the UK CF Trust for funding the project through the SRC for *Pseudomonas aeruginosa*, and for giving me rewarding opportunities to travel, share my work and meet diverse and inspiring people.

I am especially grateful to Dr Simon Cameron for his guidance, for his daily help in the REIMS laboratory and in particular for writing the scripts controlling the handling robot enabling the automation of REIMS which saved me hours of analysis. I am also very thankful to Dr Frankie Bolt who took the time to answer my microbiological questions and to review my work even after she had left the group. Both of them were very supportive and helpful in the preparation of abstracts, presentations, papers and in the writing of the present thesis.

As a chemist, I would have been overwhelmed by the amount and complexity of the data generated here; thus, I am deeply thankful to my colleague and friend Alvaro Perdones-Montero for his time, attentiveness and work on designing specific tools to process REIMS profiling and imaging data. I also thank Dr James McKenzie who adapted the current DESI-MS data workflow to the unusual analysis of skin secretion samples.

It has been a pleasure to work with Natasha Wierre-Gore, a fellow PhD student within the SRC and a friend; I want to thank her for sharing her isolates and results, and for our enriching and endless conversations. She collected and characterised hundreds of isolates by phenotypic assays and MLVA, and repeated the culturing for REIMS analysis. I am also thankful to her supervisor, Prof Huw Williams, and to my friend Dr Katie Farrant from the same group for reviewing the parts of this thesis related to this specific project.

I want to thank all the CF donors and healthy volunteers who provided sputum and skin secretion samples and I am particularly grateful to Dr Katie Bayfield, respiratory physiologist at the Royal

Brompton Hospital. She assisted me in approaching and informing CF patients at the clinic and in the compilation of the consent forms for the collection of samples. I am also thankful to Ronan Murphy, research technician, for helping out in the NHLI laboratory, to Doctors Andy Turnbull and Wynne Smith, clinicians at the Royal Brompton Hospital at the time, who provided me with the CF and bronchiectasis *P. aeruginosa* isolates, and to Neil Madden, from the clinical microbiology laboratory, who I got the other CF isolates from.

I would like to thank Kate Hardiman and Tony Rickards, microbiologists and REIMS technicians, for helping in the handling, the culturing and, from time to time, the analysis of the isolates at Charing Cross Hospital. I am also indebted to Adam Burke, REIMS technician, PhD student and friend, who always was my first point of reference when dealing with REIMS instrumental issues.

Amongst the dear friends I have met at Imperial, I particularly owe Dr Jocelyn Tillner, for being my tireless source of knowledge in English and in science and for kindly reviewing and proofreading parts of this thesis. I feel fortunate I got to share the laboratory and my PhD life with smart, dedicated and helpful people, such as Anna Mroz, Doctors Renata Filipe Soares and Luisa Doria from the DESI-MS team, and most of all, Dipa Gurung, for bearing me as her neighbour, supporting me on my good and bad days, together in high spirits and through PhD disenchantments.

And finally, I will be forever grateful to Sylvain Robbiano for his unfailing support, for helping me through the tough times and celebrating the good ones. I realise how lucky I have been to be able to count on his expertise in statistics, his listening and understanding during these four years. His help has been invaluable.

Abstract

Cystic fibrosis (CF) patients mostly succumb to respiratory failure, usually caused by *Pseudomonas aeruginosa* infection. Early, but not chronic, infection can be eradicated, mandating timely detection. However, current methods are insufficiently sensitive and unsuitable for children. This thesis explores the potential of novel ambient mass spectrometric (MS) tools for the phenotyping of *P. aeruginosa* and its detection in non-invasive samples. The hypotheses are that bacterial expression evolves as infection progresses and that the composition of host body fluids is impacted by *P. aeruginosa* acquisition.

Rapid evaporative ionisation MS (REIMS) was applied to clinical isolate cultures including common CF pathogens. The detected metabolic profiles differentiated between species and led to the identification of virulence-associated metabolites in *P. aeruginosa*: quorum sensing molecules and rhamnolipids. Repeated analysis revealed differences between *P. aeruginosa* strains that allowed isolate classification. Although classification according to genetically-defined types was not achieved, REIMS potentially provides a finer resolution on quickly evolving virulence features. Exploration of intra-species disparity showed a higher metabolic diversity in chronic respiratory than in acute infections, attributed to lengthy and site-specific adaptation. The bacterial population was more disparate between than within CF patients. Variations in virulence factor levels between early and chronic isolates provided insight into *P. aeruginosa*'s metabolism and adaptation, supporting the first hypothesis. These findings may support future clinical strategies.

REIMS and desorption electrospray ionisation MS (DESI-MS) were evaluated as direct-from-sample diagnostic tools using sputum as a reference, and urine and skin secretions as easily accessible samples. REIMS could not detect *P. aeruginosa* in sputum and urine pellets, nor could DESI-MS on skin secretions. Nevertheless, the rich sputum metabolic profile and skin lipidome contained potentially highly relevant information about host physiology that may assist clinicians in the future. Although the second hypothesis was not verified, the data reported will generate future clinically relevant hypotheses.

Table of contents

Declaration of originality	2
Copyright declaration	3
Acknowledgements.....	5
Abstract.....	7
Table of contents	8
List of figures.....	14
List of tables	18
List of appendices	19
List of abbreviations.....	21
1. General background.....	25
1.1. Cystic fibrosis and the compromised pulmonary immune system.....	25
1.2. <i>Pseudomonas aeruginosa</i> infection.....	26
1.2.1. Acquisition and prevalence.....	27
1.2.2. Establishment of chronic infection in CF patients	28
1.2.2.1. Transition to a biofilm mode of growth.....	28
1.2.2.2. Development of resistance to antibiotics and to host immune factors	30
1.2.3. Treatment of infection and its consequences	30
1.3. Current diagnostic methods and limitations	31
1.3.1. Clinical samples	31
1.3.2. Traditional phenotypic assays for pathogen detection	32
1.3.3. Genotyping methods for speciation	33
1.3.4. Bacterial classification using mass spectrometry	33
1.4. Selected characteristic metabolites of <i>P. aeruginosa</i> as potential targets	34
1.4.1. Structural phospholipids	34
1.4.2. Bacterial signalling systems and quorum sensing quinolones.....	36
1.4.3. Surfactant rhamnolipids.....	38
1.4.4. Other virulence-associated molecules.....	39
1.4.5. Metabolic impact of <i>P. aeruginosa</i> infection on the host	41
1.5. Analytical techniques	41
1.5.1. Mass spectrometry	41
1.5.1.1. Reference methods.....	43
1.5.1.1.1. Electrospray ionisation for liquid solutions	43

1.5.1.1.2.	Matrix-assisted laser desorption ionisation: application to the analysis of bacteria and surfaces	44
1.5.1.2.	Ambient mass spectrometry methods.....	45
1.5.1.2.1.	Rapid evaporative ionisation mass spectrometry	45
1.5.1.2.2.	Desorption electrospray ionisation: analysis of surfaces	47
1.5.2.	Alternative methods under development for breath analysis	48
1.5.2.1.	Gas chromatography – mass spectrometry	48
1.5.2.2.	Real-time analysis of breath using mass spectrometry	49
1.5.2.3.	Electronic noses	50
1.5.2.4.	Exhaled breath condensate	50
1.5.3.	Data processing, statistical analysis and interpretation	51
1.5.3.1.	Data pre-processing	51
1.5.3.2.	Significance of the variables.....	52
1.5.3.3.	Unsupervised analysis for data visualisation	52
1.5.3.3.1.	Principal component analysis	52
1.5.3.3.2.	Hierarchical clustering analysis and heatmap	53
1.5.3.4.	Supervised analysis for classification	54
1.5.3.4.1.	Discriminant analysis	55
1.5.3.4.2.	Random Forest classification algorithm.....	55
1.5.3.4.3.	Validation of the modelling method.....	56
2.	Hypotheses, aim and objectives	61
3.	General methods	64
3.1.	Health, safety and ethical considerations.....	64
3.2.	Collection and culturing of microorganisms	64
3.3.	DNA extraction.....	65
3.4.	REIMS analysis.....	66
3.4.1.	MS instrument	66
3.4.2.	Ionisation sources	67
3.4.2.1.	Handheld bipolar probe	67
3.4.2.2.	Automated, laser-assisted REIMS	68
3.4.3.	Tandem MS analysis.....	70
3.4.4.	Data analysis	71
3.5.	DESI-MS analysis	72
3.5.1.	DESI-MS profiling	72

3.5.2.	Data analysis	73
4.	Identification of <i>P. aeruginosa</i> using REIMS	76
4.1.	Introduction	76
4.1.1.	Strain typing through DNA analysis	76
4.1.2.	The potential of mass spectrometry for strain typing	77
4.1.3.	Objectives.....	78
4.2.	Material and methods	78
4.2.1.	REIMS analysis.....	78
4.2.2.	MLST typing.....	79
4.3.	Results.....	80
4.3.1.	Metabolic characterisation of <i>P. aeruginosa</i>	80
4.3.2.	Classification of CF pathogens at the genus level	82
4.3.3.	Exploring REIMS capacities for <i>Pseudomonas</i> speciation	83
4.3.4.	Tentative classification of strain types.....	87
4.3.4.1.	Prediction of the isolate identity	87
4.3.4.2.	Classification according to MLST type.....	92
4.3.4.3.	Classification according to VNTR type	94
4.4.	Discussion.....	96
4.4.1.	REIMS as an alternative to MALDI-ToF for pathogen identification.....	96
4.4.2.	Detection of virulence factors and potential for a direct-from-sample diagnostic tool?	96
4.4.3.	REIMS metabolic fingerprinting provides some strain level information.....	97
5.	Metabolic phenotyping of <i>P. aeruginosa</i> using REIMS.....	100
5.1.	Introduction	100
5.1.1.	Adaptation of <i>P. aeruginosa</i> – genetic and phenotypic diversity.....	100
5.1.2.	Metabolic diversity of <i>P. aeruginosa</i> detected with REIMS.....	102
5.1.3.	Objectives.....	102
5.2.	Material and methods	103
5.2.1.	Comparing various types of infection	103
5.2.2.	Investigating <i>P. aeruginosa</i> diversity in CF patients	103
5.2.2.1.	Sample collection and REIMS analysis	103
5.2.2.2.	Phenotyping and MLVA.....	104
5.2.2.3.	Data processing and statistical analysis	105
5.3.	Results.....	105

5.3.1.	Respiratory isolates showed higher (between-patient) metabolic diversity than non-respiratory infections.....	105
5.3.2.	Investigating within-patient metabolic diversity of <i>P. aeruginosa</i> in CF	107
5.3.2.1.	Preliminary observations: the majority of phenotypic diversity occurred between patients with chronic infection	108
5.3.2.2.	Comparing metabolic variability of isolates within and between patients	109
5.3.2.2.1.	Clustering analysis showed metabolic homogeneity within patients and dissimilarity between patients.....	109
5.3.2.2.2.	Colony classification suggested high reproducibility for sampling within patients.....	113
5.3.2.2.3.	Conclusion: metabolic diversity was higher between than within patients....	114
5.3.2.3.	Can we predict early and chronic infection stages?	116
5.3.2.3.1.	Chronic isolates showed metabolic profiles with higher diversity	116
5.3.2.3.2.	Correlation networks revealed different synthetic pathways according to the infection stage.....	120
5.3.2.3.3.	Correlations were found between REIMS metabolic profile and the other phenotypes	124
5.4.	Discussion.....	127
5.4.1.	REIMS as a tool to explore and monitor <i>P. aeruginosa</i> metabolism and diversity.....	127
5.4.2.	<i>P. aeruginosa</i> produces different metabolic profiles at early and chronic stage of infection.....	128
5.4.3.	Limitations and perspectives	130
6.	Direct analysis of sputum and urine samples using REIMS	133
6.1.	Introduction	133
6.1.1.	Sputum composition	133
6.1.2.	Detection of metabolites characteristic of <i>P. aeruginosa</i> in biofluids.....	134
6.1.3.	Analysis of urine	135
6.1.4.	Objectives.....	136
6.2.	Material and methods	136
6.2.1.	Clinical samples	136
6.2.2.	REIMS analysis of clinical samples	137
6.2.2.1.	Sputum	138
6.2.2.2.	Urine.....	138
6.3.	Results.....	139

6.3.1.	Detection of <i>P. aeruginosa</i> through sputum analysis.....	139
6.3.1.1.	Sputum metabolic profile	139
6.3.1.2.	Heterogeneity of the sputum samples: impact on the metabolic profile	141
6.3.1.3.	Evaluation of the reproducibility of REIMS analysis on sputum samples.....	147
6.3.1.4.	Classification of infected sputum samples using REIMS.....	148
6.3.2.	Detection of <i>P. aeruginosa</i> through urine analysis.....	150
6.3.2.1.	REIMS analysis of urine deposited on filter paper	150
6.3.2.2.	REIMS analysis of pelleted urine	152
6.4.	Discussion.....	154
6.4.1.	Limitations and perspectives for diagnosis.....	154
6.4.2.	The potential of REIMS for the direct analysis of biofluids.....	155
7.	Analysis of skin secretions using DESI-MS: a feasibility study	157
7.1.	Introduction	157
7.1.1.	Skin secretions, lipids and analysis	157
7.1.2.	Fingermarks.....	158
7.1.3.	Objectives.....	158
7.2.	Method development	159
7.2.1.	Sample collection for DESI-MS analysis and processing	159
7.2.2.	Choice of the collection material	160
7.2.3.	Choice of the sampling site	163
7.2.4.	Sample collection – Impact of IPA cleaning	166
7.2.5.	Sample storage.....	169
7.2.5.1.	Impact of freezing	169
7.2.5.2.	Impact of freeze/thaw cycles.....	170
7.3.	Evaluation of the method	172
7.3.1.	Selection of the sampling parameters	172
7.3.2.	Metabolic profile of nose skin secretions	172
7.3.3.	Evaluation of reproducibility through classification of donors.....	173
7.4.	Analysis of skin secretions from CF patients.....	175
7.4.1.	Collection and analysis of clinical samples.....	175
7.4.2.	Classification of gender	177
7.4.3.	Classification of healthy subjects and CF patients.....	178
7.4.4.	Detection of <i>P. aeruginosa</i> infection through the analysis of skin secretions	180
7.5.	Discussion.....	182

7.5.1.	Evaluation of the method	182
7.5.2.	Clinical relevance	183
8.	General discussion	186
8.1.	Summary	186
8.1.1.	Achievements.....	186
8.1.2.	Strengths of the studies and techniques	187
8.1.3.	Limitations.....	188
8.2.	Future research.....	189
8.2.1.	Follow-up work	189
8.2.2.	Perspectives for <i>P. aeruginosa</i> infection diagnostics	190
	List of communications.....	193
	Oral presentations	193
	Poster presentations.....	193
	Publications.....	194
	References	196
	Appendices.....	210
	Permissions to reprint.....	234

List of figures

Figure 1.1. Compared FEV ₁ evolution for patients with and without <i>P. aeruginosa</i> infection according to age.	26
Figure 1.2. Lung infections according to age.	27
Figure 1.3. Chemical composition of Gram-negative bacteria cell walls.	35
Figure 1.4. Chemical structure of common glycerophospholipids.	35
Figure 1.5. Chemical structure of some signalling molecules.	37
Figure 1.6. Schematic representation of the four interconnected QS systems in <i>P. aeruginosa</i>	38
Figure 1.7. Chemical structure of the four families of rhamnolipids.	39
Figure 1.8. Chemical structure and molar mass of several virulence factors from <i>P. aeruginosa</i>	40
Figure 1.9. Example of a mass spectrum.	42
Figure 1.10. Electrospray ionisation mechanism.	44
Figure 1.11. Schematic representation depicting REIMS analysis on bacterial culture using bipolar forceps.	46
Figure 1.12. Schematic representation of a desorption electrospray ionisation source.	47
Figure 1.13. Example of a PCA score plot.	53
Figure 1.14. Example of a dendrogram.	54
Figure 1.15. Simplistic representation of a decision tree.	56
Figure 1.16. Example of a confusion matrix.	58
Figure 3.1. Picture of <i>P. aeruginosa</i> strains grown as a lawn on solid LB media for 24 h on a 24-well plate.	65
Figure 3.2. Schematic representation of the REIMS setup for bacterial analysis using a handheld bipolar probe.	67
Figure 3.3. Customised Tecan Freedom EVO platform for automated culture plate handling and REIMS analysis.	68
Figure 3.4. Schematic diagram of the REIMS setup.	69
Figure 3.5. Picture showing the analysis of samples in a 24-well plate using laser-assisted REIMS. ...	70
Figure 3.6. Pictures of the set-up for DESI-MS analysis.	73
Figure 4.1. Typical mass spectrum obtained from <i>P. aeruginosa</i> isolates using REIMS.	80
Figure 4.2. REIMS analysis of common CF pathogens.	83
Figure 4.3. PCA score plot using REIMS spectral data acquired on common CF pathogens and <i>Pseudomonas</i> species.	84

Figure 4.4. HCA using REIMS spectral data acquired on common CF pathogens and <i>Pseudomonas</i> species.....	84
Figure 4.5. HCA using REIMS spectral data acquired on <i>Pseudomonas</i> species.....	85
Figure 4.6. HCA and associated heatmap built using REIMS spectral data obtained from <i>Pseudomonas</i> isolates.	86
Figure 4.7. Confusion matrix for Random Forest classification using REIMS spectral data obtained from <i>P. aeruginosa</i> and other <i>Pseudomonas</i> species.....	87
Figure 4.8. Confusion matrix for Random Forest classification of replicates amongst 45 <i>P. aeruginosa</i> isolates.	89
Figure 4.9. HCA and associated heatmap showing clustering of 45 <i>P. aeruginosa</i> isolated from CF lungs.	91
Figure 4.10. Confusion matrix for Random Forest classification of 18 <i>P. aeruginosa</i> isolates amongst eight MLST types.....	93
Figure 4.11. Confusion matrix for Random Forest classification of nine <i>P. aeruginosa</i> VNTR types. ..	95
Figure 5.1. Example of mass spectra obtained from four <i>P. aeruginosa</i> isolates using REIMS.....	102
Figure 5.2. Statistical analysis of <i>P. aeruginosa</i> isolates from CF and non-CF patients.....	106
Figure 5.3. PCA score plot using REIMS spectral data acquired on 615 <i>P. aeruginosa</i> colonies collected from chronically and early infected patients.	110
Figure 5.4. Representation of the AUC characterising the distribution of PAO1 colonies in the global HCA.....	111
Figure 5.5. Example of dendrograms showing the distribution of the colonies for selected patients.	112
Figure 5.6. Confusion matrix for Random Forest classification according to patients using REIMS spectral data obtained from <i>P. aeruginosa</i> colonies.	114
Figure 5.7. Comparison of metabolic diversity between isolates collected from early and chronically infected patients.	117
Figure 5.8. Confusion matrix for Random Forest classification of isolates between early and chronic infection.	118
Figure 5.9. Boxplots showing the distribution of selected metabolites in early and chronic isolates.	119
Figure 5.10. Correlation networks for isolates collected from (a) early infection and (b) chronic infection.	121
Figure 5.11. Betweenness centrality scores for the 20 features with the highest difference between the early and chronic networks.	124

Figure 5.12. Heatmaps representing significant correlations between metabolic features and phenotypes for (a) early and (b) chronic isolates.	126
Figure 5.13. Simplistic and theoretical representation of metabolic pathways observed in (a) early and (b) chronic isolates.	129
Figure 6.1. Picture of a sputum sample spread on a CBA plate, after laser-REIMS-imaging at a 1 mm spatial resolution.	138
Figure 6.2. Compared mass spectra of (a & b) sputum samples infected by <i>P. aeruginosa</i> and (c & d) examples of <i>P. aeruginosa</i> colonies.	139
Figure 6.3. General chemical structure of sphingolipids.	141
Figure 6.4. Pictures of sputum samples spread on a CBA plate.	141
Figure 6.5. Pictures of sputum samples and ion images obtained with laser-REIMS showing global distribution of the signal.	142
Figure 6.6. Ion distribution of sputum features obtained from sputum samples imaged with laser-REIMS.	144
Figure 6.7. Ion distribution of bacterial features obtained from sputum samples imaged with laser-REIMS.	146
Figure 6.8. Confusion matrix for Random Forest classification of replicates obtained from the REIMS analysis of 36 CF sputum samples.	148
Figure 6.9. Boxplots showing the distribution of selected metabolites in sputum samples negative and positive for <i>P. aeruginosa</i> infection.	149
Figure 6.10. Comparative examples of mass spectra obtained with laser-REIMS analysis on urine deposited on (a) filter paper and (b) NC membranes.	151
Figure 6.11. Examples of mass spectra obtained with laser-REIMS analysis on urine pellets.	153
Figure 7.1. DESI-MS analysis of a skin secretion sample.	159
Figure 7.2. Example of the pre-processing of skin secretion data acquired by DESI-MS profiling.	160
Figure 7.3. Examples of data resulting from DESI-MS profiling of fingerprints deposited on (a & b) Polysine glass and (c & d) Plexiglass.	161
Figure 7.4. DESI-MS profiling analysis of skin secretions collected from the finger (a & b) and the bridge of the nose (c & d).	164
Figure 7.5. PCA score plots using DESI-MS profiling data acquired on skin secretions collected from the finger and from the nose.	165
Figure 7.6. <i>t</i> -tests showing differentiating features between finger and nose skin secretions.	165
Figure 7.7. Examples of mass spectra acquired by DESI-MS profiling on skin secretions before cleaning the skin (a) and after IPA cleaning (b).	166

Figure 7.8. PCA score plots using DESI-MS profiling data acquired on secretions collected without cleaning the skin and after IPA cleaning.	167
Figure 7.9. <i>t</i> -tests showing differentiating features between secretions collected without cleaning the skin and after IPA cleaning.....	168
Figure 7.10. PCA score plots using DESI-MS profiling data acquired on fresh and frozen skin secretion samples.	169
Figure 7.11. <i>t</i> -tests showing differentiating features between fresh samples and after freezing.	170
Figure 7.12. PCA score plots using DESI-MS profiling data acquired on skin secretion samples after several freeze/thaw cycles.....	171
Figure 7.13. ANOVA showing differentiating features depending on the number of freeze/thaw cycles.	171
Figure 7.14. Examples of typical mass spectra acquired in (a) negative and (b) positive ionisation modes, from fresh skin secretions collected on Superfrost Plus glass slides from the bridge of the nose after cleaning the skin with IPA.	173
Figure 7.15. Score plots showing unsupervised and supervised analysis of DESI-MS profiling data acquired in positive ionisation mode on skin secretions collected from 14 subjects.	174
Figure 7.16. Confusion matrix for donor classification using DESI-MS profiling data acquired on skin secretions.....	175
Figure 7.17. Example of a glass slide with edges covered by protective silicon tubing.	176
Figure 7.18. <i>t</i> -tests showing differentiating features between the skin secretions of males and females.	177
Figure 7.19. <i>t</i> -tests showing differentiating features between the skin secretions of healthy subjects and CF patients.	179
Figure 7.20. PCA score plots using DESI-MS profiling data acquired on skin secretions collected from CF patients negative and positive for <i>P. aeruginosa</i> infection.	180
Figure 7.21. <i>t</i> -tests showing differentiating features, acquired in positive ionisation mode, between CF patients negative and positive for <i>P. aeruginosa</i> infection.	181
Figure 7.22. <i>t</i> -tests showing differentiating features, acquired in negative ionisation mode, between CF patients negative and positive for <i>P. aeruginosa</i> infection.	181

List of tables

Table 3.1. Instrumental parameters for DESI-MS analysis with a Thermo Exactive instrument.....	72
Table 4.1. Some characteristics of classical DNA-based methods for strain typing.	77
Table 4.2. Some of the most intense spectral features identified in <i>P. aeruginosa</i>	81
Table 4.3. Random Forest cross-validation results for the classification of replicates amongst <i>P. aeruginosa</i> isolates.	88
Table 4.4. Random Forest cross-validation results for the classification of 18 <i>P. aeruginosa</i> isolates amongst eight MLST types.....	93
Table 4.5. Random Forest cross-validation results for the classification of nine <i>P. aeruginosa</i> VNTR types.....	94
Table 5.1 Summary of AUC and classification results per patient.	115
Table 6.1. Number of clinical samples collected from CF patients with infection status.....	137
Table 6.2. Random Forest cross-validation results for the classification of sputum samples according to infection status.	149
Table 6.3. Random Forest cross-validation results for the classification of urine samples deposited on filter paper according to infection status.	151
Table 6.4. Random Forest cross-validation results for the classification of urine samples deposited on NC membranes according to infection status.....	152
Table 6.5. Random Forest cross-validation results for the classification of urine pellets according to infection status.	153
Table 7.1. Summary of the intensities recorded by DESI-MS in positive ionisation mode on fingerprints deposited on six materials.	162
Table 7.2. Summary of the intensities recorded by DESI-MS in negative ionisation mode on fingerprints deposited on six materials.	162
Table 7.3. Numbers of donors recruited for the pilot clinical study using skin secretions analysed by DESI-MS profiling.	176
Table 7.4. Gender prediction using skin secretion profiles.	178
Table 7.5. Prediction of the CF condition using skin secretion profiles.....	179

List of appendices

Appendix 1. Culturing conditions for the isolates analysed with REIMS in Chapter 4.	210
Appendix 2. Origins of <i>Pseudomonas</i> isolates other than aeruginosa species analysed in Chapter 4.	210
Appendix 3. Identification of significant molecules based on MS/MS spectral data acquired with direct REIMS analysis on <i>P. aeruginosa</i> colonies.	211
Appendix 4. Comparison of mass spectra obtained using REIMS in negative ionisation mode on (a) non-mucoid and (b) mucoid isolates of <i>P. aeruginosa</i>	214
Appendix 5. Examples of mass spectra obtained from various <i>Pseudomonas</i> species using REIMS in negative ionisation mode.	215
Appendix 6. Analysis dates for the 10 replicates of the 45 <i>P. aeruginosa</i> isolated from CF lungs analysed in Chapter 4.....	216
Appendix 7. List and characteristics of colonies analysed in Chapter 5.	217
Appendix 8. HCA and associated heatmaps representing (a) non-respiratory, bronchiectasis and CF isolates and (b) bronchiectasis and CF isolates only; using data collected in Chapter 5.....	218
Appendix 9. Random Forest cross-validation results for the classification of <i>P. aeruginosa</i> colonies collected from 17 chronic patients (Cxx), seven early infections (Exx) and LB, PAO1 and PA14 as controls, using REIMS spectral data collected in Chapter 5.	220
Appendix 10. Random Forest cross-validation results for the classification of colonies collected from seven early infected and 17 chronic patients in Chapter 5.	220
Appendix 11. Identified features amongst the 150 most intense <i>m/z</i> bins, with ROC curve univariate analysis results for the differentiation of early and chronic infection in Chapter 5.....	221
Appendix 12. Compiled list of the features used to build both the early and chronic networks in Chapter 5.....	222
Appendix 13. List of sputum samples and number of replicates acquired with REIMS in Chapter 6, with infection status as verified by the detection of the genes 16S rRNA, <i>exoA</i> and <i>oprL</i>	224
Appendix 14. Random Forest cross-validation results for the classification of replicates obtained from the REIMS analysis of 36 CF sputum samples in Chapter 6.	225
Appendix 15. List of the 150 most intense bins detected in sputum samples by REIMS in Chapter 6, with AUC values obtained from ROC curve univariate analysis associated to the separation of patients infected and non-infected by <i>P. aeruginosa</i>	226
Appendix 16. Picture of pelleted urine in a 96-deep well plate (a), with associated sample identification in Table (b), analysed by REIMS in Chapter 6.	229

Appendix 17. Total ion count (TIC) and mass spectra (MS) for the fingerprint and the background obtained with DESI-MS profiling in both ionisation modes on the six materials tested in Chapter 7.	230
Appendix 18. Examples of mass spectra acquired by DESI-MS profiling in (a) positive and (b) negative ionisation mode, on skin secretions collected from the bridge of the nose of one donor.	232
Appendix 19. Score plots showing unsupervised and supervised analysis of DESI-MS profiling data acquired in negative ionisation mode in Chapter 7 on skin secretions collected from 14 subjects. .	233
Appendix 20. Confusion matrix for donor classification using DESI-MS profiling data acquired in negative ionisation mode on skin secretions in Chapter 7.	233

List of abbreviations

3-oxo-C12-HSL	<i>N</i> -3-oxo-dodecanoyl-homoserine lactone
AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
AQ	Alkyl-quinolone
ASL	Airway surface liquid
AUC	Area under the curve
BAL	Bronchoalveolar lavage
C11:db-PQS	2-undecenyl-3-hydroxy-4(1 <i>H</i>)-quinolone
C11:db-UHQ	2-undecenyl-4(1 <i>H</i>)-quinolone
C11:db-UQNO	2-undecenyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide
C11-PQS	2-undecanyl-3-hydroxy-4(1 <i>H</i>)-quinolone
C11-UQNO	2-undecanyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide
C4-HSL	<i>N</i> -butyryl-homoserine lactone
C7:db-HHQ	2-heptenyl-4(1 <i>H</i>)-quinolone
C7:db-HQNO	2-heptenyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide
C7:db-PQS	2-heptenyl-3-hydroxy-4(1 <i>H</i>)-quinolone
C9:db-NHQ	2-nonenyl-4(1 <i>H</i>)-quinolone
C9:db-NQNO	2-nonenyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide
C9:db-PQS	2-nonenyl-3-hydroxy-4(1 <i>H</i>)-quinolone
C9-NQNO	2-nonyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide
C9-PQS	2-nonyl-3-hydroxy-4(1 <i>H</i>)-quinolone
CBA	Colombia blood agar
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CID	Collision-induced dissociation
COPD	Chronic obstructive pulmonary disease
cyclic-di-GMP	Cyclic-diguanosine monophosphate
DA	Discriminant analysis
DDH	DNA-DNA hybridization
DEFFI	Desorption electro-flow focusing ionisation
DESI	Desorption electrospray ionisation
DG	Diglyceride
diglyco-Cer	Diglyco-ceramide
DNA	Deoxyribonucleic acid
DPG	Diphosphatidylglycerol
EBC	Exhaled breath condensate
ELF	Epithelial lining fluid
EIC	Extracted ion count
ELISA	Enzyme-linked immunosorbent assay
ENaC	Epithelial sodium channel
EPS	Exopolysaccharide
ESI	Electrospray ionisation
ExoA	Exotoxin A
FA	Fatty acid
FAIMS	Field asymmetric ion mobility spectrometry
FDR	False discovery rate
FEV ₁	Forced expiratory volume in 1 second
Fgp	Fingerprint
FT-ICR	Fourier transform-ion cyclotron resonance detector
GC	Gas chromatography
HCA	Hierarchical clustering analysis
HCN	Hydrogen cyanide
HHQ	2-heptyl-4(1 <i>H</i>)-quinolone
HQNO	2-heptyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide
HSL	Homoserine lactone
IMS	Ion mobility spectrometry

IPA	Isopropanol
IQS	Integrated quorum sensing signal
<i>k</i> -nn	<i>k</i> -nearest neighbour classifier
LB	Lysogeny broth
LC	Liquid chromatography
LecA	Lectin A
LES	Liverpool epidemic strain
LeuEnk	Leucine enkephalin
log-transform	Logarithmic transformation
LPG	Lysophosphatidylglycerol
LPS	Lipopolysaccharides
<i>m/z</i>	Mass over charge ratio
MALDI	Matrix-assisted laser desorption ionisation
MG	Monoglyceride
MLST	Multilocus sequence typing
MLVA	Multiple loci variable-number tandem repeats
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSI	Mass spectrometry imaging
NC	Nitrocellulose
NHLI	National Heart and Lung Institute
NHQ	2-nonyl-4(1 <i>H</i>)-quinolone
NHS	National health service
NMR	Nuclear magnetic resonance
OMB	Offline model builder
PAMP	Pathogen-associated molecular pattern
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PFGE	Pulsed-field gel electrophoresis
PG	Phosphatidylglycerol
PL	Phospholipid
PLS-DA	Partial least square-discriminant analysis
PPV	Positive predictive value
PQS	<i>Pseudomonas</i> quinolone signal; 2-heptyl-3-hydroxy-4(1 <i>H</i>)-quinolone
PTFE	Polytetrafluoroethylene
PTR	Proton transfer reaction
PYO	Pyocyanin
Q	Quadrupole
QS	Quorum sensing
REIMS	Rapid evaporative ionisation mass spectrometry
Rha-Cn	Mono-rhamno-mono-lipidic congeners; when <i>n</i> =10: α -L-rhamnopyranosyl- β -hydroxydecanoate
Rha-Cn-Cm	Mono-rhamno-di-lipidic congeners; when <i>m</i> = <i>n</i> =10: α -L-rhamnopyranosyl- β -hydroxydecanoyl- β -hydroxydecanoate
Rha-Rha-Cn	Di-rhamno-mono-lipidic congeners; when <i>n</i> =10: α -L-rhamnopyranosyl- α -L-rhamnopyranosyl- β -hydroxydecanoate
Rha-Rha-Cn-Cm	Di-rhamno-di-lipidic congeners; when <i>m</i> = <i>n</i> =10: α -L-rhamnopyranosyl- α -L-rhamnopyranosyl- β -hydroxydecanoyl- β -hydroxydecanoate
rmcorr	Repeated measures correlation
RMMC	Recursive maximum margin criterion
ROC	Receiver operating characteristic
ROS	Reactive oxygen species
rRNA	Ribosomal ribonucleic acid
SESI	Secondary electrospray ionisation
SIFT	Selected ion flow tube
SPME	Solid-phase micro-extraction
SRC	Strategic research centre
TG	triglyceride
TIC	Total ion count

TLC	Thin layer chromatography
TNR	True negative rate
ToF	Time-of-flight
TPR	True positive rate
UHQ	2-undecanyl-4(1 <i>H</i>)-quinolone
UK	United Kingdom
VNTR	Variable-number tandem repeats
VOC	Volatile organic compound
WGS	Whole genome sequencing

Chapter 1

General Background

1. General background

1.1. Cystic fibrosis and the compromised pulmonary immune system

Cystic fibrosis (CF) is the most common genetic disease affecting the Caucasian population, with more than 10 000 patients reported in 2017 by the United Kingdom (UK) CF Registry and between 200-300 new cases every year in the UK (i.e. around 1 in 3000 newborns) [1]. It is an autosomal recessive disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. This gene encodes the multifunctional CFTR protein, which acts primarily as a channel for active secretion of chloride and bicarbonate ions across epithelial cell membranes. It also impacts other ion transport channels, notably the epithelial sodium channel (ENaC), which drives sodium absorption. Defective CFTR protein leads to an ion imbalance; chloride secretion is disrupted whilst sodium absorption increases, causing a secondary osmotic uptake of water and therefore the dehydration of epithelial surface fluids, e.g. respiratory mucus and digestive secretions [2]. Meanwhile, lack of bicarbonate excretion causes a pH decrease at the mucosal surface [3].

These phenomena result in impaired host defence in the airways and, whilst the digestive, reproductive and other systems are severely impacted, most of the morbidity and mortality associated with CF is due to its impacts on the respiratory system. The dehydration of airway surfaces results in fluid of increased viscosity, which impairs mucociliary clearance, one of the lung's innate defence mechanisms. Mucus accumulates in the respiratory tract, leading to respiratory distress, and airway secretions are not cleared, allowing infection to be initiated and persist. In addition, there is evidence that a reduced pH inhibits the activity of antimicrobial peptides in the airway surface liquid [3]. Also, an excessive inflammatory response to microbiological infection contributes to impaired airway defences and irreversible lung damage [4, 5]. Indeed, another hallmark of CF is the massive influx of neutrophils into the airways in response to infection. Not only do they fail to clear microbes efficiently [6], neutrophils and their products also constitute a major contributor to the tissue damage in CF patients. They secrete elastase [5, 7] and promote the production of cytokines [8] and reactive oxygen species (ROS), causing inflammation which overwhelms the physiological anti-inflammatory and antioxidant mechanisms [6]. The ability of the neutrophils to secrete antimicrobial molecules is also impaired, e.g. hypochlorous acid decreased production of which is notably due to the loss of CFTR function [9].

In 2017, median age at death for people with CF was in the 30s in the UK [1] and respiratory failure was the most common cause of mortality. In a retrospective study, Kerem et al. analysed lung function data collected from around 15 000 patients, stored in the 2007 database of the European CF Society

Patient Registry [10]. They concluded that the main factors associated with decline in lung function (as measured by forced expiratory volume in 1 second (FEV₁)) were a low body mass index and chronic airway infection with *Pseudomonas aeruginosa* (*P. aeruginosa*). This was in agreement with previous research, e.g. Kosorok et al., whose longitudinal study revealed that the acquisition of *P. aeruginosa* correlated with an accelerated and gradual decline in pulmonary function in children [11]. Jarad et al. had also further shown a systematic negative impact of *P. aeruginosa* infection on FEV₁, independent of the age of the patient (see Figure 1.1) [12].

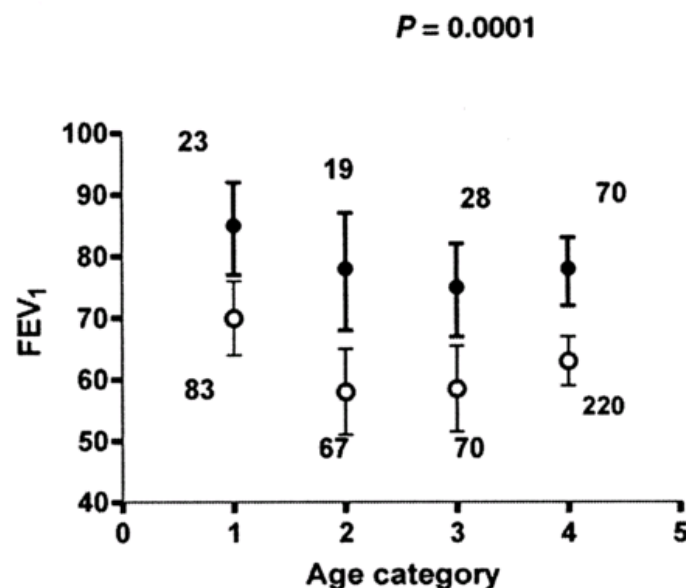


Figure 1.1. Compared FEV₁ evolution for patients with and without *P. aeruginosa* infection according to age. Shows mean and 95% confidence interval; open and closed circles representing infected and non-infected patients, respectively. Age: (1) 16-19.9 years, (2) 20-24.9 years, (3) 25-29.9 years, (4) above 30 years. Numbers represent the number of patients in each group (reproduced with permission from Ref. [12]).

In summary, CF is a life-shortening inherited condition, with most of the morbidity and mortality related to decline in respiratory health. *P. aeruginosa* is a crucial factor in disease progression from infancy and impacts life expectancy.

1.2. *Pseudomonas aeruginosa* infection

P. aeruginosa is a Gram-negative bacterium, abundant in the natural environment though readily eliminated from the healthy host. Yet, it is responsible for acute and chronic infections in a number of clinical scenarios including acute infections in intubated patients, those with burns, cancer, patients being treated with immunosuppressive drugs and chronic respiratory infections in CF and other forms of chronic suppurative lung disease [13]. *P. aeruginosa* produces a range of extracellular compounds which contribute its virulence (e.g. exotoxin A, lectin A, pyocyanin, pyoverdine, hydrogen cyanide,

alginate, rhamnolipids, enzymes)[14]. It possesses one of the largest genomes in the bacterial kingdom [15], 10% of which encodes regulators allowing environmental adaptation through versatile metabolic mechanisms [16], whereby the bacterium is able to adjust to varying environmental constraints, nutrient and oxygen availability, host immune factors and bacterial competitors. This means it can very readily lead to chronic infection of the CF airways.

1.2.1. Acquisition and prevalence

The UK CF registry reports that until adolescence, *Haemophilus influenzae* and *Staphylococcus aureus* are the most common pathogens infecting the airways of CF patients, affecting up to 35% and 28% of children, respectively (Figure 1.2). Nevertheless, 20% of children below 3 years of age have already suffered a first infection with *P. aeruginosa* and many other studies also suggest that first acquisition occurs very early in life [17-19]. The first infecting *P. aeruginosa* most likely comes from environmental sources [18]. Early treatments are often effective since the infecting strains are usually neither resistant to antibiotics nor mucoid – a phenotype overexpressing alginate, which constitutes a protection against host defences [17, 18, 20, 21]. Such an early acquisition of infection is critical since existing diagnostics are difficult to apply to infants; yet early detection offers the best hope of successful elimination of the bacteria.

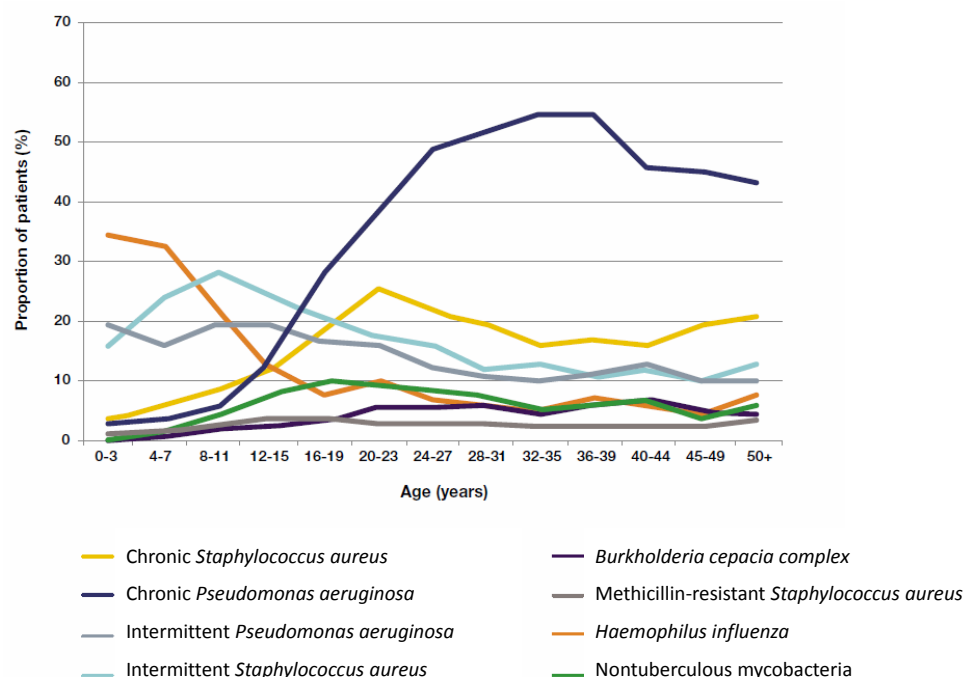


Figure 1.2. Lung infections according to age.

Reproduced with permission from the UK Cystic Fibrosis Registry, Annual Data Report 2016, published August 2017 [22].

With increasing age, intermittent infections eventually become chronic, defined as more than half of respiratory bacteriological samples positive for infection over the preceding 12 months [23]. *P. aeruginosa* then becomes dominant in the lung microflora and up to 55% of adults with CF are chronically infected (Figure 1.2).

Cross-infection is a genuine risk between patients attending the same CF clinic and through social contact (although now minimised with modern isolation precautions), or even between patients apparently unrelated, since routes of transmission remain poorly understood [24, 25]. Several studies have demonstrated that cross-infection is critical, especially in the spread of highly transmissible epidemic strains [26, 27]. These findings have motivated the reinforcement of infection control measures through patient segregation, molecular surveillance and hospital configuration. Furthermore, it has brought to light the necessity to better detect and understand these strains, at a genetic and metabolic level, in order to initiate appropriate and timely clinical responses and to understand whether patient to patient transmission has occurred.

1.2.2. Establishment of chronic infection in CF patients

The development of chronic infection with *P. aeruginosa* is enabled by a number of bacterial adaptive phenomena, including both intrinsic and acquired antibiotic resistance, which makes the infection extremely difficult to eradicate once established, and thus a major factor in morbidity and mortality [28]. This shift to chronicity is driven by the evolution of the pathogen in its environment – i.e. the airway – and its capacity to evade the immune system and chemotherapy. *P. aeruginosa* achieves this primarily by switching from a planktonic to a protective biofilm mode of growth allowing it to survive in hostile environments [29]. This is usually accompanied by a switch in motility mode (swimming, twitching or swarming), membrane modifications, and down-regulation of virulence factors. Many of these changes are made possible by the highly versatile genome of *P. aeruginosa* and are controlled by at least four quorum sensing (QS) systems, which control cell-to-cell communication and coordinate gene expression through the production of small signalling molecules [30, 31].

1.2.2.1. Transition to a biofilm mode of growth

Biofilms are elaborate communities of bacteria living in a sessile manner in a self-secreted matrix, containing exopolysaccharides (EPS), deoxyribonucleic acids (DNA) and proteins, providing protection

against the innate and adaptive host immune responses. Alginate, Pel and Psl are major EPS constituents of the matrix and there is evidence showing that this matrix contributes to hindering the action of host immune factors such as phagocytes [32]. Furthermore, alginate triggers host inflammation by inducing oxidative stress and production of pro-inflammatory cytokines [32]. It is an oxygen-radical scavenger, thereby protecting the bacterial cells from ROS liberated during phagocytosis [33]. Overproduction of alginate is often associated with the development of mucoid phenotypes, which hinder mucociliary clearance, thereby further impeding lung defence [33]. *P. aeruginosa* also release surfactants, namely rhamnolipids, to ensure the chemical stability of the biofilm matrix; they also contribute to neutrophil elimination [34]. Bacteria evolving in biofilms are characterised by a decrease in motility – notably swimming and swarming – due to the loss of flagella. Its main constituent protein, flagellin, is a pathogen-associated molecular pattern (PAMP) which triggers the release of pro-inflammatory cytokines and neutrophil recruitment. Therefore, the loss of the flagella is thought to help *P. aeruginosa* evade the immune system [35, 36]. Further chemical modifications occur at the bacterial surface, impacting lipopolysaccharide (LPS) composition, to counteract host immune responses. LPS are PAMP glycolipids constituting the external layer of the outer membrane of Gram-negative bacteria. These endotoxins play a crucial role in bacterial virulence and protection from immune biocidal factors. They are composed of an O-antigen tail (or O-polysaccharide), a core polysaccharide, and lipid A, which anchors the molecule to the outer membrane of Gram-negative bacteria (see Figure 1.3). Lipid A is the element of LPS which confers most of the pro-inflammatory activity and resistance to host innate immunity [37]. The varying composition of LPS increases the chemical asymmetry in the outer membrane of the bacteria, which impacts the global permeability of the cell membrane. The outer polysaccharide tail renders the bacterial surface lipophobic, whilst cross-linking between lipid A molecules within the membrane through divalent cations protects the cell from hydrophilic antimicrobials [38]. During chronic infection, the lipid A composition adjusts to environmental conditions [37], whilst the loss of O-antigen allows the bacterial cell to evade the host immune system [36, 39]. Biofilm growth is also accompanied by the upregulation of tissue degrading enzymes, such as alkaline protease and elastase, which also possess the capacity to inactivate components of the immune response (e.g. complements and antimicrobial peptides) [36, 40, 41]. Furthermore, positively charged antimicrobial peptides produced by the host bind to extracellular DNA, secreted in the biofilm by the bacteria, preventing them from attaching to bacterial cell walls and performing their effector functions [42].

1.2.2.2. Development of resistance to antibiotics and to host immune factors

Biofilms thus provide a physical and biochemical shield against host and synthetic antimicrobials; they also play an important role in antibiotic resistance. Several hypotheses have been proposed to explain why antibiotics are less efficient on biofilms than on planktonic cells; they mainly rely on the higher bacterial density in biofilms. Mechanisms include biochemical hindrance slowing down antibiotic penetration, localised modifications of the chemical conditions (e.g. pH) within the biofilm impeding the antibiotics' action, the development of a highly protected phenotype unique to biofilms, and the promotion of resistance proliferation through horizontal gene transfer, the kinetics of which are favoured by high cell density [43-46].

The spread of antibiotic resistance relies on the selection of genetic mutations under evolutionary pressures, which occur during prolonged treatments and is usually accompanied by changes in bacterial membranes, enzyme production and QS expression. Mechanisms to eliminate biocidal molecules are implemented [47] through the building of evacuation channels across membranes, such as porins and multi-drug efflux pumps, driven by genetic mutations during adaptation. The multidrug-efflux pumps are amongst the most common mutational targets [48], for instance, the loss of function in the *mexZ* gene, which contributes to the increased synthesis of the MexXY-OprM multidrug-efflux pump [49]. This pump provides the bacteria with higher resistance to aminoglycoside antibiotics [49]. In parallel, there is an increased production of enzymes such as β -lactamases, which catabolise biocidal molecules such as some penicillins, cephalosporins, cephamycins, and carbapenems [50].

1.2.3. Treatment of infection and its consequences

Antibiotics are still the most common treatment for *P. aeruginosa*, even though the optimal strategy remains unclear. They are thought to be more efficient during initial/early infection and there is good evidence that they can delay the emergence of chronicity [20, 21]. Early infection is treated with nebulised antibiotics, sometimes combined with oral or intravenous therapy. During pulmonary exacerbations, patients are usually administered intravenous antibiotics. Once chronicity is established, long-term "suppressive" approaches are commonly used, namely single or rotating nebulised antibiotics, and occasionally regular elective courses of intravenous antibiotics. However, there are multiple adverse effects of long-term and intensive use of antibiotics. In addition to the potential development of drug intolerances and toxicity in the patient, it may eventually result in the dominance of one or a limited number of resistant bacterial strains. Indeed, antibiotics disrupt the microflora balance and may kill the co-colonising organisms [51], stimulate genetic mutations and

favour the selection of resistant strains [52]. This is one explanation for the lower microbial diversity in the adult CF lung and the emergence of highly resistant strains, nearly impossible to eradicate.

Most patients eventually suffer from cycles of chronic inflammation and airway obstruction, which cause ongoing and irreversible lung damage. The treatment of early infections can be successful, however these usually occur early in life. Thus, appropriate methods for timely detection (particularly in non-cooperative babies and infants) and treatment are of paramount importance to prevent the establishment of chronicity and the transmission of epidemic strains.

1.3. Current diagnostic methods and limitations

Constant progress is being made towards understanding *P. aeruginosa* metabolism and innovative detection methods are continually being developed. Yet, studies have remained unsuccessful in identifying consistent markers of infection in CF patients, and diagnosis today still relies on isolation and culture of the pathogen from respiratory samples.

Firstly, it is worth defining some key terminology used throughout this thesis: samples of pure bacterial cells are referred to as isolates. Bacteria are designated by their genus and species phylogeny. At a subspecies level, strains differ from each other phenotypically and genetically. Clones are theoretical genetic copies of an isolate, generated through asexual reproduction, thus belonging to the same strain as their parent. However, random mutations are frequent and clones rarely share the exact same genome [53].

1.3.1. Clinical samples

Screening for airway infection is part of the routine surveillance of people with CF. There are a number of samples, with different attributes, which are currently used in clinics to detect infection.

Bronchoalveolar lavage (BAL) is thought to best reflect the lower airway flora, and is thus considered the reference sample for detecting infections. However, the procedure is invasive, requires a general anaesthetic (at least in children) and cannot be performed frequently [54]. Furthermore, bacterial distribution in the lungs can be spatially heterogeneous, potentially leading to false negatives when assessing one or two lobes only [55]. Sampling of multiple lobes is essential if significant bacteria are not to be missed, but this is not usual practice [56].

Spontaneously expectorated sputum is currently the most commonly collected sample in older children or adults. It provides a good means of detection of *P. aeruginosa* infection in patients who can expectorate spontaneously [57]. However, many patients, especially young children and those at milder stages of the disease, will not be able to produce sputum. Sputum induction with the inhalation of nebulised hypertonic saline solution can be performed in order to stimulate airway mucus production. Comparisons of the sensitivity of sputum induction with BAL are encouraging [56, 58, 59]. However, the procedure is time-consuming (and therefore expensive), unpopular with patients, and some studies still debate its reliability [60, 61]. Finally, strict infection control measures are required during sputum induction.

Still, these procedures are not generally used in babies and infants, in whom the detection of infection is critical. Existing alternatives use oropharyngeal (cough and throat) swabs, which lack sensitivity and specificity compared to BAL, sputum and sputum induction [56, 62, 63]. Besides, there is evidence suggesting that the upper airway microflora might not reflect the lower airway bacterial population, and bacteria can be underrepresented [63]. Hence, there is a real need for novel non-invasive samples, representative of the infection status and easily collected from young children.

1.3.2. Traditional phenotypic assays for pathogen detection

Following collection, samples are typically processed for phenotypic analyses. Culturing on selective media allows differential isolation of pathogens, and further biochemical and growth assays may be carried out for microbial characterisation, e.g. antibiotic resistance testing and sugar utilisation. In addition to the fact that microbiological assays require significant user input and skills, they have several pitfalls. There is a risk of overlooking viable but non-cultivable cells, and the high detection threshold is not adapted for the detection of low bacterial densities found at the early stage of infections [64]. In addition, misidentification can occur due to the high diversity of *P. aeruginosa* phenotypes and modes of growth [44, 65]. Besides, the whole process of isolation, purification and testing requires time for the bacteria to grow, further delaying the administration of the appropriate treatment of several days.

1.3.3. Genotyping methods for speciation

Bacterial classification was traditionally performed based on morphology and phenotypic characteristics, but DNA analysis has slowly but profoundly modified approaches to taxonomy. Since the 1960s, DNA-DNA hybridization (DDH), a non-sequencing method, has been used in molecular biology to measure the global degree of genetic similarity between two organisms. A minimum of 70% similarity in the DNA of strains is commonly accepted to define them as part of the same species [66].

Sequencing of the 16S ribosomal ribonucleic acid gene (16S rRNA) is currently considered the reference method for species identification [65]. The gene is present in all bacteria and archaea and participates in coding of the 16S ribosomal subunit. The regions coding for the binding sites are particularly stable throughout evolution and changes in this sequence are likely to be random mutations rather than due to bacterial adaptation. The hypervariable sequences in between are highly specific and differentiate species of prokaryotes, making them suited for phylogeny and taxonomy [66, 67]. Genera share above 95% sequence identity in the 16S rRNA gene and the species threshold is currently around 99% [68]. However, some strains (e.g. within the *Mycobacterium* genus) can sometimes share up to 100% similarity in the 16S rRNA sequence when their global DNA relatedness is below 50% (according to DDH) – implying they do not belong to the same species. This means that, for some species, 16S rRNA sequencing alone is not an adequate tool for speciation [66]. Whole genome sequencing (WGS) remains the ultimate means of biological identification. However facilities, costs, user skills, and time requirements for DNA sequencing preclude its use as a routine clinical test.

Therefore, approaches targeting the detection of species-specific part of the genome are being explored, e.g. the *exoA* and *oprL* genes. These genes are specific to *P. aeruginosa* and code respectively for the enzyme exotoxin A and the lipoprotein OprL. The exotoxin A is one of the most toxic enzymes and is produced by a majority of *P. aeruginosa* strains whilst OprL is a constituent of the bacterial outer membrane which contributes to cell permeability and transmembrane systems [69]. Both genes have been previously reported as reliable markers for *P. aeruginosa* [69-72].

1.3.4. Bacterial classification using mass spectrometry

During the last decade, mass spectrometry technologies have increasingly been employed in clinical laboratories for the identification of pure bacterial colonies. The main clinical platforms are the Biotyper (Bruker Daltonics) and the Vitek MS (bioMérieux), which use matrix-assisted laser desorption ionisation (MALDI) coupled to a time-of-flight (ToF) mass spectrometer. Despite a significant initial

cost, they have the advantages of being high-throughput and user-independent. They are recognised as a standard method which accurately classifies *P. aeruginosa*; yet other genera, e.g. from the *Enterobacteriaceae* family, *Acinetobacter* species or *Listeria* species, are still problematic and global accuracy is 80-90% at the species level [73].

The above methods differ in terms cost, reliability, and time requirement; yet, they all rely on the preliminary isolation of pure bacterial cells from the original samples, which remain a critical and limiting factor for the overall sensitivity of the method. In addition to their questionable quality and issues inherent to collection especially in young children, respiratory samples may not allow the detection of low bacterial loads as could be expected in the case of early infection. This means that the early stages of an infection may frequently be missed and entails two serious consequences: the delayed administration of treatment, which is potentially less likely to be successful, and the increased opportunity for infection transmission between patients. New strategies are needed to detect and monitor the infection in babies and at early stages; thus, non-invasive samples and quicker culture-free methods of analysis are under investigation.

1.4. Selected characteristic metabolites of *P. aeruginosa* as potential targets

An alternative to the isolation of the microorganisms, requiring respiratory samples and time-consuming culturing, lies in the detection of specific biomarkers, which would reveal the presence of *P. aeruginosa* in the body. Here are discussed a number of bacterial molecules, important in understanding *P. aeruginosa* metabolism and pathogenicity, that may represent relevant biomarker targets. They are well-characterised, specific to *P. aeruginosa* and/or have been previously studied in infected patients.

1.4.1. Structural phospholipids

In Gram-negative bacteria such as *P. aeruginosa*, cell walls are composed of three layers: an inner cell membrane constituted of 40% phospholipids arranged as a bilayer, a peptidoglycan layer in the periplasm, and an asymmetric outer membrane composed of 25% phospholipids plus LPS, connected via their lipid A head (Figure 1.3) [74]. The latter is characteristic of Gram-negative bacteria and constitutes an additional permeability barrier [38], which contributes to the intrinsic antibiotic

resistance of *P. aeruginosa*. Both membranes contain a number of embedded proteins and porins, which allow the passive diffusion of molecules such as nutrients and substrates or waste and toxins.

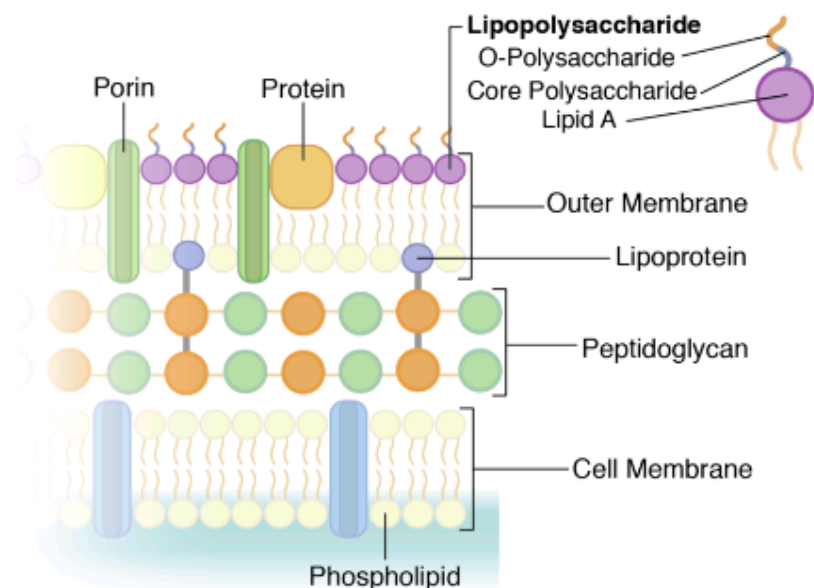


Figure 1.3. Chemical composition of Gram-negative bacteria cell walls.

(Reproduced from online open source: Linda Bruslind, <http://library.open.oregonstate.edu/microbiology/>).

The membrane chemistry is a determinant for cell survival, since the ability to adjust membrane composition allows bacteria to survive in adverse environments. Fluidity, in particular, is key in maintaining physiological activity while impeding biocidal molecules from penetrating the cell, and is determined by the phospholipid composition. Phospholipids (PLs) are the most common lipids in bacteria, constituting 90% of the cellular lipids of *Pseudomonas* species [74]. Phosphatidylethanolamines (PEs) are the most abundant [75] but membranes also contain phosphatidylglycerols (PGs), diphosphatidylglycerols (DPGs), lysophosphatidylglycerols (LPGs) and phosphatidylcholines, which can be synthesised by methylation of PEs [74, 76] (see Figure 1.4).

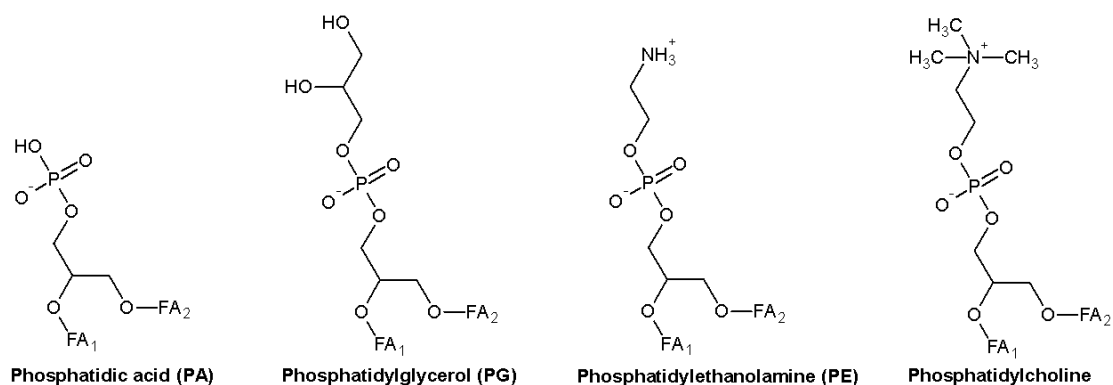


Figure 1.4. Chemical structure of common glycerophospholipids.

FA₁ and FA₂ stand for fatty acid groups; in the case of lysophospholipids, FA₁ is replaced by hydrogen.

Phospholipid nomenclature in this thesis is "PL(FA₁:n₁/FA₂:n₂)" where "PL" is replaced by the PL type (for instance PE, PG, etc.), and "n" corresponds to the number of double-bonds in the associated fatty acid (FA).

It has been suggested that *P. aeruginosa* growing in a biofilm adjusts even its inner membrane composition in order to conserve energy compared to planktonic cells [77]. Membranes were found to have phospholipids with fewer odd-numbered and branched FA, while being richer in long FA chains. This would lead to a decrease in membrane fluidity and stabilise the lipid bilayer, thereby slowing down metabolic processes [78].

1.4.2. Bacterial signalling systems and quorum sensing quinolones

The modifications occurring in *P. aeruginosa* during adaptation and establishment of chronicity, including the development of biofilm [30, 79, 80] and antibiotic resistance [81, 82], are controlled by complex signalling pathways. Bacteria adapt to population density and to their environment using small signalling molecules, namely QS molecules and cyclic-diguanosine monophosphate (cyclic-di-GMP), respectively [83]. QS is a cell-to-cell communication system relying on each cell producing QS molecules. The bacterial community uses these molecules to evaluate population density and cooperatively trigger gene expression accordingly. Cyclic-di-GMP is an intracellular second-messenger released in response to extracellular stimulation, such as local environmental conditions (e.g. light, oxygen, nitric oxide, arginine levels) and cell density as signalled by QS. Cyclic-di-GMP works conjointly with the QS systems, to coordinate gene expression, modulate phenotypes, and trigger the transition to biofilm, which is accompanied by a decrease in motility and virulence, and leads to the establishment of antibiotic and host-defence resistant populations [83]. In *P. aeruginosa*, the QS network involves at least four interconnected and self-regulated systems: *las*, *rhl*, *pqs* and *iqs* [31] (see simplified network summarised in Figure 1.6). It relies on the production of two main chemical families, acylated homoserine lactones (HSLs) (*las* and *rhl* systems) and alkyl-quinolones (AQs) (*pqs* system) (see chemical structures in Figure 1.5), although other QS molecules are being discovered. The integrated quorum sensing signal (IQS), or 2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde, belongs to a distinct class of intercellular signalling compounds that has been recently reported (*iqs* system) [84]. It has been shown to contribute to the virulence of *P. aeruginosa* and to impact the production of HSLs and AQs [31]. All together, they play an important role in regulating virulence factor production [80] such as elastase [85], pyocyanin [86] and lectin A [87], and thus impact inflammatory and immune responses in the host [88].

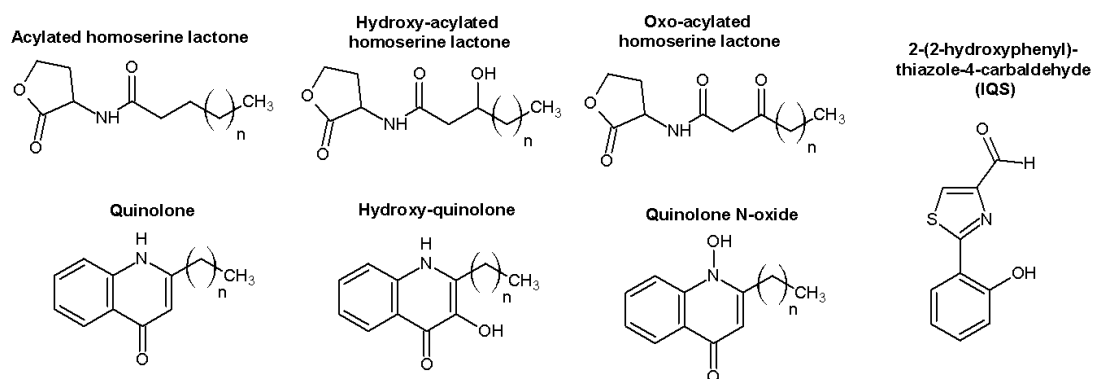


Figure 1.5. Chemical structure of some signalling molecules.

Acylated homoserine lactones, quinolones, and integrated quorum sensing signal (IQS).

Quinolones are synthesised from fatty acids and aromatic anthranilate, via the tryptophan metabolic pathway [89, 90]. They have historically been studied for their antibacterial properties and some quinolones produced by *P. aeruginosa* indeed exhibit antibiotic activity against Gram-positive bacteria such as *S. aureus*, another common CF pathogen [91]. The *Pseudomonas* quinolone signal (PQS), or 2-heptyl-3-hydroxy-4-quinolone, is one of the most studied quorum sensing quinolones. Notwithstanding the global collaborative functioning of the QS network, PQS in particular is known to participate in many bacterial metabolic processes, such as iron uptake [92] and membrane vesicle formation [93], which allows the transport of hydrophobic metabolites (including itself). It also promotes oxidative stress [94] and denitrification [95], and can provoke cell autolysis in stressful situations [96]. PQS and other quinolones have been detected in infected CF sputum, plasma, urine and BAL fluid [97, 98] and level of PQS in the airways seems to be higher at early stages of infection [99]. Quantification in liquid cultures using liquid chromatography coupled to mass spectrometry (LC-MS) has suggested that PQS is in fact produced during the logarithmic phase of growth [100]. The QS systems are common mutational targets in chronically infecting populations of *P. aeruginosa*, affecting notably *lasR*. Mutations mostly involve loss-of-function and thereby, a decrease in global virulence [48].

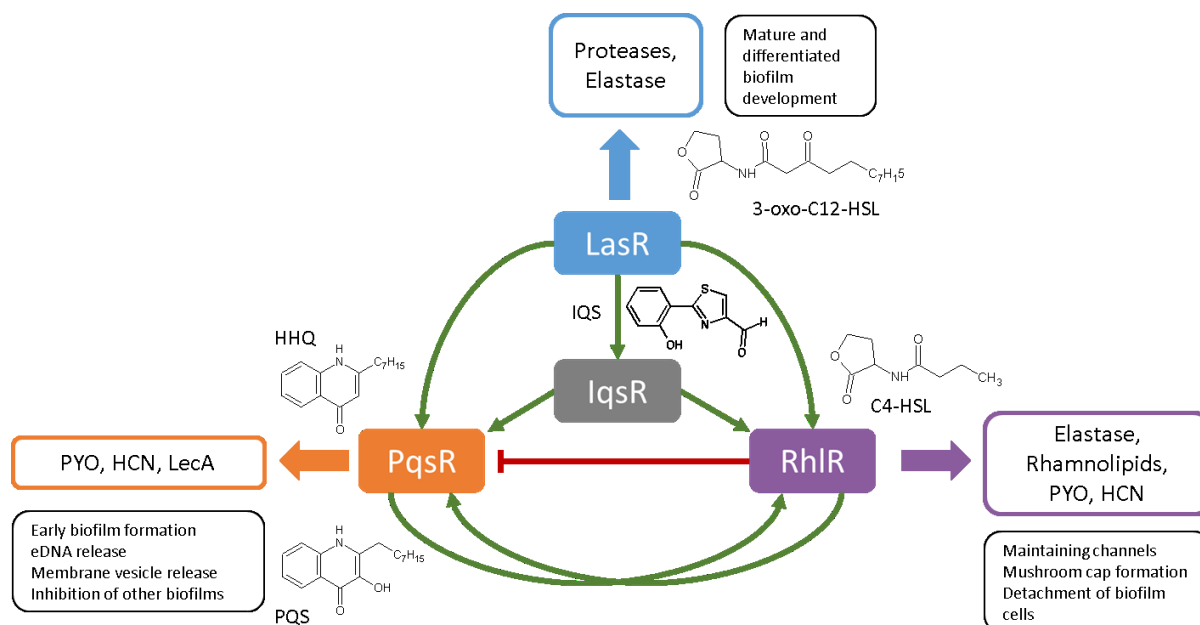


Figure 1.6. Schematic representation of the four interconnected QS systems in *P. aeruginosa*.

lasR, rhIR, pqsR and iqsR, and associated signalling molecules and virulence factors. Positive control is represented by green arrows, negative control by blunted red arrow. Virulence factor outputs and impact on biofilm are given in boxes, as reviewed by Lee et al. [80]. Abbreviations: N-3-oxo-dodecanoyl-homoserine lactone (3-oxo-C12-HSL), N-butyryl-homoserine lactone (C4-HSL), Pseudomonas quinolone signal (PQS), 2-heptyl-4(1H)-quinolone (HHQ), integrating quorum sensing signal (IQS), pyocyanin (PYO), hydrogen cyanide (HCN), lectin A (LecA).

1.4.3. Surfactant rhamnolipids

Rhamnolipids are biosurfactants composed of one or two rhamnose sugars substituted with up to three acyl chains, resulting in an amphiphilic molecule (see Figure 1.7). They are produced by several bacterial species, and *P. aeruginosa* in particular produces a mixture of about 60 rhamnolipid congeners at varying concentrations [101]. The global surfactant properties are determined by the composition of the rhamnolipid profile, which is controlled by strain variations and carbon source availability [102] as well as QS expression (*rhl* system) [103].

Their physiological functions are multilevel and continue to be explored. Their amphiphilic chemistry is important for the development, adaptability and persistence of the pathogen which synthesise them. Rhamnolipids contribute to swarming motility, and biofilm development by regulating cell-surface physical chemistry, cell detachment and dispersion [104]. *P. aeruginosa* in biofilms synthesise higher concentrations of rhamnolipids compared to planktonic cells, and there is evidence that together with alginate, rhamnolipids provide a protective shield against immune cells [32]. Due to their detergent properties, rhamnolipids destabilise cell membranes, leading to necrosis of neutrophils, which constitute the first line of host defence [34], as well as killing a wide range of

bacteria and fungi [101, 105, 106]. They trigger inflammatory reactions in the airway and inhibit airway epithelial ciliary motility [101, 107]. Rhamnolipids also make a noteworthy contribution to PQS activity by enhancing molecular solubility in aqueous solutions, thus facilitating diffusion from cell to cell [108]. Similarly, they facilitate access to hydrophobic nutrients such as linear alkanes, which represent a good substrate for *P. aeruginosa* growing in hostile environments. Overall, they are thought to be necessary for infection to become established [103, 109], and higher levels have been previously observed in CF patients with a worse clinical state [109].

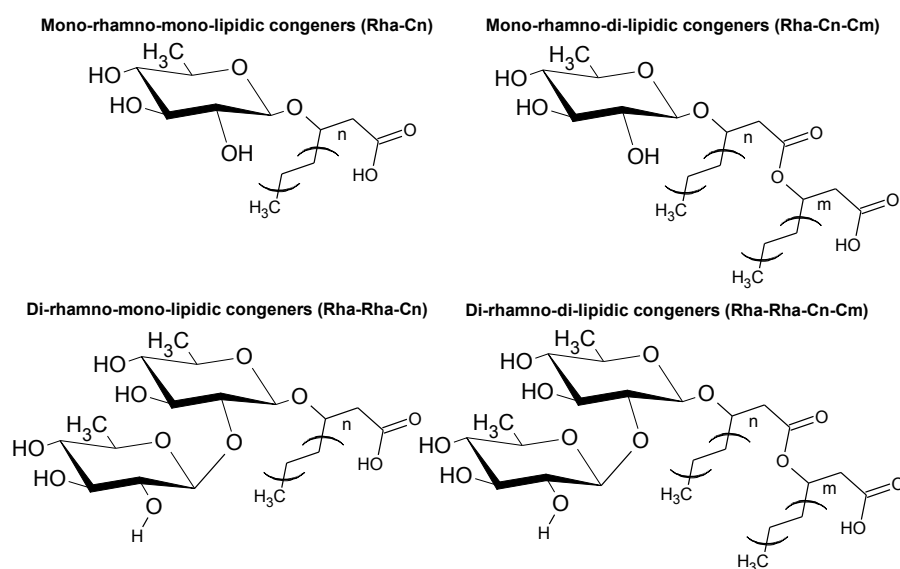


Figure 1.7. Chemical structure of the four families of rhamnolipids.

" n " and " m " represent the number of carbons in the lipid branches, e.g. the complete name of Rha-C10 is α -L-rhamnopyranosyl- β -hydroxydecanoate.

1.4.4. Other virulence-associated molecules

P. aeruginosa produces a wide variety of extracellular compounds, many of them being QS-regulated. A small number of enzymes and toxins, with high contribution to its pathogenicity, are described below as a sample of the diversity of the virulence mechanisms employed by *P. aeruginosa*.

P. aeruginosa secretes a range of toxic enzymes compromising tissue integrity in the host, such as phospholipase C, cleaving structural phospholipids, and protease and elastase that degrade structural and immune effector proteins. Exotoxin A (ExoA) is often cited as *P. aeruginosa*'s most toxic virulence factor. This enzyme inhibits protein synthesis in the cytoplasm of host cells, engendering cytopathic effects and affecting the host immune cells [110, 111].

Lectin A is a galactose-specific carbohydrate-binding protein, which mediates bacterial attachment to its substrate, e.g. the lung epithelium. It causes lung injury by inhibiting the growth of epithelial cells and also favours the absorption of ExoA by impacting cell permeability [112].

Pyoverdine is an iron-chelating metabolite allowing the metal to cross the cell membrane. It is one of the multiple systems supplying the bacteria with iron, which is required for growth and further supports biofilm formation [113]. It is also a signalling molecule regulating major virulence factors such as ExoA and protease PrpL [114].

Pyocyanin (PYO) is a multilevel virulence factor, globally weakening the host immune system. It possesses reduction-oxidation properties, which lead to the generation of toxic and pro-inflammatory ROS [115] and cause neutrophil apoptosis [116]. It is a zwitterion at blood pH (7.35), which allows it to penetrate cell membranes and kill host cells as well as other microbes [117]. Furthermore, it was shown to interfere with mucociliary clearance [118, 119] and significant quantities were previously recovered from sputum samples using HPLC-UV or immunochemical assays [120, 121].

Hydrogen cyanide (HCN) is a secondary metabolite produced by a limited number of organisms including *P. aeruginosa*. The cyanide ion blocks cellular respiration; it is deadly to eukaryotes whilst the producing bacteria have respiratory alternatives [122]. The killing action of HCN produced by *P. aeruginosa* was demonstrated in a study on nematodes [123]. Thereby, HCN compromises host cell integrity and eliminates potential competitors. Its role in human infection remains unclear but it is known to inhibit many cellular processes as well as bacterial aerobic respiration [122]. Interestingly, it was found that mucoid strains produced higher levels than non-mucoid *P. aeruginosa*, and (non-mucoid) CF clinical isolates produced more HCN than laboratory strains [124]. It was detected in CF sputum [125] and breath [126], and was associated with poorer lung function, making it an relevant virulence factor in the development of the infection *in vivo*.

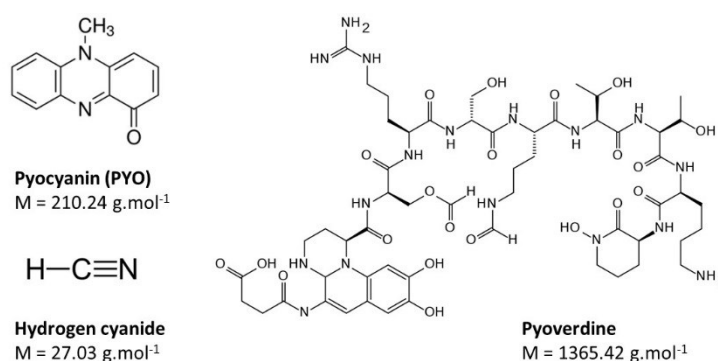


Figure 1.8. Chemical structure and molar mass of several virulence factors from *P. aeruginosa*.

1.4.5. Metabolic impact of *P. aeruginosa* infection on the host

Besides the above metabolites inherent to *P. aeruginosa*, the body itself reacts to the presence of the microorganism and metabolic changes have been observed in the host airway tissues and fluids. Whether these modifications are due to the bacteria or to the host metabolism is not always clear, nevertheless monitoring them could be helpful in the detection of an infection.

CF significantly impacts patients' metabolism and identifying further alterations due to infection by *P. aeruginosa* is likely to remain challenging. Nevertheless, various studies have unravelled variations in the host tissues and body fluids, especially in the lipid profile including fatty acids, phospholipids, sphingolipids, sterols, oxylipins and endocannabinoids [127]. In the case of *P. aeruginosa* infection, a few phosphatidylcholines and fatty acids in particular were decreased whilst some cholesterol derivatives and triglycerides increased [127, 128]. Abnormal ceramide production and accumulation in the membrane of respiratory cells was also triggered by the presence of *P. aeruginosa* [129]. Bacterial exometabolites may also reach host body fluids, resulting in elevated levels of amino acids [130]. Changes in breath composition were also reported, with regard to volatile organic compounds (VOCs) [131] and nitrite levels in exhaled breath condensate (EBC) [132].

1.5. Analytical techniques

The compounds presented above are small metabolites, many of them with lipophilic properties. Appropriate and powerful tools are needed to detect and identify them, ideally directly from sample. Advances in mass spectrometry and chemometrics in the last decade have considerably expanded their scope of application, notably in medicine. Ambient methods in particular are very promising for the quick and direct analysis of lipids in samples through innovative approaches.

1.5.1. Mass spectrometry

Mass spectrometry (MS) is an analytical technique which allows the detection, identification and quantification of ionised molecules in the gaseous phase based on the ratio of their molecular mass to their charge (m/z). Mass spectrometers are composed of three main elements: the ionisation source allows the ionisation of the sample constituents, the mass analyser separates the ions according to their m/z using electromagnetic fields and the detector records a signal – also called ion count – proportional to the ion concentration through detection of electric current or charge whilst

the ions approach or hit the detector surface. Uncharged molecules cannot be detected. A crucial requirement for this principle is a high vacuum in order to avoid chemical alterations and to allow free motion of the ions.

The results of an MS analysis are summarised in a mass spectrum (see an example in Figure 1.9) where the ion count is plotted against the m/z value of the ions. The intensity of each peak is recorded as an absolute value, proportional to the concentration of ion(s) of a given m/z , but is commonly represented as a percentage relative to the most intense peak (base peak). The list of intensities measured for each m/z value is sometimes referred to as the MS or spectral profile of a sample.

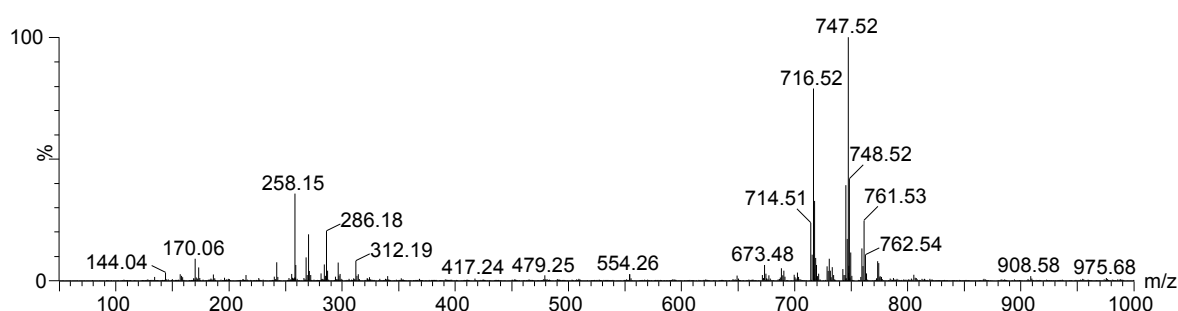


Figure 1.9. Example of a mass spectrum.

Shows signal intensity relative to the base peak (here $m/z = 747.52$) vs. m/z ratio; obtained with a Xevo G2-XS Q-ToF instrument (Waters Corporation) from *P. aeruginosa* colonies, by direct analysis with rapid evaporative ionisation mass spectrometry (REIMS), in negative ionisation mode.

Molecular formulae can be attributed to experimental m/z values by comparison with the exact mass calculated for theoretical ions. However, different ions can give the same exact m/z and one molecule can produce several ions including various adducts and fragments. As a first approach, it is usually accepted that a less than 10 ppm difference between theoretical and experimental m/z values provides a reasonable tentative identification. Yet, this degree of accuracy is not achievable with all instruments since it depends on the mass resolution of the mass analyser; magnetic sector, ToF with reflectron (e.g. Xevo G2-XS Q-ToF), Fourier transform Orbitrap and Fourier transform ion cyclotron resonance (FT-ICR) mass analysers offer a mass accuracy within 10 ppm. The growing use of MS for metabolic studies gives rise to ever larger data sets, which demand the development of bioinformatics tools for automated identification integrating strategies to deal with the false discovery rate (FDR) [133].

MS instruments commonly comprise more than one mass analyser to enable tandem mass spectrometry analysis (MS/MS). The MS/MS design allows finer identification than exact mass comparison by examining the fragmentation pattern of individual ions. When subjected to a fragmentation process under reproducible conditions, e.g. collision-induced dissociation (CID),

surface-induced dissociation or photodissociation, a precursor ion – or parent ion – always fragments following the same pattern. This MS profile of product ions can be used to elucidate the initial molecular structure. For this purpose, the concept of MS/MS often involves two mass analysers in line. The first one serves as an ion filter to isolate ions of interest with a specific m/z . The precursor ions are then subjected to fragmentation and product ions are separated through the second mass analyser and detected as before. For example, the Xevo Q-ToF from Waters (Waters Corporation, Wilmslow, U.K.) uses a quadrupole (Q) to select the precursor ions in addition to the final ToF analyser. The dissociation happens in the intermediary collision cell, where the precursor ions are accelerated by applying a voltage and collide with argon atoms thereby producing fragment ions.

Initial ionisation is a key step in mass spectrometry and varied techniques have been developed to address diverse problems. The choice of an appropriate ionisation source depends on the sample nature and the target compounds. The following section describes different ionisation techniques that are relevant to this project.

1.5.1.1. Reference methods

1.5.1.1.1. Electrospray ionisation for liquid solutions

In order to undergo mass spectrometric analysis, analytical target molecules need to be converted to gas phase ions in the so-called ionisation source. Originally, analytes were evaporated and introduced as a gas into the vacuum regime of the mass spectrometer where they underwent ionisation. However, this limited the analysis of high molecular weight compounds and compounds carrying polar or ionic moieties. It also made the analysis of condensed-phase samples difficult and necessitated sample pre-treatment often including chemical derivatisation. The advent of electrospray ionisation (ESI) facilitated the direct analysis of liquid solutions by allowing ionisation under atmospheric pressure [134]. In ESI, a high voltage is applied to the sample solution flowing through a capillary tube, causing it to form a spray of charged droplets at the exit. Solvent evaporation and analyte ion desorption from droplets then result in the formation of a desolvated ion population which is transferred to the vacuum regime of the mass spectrometer through an atmospheric pressure interface (see Figure 1.10).

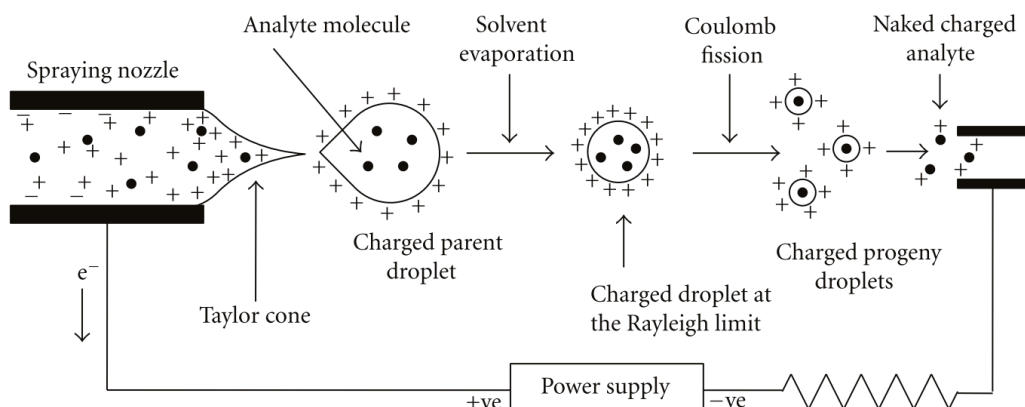


Figure 1.10. Electrospray ionisation mechanism.

(Reproduced under the Creative Commons Attribution license from Ref. [135]).

ESI remains a standard method for the analysis of thermolabile molecules since it is a soft ionisation technique. This means that chemical structures are conserved and intact molecular ions are detected. ESI also has the ability to produce multiply charged ions, which facilitates the analysis of very high molecular mass compounds (reduced m/z range).

1.5.1.1.2. Matrix-assisted laser desorption ionisation: application to the analysis of bacteria and surfaces

Around the same time that ESI made large molecules in solution amenable to MS analysis, MALDI was developed, enabling the analysis of solids [136]. This is another soft ionisation method, which allows desorption of analytes from a solid substrate and the formation of singly-charged molecular ions with no upper limit of mass. The sample is co-crystallised with a matrix compound and laser radiation is applied causing desorption of clusters. The matrix substance usually has high extinction coefficient at the wavelength of the laser applied and volatilises on irradiation. One of the possible processes of ion formation involves primary ionisation of the matrix – protecting the analytes from direct irradiation – and then proton transfer to the analytes [137]. MALDI is often used in association with a ToF mass analyser, as both are pulsed techniques, offering wide mass range capabilities. MALDI allows the analysis of a very broad range of molecules, from proteins, oligosaccharides and polymers to lipids and nucleic acids. This makes it one of the most versatile techniques with many potential applications, especially since it has been adapted to operate at atmospheric pressure [138].

Both ESI and MALDI have contributed greatly to establishing mass spectrometry as a reference method for biological analysis [139]. MALDI-ToF is now well-established in clinical microbiology laboratories as a quick and reliable standard method for microbial identification. Clinical platforms are

widely available and overall accuracy reaches 80-90% at the species level [73]. Analyses are conducted on bacterial cells embedded in a UV-absorbent matrix. The speciation of microorganisms relies on the detection of protein profiles, primarily of ribosomal origin [140]. However, most of the embedding matrices induce a high signal below $m/z = 400$ which restrains the possibility of studying lower mass compounds. MALDI-ToF previously allowed the detection of QS molecules [141] and rhamnolipids [142], but protocols necessitated preliminary extraction from pure bacterial cells and included a chemical derivatisation or labelling step.

MALDI is commonly used for MS imaging (MSI) of solid surfaces. Samples are coated with matrix and analysis is performed pixel by pixel, to cover the area of interest. MSI provides information about chemical distribution on the sample surface. Spatial resolution can be as low as 1 μm [143], i.e. at a sub-cellular level, in clinical tissues or bacterial colonies.

1.5.1.2. Ambient mass spectrometry methods

Efforts are now being made to extend MS to the analysis of native samples in the ambient environment by methods generally termed “ambient mass spectrometry”. Amongst them, desorption electrospray ionisation (DESI) [144] and rapid evaporative ionisation mass spectrometry (REIMS) [145] have demonstrated great potential for direct analysis of biological samples as discussed in the following sections. Lipids in particular are readily detected in biological tissues by these methods, due to their high abundance, low solubility in aqueous fluids and their favourable desorption from charged droplets [146].

1.5.1.2.1. Rapid evaporative Ionisation mass spectrometry

REIMS allows near-real-time MS analysis of aerosols and was developed for the *in situ* analysis of biological samples. Aerosols are generated by applying a thermal stress on the biomass which causes it to rapidly heat and results in sample evaporation. The evaporated particles contain ionised analytes and are introduced into the mass spectrometer using the inherent instrument vacuum. This is an aggressive process, which frequently causes molecules to fragment or react, and destroys the sample. It also favours deprotonation, making ionisation usually more efficient in negative ion mode.

REIMS was originally developed as a surgical tissue identification tool for tumour margin detection. The “Intelligent Knife”, or iKnife, was designed to analyse the aerosol released *in vivo* during electrosurgical dissection in order to discriminate healthy from cancerous tissues [139, 145, 147].

Other setups have been developed to allow the direct analysis of microbial isolates cultured on agar and were successful in classifying clinically relevant yeast and bacteria at species level. First, electrosurgical bipolar forceps were connected to an MS instrument; bacterial colonies were then manually sampled and subjected to a radiofrequency electrical current (100kHz-4MHz) to vaporise particles (see Figure 1.11) [148, 149].

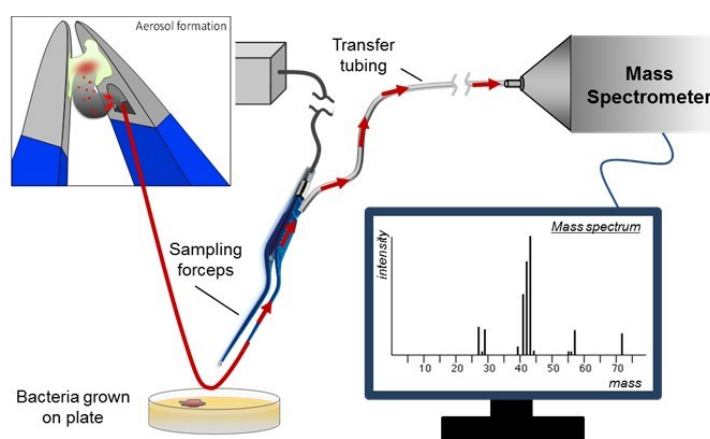


Figure 1.11. Schematic representation depicting REIMS analysis on bacterial culture using bipolar forceps. (Adapted with permission from Ref. [148]).

Typically, the analysis of microorganisms in negative ionisation mode leads to the detection of an intense signal from phospholipids, released by the cell membranes, and fatty acid ions, originating from free fatty acids and lipid fragmentation. Additionally, rhamnolipids and a range of products deriving from QS molecules have previously been reported [149]. Fatty acids and lipids have been long known to carry taxonomic information [150]. The phospholipid profile detected by REIMS in the 600 to 900 m/z range allows robust and accurate classification to genus, and species level [149, 151].

On-going developments recently led to the creation of an automated platform with the adjunct of a customised robot carrying out plate handling, colony selection and REIMS analysis [152]. With increased productivity, reduced costs and user-independent assays, this new platform better meets the modern needs of clinical microbiology laboratories. The system uses reusable monopolar probes to apply current upon contact with the sample. It is currently being adapted and tested with a laser system as a cleaner source of ionisation, since no contact is needed to apply thermal stress; the method is termed “laser-assisted REIMS”. With the integration of a robot and a laser source, the sample load could be significantly increased allowing high-throughput data acquisition, e.g. on multi-

well plates or for imaging purposes. Indeed, the “burn” caused by the laser to the sample surface is less than 1 mm diameter and could thus be repeated every 1 mm, providing low-resolution images.

1.5.1.2.2. Desorption electrospray ionisation: analysis of surfaces

DESI combines ESI with surface desorption, allowing the analysis of a broad range of samples such as solids, adsorbed gases or deposited liquids. Charged solvent droplets generated by a pneumatically assisted electrospray probe are directed towards the sample surface under atmospheric pressure. The solvent dissolves the analytes on the surface and secondary charged microdroplets are generated. A subsequent ESI-like process occurs, leading to the formation of gaseous ions of analytes, which are directed towards the MS atmospheric interface. By moving the sample (or alternatively the instrument), signal can be acquired continuously across the surface and ion images can be reconstituted with a spatial resolution as low as 20 μm (see Figure 1.12) [153, 154]. DESI-MSI is currently being extensively examined as a potential tool for cancer margin detection *ex vivo* in histopathological laboratories [146, 155-157].

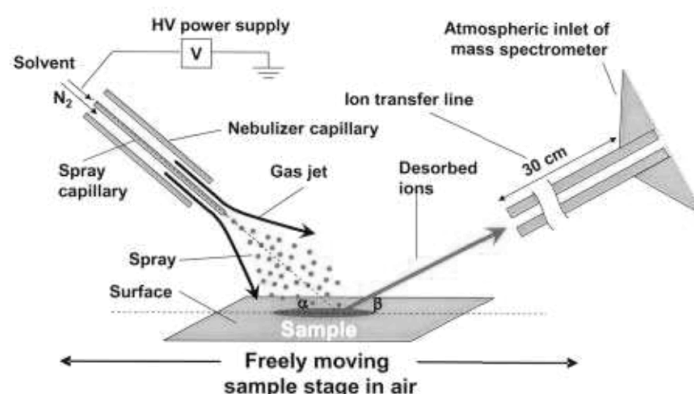


Figure 1.12. Schematic representation of a desorption electrospray ionisation source.
(Reproduced with permission from Ref. [144]).

There are only few compounds below 1000 Da that cannot be detected by DESI. Indeed, the composition of the solvent can be adjusted to react selectively with molecules or chemical classes of interest and enhance their desorption and/or ionisation. This is called reactive DESI [158]. Cholesterol or steroids for example, are not readily ionised, due to the lack of functional ionisable groups, but can be with the aid of betaine aldehyde or hydroxylamine, respectively [158, 159]. The use of extended DESI ion transfer tubes can also allow remote ionisation of samples. Extended and flexible inlet capillaries are used to conduct ions toward the MS interface when the ionisation is performed at some

distance from the instrument [160]. This could lead to an even greater versatility of the technique and a wider scope of application. Flexibility is one of the greatest strengths of DESI.

Unlike MALDI-MSI, DESI is an ambient, minimally destructive technique, but its sensitivity, spatial resolution and dynamic range can still be improved [137]. It is better suited to the analysis of low mass metabolites, and especially lipids, but quantification is limited. Firstly, quantification would require the incorporation of a standard, and secondly, the desorption and ionisation of analytes are not necessarily comparable. Intrinsic compounds can be used as internal standards but quantification is likely to be less reliable [158].

1.5.2. Alternative methods under development for breath analysis

Breath represents an attractive non-invasive sample which can be obtained even in babies. Hence, breath analysis has been the subject of studies since the 1970s, exploring the capacities of mass spectrometry in particular as a diagnostic tool. The original focus was on a coupling with gas chromatography (GC) which then shifted towards real-time techniques. These techniques have been tested in a CF context, since it is known that some bacterial communities, including *P. aeruginosa*, release characteristic smell and volatiles [161, 162].

1.5.2.1. Gas chromatography – mass spectrometry

Scott-Thomas *et al.* concentrated breath samples prior to GC-MS analysis using solid-phase micro-extraction (SPME) fibres and significantly correlated 2-aminoacetophenone with *P. aeruginosa* infection in both healthy and CF patients [163]. However, this molecule can also be detected in different food products and cannot be relied upon as a single biomarker [164]. In another study, exhaled breath samples were collected in Tedlar® plastic bags and analysed by GC-MS, allowing discrimination between CF patients with and without *P. aeruginosa* infection using 22 VOCs. Amongst the ten most discriminating markers, many were hydrocarbons which are not specific to *P. aeruginosa* [131]. Drawbacks of the method include a low sensitivity to very small molecules and the use of time-consuming chromatography which precludes its use for real-time analysis.

1.5.2.2. Real-time analysis of breath using mass spectrometry

Selected ion flow tube (SIFT), proton transfer reaction (PTR) and secondary electrospray ionisation (SESI) are examples of indirect ionisation strategies that have been developed for the real-time analysis of breath with MS. They rely on a two-step ionisation process with primary ions reacting with the gaseous analytes in the breath sample to form secondary ions through chemical ionisation.

Several studies employing SIFT-MS have been carried out in order to identify breath biomarkers for CF patients infected with *P. aeruginosa*. One example proposed by Enderby *et al.* is HCN, a secondary metabolite of *P. aeruginosa* [126]. However, studies have shown that HCN is not sensitive enough for diagnostic purposes [165]. Indeed, HCN is a very small molecule, which can have multiple different origins, such as the oral microflora [166], and cannot be relied upon alone. Further research is being carried out and other markers have been suggested; yet recent attempts using a combination of selected VOCs to predict infection remained insufficiently sensitive [167, 168]. It has also been pointed out that breath is not only constituted of vapours and volatile species, but is actually an aerosol containing non-volatile compounds dissolved in microdroplets, such as amino acids or sugars [169], which cannot be analysed by SIFT- or PTR-MS.

Other strategies have focused on developing external ionisation sources for ambient applications. SESI is a general term referring to the ionisation of analytes in a gaseous state by charged microdroplets generated by ESI [170-172]. In the last ten years, SESI techniques have been applied to many kinds of matrices and even living subjects, including skin vapour [173], breath [169, 174] and bacterial culture headspace [175]. They allow the detection of volatiles, proteins and metabolites. Accordingly, urea, amino acids and glucose were detected in heated breath [169], while bacterial cultures were shown to produce a unique volatile signatures [175]. Furthermore, the analysis of breath of artificially ventilated mice with SESI-MS allowed the differentiation of lung infections, including *P. aeruginosa*, using multiple *m/z* peaks [176, 177]. Benefits of SESI-MS compared to SIFT and PTR include the detection both volatile and non-volatile species in liquid, gaseous or mixed samples [169, 174]. Nevertheless, SESI techniques demand a sophisticated configuration which is critical for the quality and reproducibility of the results. Expertise and sometimes fastidious optimisation of experimental parameters are required prior to analysis [174].

1.5.2.3. Electronic noses

Electronic noses, or e-Noses, are usually portable devices with GC, ion mobility spectrometry (IMS) or optical or electrochemical sensors [178]. Sensors can be conducting polymers, semi-conductors or proteins and are specific to a class of compounds. VOCs interact differently with each sensor, e.g. adsorption on a polymer affecting the material resistance, which is monitored [179]. Compounds cannot be identified but odour patterns are recognised after training the instrument [180].

An e-Nose comprising 14 conducting polymers was able to differentiate headspaces from different mycobacteria and *P. aeruginosa* and successfully detected *P. aeruginosa* in spiked sputum samples [179]. The SpiroNose, which is being developed at the University of Amsterdam by Peter Sterk's team, combines a spirometer with an e-Nose that uses an array of chemical detectors. It has been tested for diagnosing asthma, chronic obstructive pulmonary disease (COPD) and lung cancer [181]. The Lonestar® developed by the Owlstone Company (Cambridge, UK) is based on field asymmetric ion mobility spectrometry (FAIMS) technology. This system achieves the separation of ions on the basis of their mobility in an asymmetric high electric field. The technology has been applied to the analysis of volatiles for the diagnosis of various gastro-intestinal diseases [182-184].

Overall, e-Noses constitute a promising tool to diagnose and monitor various diseases although they do not allow identification of biomarkers. Limitations also include validation and standardisation of the methodology, and most sensors have not reached the level of sensitivity of other techniques [164, 180].

1.5.2.4. Exhaled breath condensate

Analysis of EBC was seen as having great potential for diagnosis, providing information about the epithelial lining fluids (ELF) and being more convenient than BAL or induced sputum procedures. However, little is known about the process that drives secretion of ELF and it can be extremely diluted in water vapour in varying proportions. Markers of dilution must therefore be used and salivary contamination should also be considered, so volatile components must be assessed carefully. Advanced or novel technologies providing increased sensitivity may prove the potential of EBC, such as nuclear magnetic resonance (NMR) [185] and LC-MS [186], which have been tested to evaluate airway inflammation. However, so far, analysis of breath condensate has not demonstrated its efficiency as a diagnostic or monitoring tool [164, 187].

1.5.3. Data processing, statistical analysis and interpretation

Data obtained from a single MS analysis can contain thousands of peaks and the analysis of large sample sets with MS leads to the generation of huge data matrices. Therefore, statistical strategies are necessary to visualise, interpret and interrogate the data. Several approaches exist, which can be complementary, and should be chosen according to the size, the nature and the structure of the data, as well as the objective of the study. In this project, the main focus was on classifying samples and identifying patterns and molecules separating groups.

The first step in data processing is to prepare the raw data and to make it fit for statistical analysis. Individual variables are classically compared using univariate tools briefly presented below. When it comes to evaluate the differences between samples, taking into account multiple variables, multivariate analyses are better suited. The most common tools for visualisation and some important algorithms in the field of classification are presented.

1.5.3.1. Data pre-processing

Raw data obtained from MS analysis can be noisy, incomplete or inconsistent, mainly due to analytical variability [146, 188, 189]. In order to allow comparisons and statistical processing, the data need to be cleaned up, corrected and scaled. Pre-processing usually includes at least some of these operations:

- Averaging replicates: accounts for analytical reproducibility.
- Calibration: accounts for the instrumental error, using internal or external standards.
- Signal filtering: eliminates noise (usually through background subtraction), isotopes, fragments, etc.
- Normalisation: re-balances the importance of the variables, e.g. total ion count (TIC), mean or median-fold change normalisations.
- Transformation: improves the fitting of the data to the model hypotheses (e.g. to match a Gaussian distribution), e.g. logarithmic transformation (log-transform), mean-centring, Pareto scaling.
- Alignment of the MS profiles, binning or peak matching: eliminate instrumental shifts over time and between batches.

1.5.3.2. Significance of the variables

A Student's t -test is generally used to determine if two sets of samples, characterised by a single variable normally distributed, are significantly different. It calculates the probability to be mistaken when rejecting the hypothesis stating that the means of two classes are equal (null hypothesis). The resulting probability is called p -value and corresponds to the risk to be mistaken when saying that the groups are different. The threshold is usually chosen at 5%; below, the risk is considered small and the means are likely to be significantly different. When comparing two classes of samples, the list of differentiating features (if any) can be established by calculating their individual p -values. Yet, in the case of datasets with high numbers of variables, e.g. MS profiles, a threshold of 5% leads to the identification of numerous falsely significant features. In order to limit this number, the p -value can be adjusted to take into account the false discovery rate (FDR) (expected proportion of false positives) by dividing the p -value by the number of tests – Bonferroni correction – or through the calculation of an alternative statistics, e.g. the q -value.

The receiver operating characteristic (ROC) curve is another univariate tool, applicable to non-Gaussian series, which takes into account global differences in data distribution. It is classically used to measure the performance of a binary classifier. The ROC curve plots the sensitivity against (1- specificity) as the discrimination threshold between the two groups varies. The area under the curve (AUC) summarises the ability of the variable to separate two classes.

When more than two classes are considered, an analysis of variance (ANOVA) is implemented. ANOVA is a generalisation of the t -test; the calculated p -value evaluates the risk to be mistaken when saying that the mean of at least one group is significantly different from the others' mean. The Kruskal-Wallis test is an adaptation of ANOVA, comparing the medians instead of the means. It is suited when the distribution of variables is non-standard (i.e. non-Gaussian).

1.5.3.3. Unsupervised analysis for data visualisation

1.5.3.3.1. Principal component analysis

Principal component analysis (PCA) is an unsupervised procedure aiming to reduce the dimensionality of a dataset by combining and weighing variables, i.e. m/z features in the case of MS data matrices. The technique identifies the theoretical orthogonal axes in an N dimensional space (N being the number of variables) which best explain the variance within the dataset – i.e. the differences between the samples. These axes are called principal components (PCs). PCA is a useful and objective tool to

observe general trends and eliminate potential outliers. Results of a PCA are represented as a score plot and an associated loading plot for given principal components. An example of a score plot is showed in Figure 1.13, where each dot contains the mass spectral information obtained from one sample. This scatter plot shows the samples' position compared to the others according to the first two PCs, indicating spectral similarity between the samples.

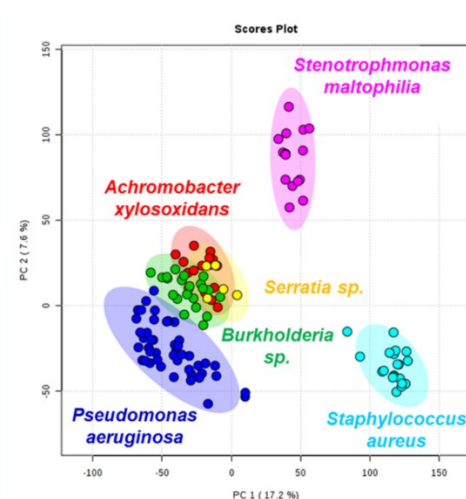


Figure 1.13. Example of a PCA score plot.

Shows the first two principal components; built using the REIMS data obtained from bacterial cultures (Xevo instrument (Waters Corporation), negative ionisation mode). Shaded areas show 95% confidence intervals of the sample groups.

In this example, the two-dimensional graph explains only 25% of the variance (17.2% and 7.6% for PC1 and PC2, respectively). Hence, a large part of the variance is supported by the following principal components, potentially allowing the separation of overlapping groups. The loading plot (not shown here) represents the contribution of each m/z feature to the overall variance for given PCs.

1.5.3.3.2. Hierarchical clustering analysis and heatmap

Hierarchical clustering analysis (HCA) is another method to assess the similarity of samples within a dataset. Samples are displayed on a tree-like diagram called dendrogram; similarity is translated into distance on the tree and levels of branching nodes (see example in Figure 1.14). Dendrograms are built by measuring the distance (e.g. Euclidean) between each pair of samples – i.e. between MS profiles – and by applying a clustering algorithm to distribute them in the tree according to their relatedness.

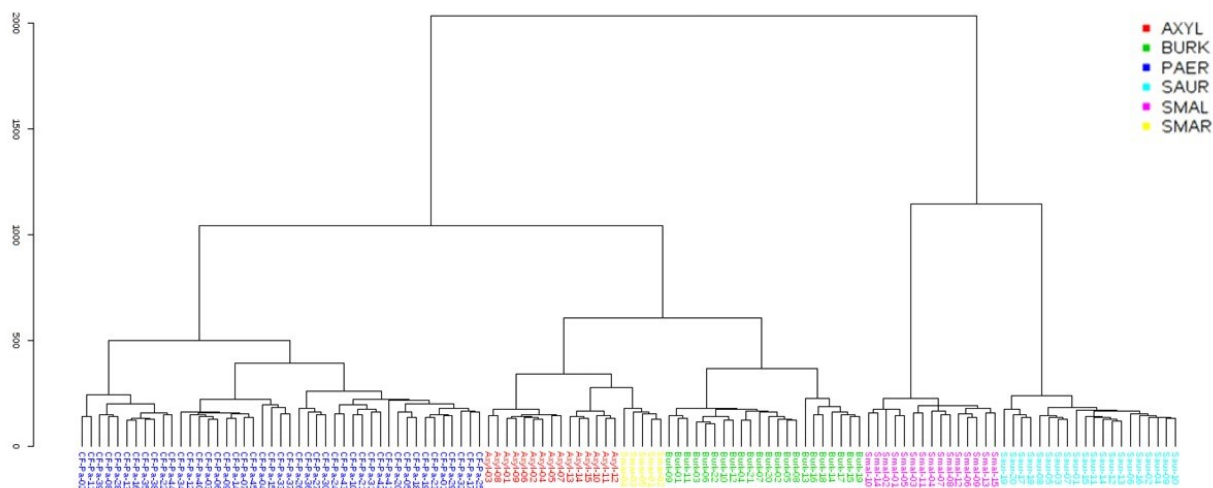


Figure 1.14. Example of a dendrogram.

Built using reduced REIMS data obtained from bacterial cultures (Xevo instrument (Waters corporation), negative ionisation mode). Colour coding indicates *P. aeruginosa* in dark blue; *A. xylosoxidans* in red; *Burkholderia* species in green; *S. aureus* in light blue; *S. maltophilia* in pink; *Serratia* species in yellow.

HCA offers another view of the samples' repartition that can be helpful when PCA visualisation with two or three PCs is not satisfactory. However, this representation is neither suitable for high numbers of samples – which render readability difficult – nor very large numbers of variables. Indeed, with the number of variables, the noise increases, which can bias the evaluation of the relatedness between the samples. For this reason, variables are usually reduced before HCA, through filtering (e.g. choice of the most differentiating features with a *t*-test) or arbitrary selection (e.g. reduction of the *m/z* range, selection of the most intense peaks). In this case, information about the selected variables is sometimes appended to the dendrogram as a table where colour is proportional to intensity; this representation is called a heatmap.

1.5.3.4. Supervised analysis for classification

Whereas unsupervised analysis summarises the information for an impartial visualisation of the data, supervised analysis aims to answer the problem of prediction. The procedure intends to identify the variables that will best separate groups defined by the user. The ultimate objective is to create a mathematical model based on a dataset (training set) which can then predict the class of unknown samples (validation set) using the values of their variables, i.e. their MS profile. However, by imposing hypotheses on the data, there is a risk of bias and overfitting of the model.

1.5.3.4.1. Discriminant analysis

The principle of discriminant analyses (DA) is based on the selection of components which will best discriminate between defined classes, by minimising intra-class variances while maximising inter-class variance. Similarly to PCA, results are presented in a score plot.

Linear discriminant analysis (LDA) seeks to maximise the separation of the means of two or more classes using linear combinations of variables. When the dataset comprises a large number of variables, it is likely to identify several discriminating combinations and therefore, there is a higher risk of determining an incorrect model. This is why, based on the hypothesis that differences between the groups are the main source of variance within the dataset, PCA is often used beforehand to reduce the number of dimensions.

Recursive maximum margin criterion (RMMC) is another supervised model which has been adapted by Veselkov et al. [188], and is particularly suited to processing complex mass spectral imaging data. It differs from LDA in the criterion to be optimised, which is the margins between the groups, i.e. the distance between the samples from one class and the splitting hyperplane, as opposed to the distance of the means.

Partial least square-discriminant analysis (PLS-DA) is a method commonly employed in biology, especially as it is well suited for the analysis of datasets where variables outnumber the samples and with a high level of colinearity. It takes a slightly different approach in that it seeks the theoretical axes which maximise the covariance of projected samples and classes and overlook the dimensions which do not contribute to the variance.

1.5.3.4.2. Random Forest classification algorithm

The Random Forest algorithm is a machine learning method which allows more complex separations than the previously discussed discriminant analyses thanks to its non-linear nature [190]. It is based on the utilisation of decision trees, formed of nodes – defined by a couple variable/threshold – and ending with leaves populated with samples belonging to a unique class (see schematic representation in Figure 1.15). At the beginning of the tree are the initial set of samples and the first decision node called the root. All possible combinations of [variable; threshold value] are tested to divide the sample set into two groups and the combination which gives the best separation of the samples according to their class is selected. The operation is repeated on each branch until they end up as pure leaves, i.e. where all the samples belong to the same class. When it comes to the prediction of an unknown

sample, its variables are evaluated at each node it encounters and it is classified with the same label as the pure leaf it eventually falls in. In the case of Random Forest, many trees are built and diversity is introduced through the random selection of subgroups of samples and/or variables. In the present project, subsets of samples were randomly selected to build a user-defined number of trees (usually between 400-1000). In addition, at each node, only a reduced number of variables were tested (e.g. the square root of the total number of variables), randomly selected from the initial set. Every single tree is thus optimal for its subset of samples. Accordingly, a new sample will be tested through each tree, and the predictions will be averaged to conclude its final classification. The aggregation of trees – the “forest” – built on subsets of data allows the smoothing of the classification and the method is less likely to overfit the training set than a single tree.

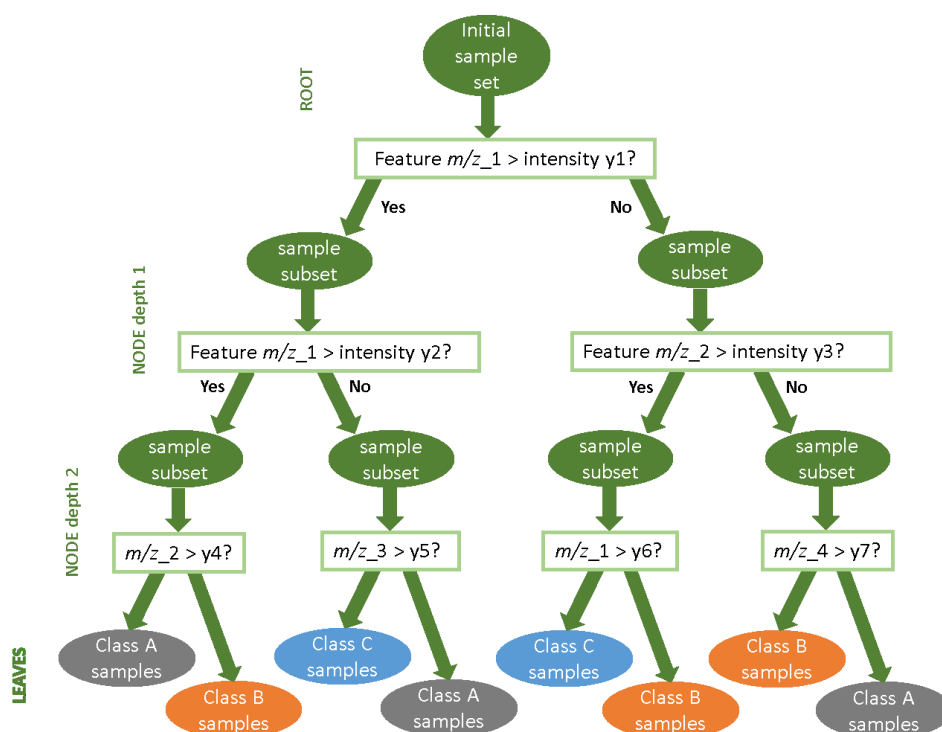


Figure 1.15. Simplistic representation of a decision tree.
Separates three classes based on their m/z profiles (comprising four m/z features).

1.5.3.4.3. Validation of the modelling method

Validation procedures are used to evaluate the quality of these prediction models. Classically, samples are divided into a training set, used to build the model, and a test set of anonymised samples to challenge the model. The accuracy of the modelling method is estimated by calculating the proportion of correct predictions. This approach is well suited for large sample sets.

When the number of samples is not sufficient to assure reliable evaluation of the modelling method, cross-validation techniques are employed where the same set of samples is used to train and test the model. In the “ k -fold” cross-validation, training and test sets are randomly generated, by dividing the original dataset in k subsets. One of these is left out, while the $(k-1)$ other subsets serve to build a model. The removed samples are used to estimate the accuracy of this model. The operation is repeated k times and the accuracy of the modelling method is the averaged accuracy of the k models.

In the particular case where k equals the total number of samples, only one sample is removed at a time and there is one model built for every single sample using the rest of the samples as a training set. This is called a “leave-one-out” cross-validation, and the performance of the modelling method is assessed by the accuracy of the predictions. In some cases, it is desirable to leave out a specific subset of samples at once, notably when samples are not independent, e.g. replicates or if taken from the same source. For instance, in the case of patients providing multiple samples, it would be more accurate to remove all of them at once when trying to diagnose a condition; the cross-validation can then be qualified as “leave-one-patient-out”.

For linear discriminant analyses (i.e. not Random Forest), classifying an unknown sample equates to computing its score according to the prediction model. Another calculation is then required to determine which is the closest group and thereby assigning it with a label. This second calculation employs a classifier method, for example “ k -nearest neighbours” (k -nn) (assigns a label according to the identity of the k closest samples from the training set), quadratic or mahalanobis distances (assign a label based on the smallest distance between the tested sample and the “centre” of the classes).

The results of a validation procedure for a classification method are presented in a table called a confusion matrix. In the example Figure 1.16, actual classes are displayed on the left-hand side and predicted classes at the bottom. Each value in the table gives the frequency of the samples from one class being predicted as another. In Figure 1.16, it is represented on a blue scale, darkening from zero to one. If the modelling method is accurate, the samples are predicted as belonging to their true class, and a dark blue diagonal shows in the confusion matrix.

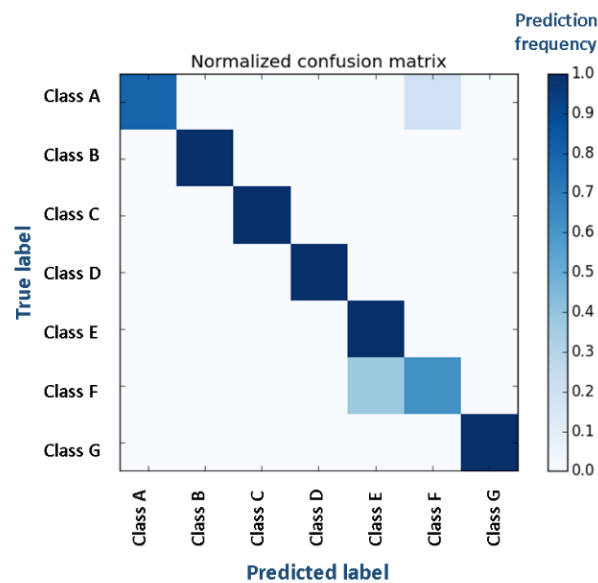


Figure 1.16. Example of a confusion matrix.

Obtained when validating a Random Forest model with a leave-one-out cross-validation.

There are a number of statistical measures which can be used to describe the prediction accuracy of a modelling method. These include:

- Accuracy: Fraction of samples from the initial set that are correctly classified.
- Sensitivity, recall or true positive rate (TPR): When considering one true label, fraction of samples that are correctly predicted as belonging to this class (i.e. accuracy per class).
- Specificity, selectivity or true negative rate (TNR): When considering one true label, fraction of samples that are correctly predicted as not belonging to this class.
- Precision or positive predictive value (PPV): When considering one predicted label, fraction of samples that actually belong to this class, i.e. fraction of the prediction which is correct.
- F1-score: harmonic mean of the sensitivity and the precision.

In 1997 Wolpert and Macready presented a number of theorems, the “no free lunch” theorems, demonstrating that, for a given problem, there is no optimal modelling method universally superior to any other [191]. Therefore, it is not possible to foretell which method will be the most suited and give the best results, and it is important to consider various classification strategies.

In summary, CF is a life-threatening disease, the morbidity and mortality of which is mostly due to respiratory infection especially caused by *P. aeruginosa*. The infection can be eradicated when detected at an early stage. But first infection usually occurs in babies and infants, when diagnosis is

compromised by the difficulty of obtaining adequate respiratory samples and the low infecting bacterial load. Current methods are neither suitable for young children, nor sufficiently sensitive. However, modern powerful analytical techniques, reviewed above, offer a new opportunity to address these issues. Firstly, they may enhance understanding of the bacterial metabolism which may improve management of patients by providing clinically relevant information, e.g. about virulence, transmissibility or infection progression. Secondly, they may obviate the need for conventional respiratory samples such as sputum and offer novel, non-culture based means of detecting early infection. Therefore, based on the background presented here, I propose to explore in this thesis the potential of novel mass spectrometry technologies to answer these questions.

Chapter 2

Hypotheses, aim and objectives

2. Hypotheses, aim and objectives

Based on the background described previously, this project rests upon this specific hypothesis: mass spectrometry techniques, together with multivariate statistical tools, can detect metabolic changes and alterations occurring (a) in the bacteria *P. aeruginosa*, e.g. as infection progresses, and (b) in the host body fluids, such as sputum, urine and skin secretions, in case of infection.

The overarching aim of this project is to provide insight into *P. aeruginosa* metabolism and community behaviour and to identify biomarkers in order to support (a) understanding of the infection and (b) increased accuracy and timely detection. To achieve this, novel ambient MS tools will be tested for the metabolic characterisation of *P. aeruginosa* and for non-invasive culture-free diagnosis.

The specific objectives are:

1. Use REIMS to characterise *P. aeruginosa* at the strain level and identify key metabolites.

REIMS analyses will be applied on common CF pathogens including *P. aeruginosa* isolates with various MLST and VNTR types. MS/MS will be performed to identify specific features relevant to the understanding and assessment of the virulence of the bacteria, including detection of rhamnolipids and quorum sensing molecules. Random Forest classification will be performed based on the detected metabolic profiles and challenged for strain type prediction.

2. Use REIMS to examine the intra-species metabolic diversity of *P. aeruginosa* and to investigate its adaptation mechanisms.

Isolates collected from various types of infections, i.e. early and chronic respiratory infection and acute non-respiratory infection, will be analysed. Metabolic information will be compared, seeking correlations, key metabolites and differences between the groups, in order to characterise respiratory vs. non-respiratory infections and to differentiate early and chronic stages in CF patients. Multiple samples will be collected and analysed per CF patient; within- and between-patients variability will be evaluated in order to study bacterial community structure.

3. Explore REIMS capacities as a direct-from-sample diagnostic tool.

REIMS analysis will be conducted directly on clinical samples. Raw sputum will be tested as part of a culture-free method, and urine as a non-invasive sample. Samples will be collected from CF patients

with and without infection and the detected metabolic fingerprints will be compared in order to identify differences and, ideally, biomarkers for *P. aeruginosa* Infection.

4. Test DESI-MS as a diagnostic tool.

A protocol will be developed for the collection, the DESI-MS profiling analysis and the data processing of skin secretions, as a culture-free, non-invasive alternative technique. The data acquired from clinical samples provided by CF patients with and without infection will be interrogated, seeking metabolic manifestations of the infection.

To achieve this, several techniques will be employed with the support of collaborators from various fields, including mass spectrometry, statistics and microbiology. General methods are presented in Chapter 3, followed by the chapters resulting from each objective and a concluding discussion. Individual results chapters include specific introduction, methodological details and discussion.

Chapter 3

General methods

3. General methods

The specific methods employed in this thesis are presented, including conditions and technical aspects regarding the handling of the microorganisms, REIMS and DESI-MS analyses, and associated data processing.

3.1. Health, safety and ethical considerations

Culture, handling and analysis of bacterial cells and bacteria-containing samples were performed following standard operating procedures written in accordance with the risk assessment approved by the Imperial College Health and Safety department (Bio-1 form, I.D.: IC-4987). All microbiological samples were managed as hazard group 2 material and processed within containment level 2 facilities to minimise exposure, as described in safety procedures.

Collection and analysis of sputum, urine and skin secretion clinical samples were approved by the institutional review board at Imperial College Healthcare National Health Service (NHS) Trust (project JD18a, ethics reference number 15/SC/0101). Before signing consent forms, patients were informed of the aim of the project and the procedures to collect, anonymise, and discard the samples.

3.2. Collection and culturing of microorganisms

Sputum and cough swabs were collected from CF patients attending routine clinical visits and initially cultured by the clinical microbiology laboratory of the Royal Brompton Hospital (London, UK). Upon collection, sputum was mixed with Sputasol, a saline solution of dithiothreitol (Oxoid, Basingstoke, UK), in a one-to-one ratio and incubated at 37°C for 15 min for protein denaturation. A volume of 20 µL of the mix was plated and cultures were incubated at 37°C for 48 h. The bacteria isolated included *P. aeruginosa*, *Pseudomonas putida*, *Staphylococcus aureus*, *Serratia* species, *Burkholderia* species, *Achromobacter xyloxydans* and *Stenotrophomonas maltophilia*. Bacterial identification was done through culturing on selective growth media (cetrimide agar for *P. aeruginosa*). Pure colonies were then passed on and stored in the CF repository of the CF & Chronic Lung Infection Group, National Heart and Lung Institute (NHLI), Imperial College London, at -80°C on beads with glycerol in CryoVials (Microbank™, Pro-Lab Diagnostics, Bromborough, UK) following the manufacturer's guidelines. All isolates used in this thesis were also kept in the same conditions at the Imperial College Healthcare

NHS Trust laboratory at Charing Cross Hospital (London, UK). More isolates collected from non-CF patients were obtained from Charing Cross Hospital microbiological bank (London, UK), including *P. aeruginosa*, *Pseudomonas alcaligenes*, *Pseudomonas koreensis*, *Pseudomonas monteilii*, *Pseudomonas mosselii*, *Pseudomonas oryzae*, *Pseudomonas putida* and *Pseudomonas stutzeri*.

All microbial identities were verified by Kate Hardiman, technician at the Charing Cross Hospital clinical microbiology laboratory, at the species level (when possible) using a Microflex LT MALDI-ToF instrument (Bruker Daltonics, Bremen, Germany) following the standard procedure used in clinical diagnostics. Single colonies were placed onto a 24-spot steel plate and overlaid with an acidic solution of α -cyano-4-hydroxycinnamic acid as a UV-adsorbent matrix. Samples were analysed after drying following the manufacturer's instructions. The Bruker Biotyper software (version 3.0) and library (version 5.0) were used to process the data and for taxonomic classification of the isolates.

Prior to REIMS analysis, isolates were re-cultured at 37°C for 24 h in appropriate atmospheric conditions (see Appendix 1). They were grown either from beads on Colombia blood agar (CBA) plates (Oxoid, Basingstoke, UK) or on 24-well plates (Corning, Flintshire, UK) filled in with 2 mL of solidified lysogeny broth (LB) media (see Figure 3.1). Natasha Wierre-Gore, PhD student at the Life Sciences department, Imperial College London (UK), prepared the 24-well plates and inoculated each well with 10 μ L of liquid cultures incubated for 24 h beforehand.

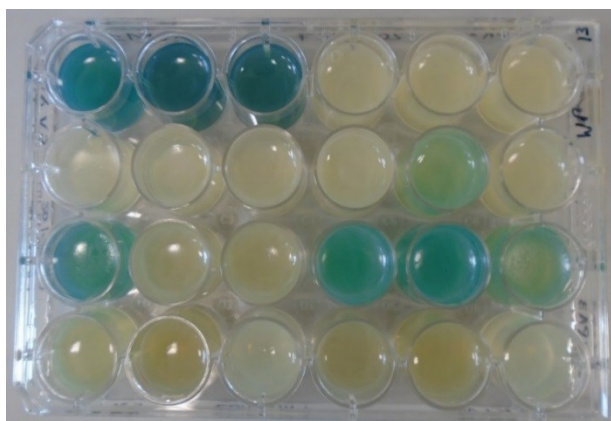


Figure 3.1. Picture of *P. aeruginosa* strains grown as a lawn on solid LB media for 24 h on a 24-well plate.

3.3. DNA extraction

For pure bacteria, single colonies were collected from cultures grown on CBA and used to inoculate 10 mL of brain heart infusion broth (Oxoid, Basingstoke, UK). The liquid cultures were incubated overnight at 37°C, under a normal aerobic atmosphere. Afterwards, a 2 mL aliquot of culture media

was subjected to centrifugation for 3 min at 21 000 x g. DNA extraction was carried out on the cell pellet, using a Fast DNA SPIN kit (MP Biomedical, Eschwege, Germany), according to the manufacturer's directions. A FastPrep 24 instrument (MP Biomedical, Eschwege, Germany) was used for bead beating applying three consecutive cycles of 30 s at 6.0 m.s⁻¹, with an intermediary cooling step at -20°C for 3 min. DNA extracts were eluted into 100 µL of PCR grade water (Roche, UK) and kept at -20°C awaiting downstream applications. DNA extraction from sputum was performed similarly, using 200 µL of defrosted sputum sample, and a Fast DNA SPIN kit for soil (MP Biomedical, Eschwege, Germany).

The quality of the DNA extracted from the single colonies was checked by running 5 µL of eluted DNA on an 1.0% (w/v) agarose gel for 1 h at 100 volts in 1X tris acetic acid EDTA buffer. Visualisation of the gel confirmed the presence of an intense band of high molecular weight DNA (> 20 kbp) for all isolates. Eluted DNA was further diluted (one in ten) in PCR grade water (Roche, UK); 1 µL was used to determine DNA concentration through Qubit fluorometry using the dsDNA Broad sensitivity kit (Invitrogen, UK) and a Qubit 3.0 fluorometer (Invitrogen, UK).

3.4. REIMS analysis

3.4.1. MS instrument

REIMS analyses were all performed on a Xevo G2-XS Q-ToF instrument (Waters Corporation, Wilmslow, UK), interfaced with MassLynx software. The instrument was operated in sensitivity mode, acquiring 1 scan per second from $m/z = 50$ to 2500. Ions were detected in negative ionisation mode. Instrumental calibration was carried out routinely every other day following the manufacturer's instructions, using sodium formate solution (Sigma-Aldrich, Dorset, UK) at 5 mM with an injection rate of 0.25 mL.min⁻¹.

The instrument was fitted with a custom-made atmospheric interface (Waters Research Centre, Budapest, Hungary) including an inlet capillary, a T-piece allowing mixing of the sample with a solvent matrix, and a heated collision surface (current set at 2.8 A) for declustering the evaporated sample particles (see setup in Figure 3.4). Throughout REIMS analyses, Isopropanol (IPA) (LC-MS grade, Romil, Cambridge, UK) containing 0.01 g.L⁻¹ of leucine enkephalin (LeuEnk) (VWR, Radnor, PA, USA) was introduced via the T-piece and infused at a flow rate of 0.25 mL.min⁻¹. IPA helps to break up the ion clusters whilst LeuEnk serves as a lock mass compound for internal re-calibration. The LeuEnk signal

would be impacted by tubing blockages, instrumental variations or drift; a steady signal thereby guarantees the proper functioning of the system.

3.4.2. Ionisation sources

3.4.2.1. Handheld bipolar probe

Manual REIMS analysis of bacteria grown on plate has been described previously [149]. Electrosurgical bipolar forceps (Erbe Medical Ltd., Leeds, UK) connected to an electrosurgical generator (ERBE ICC 300, Erbe Medical Ltd., Leeds, UK) were used as a sampling probe. A small amount of biomass was collected directly from the plate so as to cover ca. 2 mm² of one tip (Figure 3.2). As shown in Figure 3.2b, upon bringing the tips together, an electric power of 70 W was applied for about 3 s. The heated biomass released analyte-containing vapours which were conveyed through the inbuilt aspiration line of the forceps and then through to a polytetrafluoroethylene (PTFE) tubing line (1.5 m, 1/16 in inner diameter, Sigma-Aldrich, Dorset, UK) leading to the inlet capillary of the instrument.

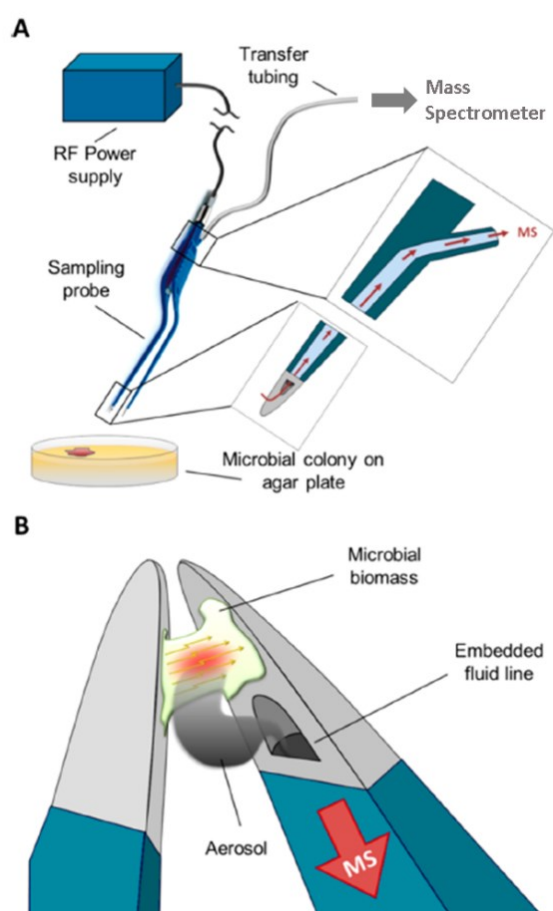


Figure 3.2. Schematic representation of the REIMS setup for bacterial analysis using a handheld bipolar probe. (Reproduced with permission from Ref. [149]).

MS data were acquired using MassLynx for a minimum of three individual sampling points per sample with a waiting time of 5 s in between. Biomass was collected from several colonies growing on the plates to account for the natural diversity within each isolate.

3.4.2.2. Automated, laser-assisted REIMS

Automated REIMS analysis was enabled by the integration of a customised liquid handling robot, a Tecan Freedom EVO 75 (Tecan Group Ltd., Männedorf, Switzerland & Waters Research Centre, Budapest, Hungary) (see Figure 3.3) controlled through the Freedom EVOware software (Tecan Group Ltd., Männedorf, Switzerland).

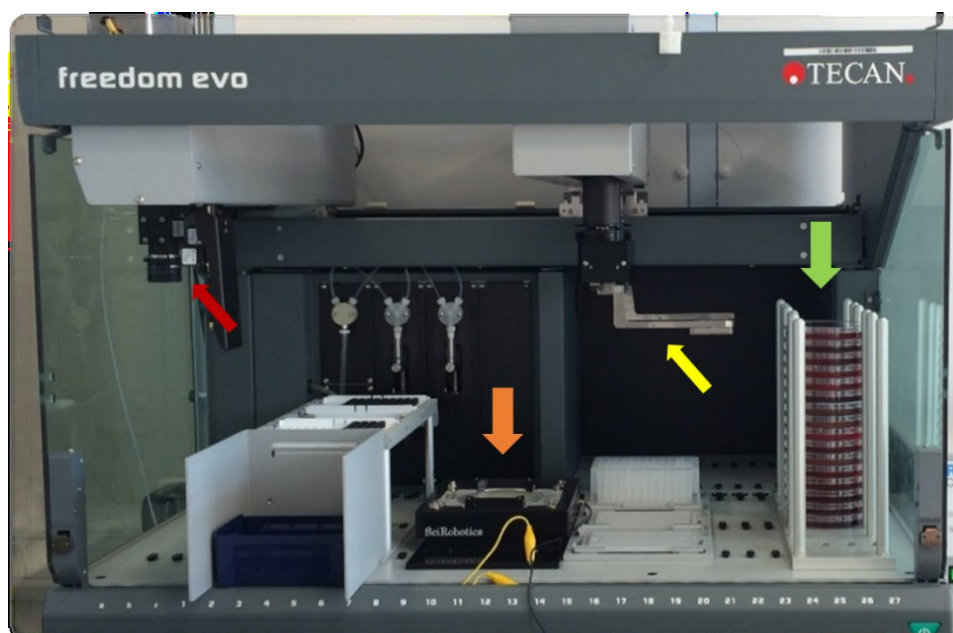


Figure 3.3. Customised Tecan Freedom EVO platform for automated culture plate handling and REIMS analysis.

The plate rack stands on the right hand side (green arrow), the sample handling arm awaits on its left (yellow arrow), in the middle lies the SciRobotics light stage (orange arrow) and the second arm holding the camera and the ionisation probe is situated at the top left-hand corner (red arrow).

The sample plates were placed in racks (green arrow in Figure 3.3) and the operator entered the number of plates on the Freedom EVOware software before launching the experiments. The arm of the robot (yellow arrow, Figure 3.3) then picked up the first plate and positioned it onto an illuminated stage (orange arrow, Figure 3.3). There, a camera took a picture of the sample plate and the image was retrieved by the computer where the user could select the local areas to be analysed – e.g. microbial colonies – thanks to the Pickolo software (SciRobotics, Israel). Another robotic arm (red arrow, Figure 3.3) was holding the ionisation source which then descended to the analysis site where rapid heating was applied for a few seconds. A PTFE tubing line (1.5 m, 1/16 in inner diameter, Sigma-

Aldrich, Dorset, UK) connected to the interface of the MS instrument was fitted alongside the ionisation probe to aspirate the produced aerosol and MS data was acquired via the MassLynx software. This was repeated for as many points as had been selected by the user, with a waiting period (depending on the sample) between each “burn” to allow aerosol transfer to the instrument. Finally, the first arm returned the plate to the rack and the following sample was processed.

For laser-assisted REIMS, a CO₂ laser of 10.6 µm wavelength (Aesculight AE-10, Bothell, WA, USA) was used as an ionisation source. The laser module was supplied with a 30 psi pressure of helium gas to avoid overheating. The laser beam was conveyed through an optic fibre to a laser lens assembly which constituted the ionisation probe, directed by the second robotic arm (see picture of the laser head in Figure 3.5 and connection to the REIMS setup in Figure 3.4). The sample to probe distance was set to 2-2.5 mm, resulting in a laser beam of 485-520 µm diameter for an optimised beam focusing. The laser was usually operated with 40 ms pulse width in SuperPulse mode and power was adjusted to fit sample properties (e.g. water content). Parameters of the Freedom EVOware script could be modified to accommodate the analysis of different sample formats (see the example of a 24-well plate in Figure 3.5), and included three dimensional positioning of the robotic arms, “burning” and waiting times, and MassLynx parameters (acquisition time, file names).

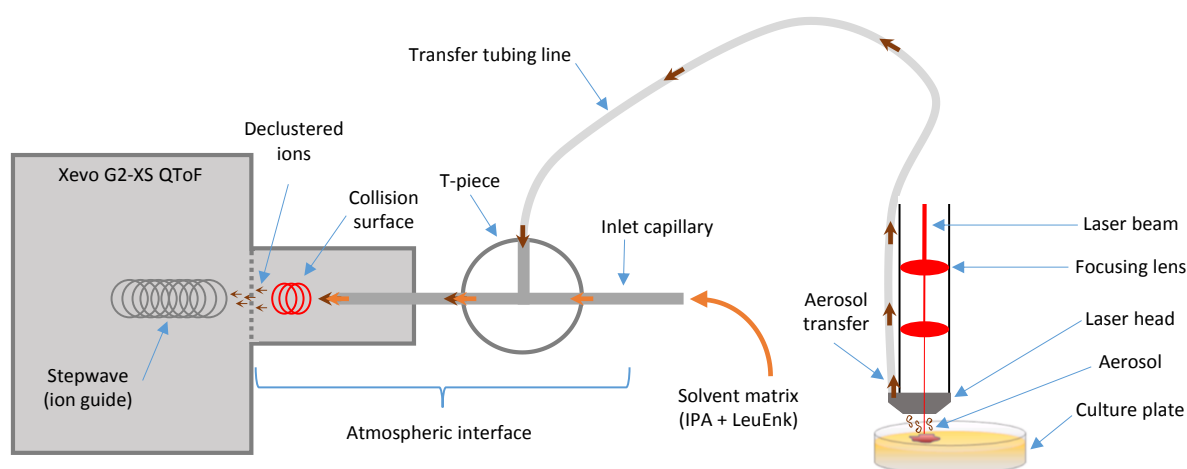


Figure 3.4. Schematic diagram of the REIMS setup.

Shows the custom-made atmospheric interface from Waters (Research Centre, Budapest, Hungary) connected to the laser ionisation probe.

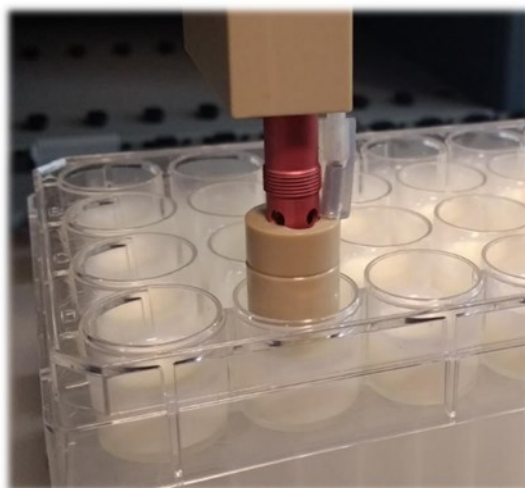


Figure 3.5. Picture showing the analysis of samples in a 24-well plate using laser-assisted REIMS.
The sampling probe includes the laser lens assembly (in red) and the extracting line on the right-hand side to channel the vapour to the instrument.

A specific script was written for REIMS-imaging by Simon Cameron, research associate at the department of Surgery and Cancer, Imperial College London (UK). The area to be imaged (typically ca. 4 cm²) was manually selected on the Pickolo software and REIMS-imaging analysis was performed with 1 mm spatial resolution. One raw data file was acquired for each line.

The Tecan platform was enclosed in a bespoke Class II biological safety cabinet (Bigneat, Waterloooville, UK) adapted to enable safe laser operations.

3.4.3. Tandem MS analysis

MS/MS analysis was performed manually on bacteria grown on CBA plates using bipolar forceps, working with an argon pressure of 1 bar. Parameters for ion selection and fragmentation were tuned through MassLynx set in the MS/MS mode. Operating values varied depending on the chemistry and concentration of the ions and included CID energy and quadrupole resolving settings. Optimisation was done intermittently since it required the manual analysis of sample. The quadrupole settings were first refined at 0 eV collision energy in order to obtain a pure and intense signal for the precursor ion. The CID energy was progressively increased until the parent ion's intensity decreased to reach half of the most intense product ion peak. Data were acquired throughout the process, and MS/MS profiles were manually extracted from the raw data. Fragmentation profiles were compared with online databases such as Lipidmaps (<https://www.lipidmaps.org/>) and Metlin (<https://metlin.scripps.edu/>). When tentative identification had been previously made based on exact m/z comparison, the putative chemical structure could be confirmed (or not) by elucidating the fragmentation pattern using the

mass of the product ions. Spectral features identified in this manner are reported with product ions and individual MS/MS settings in Appendix 3.

3.4.4. Data analysis

The Offline Model Builder (OMB) (version 1.1.29.0) software (Waters Research Centre, Budapest, Hungary) was used for data pre-processing. Preliminary processing included combining scan measurements per sample, background subtraction, correction for mass drift using LeuEnk and binning to 0.1 m/z bins. The LeuEnk ion used for lock mass correction in negative ionisation mode was m/z 554.2516. The LeuEnk ion and its negative isotopes (correlation above 0.9) were excluded before performing TIC-normalisation.

Statistical analysis was carried out either using the open source R language and environment (R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>) or the Metaboanalyst 3.0 free online platform (<http://www.metaboanalyst.ca/>) [192]. Data were classically filtered, log-transformed and scaled before completing PCA and HCA (Euclidean distance) to get an overview of the data.

Supervised classification was usually performed on TIC-normalised data using the Random Forest algorithm [190] and the machine learning scikit-learn package [193]. The model was assessed by cross-validation, using leave-one-out or leave-one-patient/sample-out methods. The F1-scores were calculated for each class to attest to the prediction correctness; global accuracy, defined as the overall proportion of samples accurately classified, was used to evaluate the overall power of the classification model.

Imaging data were pre-processed line by line on OMB and included background subtraction, lock mass correction using LeuEnk, and 0.1 m/z binning. A TIC threshold was determined to detect the REIMS signal which constituted the pixels; scans were combined per pixel. Features of interest, e.g. N most intense peaks or specific m/z , were selected to generate ion images. Images were plotted per feature – i.e. ion maps – using the intensity of individual pixels normalised according to the most intense pixel and log-transformed.

3.5. DESI-MS analysis

3.5.1. DESI-MS profiling

DESI-MS analyses were performed on an Exactive Orbitrap MS instrument (Thermo Fisher Scientific Inc., Waltham, MA, USA) operated in positive and negative ionisation modes, with the parameters listed in Table 3.1 and controlled by XCalibur 2.1 software. The instrument was equipped with a computer-controlled, customised, 2D moving stage (Newmark Systems Inc., Rancho Santa Margarita, CA, USA) previously described by Strittmatter et al. and presented in Figure 3.6a [194]. The ionisation source employed the electro-flow focusing technology (Ingeniatrics Technologies, Seville, Spain) which is sometimes referred to as desorption electro-flow focusing ionisation (DEFFI) [195] (sprayer shown in Figure 3.6b). Instrumental calibration was carried out routinely every other day, using the manufacturer's ESI positive and negative ion calibration solutions (Thermo Fisher Scientific Inc., Waltham, MA, USA) with an injection rate of 5 $\mu\text{L}\cdot\text{min}^{-1}$.

Table 3.1. Instrumental parameters for DESI-MS analysis with a Thermo Exactive instrument.

Ionisation mode	Positive	Negative
Resolution		100000
Orbitrap injection time		1000 ms
Inlet capillary temperature		250°C
Spray voltage	+4 kV	-3.5 kV
Inlet capillary voltage	+50 V	-50 V
Tube lens voltage	+140 V	-150 V
Skimmer voltage	+24 V	-40 V

The DESI parameters, mainly sprayer angle, sprayer distance to surface and sprayer distance to capillary, were initially fixed to respectively 75 degrees, 2 mm and 10 mm, but needed daily adjustments to maximise the signal. Home-prepared pork liver sections deposited on microscope glass slides were used as a reference material for instrument monitoring and optimisation. The spray solvent (methanol/water (UPLC grade solvents, Sigma-Aldrich, Dorset, UK), 95:5 v/v) was supplied by a syringe pump at a flow rate of 1 $\mu\text{L}\cdot\text{min}^{-1}$ and a nebulising gas pressure of 2 bars (N_2). Samples were fixed on the table stage using double-sided adhesive tape (3M, San Louis, MO, USA). For profiling DESI-MS analysis, data was acquired continuously across the sample whilst the stage was moving at a speed of 0.1 $\text{mm}\cdot\text{s}^{-1}$ until the sample section was over.

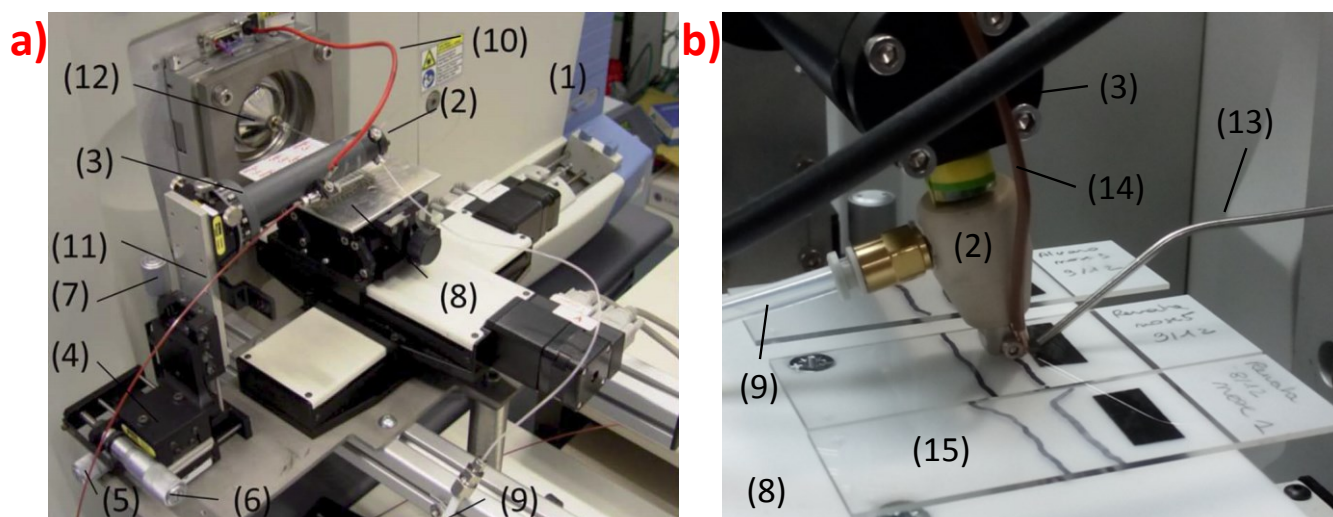


Figure 3.6. Pictures of the set-up for DESI-MS analysis.

(a) Customised moving sample stage mounted onto an MS instrument (Orbitrap Discovery hybrid, Thermo Fisher Scientific Inc., Waltham, MA, USA) (1), the DESI source (2) hold by a rotating arm (3) mounted on the moving stage (4), screws for adjustment (5) – (7) of the sample table (8) in the x-, y-, and z-directions, gas supply (9), high voltage supply (10), solvent supply (11), atmospheric interface of MS (12); (b) Zoom on a DESI source (DEFFI sprayer, Ingeniatrix Technologies, Seville, Spain) (2), showing the MS inlet capillary (13), the earth cable (14) and samples deposited on glass slides (15) (adapted with permission from Ref. [194]).

3.5.2. Data analysis

Raw files were converted into an mzXML format using ProteoWizard msConvertGUI (Vanderbilt University, Nashville, TN, USA) so that they could be imported and processed in a MatLab environment (MathWorks, Nattick, MA, USA). Statistical analysis was performed using a customised programme developed by James McKenzie, research associate at the department of Surgery and Cancer, Imperial College London (UK). Depending on the analytical needs, the spectral data could be averaged per sample or a subset of spectra could be randomly selected and used as analytical replicates of one sample. Further data pre-processing systematically included alignment of the spectral profiles to correct for small mass shifts, normalisation (TIC, mean- or median-fold change normalisations) and log-transform. Statistical tools available included PCA, ANOVA, supervised LDA and RMMC, and various cross-validation options.

The above methods were first used on pure colonies of *P. aeruginosa* in order to identify characteristic molecules and to study bacterial metabolism; then they were applied to clinical samples in an attempt to detect infection biomarkers. Technical details specific to each study is given throughout this thesis.

Chapter 4

Identification of *P. aeruginosa* using REIMS

4. Identification of *P. aeruginosa* using REIMS

In this chapter, REIMS was used as a direct-from-culture analytical tool to characterise *P. aeruginosa* and to identify key virulence metabolites, which differentiate it from other common CF pathogens. Based on the assumption that REIMS is sensitive enough to detect strain level metabolic differences with the help of powerful statistical tools – namely the Random Forest algorithm –, its utility for typing was tested (Chapter 2, objective 1).

4.1. Introduction

In clinical environments, reliable methods to identify pathogens at the strain level are necessary to map and control any cross-infection. They are also key to understanding microbial evolution and epidemiology [196]. Furthermore, bacterial resistance to antibiotics is thought to be concentrated in a few strains that generally develop in hospital environments [197]. These strains are commonly referred to as “high-risk clones” and can disperse worldwide [198].

The dominant methods for determining strain level identifications rely on genomic analysis. Common methods currently used include pulsed-field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), variable-number tandem repeats (VNTR) and multiple loci VNTR analysis (MLVA), multilocus sequence typing (MLST) and WGS [199, 200] (see summary in Table 4.1).

4.1.1. Strain typing through DNA analysis

PFGE is sometimes considered a gold standard for DNA-fingerprinting of pathogens [198-200]. This electrophoretic method has been adapted for the complex separation of very high mass DNA fragments, lysed beforehand by infrequent cutter restriction enzymes specific to each pathogen. The fragment size patterns of each isolate are then compared to determine their relatedness within one species.

AFLP also requires DNA digestion, using a rare cutter in combination with a frequent cutter, specifically chosen for the organism of interest [201]. The generated restriction fragments are amplified using polymerase chain reaction (PCR) and separated on agarose gel. Similarly to PFGE, each strain will display a different fragment size pattern, although the fragments are smaller in the case of AFLP.

MLVA focuses on amplifying by PCR a reduced number of known loci which comprise VNTR. Tandem repeat genomic regions can be strain specific, thus MLVA determines the number of repeats by calculating the molecular size of each locus. To improve the resolving power and reproducibility, multiple loci are examined. In the case of *P. aeruginosa*, nine loci have been shown to offer a discrimination of strains similar to that obtained with PFGE [202].

DNA-fingerprinting methods in particular, are subject to many potential difficulties, including enzymatic digestion and pattern reading. WGS and MLST methods rely on the DNA sequencing of either the whole genome for the former, or selected housekeeping genes with a high degree of conservation for the latter [203]. These are expensive techniques with long turnaround times, which limits their usefulness in tracking disease outbreaks.

Table 4.1. Some characteristics of classical DNA-based methods for strain typing.

Method	Pulsed-field gel electrophoresis (PFGE)	Amplified fragment length polymorphism (AFLP)	Multiple loci VNTR analysis (MLVA)	Multilocus sequence typing (MLST)	Whole genome sequencing (WGS)
	DNA fingerprinting			DNA sequencing	
	Rare cutter DNA digestion	Rare + frequent cutter DNA digestion + PCR	VNTR loci extraction + PCR	Amplification of housekeeping genes	
Target	High mass DNA fragments	Low mass DNA fragments	VNTR loci	Housekeeping genes	Whole genome
Advantages*	High discrimination	Rapid, requires little amount of DNA	Rapid, portable	Portable, high discrimination	Portable, high discrimination
Drawbacks*	Reproducibility, interpretation, time-consuming, sensitive to minor mutations and laboratory conditions	Reproducibility, interpretation, requires pure DNA	Sensitive to “quickly” evolving VNTR loci	Expensive, time-consuming, low sensitivity for closely related strains	Expensive, time-consuming, management of important data load
Well-suited for*	Epidemiological studies		Outbreak surveillance	Evolutionary studies	Individual transmission tracking

*Non-exhaustive lists.

4.1.2. The potential of mass spectrometry for strain typing

The commercially available clinical Microflex LT and Vitek MS platforms are currently unable to provide strain level resolution. To some extent, MALDI-ToF analysis has shown promise in classifying bacteria using their protein extracts. *Acinetobacter baumannii* outbreak strains [204] and *Staphylococcus aureus* lineages [205] could thereby be differentiated. In another study, biomarker peaks were identified which allowed the identification of five high-risk clones amongst 157 MLST types

with a relative accuracy (PPV = 40-90%) [206]. Fleurbaaij et al. achieved the classification of 18 strains of *P. aeruginosa* in accordance with AFLP clusters using a FT-ICR detector [207]. Indeed, the FT-ICR technology stands out compared to other detectors thanks to an ultrahigh resolution and increased dynamic range, which enables the detection of smaller proteins with better isotopic resolution. This leads to finer MS profiles which can then be more easily differentiated. However, the FT-ICR systems are not yet suited for clinical implementation, since the instruments remain excessively costly and bulky. A few years ago, while demonstrating the utility of REIMS for bacterial speciation, Strittmatter et al. managed to distinguish between seven strains of *Escherichia coli* with 88% of accuracy, revealing the potential of REIMS for subspecies characterisation [149].

4.1.3. Objectives

The DNA-typing methods are time consuming which limits their use to a retrospective surveillance rather than a real-time monitoring to detect and control outbreaks. In addition, the current techniques mainly target slowly evolving characteristics of the pathogens, e.g. housekeeping genes or ribosomal proteins, which are more suited to phylogenetic comparisons. REIMS, on the other hand, is able to detect lower mass metabolites, whose production is subjected to quick changes and reflect the bacterial actual metabolism.

Based on the hypothesis that it is capable of distinguishing variations in the metabolome expressed by different species and even between different strains, REIMS was tested as a tool to characterise *P. aeruginosa* with the objectives of (1) identifying metabolites specific to *P. aeruginosa*, (2) differentiating *P. aeruginosa* from common CF pathogens and other *Pseudomonas* species, and (3) typing *P. aeruginosa* strains.

4.2. Material and methods

4.2.1. REIMS analysis

The clinical isolates retrieved from the Royal Brompton Hospital (London, UK) were collected from different CF patients and included 201 isolates of *P. aeruginosa*, amongst which 156 had VNTR type information, one *P. putida*, 20 *S. aureus*, five *Serratia* species, 22 *Burkholderia* species, 15 *A. xyloxydans* and 15 *S. maltophilia*. Isolates from other *Pseudomonas* species were obtained from Charing Cross Hospital microbiological bank (London, UK) and numbers were subjected to availability.

They included two *P. alcaligenes*, one *P. koreensis*, five *P. monteilii*, one *P. mosselii*, three *P. oryzihabitans*, two *P. putida* and two *P. stutzeri*. They originated from toilets, disposable bins, urine, blood, vagina, nose, wound and rectal infections (see details in Appendix 2).

Isolates were cultured on CBA plates and analysed according to the workflow described in Chapter 3, sections 3.2. *Collection and culturing of microorganisms* and 3.4 *REIMS analysis*. The 156 VNTR-typed *P. aeruginosa* were analysed using laser-assisted REIMS and the automated Tecan platform (section 3.4.2.2. *Automated, laser-assisted REIMS*), applying three “burns” of 10 s per isolate at a laser power of 2 W, with 6 s of waiting time in between. All the other isolates were analysed manually using a handheld bipolar probe (section 3.4.2.1. *Handheld bipolar probe*) with a minimum of three analyses each time. Culturing and analysis of the 45 *P. aeruginosa* isolates without VNTR information was repeated 10 times on different days over the course of six months (see dates of analysis in Appendix 6). MS/MS was performed on 49 *m/z* values detected in *P. aeruginosa* for structural identification following the instructions provided in section 3.4.3. *Tandem MS analysis*.

Statistical analyses included *t*-test, PCA, HCA, PLS-DA and Random Forest classification. The Random Forest algorithm used the whole TIC-normalised *m/z* profile. Implementation on the Metaboanalyst platform was performed on the 5000 most intense median bins and included log-transform and Pareto scaling.

4.2.2. MLST typing

DNA of the 45 *P. aeruginosa* isolates which did not have VNTR information (extracted according to section 3.3. *DNA extraction*) was sent out to the University of Birmingham (Birmingham, UK) where the MicrobesNG service completed WGS. The HiSeq 2500 platform with 2 x 250 bp sequencing was used. Preparation of the DNA sequencing libraries used the NexteraXT kit (Illumina, Cambridge, UK). The HiSeq 2500 platform sequenced the 45 *P. aeruginosa* libraries alongside 339 other DNA libraries (from other customers) leading to a total of 384 libraries on one run. Resulting DNA sequences were merged, trimmed for quality, and aligned to match the *P. aeruginosa* reference genome following the MicrobesNG sequence analysis workflow. DNA sequences relevant to the seven housekeeping genes of interest in the MLST approach [208] were examined by Alvaro Perdonés-Montero, research assistant at the department of Surgery and Cancer, Imperial College London (UK). Strain match searches were completed on the PubMLST website using standard operating parameters and led to the identification of five novel MLST types, which were assigned new ST numbers.

4.3. Results

4.3.1. Metabolic characterisation of *P. aeruginosa*

A set of 45 isolates from individual CF subjects were selected from the thousands of isolates stored at the CF repository of the NHLI (London, UK) in an attempt to represent the diversity of infecting strains. They originated from patients with varying age, clinical status, infection stage, mucoid phenotype, and antibiotic susceptibility.

A representative mass spectrum obtained with REIMS on *P. aeruginosa* is shown in Figure 4.1; m/z regions of metabolic interest are highlighted.

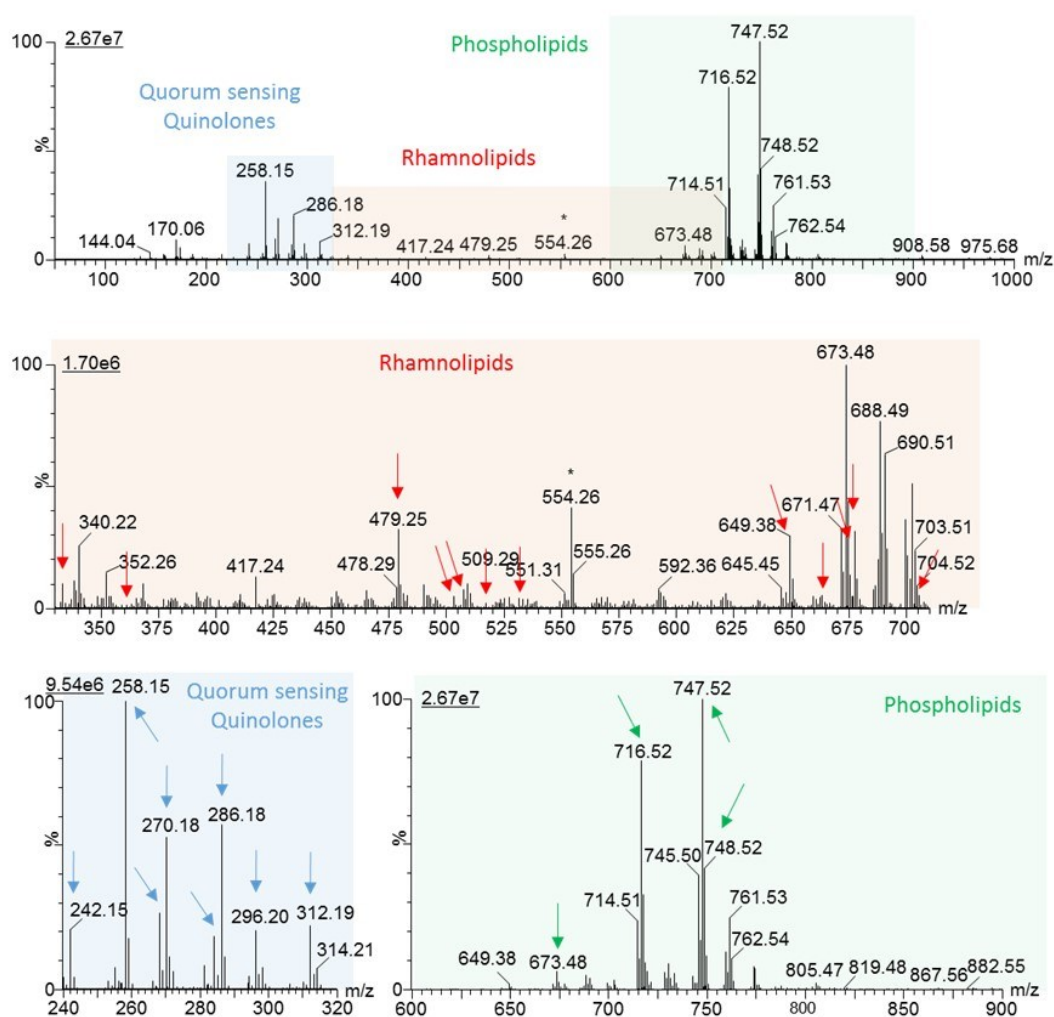


Figure 4.1. Typical mass spectrum obtained from *P. aeruginosa* isolates using REIMS.

Acquired with manual REIMS, in negative ion detection mode on CF_01, grown for 24h on CBA. After background removal and mass drift correction using the LeuEnk ion ($m/z = 554.26$); highlighted spectral regions include (a) the 50-1000 m/z range showing the most intense features; (b) the rhamnolipid m/z range of 330-710; (c) the AQ m/z range of 240-320; and (d) the phospholipid m/z range 600-900. Arrows highlight some of the annotated spectral features that were determined based on MS/MS spectral data (reproduced under the Creative Commons Attribution license from Ref. [209]).

Tandem MS was used to identify 49 spectral features based on their fragmentation pattern (Appendix 3). Some of the most intense peaks are listed in Table 4.2. Most of the phospholipids were identified as PEs and PGs. Twelve m/z values corresponded to AQs, including quinolones, hydroxy-quinolones and their structural isomer quinolone-*N*-oxides. These two isomers have similar masses and fragmentation patterns and could not be differentiated. Identified AQs included alkyl chain of seven, nine or eleven carbons, with and without unsaturation. They constitute part of the quorum sensing communicating system of *P. aeruginosa* and contribute to the coordination of its virulence mechanisms. The widely reported PQS, mixed with 2-heptyl-4(1*H*)-quinolone-*N*-oxide (HQNO), usually showed the most intense signal at $m/z=258.1472$. It should be noted that AHLs, usually detected as positive ions, have not been observed when testing the positive ionisation mode. Likewise, the identity of 17 rhamnolipids was confirmed. They make a multilevel contribution to the bacterial adaptation to its environment by modifying surface tension. This is modulated through the production of various combinations of rhamnose sugars with lipidic acyl chains. The secreted profiles rely on carbon source availability but are also strain-dependant [102]. Rha-C10-C10 and Rha-Rha-C10-C10 are the most commonly reported compounds [102]. Within the present population of CF *P. aeruginosa* isolates, the di-rhamnolipids Rha-Rha-C10, Rha-Rha-C10-C10, and Rha-Rha-C10-C12 were highly prevalent. Disparity in the produced metabolome may be due to genetic variations and/or differences in growing conditions.

Table 4.2. Some of the most intense spectral features identified in *P. aeruginosa*.
Identified by MS/MS using REIMS in negative ionisation mode.

	Experimental m/z	Molecules identified through MS/MS
Quorum sensing molecules	242.15	2-heptyl-4(1 <i>H</i>)-quinolone (HHQ)
	258.15	PQS and/or 2-heptyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide (HQNO)
	270.18	2-nonyl-4(1 <i>H</i>)-quinolone (NHQ)
	286.18	2-nonyl-3-hydroxy-4(1 <i>H</i>)-quinolone (C9-PQS) and/or 2-nonyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide (C9-NQNO)
	296.20	2-undecenyl-4(1 <i>H</i>)-quinolone (C11:db-UHQ)
	312.19	2-undecenyl-3-hydroxy-4(1 <i>H</i>)-quinolone (C11:db-PQS) and/or 2-undecenyl-4(1 <i>H</i>)-quinolone- <i>N</i> -oxide (C11:db-UQNO)
Rhamno-lipids	479.25	Rha-Rha-C10
	649.38	Rha-Rha-C10-C10
	677.41	Rha-Rha-C10-C12
Phospho-lipids	673.48	PA(18:1/16:0)
	714.51	PE(18:1/16:1)
	716.52	PE(18:1/16:0)
	745.50	PG(18:1/16:1)
	747.51	PG(18:1/16:0)
	761.53	PG(19:1/16:0); PG(18:1/17:0); PG(18:0/17:1)
	773.53	PG(18:1/18:1); PG(19:1/17:1)

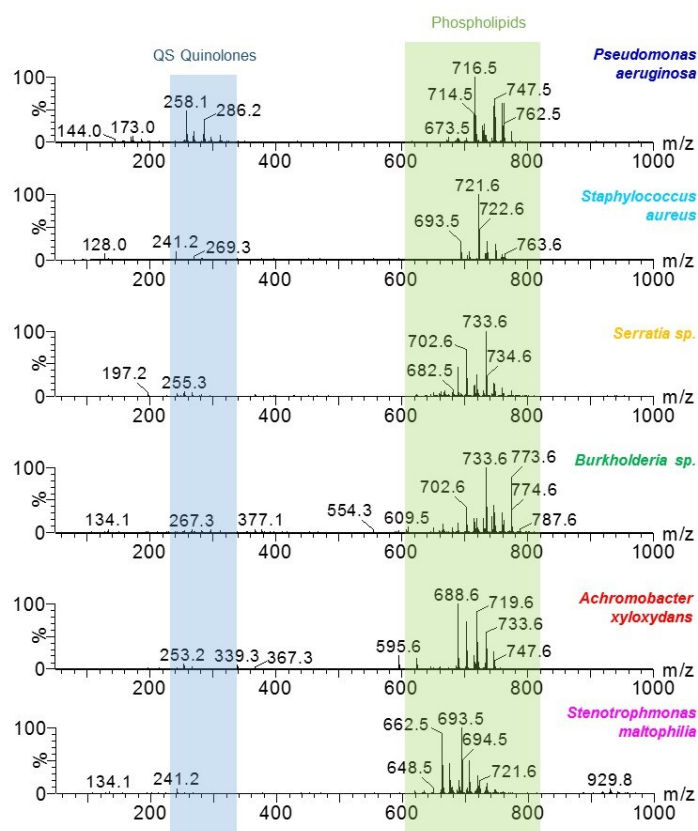
It is worth noting that amongst the 45 isolates, 19 had been described as mucoid when isolated at the clinic. Whilst isolates were cultured 10 times to evaluate REIMS reproducibility, only six grew mucoid more than once and were considered mucoid in our study (CF_02, CF_08, CF_12, CF_28, CF_35, CF_39). Indeed, the same isolate can display variable mucoid characteristics when re-cultured. This is important since the higher water content of mucoid strains impacts the profile detected by REIMS leading to lower reproducibility, higher background, global lower intensities especially in AQs, and different phospholipid profiles (see an example of a mass spectrum obtained on a mucoid isolate in Appendix 4). Although it was not possible to establish whether this actually reflects the bacterial metabolism or is an analytical effect, the physical nature of the sample most likely impacts the results.

4.3.2. Classification of CF pathogens at the genus level

REIMS profiles were acquired on six pathogens commonly isolated in CF patients: *P. aeruginosa*, *S. aureus*, *Serratia* species, *Burkholderia* species, *A. xyloxydans*, and *S. maltophilia*. Typical mass spectra for each microbe are presented in Figure 4.2a, and the metabolic information has been summarised using PCA in the graph Figure 4.2b.

The spectral profiles showed differences mainly in the phospholipid mass range, $m/z=600-900$, and *P. aeruginosa* exhibited a distinctive signal in the lower mass range, $m/z=240-320$, which was shown to comprise AQs. The first principal component in Figure 4.2b opposes the Gram-negative bacteria to the Gram-positive *S. aureus*, and a clear separation of *P. aeruginosa* was observed using the two first principal components. Although not visible on a two-dimensional graph, the other species are well separated on the following principal components. When using the Random Forest algorithm, the accuracy of microbial classification was 100%, confirming the previous findings of Strittmatter et al. [149].

a. Mass spectra



b. Scores plot

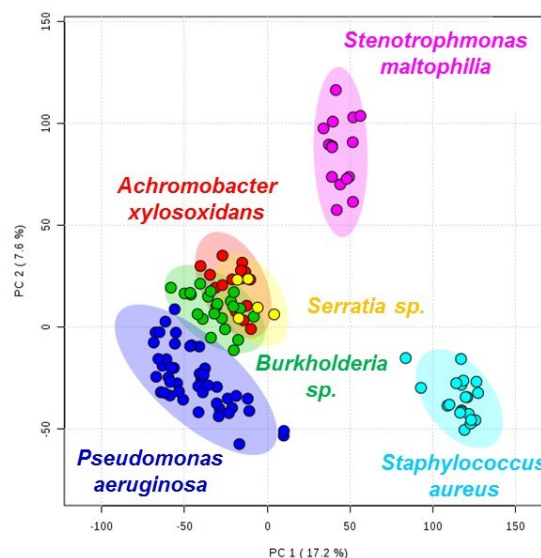


Figure 4.2. REIMS analysis of common CF pathogens.

(a) Typical mass spectra obtained in the negative ionisation mode, highlighting high intensity regions relevant to *P. aeruginosa*. (b) PCA score plot showing the two first principal components using the REIMS spectral data. Shaded areas show 95% confidence intervals of the sample groups (built on the Metaboanalyst platform).

4.3.3. Exploring REIMS capacities for *Pseudomonas* speciation

In order to investigate the differences between *Pseudomonas* species, supplementary data were acquired on *Pseudomonas alcaligenes*, *koreensis*, *monteilii*, *mosselii*, *oryzihabitans*, *putida*, and *stutzeri*. These are not common in CF patients and numbers were limited by sample availability at Charing Cross Hospital microbiological bank. Spectral data were added to the previous dataset acquired on the CF pathogens exploited in section 4.3.2. *Classification of CF pathogens at the genus level*. Since the number of isolates per species was insufficient to be predicted separately, they were grouped under a *Pseudomonas non-aeruginosa* class. In this configuration, we are evaluating REIMS ability to differentiate the *P. aeruginosa* profile from other pathogens – including other *Pseudomonas* species – rather than its capacity to classify at the individual species level.

The PCA in Figure 4.3 shows an overlap of the *Pseudomonas* species on the two first principal components. A prediction model using Random Forest reached 95% of global accuracy, yet a significant number of *Pseudomonas* species were mistakenly predicted as *aeruginosa*. As a consequence, sensitivity was 100% but precision (or PPV) was 88% for *P. aeruginosa*.

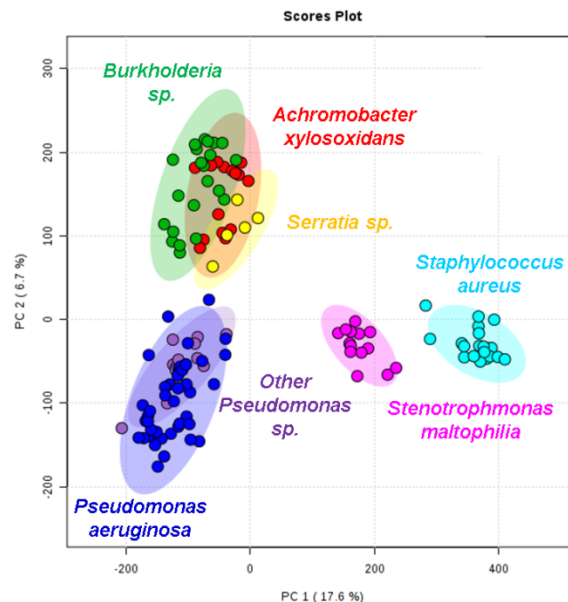


Figure 4.3. PCA score plot using REIMS spectral data acquired on common CF pathogens and *Pseudomonas* species.

Shows the two first principal components; shaded areas show 95% confidence intervals of the sample groups (built on the Metaboanalyst platform).

The dendrogram in Figure 4.4 enables a more global view of the metabolic similarity of the samples. The *Pseudomonas* species appeared mixed again whilst being well separated from the other genera.

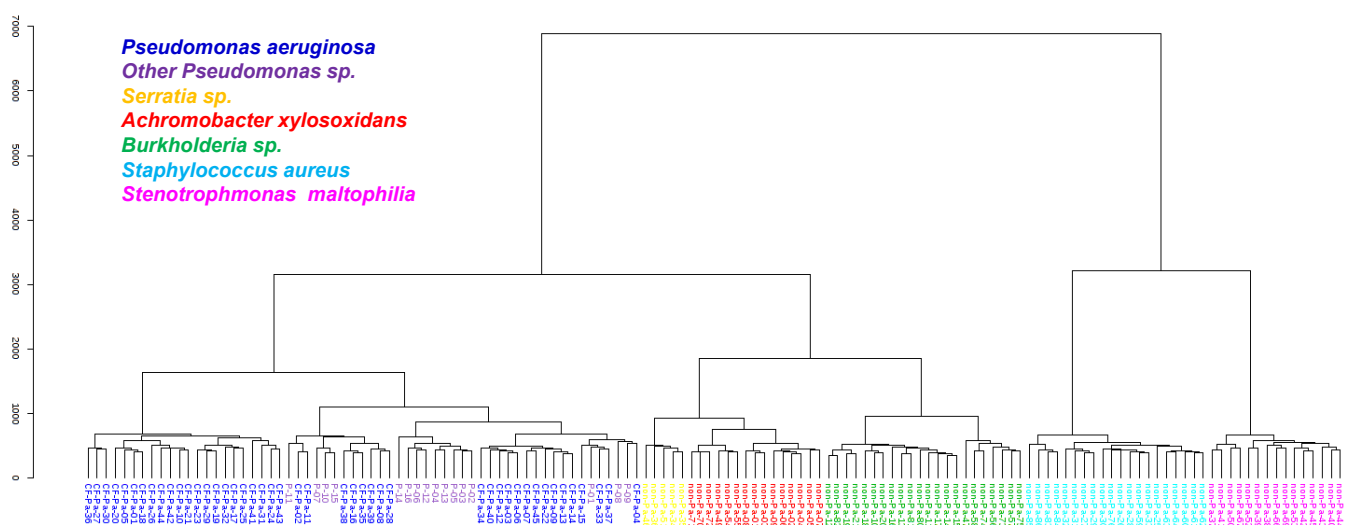


Figure 4.4. HCA using REIMS spectral data acquired on common CF pathogens and *Pseudomonas* species. (Built on the Metaboanalyst platform).

Further considering the *Pseudomonas* genus only, HCA showed that isolates from the same species tended to cluster close to each other (see dendrogram in Figure 4.5), i.e. they exhibited similar metabolic profiles. Notably, most of the *P. aeruginosa* clustered under one main branch. Examination of the spectral data revealed differences in the phospholipid profile and in the lower mass range including the AQs; examples of mass spectra are available in Appendix 5.

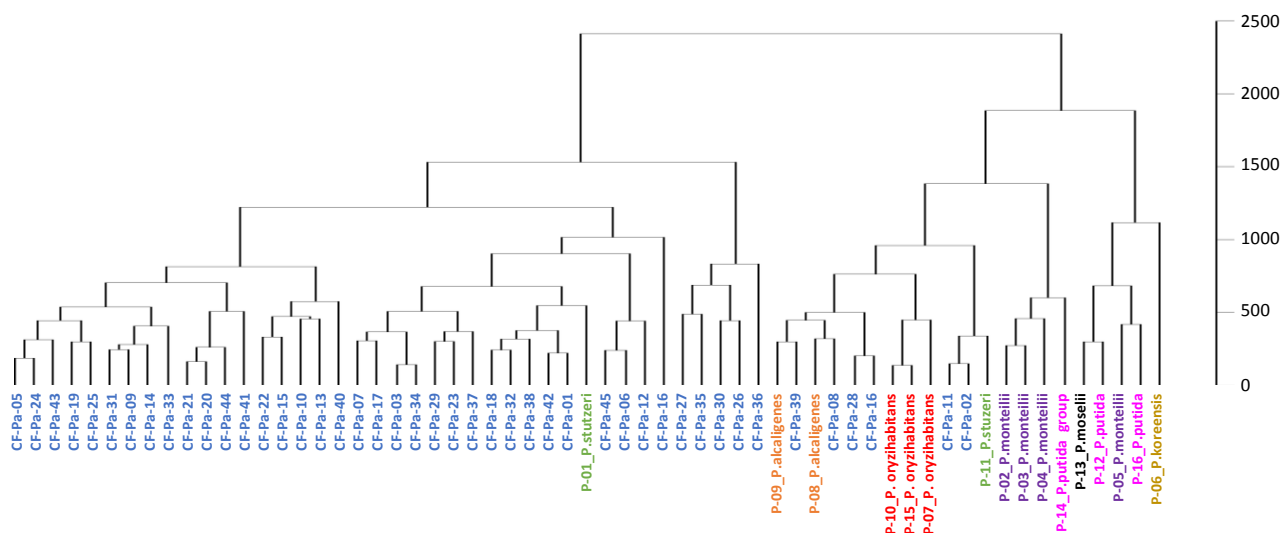


Figure 4.5. HCA using REIMS spectral data acquired on *Pseudomonas* species.

Built in R using the 100 most intense features; *P. aeruginosa* is in dark blue.

A heatmap was built using the 50 most discriminant spectral features, selected using a *t*-test (Figure 4.6). These include mainly features from the phospholipid *m/z* range and some lower mass metabolites, and were generally less intense in the non-*aeruginosa* species. In this configuration, only five *P. aeruginosa* were mixed with the other *Pseudomonas* species (CF_02, CF_08, CF_11, CF_28 and CF_39), which could be explained by the fact that four of these grew with a mucoid phenotype, therefore displaying a less intense profile.

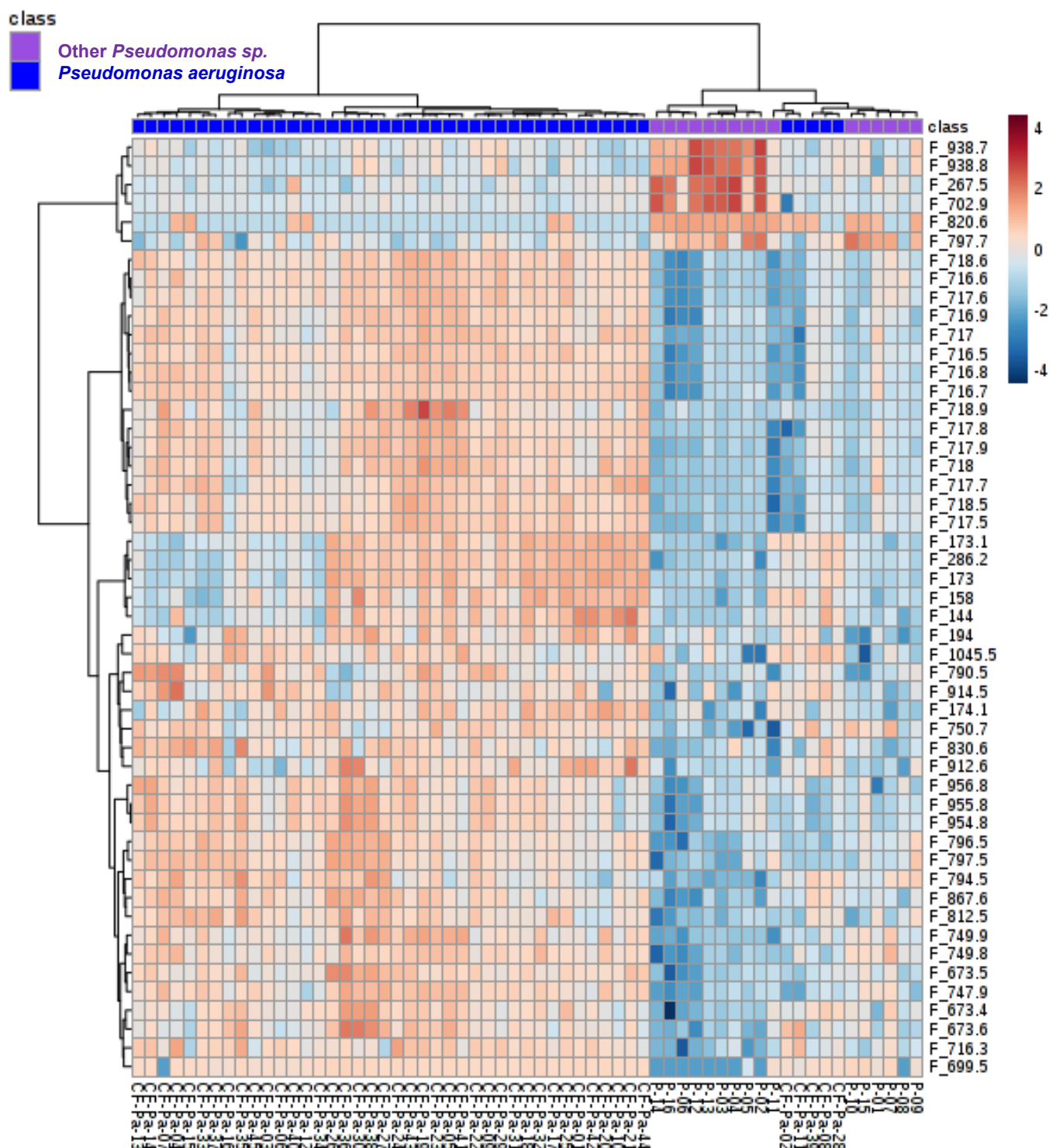


Figure 4.6. HCA and associated heatmap built using REIMS spectral data obtained from *Pseudomonas* isolates. Uses the top 50 most differentiating mass spectral bins, based on p-values of univariate t-tests. *P. aeruginosa* isolates collected from CF patients are shown in dark blue, other *Pseudomonas* species collected from hospital acute infection are in purple (built on the Metaboanalyst platform).

The detection of *P. aeruginosa* amongst the *Pseudomonas* genus using random forest was 90% accurate, with six isolates mistakenly classified as *P. aeruginosa* (cross-validation in Figure 4.7).

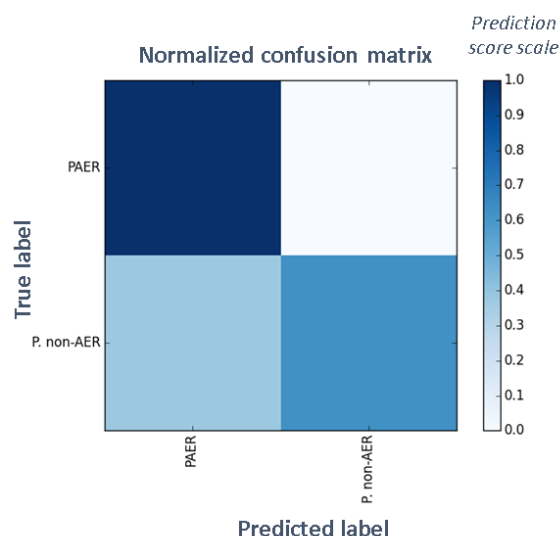


Figure 4.7. Confusion matrix for Random Forest classification using REIMS spectral data obtained from *P. aeruginosa* and other *Pseudomonas* species.

These are abbreviated PAER and P. non-AER, respectively. Real and predicted identities are displayed respectively on the left-hand and the bottom sides. Frequency of prediction for each isolate is represented in darkening blue. Overall classification was 90% accurate.

Although the small number of isolates available for each species restricted the assessment of REIMS as a speciation tool, the differences observed in the detected metabolic profiles indicated that REIMS has the power to differentiate *Pseudomonas* up to the species level, and led to an acceptable identification of *P. aeruginosa* within its genus.

4.3.4. Tentative classification of strain types

4.3.4.1. Prediction of the isolate identity

Following the observation that CF isolates demonstrated a high metabolic diversity, analyses were repeated 10 times on different days over the course of six months in order to assess REIMS reproducibility and its ability to recognise *P. aeruginosa* clones. The dataset comprised 45 CF isolates, each with 10 replicates which can be considered clones of the original isolate, although adaptation may occur when re-culturing. PCA showed that the date effect was the first contributor to variance (first principal component of 9.4%, data not shown). A prediction model was built employing the Random Forest algorithm. The results of the cross-validation using a leave-one-out procedure are presented in Table 4.3 and in Figure 4.8.

Table 4.3. Random Forest cross-validation results for the classification of replicates amongst 45 *P. aeruginosa* isolates.

Leave-one-out cross-validation applied on a REIMS dataset including 10 replicates per isolate. MLST type information are included; isolates which grew mucoid more than once are indicated with a star. Overall accuracy of the prediction was 81%.

Isolate	Precision	Recall	F1 score	MLST type
CF_01	0.91	1.00	0.95	ST_0395 (VNTR cluster D)
CF_02*	0.64	0.90	0.75	ST_1207
CF_03	1.00	0.90	0.95	ST_3018
CF_04	0.70	0.70	0.70	ST_0110
CF_05	0.00	0.00	0.00	ST_0390
CF_06	0.91	1.00	0.95	ST_3017
CF_07	0.77	1.00	0.87	ST_0146 (LES)
CF_08*	0.69	0.90	0.78	ST_0195
CF_09	0.88	0.70	0.78	ST_0954
CF_10	1.00	1.00	1.00	ST_0298
CF_11	0.73	0.80	0.76	ST_2362
CF_12*	1.00	0.60	0.75	ST_1641
CF_13	0.77	1.00	0.87	ST_2951
CF_14	1.00	1.00	1.00	ST_2951
CF_15	0.91	1.00	0.95	ST_0560
CF_16	0.64	0.70	0.67	ST_0633
CF_17	1.00	1.00	1.00	ST_0253 (PA14)
CF_18	0.00	0.00	0.00	ST_0390
CF_19	0.73	0.80	0.76	ST_0132
CF_20	0.12	0.10	0.11	ST_0390
CF_21	1.00	1.00	1.00	ST_0612
CF_22	0.83	1.00	0.91	ST_0009
CF_23	1.00	1.00	1.00	ST_0235 (VNTR complex B)
CF_24	0.88	0.70	0.78	ST_2859
CF_25	1.00	0.90	0.95	ST_0017 (VNTR clone C)
CF_26	0.78	0.70	0.74	ST_0017 (VNTR clone C)
CF_27	0.82	0.90	0.86	ST_0175 (VNTR complex B)
CF_28*	0.86	0.60	0.71	ST_0014
CF_29	1.00	0.80	0.89	ST_1248
CF_30	1.00	1.00	1.00	ST_0244
CF_31	1.00	0.90	0.95	ST_0252
CF_32	1.00	1.00	1.00	ST_0253 (PA14)
CF_33	0.91	1.00	0.95	ST_0298
CF_34	0.88	0.70	0.78	ST_0381
CF_35*	1.00	0.70	0.82	ST_1641
CF_36	1.00	0.90	0.95	ST_0244
CF_37	0.80	0.80	0.80	ST_3020
CF_38	1.00	0.70	0.82	ST_3019
CF_39*	0.82	0.90	0.86	ST_0395 (VNTR cluster D)
CF_40	0.56	0.50	0.53	ST_0274
CF_41	1.00	1.00	1.00	ST_3016
CF_42	0.73	0.80	0.76	ST_0017 (VNTR clone C)
CF_43	0.77	1.00	0.87	ST_1225
CF_44	0.90	0.90	0.90	ST_0179
CF_45	0.82	0.9	0.86	ST_0217
Average / Total	0.82	0.81	0.81	

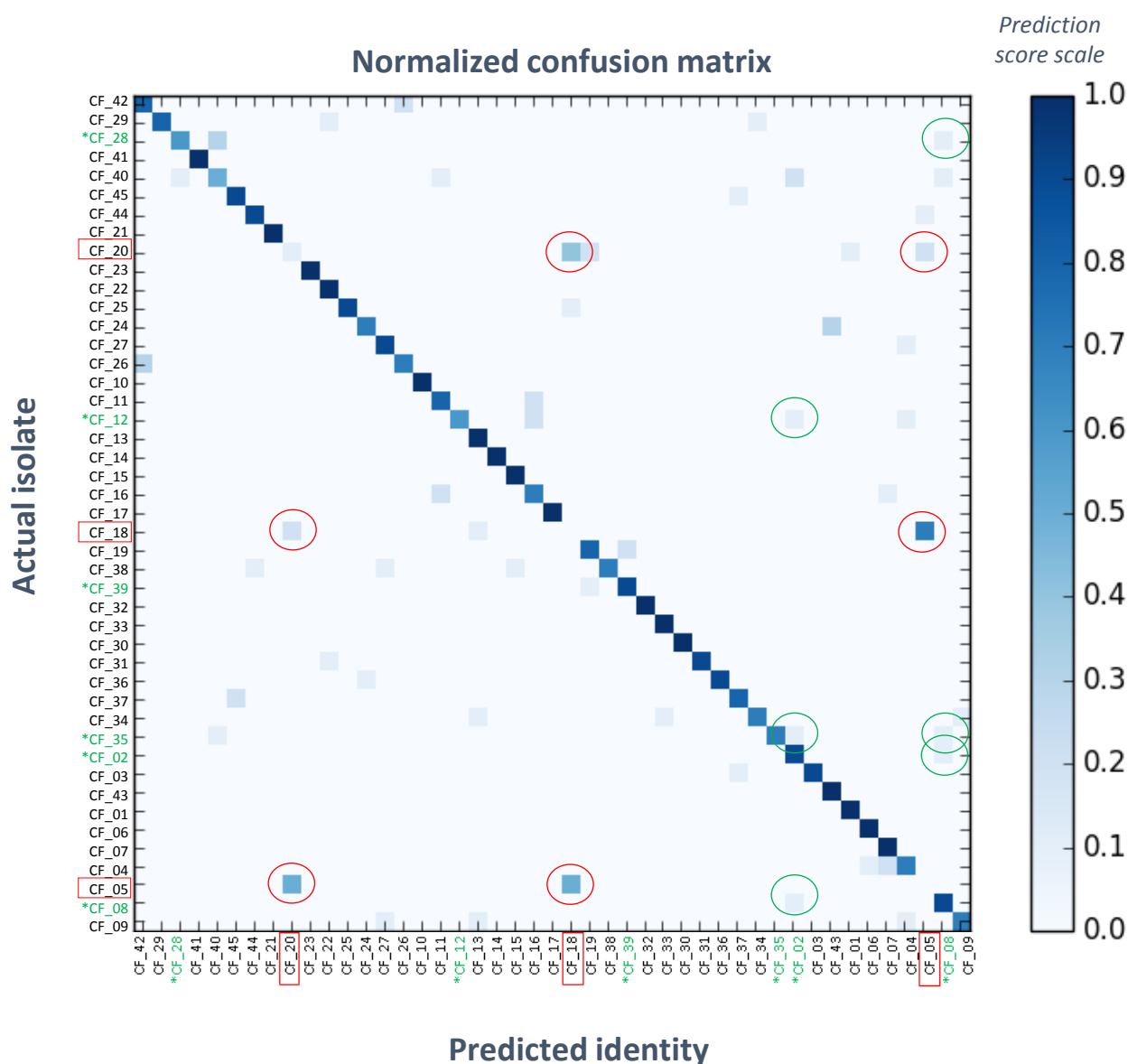


Figure 4.8. Confusion matrix for Random Forest classification of replicates amongst 45 *P. aeruginosa* isolates. Leave-one-out cross-validation applied on a REIMS dataset including 10 replicates per isolate. Real and predicted identities are displayed respectively on the left-hand and the bottom sides. Frequency of prediction for each isolate is represented in darkening blue. Overall prediction was 81% accurate. Isolates circled in red (CF_05, CF_18, CF_20) were predicted as each other; isolates circled in green are mucoid isolates predicted as one another (adapted from Ref. [209] under the Creative Commons Attribution license).

The overall accuracy of the prediction model, calculated as the proportion of replicates correctly classified, was 81%. Notably, three isolates were predicted as each other, being the only ones with a F1-score below 50%; namely CF_05, CF_18, and CF_20 scoring 0%, 0% and 11% respectively. On the other hand, more than a third of the isolates had F1-score equal or above 90%. This suggests that the metabolic profile detected with REIMS, i.e. AQs, rhamnolipids and phospholipids, is highly specific to individual isolates, and that REIMS reproducibility is good enough to detect it, despite the day/batch effect. Yet, the mucoid isolates (circled in green in Figure 4.8) globally showed a lower reproducibility,

and tended to be predicted as each other (particularly as CF_02 and CF_08). It is worth emphasizing that the mucoid phenotype itself was not reproducible.

In order to compare the samples' metabolic profiles, and identify similarities which could explain prediction errors, a mean MS profile was calculated for each isolate – i.e. intensities were averaged for each m/z bin using the ten replicates. Intensities were averaged again per bin across all the isolates in order to select the 100 most intense mean bins to perform HCA. The resulting heatmap shown in Figure 4.9 confirms the dominance of phospholipids and quorum sensing molecules in the metabolic profile of *P. aeruginosa*. Notably, the three isolates which were previously misidentified as each other clustered together. Their metabolic profiles were so similar that the model predicted them as one another. Interestingly, a subgroup of isolates displayed higher intensities of AQs and of some specific lipids (i.e. PG(35:2) and PG(35:1)), whilst other lipids exhibited lower levels (i.e. PE(32:1), PE(32:0) and PG(32:1)). This might be an illustration of the involvement of AQs in the regulation of the biochemical structure of bacterial cell walls. These isolates may have had to adapt to comparable conditions, or faced analogous stress, resulting in similarities in their membrane composition. Further study may unravel clinical common traits and metabolic pathways linking AQs to the synthesis of particular lipids.

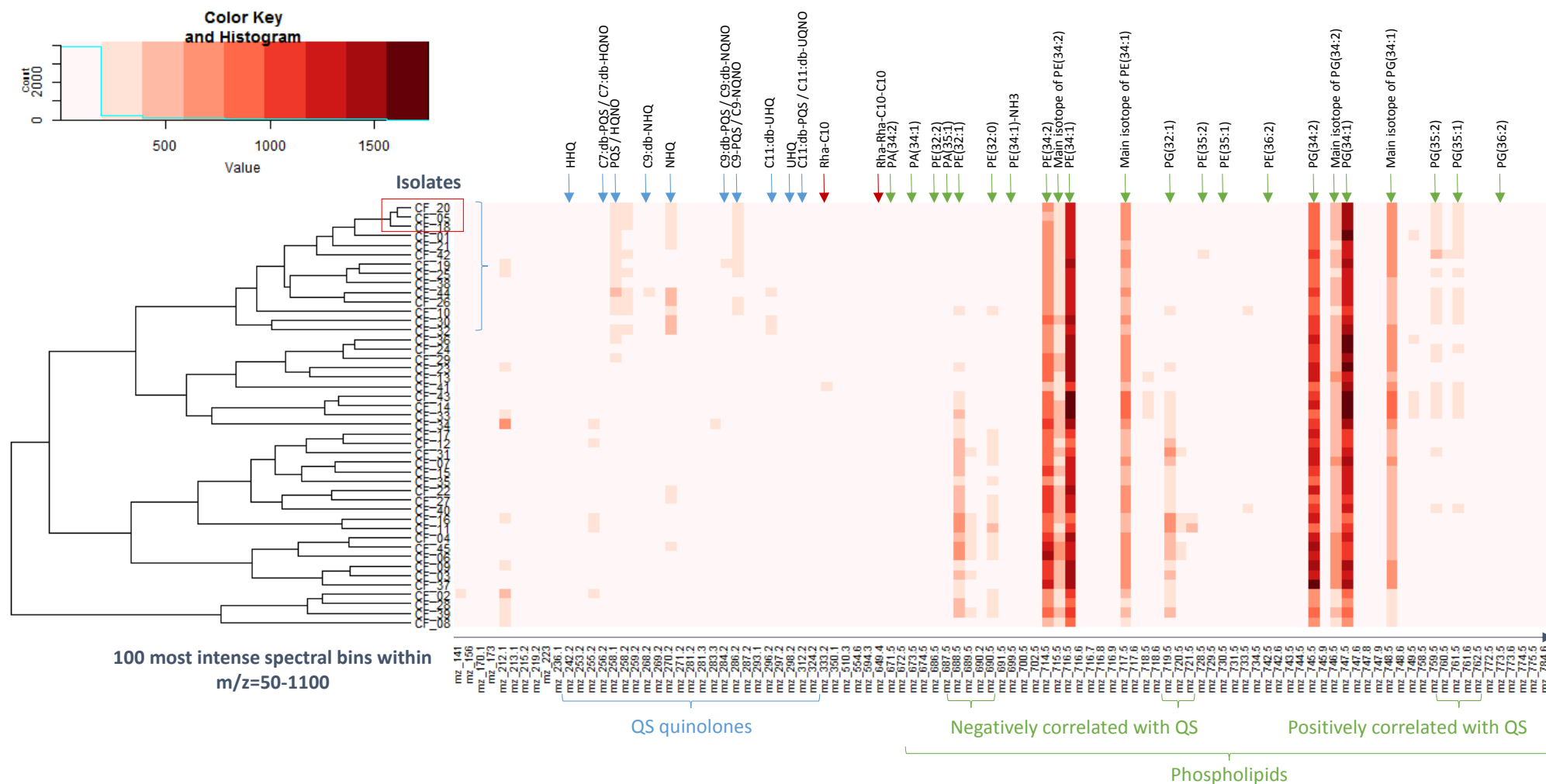


Figure 4.9. HCA and associated heatmap showing clustering of 45 *P. aeruginosa* isolated from CF lungs.

Built in R using the top 100 most intense features detected with REIMS within the 50-1100 m/z range, and spectra averaged on 10 replicates for each isolate. Column annotations depict identified metabolic features. Isolates in the red box that cluster tightly together correspond to the isolates that were previously misclassified as each other (circled in red in Figure 4.8) (reproduced under the Creative Commons Attribution license from Ref. [209]).

4.3.4.2. Classification according to MLST type

Following isolate classification, the 45 *P. aeruginosa* samples isolated from CF patients were typed by MLST in order to evaluate the potential of REIMS to assign sub-species strain types to *P. aeruginosa*. MLST information for each isolate is provided in Table 4.3. A number of 35 different STs were identified, notably including four recognised high-risk strains of *P. aeruginosa* [204-206] – ST_0175, ST_0235 (VNTR complex B), ST_0253 (CF specific PA14), ST_0395 (VNTR cluster D) [206, 210] – , the Liverpool epidemic strain (LES) – ST_0146 [210] – and five novel types. Amongst these 35 types, eight comprised at least two isolates: ST_0017 (corresponding to the ubiquitous VNTR clone C [211], 3 isolates), ST_0179 (2 isolates), ST_0244 (2 isolates), PA14 (2 isolates), ST_0298 (2 isolates), ST_0390 (3 isolates), ST_0395 (VNTR cluster D, 2 isolates), ST_1641 (2 isolates); the remaining 27 MLST types had unique representatives. The previously poorly predicted isolates (CF_05, CF_18, and CF_20) belonged to the same strain (ST_0390). Additionally, isolates CF_26 and CF_42 which were previously confused as each other shared the same MLST sequence (ST_0017 or clone C) together with CF_25, highlighting that REIMS is able to detect some strain level similarities.

The eight MLST types with more than one representative were retained for further analysis (leading to the selection of 18 isolates). Since the number of samples per MLST type was limited (maximum 3), the 10 replicates of the 18 isolates were used for statistical analysis and the 180 analyses were grouped according to MLST type. The Random Forest algorithm was used to build the classification model, and the prediction was tested applying a leave-one-isolate-out procedure. This entails that the replicates belonging to the same isolate as the tested sample were excluded when building the model to avoid confounding effects (i.e. prediction driven by isolate identity rather than MLST type). The model employed the whole *m/z* range, including phospholipids as well as the lower mass molecules AQs and rhamnolipids, which have been shown to vary between strains of *P. aeruginosa* [212]. The accuracy of the MLST prediction for the 180 replicates, using leave-one-isolate-out cross-validation, is presented in the confusion matrix Figure 4.10, individual F1-scores are provided in Table 4.4.

Table 4.4. Random Forest cross-validation results for the classification of 18 *P. aeruginosa* isolates amongst eight MLST types.

Leave-one-isolate-out cross-validation applied on a REIMS dataset of 18 isolates with 10 replicates each; isolates which grew mucoid more than once are indicated with a star. Overall accuracy of the prediction was 33%.

MLST type	Number of isolates	Isolates	Precision	Recall	F1-score	Number of replicates
ST_0017 (VNTR Clone C)	3	CF_25 ; CF_26 ; CF_42	0.97	0.87	0.91	30
ST_0179	2	CF_14 ; CF_44	0.00	0.00	0.00	20
ST_0244	2	CF_30 ; CF_36	0.00	0.00	0.00	20
ST_0253 (PA14)	2	CF_17 ; CF_32	0.00	0.00	0.00	20
ST_0298	2	CF_10 ; CF_33	0.44	0.20	0.28	20
ST_0390	3	CF_05 ; CF_18 ; CF_20	0.49	0.93	0.64	30
ST_0395 (VNTR Cluster D)	2	CF_01 ; CF_39*	0.00	0.00	0.00	20
ST_1641	2	CF_12* ; CF_35*	0.05	0.05	0.05	20
Average / Total	18	-	0.30	0.33	0.30	180

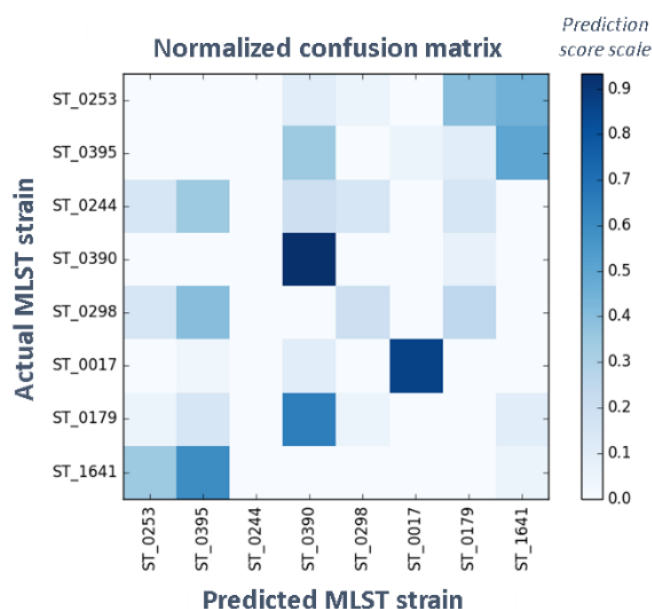


Figure 4.10. Confusion matrix for Random Forest classification of 18 *P. aeruginosa* isolates amongst eight MLST types.

Leave-one-isolate-out cross-validation applied to a REIMS dataset of 10 replicates for each isolate, grouped per MLST type. Real and predicted identities are displayed respectively on the left-hand and the bottom sides. Frequency of prediction for each isolate is represented in darkening blue. Overall prediction was 33% accurate.

The global accuracy of the prediction model of 33% and, in most of the cases, it was not possible to predict the MLST type of one sample based on the replicates of one or two other isolates. Nevertheless, there were two types encompassing three isolates each that achieved promising prediction scores of 91% (ST_0017 or clone C) and 64% (ST_0390). Both of them were globally enhanced when compared to the previous prediction of individual isolates: CF_25 (F1-score 95%), CF_26 (74%) and CF_42 (76%) belonging to ST_0017 (91%); CF_05 (0%), CF_18 (0%), and CF_20 (11%) belonging to ST_0390 (64%). This indicates there may be some correlations between the REIMS profile

and the MLST type. However, the small number of independent isolates does not allow conclusions to be drawn on the typing power of REIMS.

The MS data have been examined, seeking for potential causes of low F1-scores. Spectral differences were observed between the isolates, in the quorum sensing profile for ST_0179, and the phospholipid region for ST_0395. The type ST_1641 comprises two mucoid isolates, which have demonstrated lower reproducibility. On the whole, this illustrates that REIMS can sometimes differentiate isolates within an MLST type, thereby providing potentially higher resolution information than methods based on slowly evolving genomic regions such as MLST typing.

4.3.4.3. Classification according to VNTR type

To further compare REIMS classification with DNA-typing, different CF isolates which had been typed by MLVA were obtained from the Royal Brompton biobank. The 156 isolates were distributed amongst nine reputed VNTR types, each comprising between five and 35 samples. The prediction model was built as before, processing the full m/z range with the Random Forest algorithm, and tested through leave-one-out cross-validation. The consequent confusion matrix is presented in Figure 4.11 and detailed scores are in Table 4.5; the overall accuracy was just above 50%.

Table 4.5. Random Forest cross-validation results for the classification of nine *P. aeruginosa* VNTR types.
Leave-one-out cross-validation applied on a REIMS dataset of 156 *P. aeruginosa* isolates, grouped per VNTR type. Overall prediction was 51% accurate.

VNTR type	Precision	Recall	F1-score	Number of isolates
Clone C	0.46	0.74	0.57	31
Cluster A	0.29	0.15	0.20	13
Cluster D	0.33	0.12	0.18	16
Cluster E	0.00	0.00	0.00	10
LES	0.64	0.83	0.73	35
Manchester strain	0.42	0.74	0.54	19
Midlands_1 strain	0.00	0.00	0.00	10
Nottingham Trent strain	0.00	0.00	0.00	5
PA14	0.64	0.53	0.58	17
Average / Total	0.42	0.51	0.44	156

Whilst concordance between VNTR and REIMS classification was low, some types demonstrated promising results. For instance, the highly contagious LES – which encompassed the highest number of isolates – had an F1-score of 73%, with a sensitivity (recall) of 83%. This implies that a significant number of the LES isolates were correctly identified using their REIMS profiles (few false negatives). However, the precision was 64%, meaning that there was an important risk of false positives, including

in particular some Clone C and Midlands_1 isolates (as visible in Figure 4.11). The VNTR profile of Cluster E is very close to that of clone C, since only 2 loci can differ [210]. They must be therefore genetically very similar, potentially producing comparable metabolomes, which could explain why most of the cluster E strains have been predicted as clone C using their REIMS profiles (Figure 4.11). The other confusions in prediction could not be explained through the study of the individual VNTR profiles.

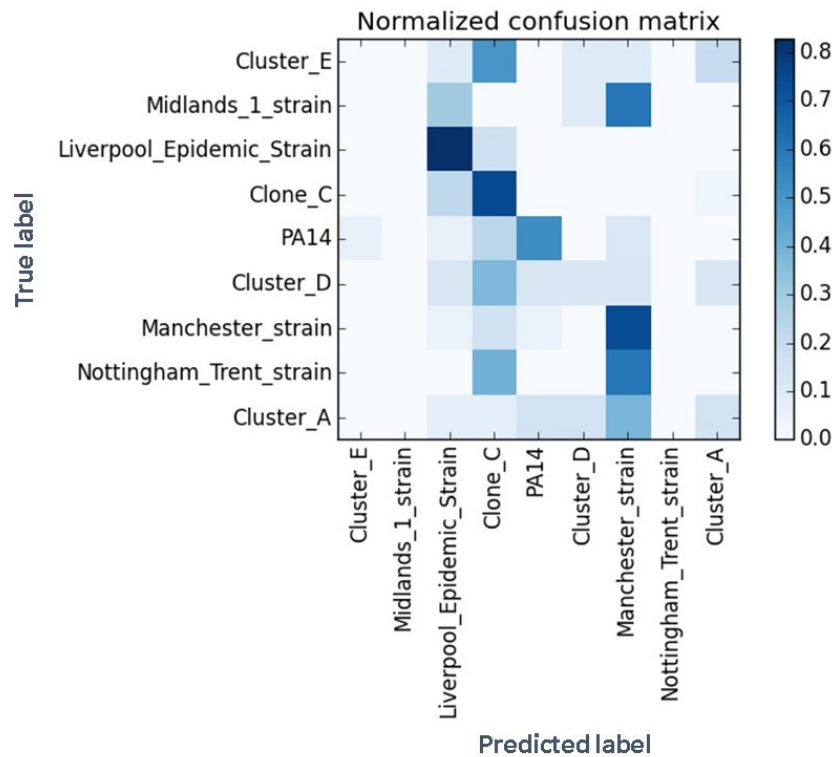


Figure 4.11. Confusion matrix for Random Forest classification of nine *P. aeruginosa* VNTR types.

Leave-one-out cross-validation applied on a REIMS dataset of 156 *P. aeruginosa* isolates, grouped per VNTR type. Real and predicted identities are displayed respectively on the left-hand and the bottom sides. Frequency of prediction for each isolate is represented in darkening blue. Overall prediction was 51% accurate.

Of interest are the results obtained for Clone C, Cluster D and PA14 (57%, 18% and 58% respectively), since these were also included in the MLST classification (F1-scores of 91%, 0%, 0%, respectively). The differences between the two experimental plans, and most of all the limited number of MLST-typed samples, prevent any direct comparison between the MLST and VNTR classifications. It can only be observed that there was some concordance in prediction accuracy for Clone C and Cluster D, but not for PA14, which was amongst the highest prediction scores in the VNTR classification and could not be classified according to MLST.

4.4. Discussion

4.4.1. REIMS as an alternative to MALDI-ToF for pathogen identification

The introduction of MS-based platforms such as MALDI-ToF have revolutionised the clinical workflows within microbiology laboratories, most notably for speciation. Clinical instruments (Microflex LT, Bruker Daltonics; Vitek MS, bioMérieux) are now widely used and provide a high-throughput and user-independent species identification. The accuracy is 80-90%, since some genera have proven problematic, although this is not the case of *Pseudomonas*. Speciation accuracy with laser-assisted REIMS and the Tecan-automated platform is now around 90%, however the libraries currently available for MALDI-ToF cover more species [73, 152]. Here, REIMS' capacity to differentiate *P. aeruginosa* from other common CF pathogens was evaluated. Although classification at the species level could not be assessed, *P. aeruginosa* was accurately detected amongst other *Pseudomonas* species at 90%. Metabolic differences were visible and it is likely that an increased number of isolates would allow the differentiation of species, as previously achieved for the Gram-negative genera *Enterobacter* and *Klebsiella*, and the Gram-positive *Staphylococcus*, *Enterococcus*, and *Streptococcus* [149, 152]. Furthermore, REIMS analysis can be carried out on selected colonies directly after the first sputum culture, and may allow the detection of co-infection at once, similarly to what is currently implemented in clinical laboratories with MALDI-ToF. Therefore, REIMS offers a quick and reliable alternative for identifying pathogens commonly isolated in CF patients.

4.4.2. Detection of virulence factors and potential for a direct-from-sample diagnostic tool?

MALDI-ToF and REIMS distinguish in terms of capacities and applications notably due to differences in the methodology and analytical process. In the case of MALDI-ToF, speciation relies on fingerprints of highly conserved ribosomal proteins in the high m/z range, which do not bring additional biochemical knowledge. Most of the time, peaks originating from the matrix in addition to the poor resolution of the linear ToF, restrains the study of low m/z metabolites, or require the addition of extensive preparative steps. In contrast, REIMS can be applied directly on bacterial cultures, and allows the detection of the more versatile lower m/z metabolic profile. This encompasses the phospholipid region, which mainly drives pathogen classification, and virulence-associated AQs and rhamnolipids. These are key factors in the evolution of an infection towards chronicity and may be useful indicators if they could be further associated to disease progression or exacerbations. Yet, their study is restricted by the technologies currently available, which remain costly, complex or not sensitive

enough. Clinical MALDI-ToF platforms do not allow routine detection of these metabolites, whereas REIMS may provide useful information regarding virulence alongside microbe identification. Speculatively, it could offer support to clinicians in the management of infections by allowing risk stratification.

Since the specific metabolites described here were detected without sample extraction, they may be present and detectable in native samples such as sputum and represent potential diagnosis targets for culture-free methods. Direct analysis of samples would exempt from preliminary culturing, thereby eliminating additional bias and reducing time-lag for diagnosis.

4.4.3. REIMS metabolic fingerprinting provides some strain level information

Differences between *P. aeruginosa* isolates were observed in the profiles of phospholipids, AQs and rhamnolipids. These in particular have been previously reported to vary between strains [212]. Despite an important batch effect, REIMS analysis enabled the identification of 45 individual *P. aeruginosa* isolates with 81% of accuracy, using 10 replicates and the Random Forest algorithm. Notably, some isolates which demonstrated close metabolic profiles belonged to the same MLST type. This indicates that the metabolic profile detected with REIMS contains information which could be strain-dependant. Conversely, some isolates sharing the same MLST type sometimes displayed spectral differences which would confuse the type prediction. If we consider that the MLST is based upon the assessment of slowly evolving housekeeping genes whilst REIMS analyses rapidly evolving components, it is clear that the latter would offer higher resolution that may differentiate isolates considered the same strain by MLST. The present experimental design does not allow to conclude on the classification of MLSTs offered by REIMS, due to the limited number of isolates per type.

The further analysis of independent VNTR-typed isolates did not reveal correlations between the VNTR and metabolic profiles, although the detection of the LES type was promising. It was therefore concluded that REIMS classifications did not concord with VNTR classification of *P. aeruginosa*.

The MLST and VNTR typing methods rely on the analysis of stable regions of the genome. *P. aeruginosa* is a versatile organism with an elastic genome whose expression adapts according to its environment, through the production of virulence-associated molecules – e.g. AQs and rhamnolipids – or structural modifications of the cell membrane – e.g. involving phospholipids. Hence, although not providing phylogenetic information, REIMS might offer – through the detection of those compounds – a finer and more representative picture of the real-time pathogenicity of a strain, potentially more

relevant to the clinicians facing treatment decisions. Further research should investigate whether these virulence factors could serve to rapidly detect high-risk clones, thereby leading to improved clinical management.

Finally, it is essential to remember that, like most of the other methods, REIMS is an *in vitro* procedure, providing information on microorganisms growing on agar plates. Preliminary assays confirmed that growing conditions impacted the detected metabolic profile, as previously observed [149] (data not shown). Therefore, experimental conditions must be carefully considered beforehand and standardised to allow pertinent and reliable comparisons. Notwithstanding, how much the REIMS profile is representative of the *in vivo* behaviour of a given pathogen remains uncertain.

REIMS was successful in differentiating CF pathogens and important metabolites and virulence factors were detected and identified. Strain level disparity was observed, enabling the separation of individual isolates and proving our initial hypotheses right. However, this metabolic divergence did not lead to strain type classification according to VNTR, and further testing with more samples is required to conclude on MLST classification. Nevertheless, the metabolic profile detected with REIMS may provide information relevant to the behaviour and virulence of individual strains; this subject will be explored in the next chapter.

Chapter 5

Metabolic phenotyping of *P. aeruginosa* using REIMS

5. Metabolic phenotyping of *P. aeruginosa* using REIMS

Understanding the phenotypes of infecting bacteria and how they differ between patients, bacterial strains and stages of infection, is important to ensure optimised personalised treatments and clinical management, e.g. appropriate antibiotic treatment or patient segregation. Hence, there is a need for novel experimental solutions to further investigate phenotypic differences in *P. aeruginosa*'s metabolism, virulence and their evolution during the course of an infection.

Following up on the potential of REIMS to characterise individual isolates (Chapter 4), REIMS was next challenged to differentiate types of *P. aeruginosa* infection, and in particular between early and chronic airway infections in CF. This study relies on the initial hypothesis that REIMS is capable of detecting changes occurring in the metabolome of *P. aeruginosa* during its evolution, as it did between strains. The data were further interrogated, seeking insight about *P. aeruginosa* community behaviour (Chapter 2, objective 2).

5.1. Introduction

5.1.1. Adaptation of *P. aeruginosa* – genetic and phenotypic diversity

P. aeruginosa is known for its capacity to adapt to various challenging environments, thanks to an extensive and flexible genome. This versatility leads to the development of varied phenotypes, suited to specific niches such as different infection sites, e.g. urinary or respiratory tracts. This heterogeneity may also be observed within communities sharing the same evolving environment, as variable phenotypes have been detected from the lung of a given CF patient, and even in bacteria isolated from one sputum sample apparently infected by a single strain [213, 214]. The variable phenotypes most commonly described include motility, production of toxins, siderophores and virulence factors, nutrient utilisation, cell surface modification, colony morphology, hypermutability, and antibiotic susceptibility [46, 211].

These phenotypic changes mainly occur due to passive accumulation of genetic mutations during prolonged exposure to particular conditions [211]. After initial acquisition by the patient, most commonly from the environment, and a first stage of acute infection, the bacteria adapt in order to persist in the CF airway. Environmental pressures include availability of nutrients, host immune factors, antibiotics, and the presence of other microorganisms [46]. The selection of mutations

favouring survival in the CF airways leads the bacterial evolution towards chronic infection. Frequently, mutations involve loss of function, e.g. loss of motility, and imply decrease of virulence, notably by impacting the type III secretions and QS systems [48]. The latter affects the production of PYO, HCN, elastase and rhamnolipids (see Figure 1.6) whose production is commonly down-regulated by mutations on the *lasR* gene [48]. Inactivation of this key gene in the QS system also (indirectly) leads to modification in the secreted QS profile, e.g. cessation of the conversion of 2-heptyl-4(1H)-quinolone (HHQ) to PQS. It is accompanied by enhanced catabolism of amino acids, abundant in the CF lungs, and increased tolerance to β -lactam antibiotics [81].

The evolution from acute to chronic infection in CF usually manifests as a transition from a motile virulent stage, involving the secretions of toxins, enzymes and rhamnolipids, to a sessile, self-protected biofilm mode of growth. Notably, the development of chronicity can be accompanied by the development of a mucoid phenotype, type VI secretions, and the emergence of hypermutators and antibiotic resistance. The establishment of chronic infection favours the further accumulation of mutations. This is notably due to a high cell density accelerating gene transfer [43], and to the continuous change in the bacteria's environment, due to disease progression, clinical treatments and tissue deterioration. It leads to the evolution and coexistence of different mutants infecting the same patient, resulting in phenotypic diversity [46]. Social behaviour and distribution of the metabolic costs amongst the mutants, e.g. QS and toxin secretions, may contribute to the long-term establishment of *P. aeruginosa* [46]. This provides the bacterial community the capacity to resist and respond to a large array of adverse factors, in particular antibiotics [214]. Nevertheless, with decreasing virulence, loss-of-function mutations may also generate vulnerability in bacterial metabolism, e.g. impacting DNA-repair or biofilm formation [48].

In recent decades, there has been a worrying emergence of transmissible strains, associated with high antibiotic resistance and poorer prognosis [27]. Strains such as the LES have emerged due to the occurrence of mutations favouring transmission and competitiveness in the airways [46]. In patients infected by LES, the majority of phenotypic variability was observed within individual patients rather than between patients [214]. Not only was the haplotype composition varied within a single patient, it also evolved over time, providing the bacteria with a pool of antibiotic-resistant haplotypes [214].

5.1.2. Metabolic diversity of *P. aeruginosa* detected with REIMS

In Chapter 4, REIMS analyses showed that *P. aeruginosa* had a distinctive metabolic profile compared to other pathogens. It was observed that, within the *P. aeruginosa* species itself, isolates may present significant differences that were evident in the quorum sensing, rhamnolipid and phospholipid profiles (see spectra in Figure 5.1). These variations allowed the prediction of individual isolates with 81% of accuracy.

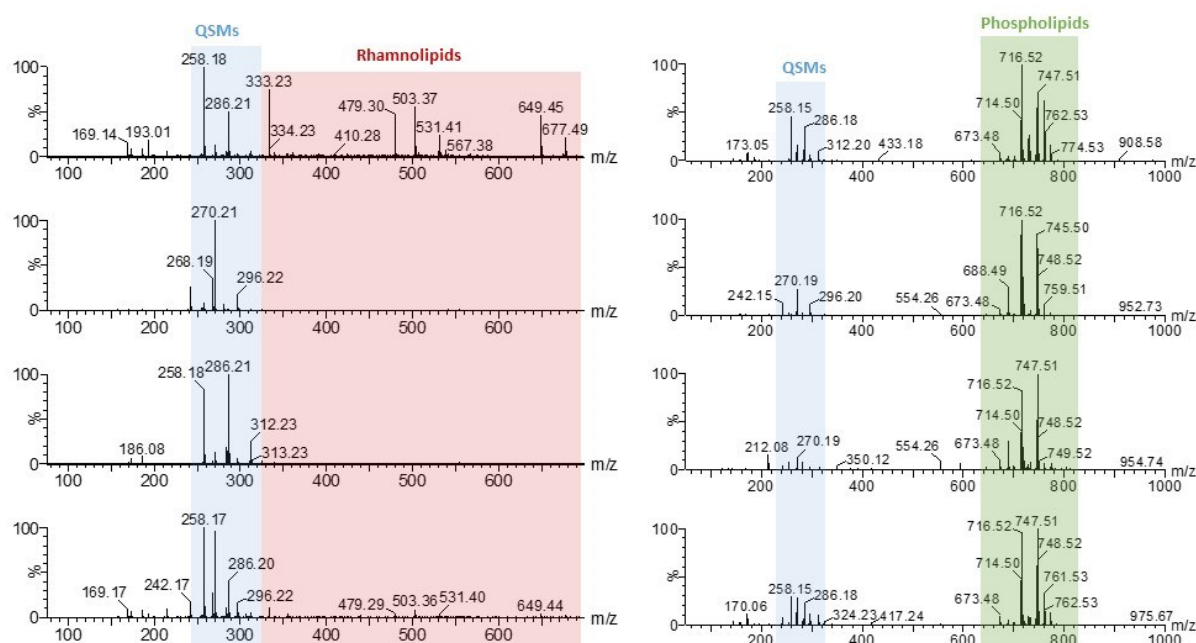


Figure 5.1. Example of mass spectra obtained from four *P. aeruginosa* isolates using REIMS.

Acquired with manual REIMS, on bacteria grown on CBA. Spectra on the left show a zoom on the 80-700 m/z range, highlighting the AQ (blue) and rhamnolipid (red) regions; spectra on the right show the 50-1000 m/z range, highlighting the phospholipid region in green (same data).

5.1.3. Objectives

The aim of this study was to use REIMS to gain insight on the adaptation mechanisms of *P. aeruginosa*. The objective was to explore and characterise *P. aeruginosa* intra-species diversity (1) between different infections, (2) within individual CF patients, (3) between early and chronic infections. For this purpose, I intended to answer the following specific questions:

- Does *P. aeruginosa* produce a similar metabolic profile in acute (non-respiratory) and respiratory infections?
- Does the metabolic profile of *P. aeruginosa* vary within and between CF patients, and how does it relate to phenotypic diversity?

- Can we predict or differentiate the infection stage in CF patients using the metabolic profile detected with REIMS?
- Does the REIMS metabolic profile correlate with *in vitro* phenotypes?

To achieve this, isolates from respiratory and acute non-respiratory infection have been analysed and compared. Multiple colonies were collected from each sputum sample provided by CF patients with early and chronic *P. aeruginosa* infection; their metabolic profiles were compared, clustered and classified; differences and correlations at the two stages of infection were sought.

5.2. Material and methods

5.2.1. Comparing various types of infection

The Royal Brompton Hospital (London, UK) provided 45 CF *P. aeruginosa* isolates. Additional isolates from non-CF patients were included in the study. There were 35 chronic *P. aeruginosa* isolates obtained from the respiratory samples (sputum and swabs) of bronchiectasis patients attending routine clinical visits at the Royal Brompton Hospital (London UK), and 45 acute isolates cultured from non-respiratory samples (wounds, blood cultures, urine, drain fluid, ear, vaginal, perineum, rectum, and skin swabs) as part of routine clinical identification at Charing Cross Hospital (London, UK). All isolates were cultured on CBA plates and processed according to the workflow described in Chapter 3, section 3.2. *Collection and culturing of microorganisms*. Analyses were performed with a handheld bipolar probe with a minimum of three sampling points each time (section 3.4.2.1. *Handheld bipolar probe*).

Data processing used a reduced *m/z* range (*m/z* = 50-1200) and was implemented on the Metaboanalyst platform. It involved filtering of the 5000 most intense median bins, log-transform and Pareto scaling; statistical analyses included PCA, HCA, *t*-test, ANOVA and PLS-DA classification.

5.2.2. Investigating *P. aeruginosa* diversity in CF patients

5.2.2.1. Sample collection and REIMS analysis

One sputum sample was provided by 25 CF patients attending routine clinical visits at the Royal Brompton Hospital (London UK). For seven patients, it was the first record of *P. aeruginosa* positive culture (considered early infection stage, patient nomenclature is “Exx”); 18 patients were chronically

infected (patient nomenclature “Cxx”). Sputum samples were selectively cultured on ceftrimide agar, and between 12 and 60 colonies were collected from each sample, resulting in a total of 615 isolates. If colonies appeared morphologically different, examples of all morphotypes were selected. Samples were processed at the Royal Brompton microbiology laboratory and cultured by Natasha Wierre-Gore, PhD student at the Life Sciences department, Imperial College London (UK), as part of a collaboration within the CF Trust Strategic Research Centre (SRC) for *P. aeruginosa*.

REIMS analysis was performed on the 615 isolates grown on solid LB media in 24-well plates, together with 32 replicates of LB and of *P. aeruginosa* controls (PAO1 and PA14), leading to a total of 30 well-plates including 711 samples (see detailed numbers in Appendix 7). Analysis required two days, using laser-assisted REIMS and the automated Tecan platform (see section 3.4.2.2. *Automated, laser-assisted REIMS*), applying one “burn” per well at a laser power of 2 W for 8 s, with 20 s of waiting time in between. The experiment required one week for culturing and analysis, and was carried out twice. Data processing led to the same conclusions both times; the results presented here come from the second run of analyses.

5.2.2.2. Phenotyping and MLVA

Phenotypes of the 615 isolates were characterised by Natasha Wierre-Gore following previously reported protocols with a minimum of three replicates and included quantitative measures of:

- Swimming, swarming and twitching motilities [215],
- Biofilm formation [216],
- HCN production [217],
- PYO production [218],
- Protease and elastase productions [219].

MLVA was carried out by Natasha Wierre-Gore at the Public Health England facilities (Colindale, UK) with the assistance of Dervla Kenna and Jane Turton’s teams using nine loci as described by Turton et al. [202]. Due to financial constraints, only a subset of isolates were typed: 12 isolates were randomly selected amongst the patients who demonstrated the highest phenotypic variability, i.e. nine chronic patients (see patients and VNTR types in Appendix 7).

5.2.2.3. Data processing and statistical analysis

Data were processed within the range $m/z = 50-1200$ and underwent the pre-processing steps on the OMB Waters software as described in Chapter 3, section 3.4.4. *Data analysis*. Additional operations were implemented in R:

- a) Neighbour bins exhibiting a correlation in intensity higher than 0.9 (within the same day of analysis) were merged. This was to correct for the instrumental shift observed between the two days of analysis which caused peak splitting when applying binning.
- b) $m/z + 1$ isotopes with a correlation higher than 0.9 were excluded.
- c) The 150 most intense features were selected for most of the following statistical workflow.

PCA and ROC curve analysis were implemented on the Metaboanalyst platform after log-transform and auto-scaling. HCA and correlations were calculated in R. Correlation networks were built with the open source software Cytoscape [220]. Classification was performed using the Random Forest algorithm.

5.3. Results

5.3.1. Respiratory isolates showed higher (between-patient) metabolic diversity than non-respiratory infections

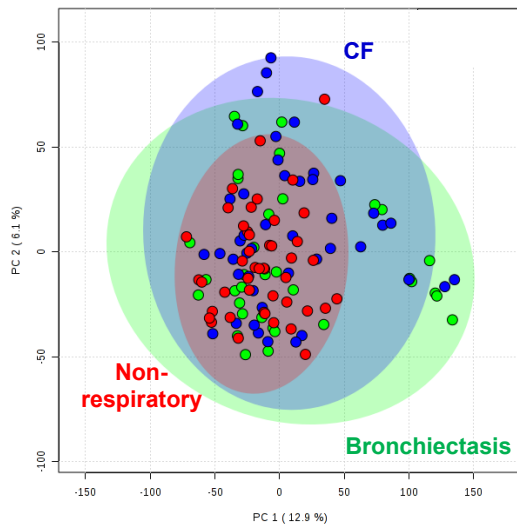
The aim of the work in this section was to compare the metabolic profile produced by *P. aeruginosa* in various cases of infection in order to investigate the potential specificities of CF airway infections. Ultimately, it may orientate the development of more targeted treatment depending on the type of infection.

A set of 45 *P. aeruginosa* cultured from CF samples was compared to 35 isolates collected from bronchiectasis patients and 45 isolates originating from non-respiratory infections. Metabolic profiles were recorded using REIMS, and PCA and PLS-DA classification were used to visualise the diversity of the data (Figure 5.2).

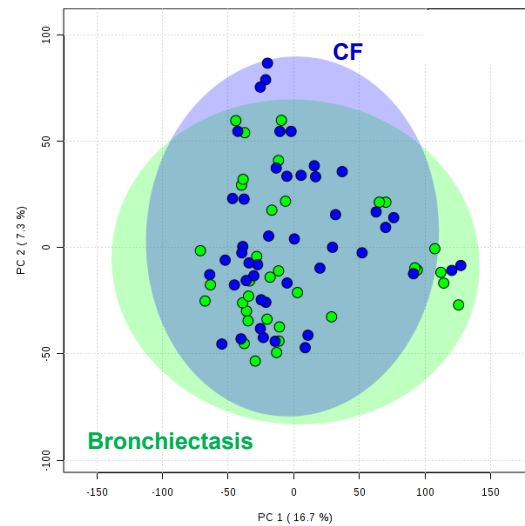
Figure 5.2a and Figure 5.2c show a substantially lower diversity for the non-respiratory isolates compared to the respiratory *P. aeruginosa*, despite the variety of sites of infection. Chronic isolates of *P. aeruginosa* are indeed well-known for being genetically and phenotypically very variable. The passage of time, in addition to repeated and disparate antibiotic treatments, leads to evolutionary selection specific to infection environment, therefore likely explaining the greater diversity in the

metabolome of respiratory isolates. Thus, it is plausible that the *P. aeruginosa* strains found in our close environment and causing acute infections, e.g. wound or early lung infections, would produce at first a similar metabolome, and would then diversify during the establishment of a chronic infection, as suggested by previous genetic studies [48]. This was further verified with CF isolates, in section 5.3.2.3.1. *Chronic isolates showed metabolic profiles with higher diversity.*

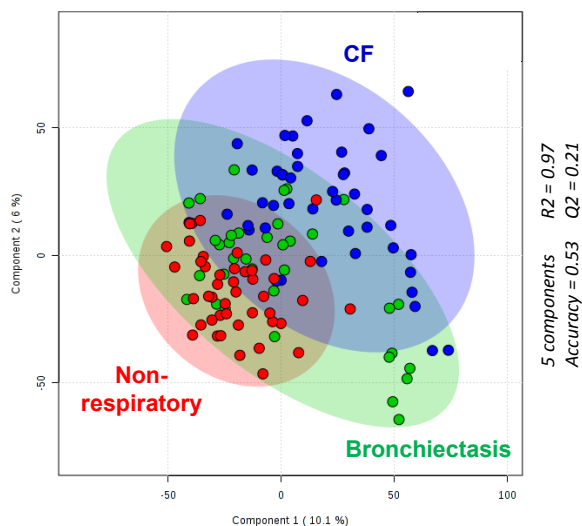
a. Scores plot with all *P. aeruginosa* isolates



b. Scores plot for the respiratory isolates



c. Scores plot with all *P. aeruginosa* isolates



d. Scores plot for the respiratory isolates

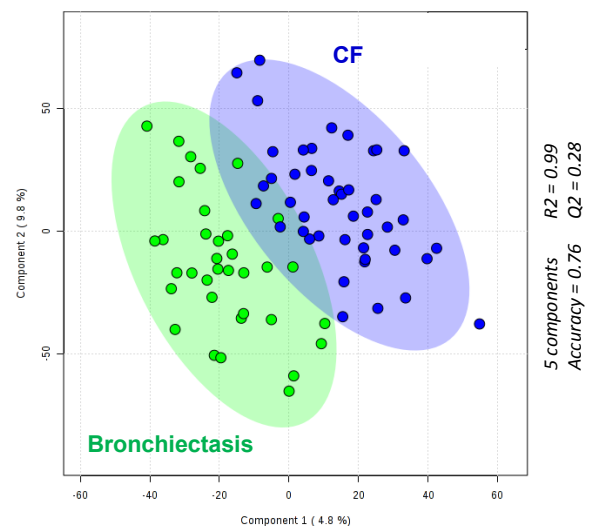


Figure 5.2. Statistical analysis of *P. aeruginosa* isolates from CF and non-CF patients.

Shows components one and two for PCA (a & b), and PLS-DA (c & d). 95% confidence intervals are shown by shaded areas. Cross-validation results for PLS-DA are given on the right-hand side (built on the Metaboanalyst platform, reproduced under the Creative Commons Attribution license from Ref. [209]).

The two first principal components of the PCA did not show any separation between the CF and the bronchiectasis isolates, yet, PLS-DA was able to separate the two groups (Figure 5.2d, Q2 = 0.28). This suggests that bacterial adaptation may differ depending on airway pathophysiologies and/or clinical

management, e.g. antibiotic treatment (clinical data not available due to data collection method), as observed in previous studies comparing phenotypes of bronchiectasis and CF isolates [221]. In that particular study by Woo et al., the only difference observed was a higher lipase production in CF isolates, suggesting a distinct lipid metabolism. It was attributed to the abnormal composition the CF airway secretions, particularly rich in phospholipids, which constitute a source of fatty acid nutrients for *P. aeruginosa* [222]. Altogether, these observations suggest that bacterial adaptation to different abnormal airway environments has metabolic and phenotypic consequences.

Further statistical processing was performed to investigate what was driving these partitions. HCA was used to visualise the repartition of the isolates without the overlap confounding the PCA. It was conducted on a reduced number of features, selected as the 50 most discriminant through *t*-test and ANOVA, in order to focus on the separating trends (heatmaps available in Appendix 8). This operation did not allow group separation and no features specific to any class could be identified. One of the difficulties of separating these three classes may lie in the fact that the respiratory isolates analysed here potentially included isolates from various stages of infection. Notwithstanding, subgroups of CF isolates were observed. Most of the selected significant bins were within the 600-900 *m/z* range, and were likely to be related to the phospholipid chemical family, which are components of the cytoplasmic and outer membranes of the cell envelope. Other studies have previously detected variations in the bacterial membranes, notably at the level of the lipid A, component of the outer membrane of *P. aeruginosa*. Indeed, this pro-inflammatory part of LPS showed a different structure in CF patients compared to other infections, including bronchiectasis (e.g. hexa-acetylated instead of penta-acetylated) [37]. This additional acetylation led to an increased pro-inflammatory response and resistance to the host antimicrobial peptides. Moreover, alterations in the phospholipid profile were observed in the bacteria growing as a biofilm [77] and suggested a decrease in membrane fluidity. An increased rigidity of the cell membrane may induce a slowed down metabolic activity [78], which would be consistent with chronicity. These observations support the fact that microbes adapt to their environment through membrane modifications, either because of nutrient availability, and/or as a protective strategy.

5.3.2. Investigating within-patient metabolic diversity of *P. aeruginosa* in CF

This section focused on exploring the metabolic diversity of *P. aeruginosa* in CF, testing multiple colonies from a single sputum sample from early and chronically infected patients. Indeed,

understanding the diversity and the evolution of *P. aeruginosa* in CF may support further personalisation of treatments and clinical management.

5.3.2.1. Preliminary observations: the majority of phenotypic diversity occurred between patients with chronic infection

Colonies were collected to determine whether the phenotypic diversity of the *P. aeruginosa* population evolving within the lungs of a single patient was more significant than diversity between patients, as has been observed before in patients infected by the LES [214].

The sample set included a total of 615 isolates collected from seven early and 18 chronically infected patients. Using classical phenotypic assays, it was observed that:

- Global inter-patient variability was higher than intra-patient variability;
- Diversity was higher within chronic isolates than within early isolates;
- Patients E04, C07, C09 and C15 displayed the highest phenotypic diversity, followed by C01, C08, C10 and C13 to a lesser extent;
- All the tested phenotypes were suppressed in the chronic isolates compared to the early isolates, except biofilm formation, which remained constant.

Next, 12 isolates were randomly selected for MLVA from each of nine chronic patients. According to these analyses, the tested patients were infected by a single VNTR type except patient C08. Patient C08 was the only one infected by two different strains of *P. aeruginosa*: two colonies shared one VNTR type, 10 colonies had another type. Notably, patient C13 was infected by LES.

REIMS profiling was performed as an additional phenotypic assay since it provides information about metabolism and virulence of *P. aeruginosa* through the detection of molecules that are potentially relevant to the diversity and adaptation of the bacteria, i.e. QS molecules and rhamnolipids. The study included biological replicates of PAO1, PA14 and LB controls. Various statistical tools – detailed thereafter – were implemented to compare the metabolic variability within and between patients. Then, the data was interrogated in order to examine the differences between the metabolic profiles of early and chronic isolates.

5.3.2.2. Comparing metabolic variability of isolates within and between patients

This section aimed to elucidate whether the majority of bacterial metabolic diversity occurred within an individual CF patient or amongst them, and whether this was in agreement with the other phenotypic findings.

The comparison of profiles and their variability presents some difficulties, especially when a high number of samples is involved. Here, the metabolic diversity of the *P. aeruginosa* colonies was evaluated using two approaches, in order to take into account both the similarity of the colonies within patients and the dissimilarity between patients:

- Similarity of the colonies within patients, and dissimilarity between the patients, were estimated using clustering analysis (HCA).
- The similarity of the colonies within patients – i.e. colonies collected from one sputum sample produced by a single patient – was further assessed by testing their classification according to the patients from whom they were collected, using the Random Forest algorithm. This was equivalent to checking the reproducibility of sampling per patient (or rather, per sputum sample).

5.3.2.2.1. Clustering analysis showed metabolic homogeneity within patients and dissimilarity between patients

PCA (in Figure 5.3) and HCA were built using the 150 most intense features to get an overview of the distribution of the isolates.

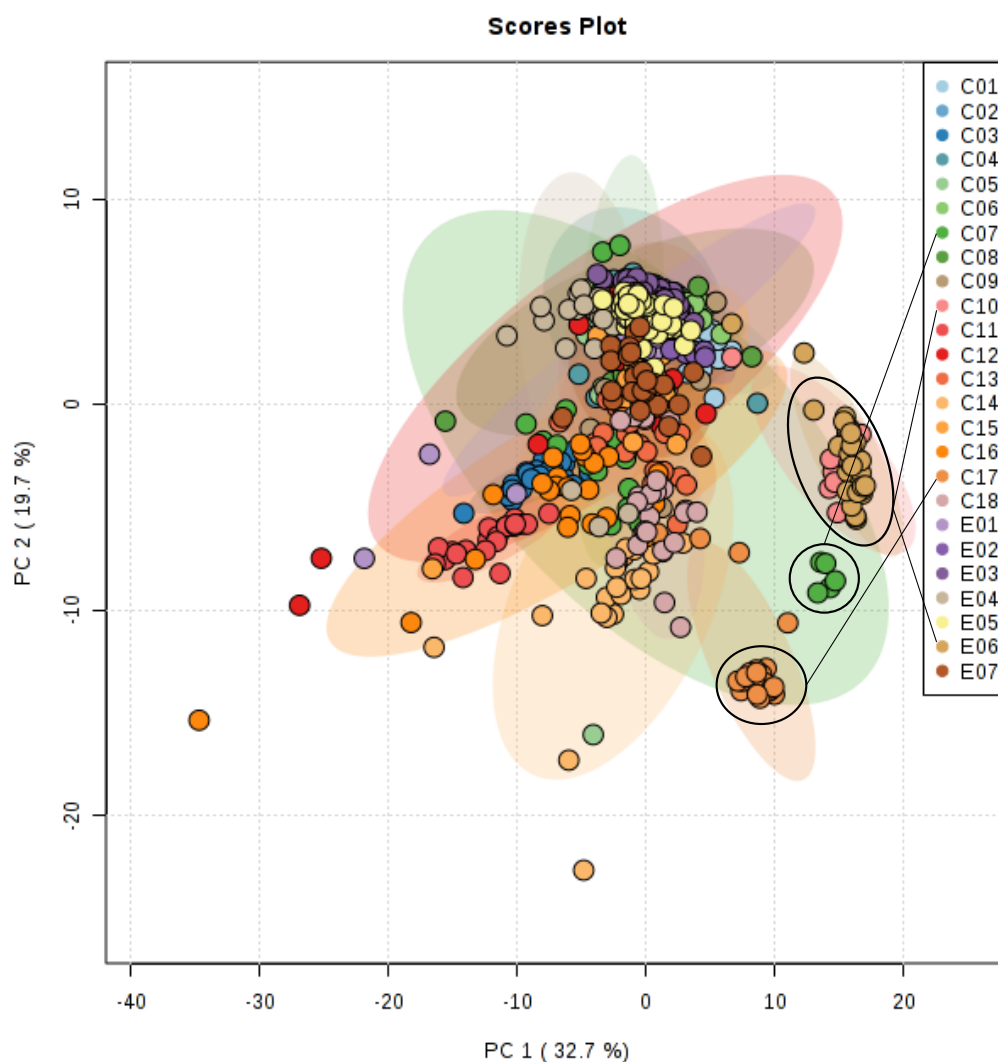


Figure 5.3. PCA score plot using REIMS spectral data acquired on 615 *P. aeruginosa* colonies collected from chronically and early infected patients.

Includes 18 chronic patients (noted Cxx) and seven early (Exx). Shows the two first principal components; shaded areas show 95% confidence intervals of the patient groups (built on the Metaboanalyst platform).

The colonies were coloured by patient and, although overlap was important due to the two-dimensional representation, it suggested patient grouping. Notably, some colonies from patients C07, C10, C17 and E06 stood out of the pool, implying significant metabolic differences from the other samples and therefore from the other patients.

If the interpretation of a PCA is difficult when dealing with high numbers of samples, for a HCA it is nearly impossible (dendrogram not shown). Therefore, for each patient and control, the proportion of clustered colonies was represented as a function of the distance in the HCA tree. The area under this curve (AUC) (no unit) characterises the distribution of the considered samples in the HCA: the higher the AUC, the tighter the clustering of the colonies within the global HCA. This would suggest that the colonies from the considered patient are metabolically similar, and/or significantly different from the other samples. An explanatory example is given for PAO1 in Figure 5.4.

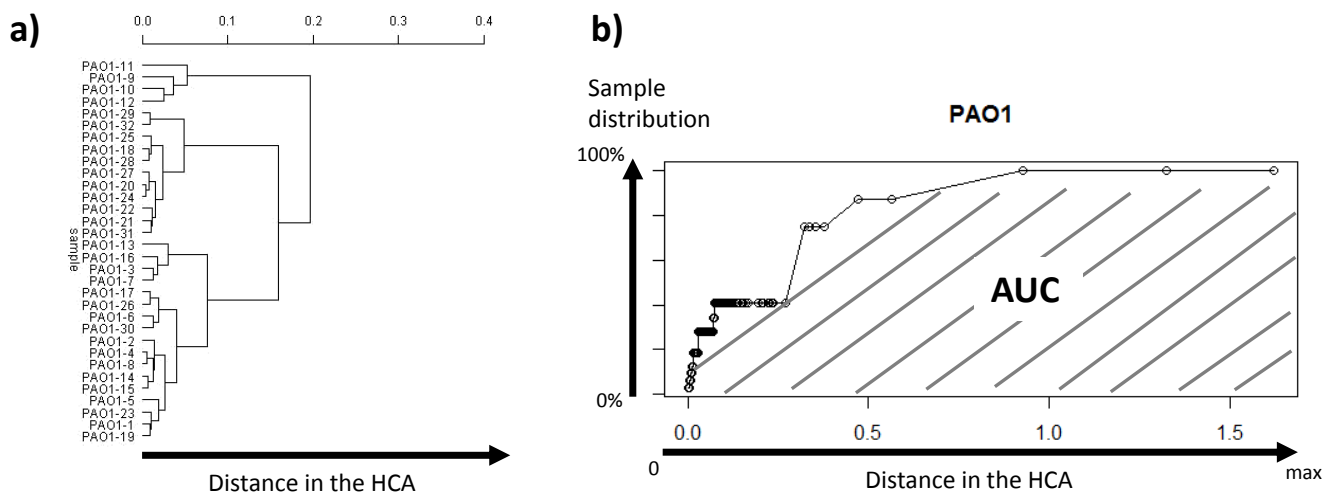


Figure 5.4. Representation of the AUC characterising the distribution of PAO1 colonies in the global HCA.

(a) Dendrogram representing the distribution of the PAO1 clones taken alone; (b) Distribution curve associated to PAO1 according to the distance in the global HCA including all the samples (built in R). The shorter the distance on the HCA to regroup all the samples (e.g. 0.2 here), the higher the AUC (1.30 here).

The AUC values are reported in Table 5.1 for each patient. Notable HCAs are displayed in Figure 5.5: LB (a) and patient E06 (b) demonstrated high AUCs, whilst patients C07 (c) and C08 (d) had low AUCs. The same HCA distance scale was used in the four graphs to allow visual comparison of the distribution: samples of LB and E06 are clustered at a distance of 0.1, whereas C07 and C08 complete the cluster at 0.4 and 0.2, respectively.

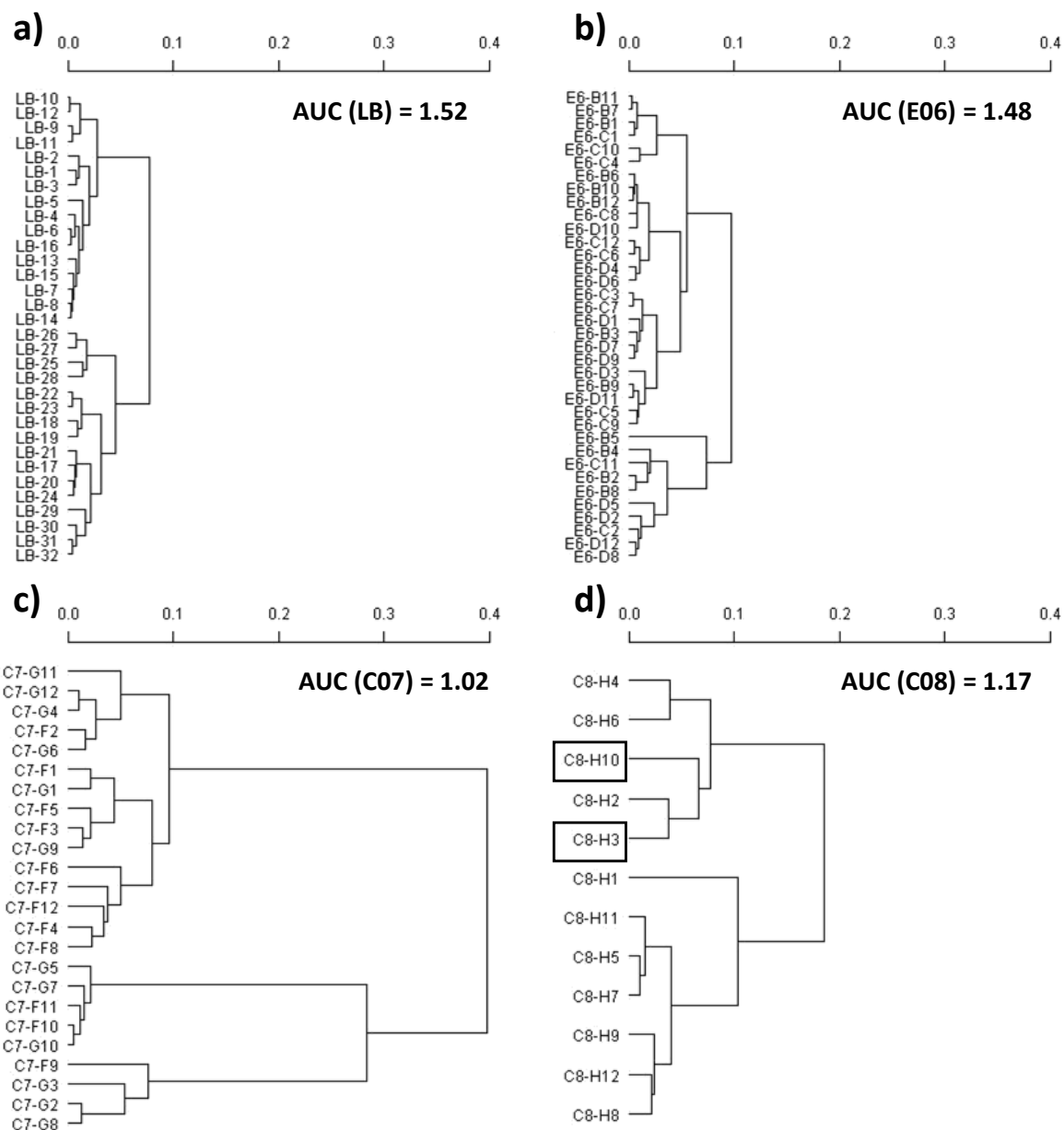


Figure 5.5. Example of dendrograms showing the distribution of the colonies for selected patients. Distribution of the analyses of (a) the LB media control, (b) patient E06, (c) patient C07 and (d) patient C08, taken alone, represented on the same distance scale (0.4) (built in R). Isolates grouping at a short distance on the HCA (e.g. (a) LB and (b) E06 cluster at 0.1) show lower variability than (c) C07 and (d) C08 (clustering at distances of 0.4 and 0.2, respectively). AUCs related to the distribution in the global HCA including all the samples are indicated. Colonies with a different VNTR type in patient C08 are shown in boxes.

The LB growing media led to one of the highest AUC = 1.52, due to the fact that its metabolic profile was particularly different from that of *P. aeruginosa*. Patients E06, C11 and C17 demonstrated the highest AUCs of 1.48, 1.50 and 1.54, respectively, meaning they displayed a metabolic profile as different from the other patients as LB. Spectral data were therefore closely examined and revealed notable differences. Most of E06 isolates grew with a mucoid phenotype, which is typically accompanied by a higher signal in rhamnolipid and lower levels of AQs, which may explain their clear

differentiation from the rest of the samples. Interestingly, patient C11 showed a distinctive metabolic profile; rhamnolipid intensities were low, and AQ and phospholipid profiles notably differed from the other *P. aeruginosa*, with high intensities in quinolones (2-nonyl-4(1*H*)-quinolone (NHQ), 2-undecanyl-4(1*H*)-quinolone (UHQ), 2-undecenyl-4(1*H*)-quinolone (C11:db-UHQ)), and specific phospholipids (PE(35:1) and PG(35:1)). In the case of patient C17, the presence of non-specific *P. aeruginosa* peaks was observed, and attributed to contamination with another microbe. The Royal Brompton clinical microbiology lab later confirmed contamination with *S. maltophilia* (by VNTR typing) and patient C17 was excluded from further processing.

5.3.2.2.2. Colony classification suggested high reproducibility for sampling within patients

In Chapter 4, isolate replicates were analysed to demonstrate REIMS' reproducibility: replicates could be classified according to isolate using their unique metabolic profile. Conversely here, assuming REIMS' high reproducibility, it is the reproducibility of the colonies collected from each patient that was evaluated. Again, this was to evaluate the metabolic variability of colonies collected from a single sputum sample. The Random Forest algorithm was applied to test whether the colonies would classify according to patient, thereby supporting the hypothesis that they would be the same isolate. The cross-validation matrix with controls is presented in Figure 5.6, F1-scores are presented in Table 5.1 and complete prediction scores are available in Appendix 9.

The overall classification of the colonies per patient was 93% accurate, showing that the majority of the colonies were highly similar within patient (i.e. sputum sample). The controls scored around 90%; 79% of the patients were predicted above this score, meaning that their colonies were likely to be clones. The chronic patients C07, C08, C12 (and C02) on the other hand, were the only scores below 90%, reaching 83%, 38%, 77% (and 89%), respectively. This suggested some metabolic variability within the colonies of these patients. Notably, the only score below 50% was obtained for C08, who was the only patient known to be infected by two distinct strains of *P. aeruginosa*. The exclusion of C08 increased patient classification up to 95% accuracy.

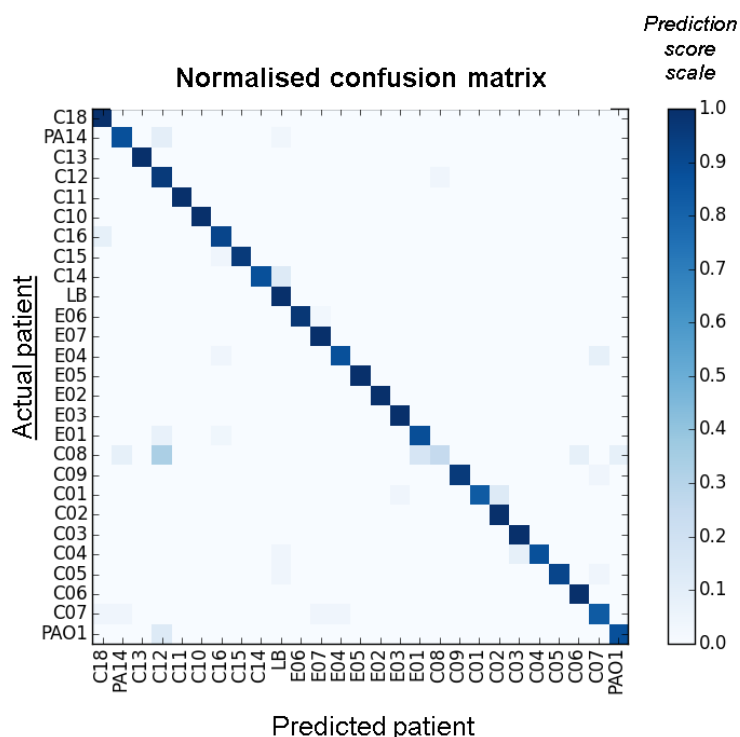


Figure 5.6. Confusion matrix for Random Forest classification according to patients using REIMS spectral data obtained from *P. aeruginosa* colonies.

Colonies collected from 17 chronic patients (Cxx) and seven early infections (Exx). Controls were LB, PA01 and PA14. Leave-one-colony-out cross-validation; real and predicted identities are displayed respectively on the left-hand and the bottom sides. The model was built using all the bins within $m/z = 50-1200$. Frequency of prediction for each patient is represented in darkening blue, overall classification per patient was 93% accurate.

The higher classification accuracy obtained in this study compared to the previous results of Chapter 4 may be attributed to the date effect. In the present study, the samples were analysed within two days, whereas repetitions of Chapter 4 were carried out over several months, in order to evaluate REIMS' reproducibility over the long term.

5.3.2.2.3. Conclusion: metabolic diversity was higher between than within patients

Random Forest classification does not highlight differences between patients, but may reveal heterogeneity within a group. However, this is not necessarily true, if the samples are numerous enough to allow a "grouped prediction". Nevertheless, it remains an efficient tool to evaluate reproducibility within a group of samples, and supported the results obtained with the AUCs from the HCA. Results per patient from the two statistical approaches are summarised in Table 5.1. The values in blue, calculated for the controls, were chosen as references to evaluate the variability within the patients: F1-scores for classification below 0.90 and clustering AUCs below 1.28 were considered more

variable than the controls and are shown in red. AUCs around that of the LB control (1.52) indicated patients with a metabolic profile very different from the other samples and are shown in green.

Table 5.1 Summary of AUC and classification results per patient.

(a) AUC as a measure of the distribution of the colonies throughout the global HCA, (b) F1-scores for Random Forest classification using leave-one-colony-out cross-validation (prediction as in Figure 5.6). Reference groups are in blue, values showing higher variability are in red, the AUC values showing the three most different patients are in green.

a) Clustering analysis		b) Classification
Patient	AUC	F1-score
C01	1.31	0.91
C02	1.43	0.89
C03	1.42	0.96
C04	1.30	0.93
C05	1.35	0.96
C06	1.43	0.98
C07	1.02	0.83
C08	1.17	0.38
C09	1.14	0.98
C10	1.41	1.00
C11	1.50	1.00
C12	1.20	0.77
C13	1.46	1.00
C14	1.42	0.93
C15	1.40	0.98
C16	1.05	0.90
C17	1.54	-
C18	1.21	0.94
E01	1.42	0.90
E02	1.31	1.00
E03	1.44	0.99
E04	1.23	0.91
E05	1.41	1.00
E06	1.48	0.99
E07	1.34	0.96
LB	1.52	0.91
PA14	1.28	0.90
PA01	1.30	0.92

Classification scores and AUC values were consistent for the chronic patients C07, C08 and C12, whose colonies showed the highest metabolic diversity. Closer inspection of the data from patient C12 suggested a batch effect, which separated the isolates according to the two dates of analysis. Patient C08 was co-infected with two strains, which explains the metabolic variability of his colonies. Metabolic variability of C07 was in agreement with phenotypic observations although VNTR typing indicated that the patient was infected by a single strain. The calculation of AUC revealed that the metabolic profile of patients C11 and E06 were notably different from the other patients, whilst C17 was identified as contaminated. Of interest are the results obtained for patient C13, infected by LES. The phenotypic diversity of his colonies was high, in agreement with previous observations [214]. By

contrast, the metabolic profile detected with REIMS appeared very reproducible (F1-score = 1), and notably different from the other strains (AUC = 1.46). This was consistent with the results obtained in Chapter 4, section 4.3.4.3. *Classification according to VNTR type*, where LES scored the highest classification accuracy (73%) compared to the other VNTR types. This supports the hypothesis of some metabolic specificities of LES, which could be helpful in the early identification of this highly transmissible strain.

Most of the patients showed a lower variability than PAO1 and PA14 clones, and could be easily differentiated from each other. This led to the conclusion that metabolic diversity was higher between than within patients. Moreover, the mean AUCs and F1-scores for early infected patients (1.38 and 0.96, respectively) were higher than for chronic patients (1.31 and 0.90, respectively, C17 excluded). This further confirms that intra-patient diversity increases whilst the infection progresses. VNTR typing proved to be a poor method to study variability within individual patients, since most of them harboured a single strain, whereas REIMS and phenotypic assays offered some visibility on quickly evolving bacterial characteristics.

5.3.2.3. Can we predict early and chronic infection stages?

The metabolic profile detected with REIMS was compared between the two stages of infection, seeking separating trends that would allow the prediction of the infection stage in CF patients. Correlations within metabolites and with phenotypes were also calculated in order to support better understanding of *P. aeruginosa* metabolism and behaviour.

5.3.2.3.1. Chronic isolates showed metabolic profiles with higher diversity

The metabolic profile of chronic isolates were more diverse than those collected at an early stage of infection as visible on the PCA and clustering distributions displayed in Figure 5.7. The early and the chronic isolates did not separate on the PCA and their AUCs were way below the controls, also suggesting that neither displayed a distinctive profile.

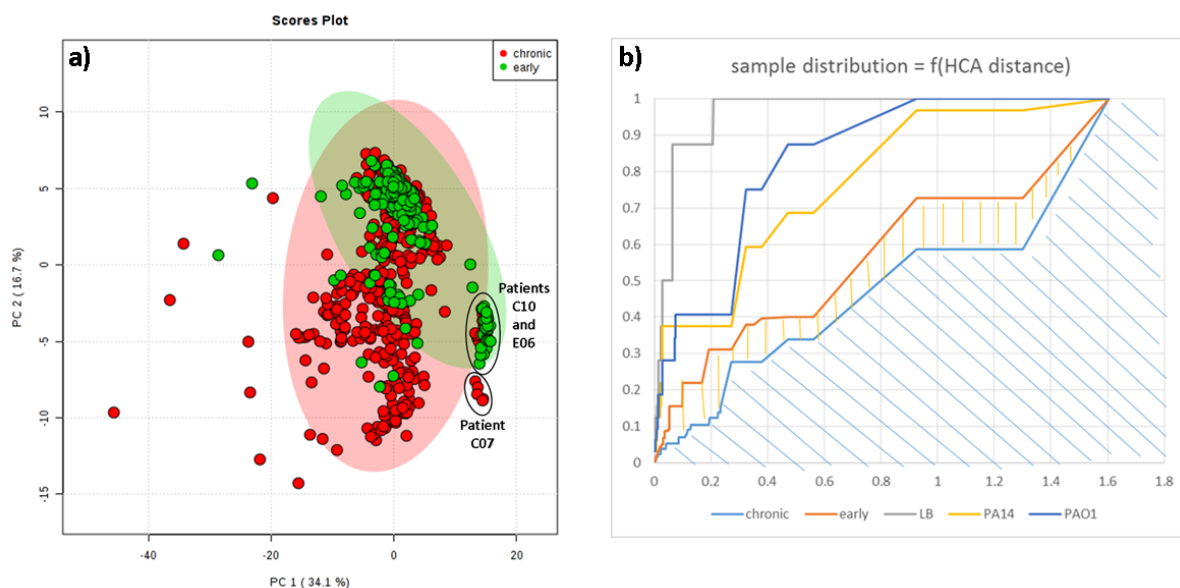


Figure 5.7. Comparison of metabolic diversity between isolates collected from early and chronically infected patients.

(a) PCA score plot showing the two first principal components, shaded areas show 95% confidence intervals. Chronic patients in red showed higher variability but did not separate from early isolates in green (built on the Metaboanalyst platform); (b) Distribution curves according to the distance in the global HCA; early isolates (orange curve, AUC = 0.91) and chronic isolates (light blue, AUC = 0.74) were more variable than the controls PAO1 (dark blue, AUC = 1.30), PA14 (yellow, AUC = 1.28) and LB (grey, AUC = 1.52) (data generated on R). Analyses included seven early patients and 17 chronic patients (C17 excluded).

Random Forest classification was tested as a tool to diagnose the infection stage using the metabolic profile detected with REIMS. The model used all the bins and a leave-one-patient-out cross-validation, to avoid confounding effects with the patients. Overall classification accuracy was 65%, due to the low prediction rate for the early isolates (20% sensitivity), most of them being predicted as chronic isolates (confusion matrix in Figure 5.8, detailed scores in Appendix 10), as a result of the overlap of their metabolic profiles (Figure 5.7a).

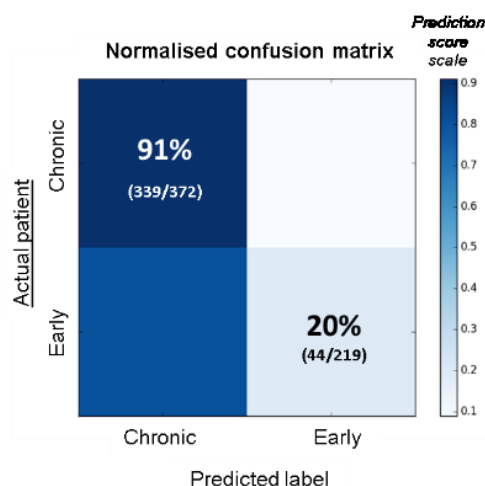


Figure 5.8. Confusion matrix for Random Forest classification of isolates between early and chronic infection. Colonies collected from seven early and 17 chronically infected patients. The model used all the bins within $m/z = 50-1200$ and a leave-one-patient-out cross-validation; real and predicted classes are displayed respectively on the left-hand and the bottom sides. Recall (sensitivity) is shown on the matrix for chronic and early isolates; F1-scores were 0.76 and 0.29, respectively. Overall classification was 65% accurate.

Random Forest classification did not allow the prediction of the stage of infection. Nevertheless, levels of a few relevant molecules were examined in the search for differentiating trends. Amongst the 150 most intense features, 27 had been previously identified using MS/MS (in Chapter 4), including 11 AQs, eight rhamnolipids and eight fatty acids and phospholipids (see Appendix 11). Amongst these, univariate analyses (ROC curves) revealed some significant differences between the two infection stages. The distribution of those metabolites identified with $AUC > 65\%$ is presented in the form of boxplots in Figure 5.9.

Overall, there was a more heterogeneous distribution in the chronic isolates, in agreement with their higher variability observed previously. According to the medians, the relative levels of the differentiating metabolites seemed to be decreased in the chronic compared to the early isolates. This was in concordance with the global decline in virulence described in the literature when switching from an early to a chronic stage.

Higher in early isolates

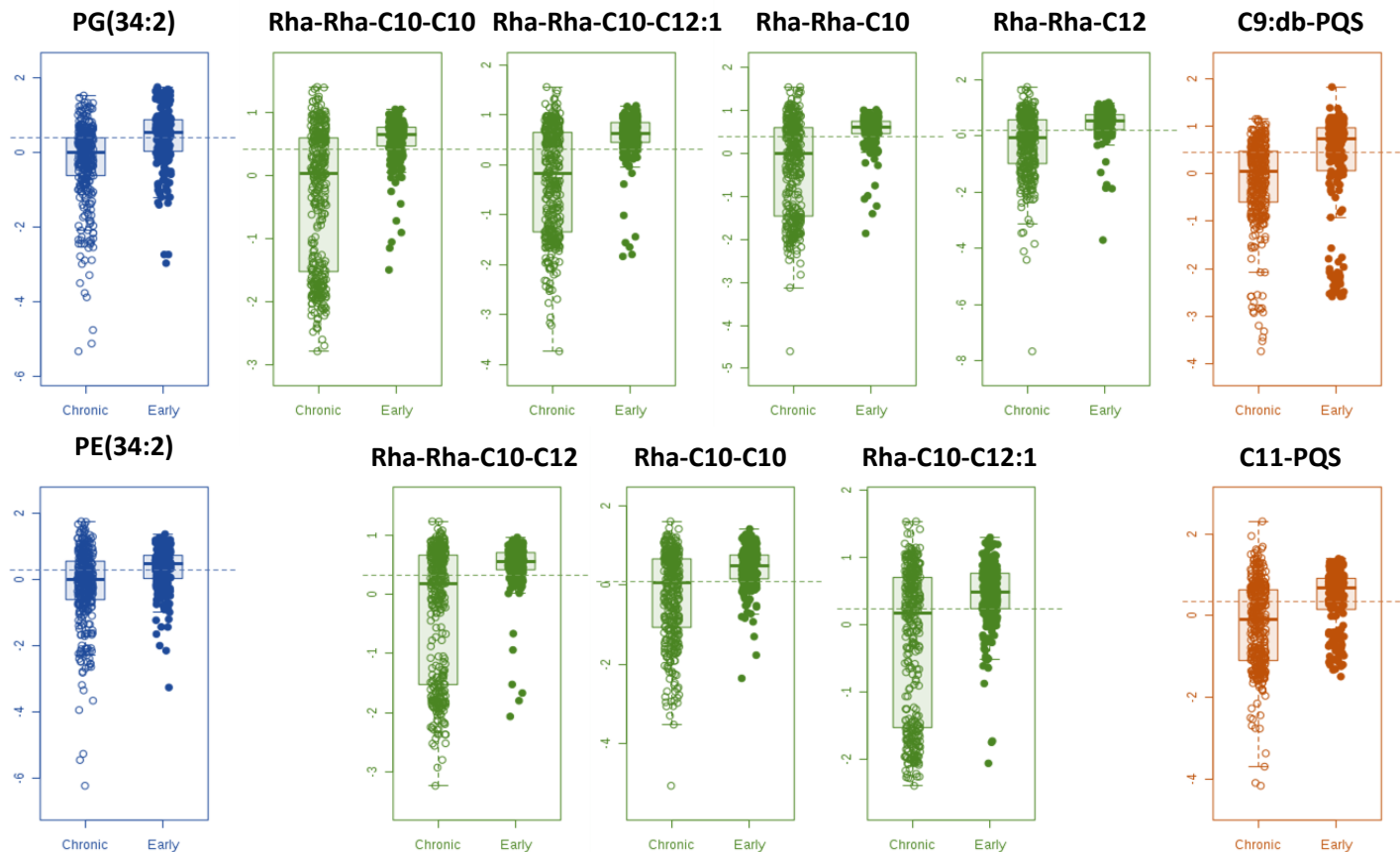


Figure 5.9. Boxplots showing the distribution of selected metabolites in early and chronic isolates.

Levels as detected by REIMS and relative to the TIC signal (TIC-normalisation). Early isolates are plotted on the *right*-hand side of each box, chronic isolates on the *left*-hand side. Phospholipids boxplots are in blue, rhamnolipids in green, AQs in orange. The dashed lines show the optimal cutoff to separate the two classes for each compound (generated on the Metaboanalyst platform).

Despite the limited number of identified metabolites, some trends should be highlighted. First, median levels of rhamnolipids were systematically higher in the early isolates. Secondly, the profile of AQs seemed to change as the infection progresses: globally, levels of hydroxy-quinolones (2-nonenyl-3-hydroxy-4(1*H*)-quinolone (C9:db-PQS), 2-undecanyl-3-hydroxy-4(1*H*)-quinolone (C11-PQS)) were higher in the early isolates, whilst quinolones (NHQ, UHQ, HHQ) were higher in the chronic isolates. This may be a consequence of the *lasR* loss-of-function mutation, which is common in chronic isolates [48]. Mutations in this gene suppress the hydroxylation of HHQ that produces PQS [81]. It is therefore plausible that the entire quinolone family would be impacted in a similar manner, leading to the accumulation of quinolones in the metabolome of chronic isolates, and a decrease in the production of hydroxy-quinolones. In addition, the *las* QS system controls the *rhl* QS system through PQS [223]. Since the *rhl* system regulates the production of rhamnolipids, it is also consistent with the decrease observed in rhamnolipids. However, it should be remembered that this method did not allow the separation of the signal of hydroxy-quinolones from that of their quinolone-*N*-oxide isomers. Although little is known about the latter, quinolone-*N*-oxides have distinct metabolic functions. They do not act as QS signalling molecules but have some antimicrobial properties that seem to contribute to *P. aeruginosa* competitiveness [91, 224]. They also follow production kinetics different from that of PQS during bacterial growth, which may confuse observations [100].

Finally, the decrease observed in specific PG(34:2) and PE(34:2) is in accord with previous observations by Benarama et al. who studied the evolution of membrane composition during biofilm growth [225]. They found that PEs and PGs, with fatty acids totalling 32 to 36 carbons were predominant, although they tended to decrease in favour of phospholipids containing both shorter and longer fatty acids. Unfortunately, most of the phospholipids identified with REIMS were within the 32 to 36-carbon range, which prevented from observing variations in lower and higher mass phospholipids. Furthermore, even numbered phospholipids with two unsaturations – such as PG(34:2) and PE(34:2) – have shown a steady decrease as the biofilm matured [77]. Globally, the degree of unsaturation of the lipids tended to decrease, which would lead to a diminution in cell fluidity, as commented on before.

5.3.2.3.2. Correlation networks revealed different synthetic pathways according to the infection stage

The most intense features were used to calculate correlations within early and within chronic isolates separately; Spearman correlations were used to account for the non-Gaussian distribution of the data (as visible in boxplots in Figure 5.9). Networks were built connecting the metabolites with positive

correlation above 0.5 (Figure 5.10). The objective was to visualise the links between the metabolites in order to possibly:

- Reveal metabolic pathways;
- Observe differences between the metabolic profiles of early and chronic isolates;
- Identify molecules with pivotal roles;
- Detect analytical effects, due to vaporisation, ionisation, fragmentation, etc.

The most intense features detected in the early (120 bins) and in the chronic isolates (120 bins) were compiled in a common matrix totalling 128 bins that was used to build the networks (see list in Appendix 12). Some had been previously characterised (Chapter 4); others were tentatively identified using exact mass comparison with theoretical values and preliminary MS/MS data.

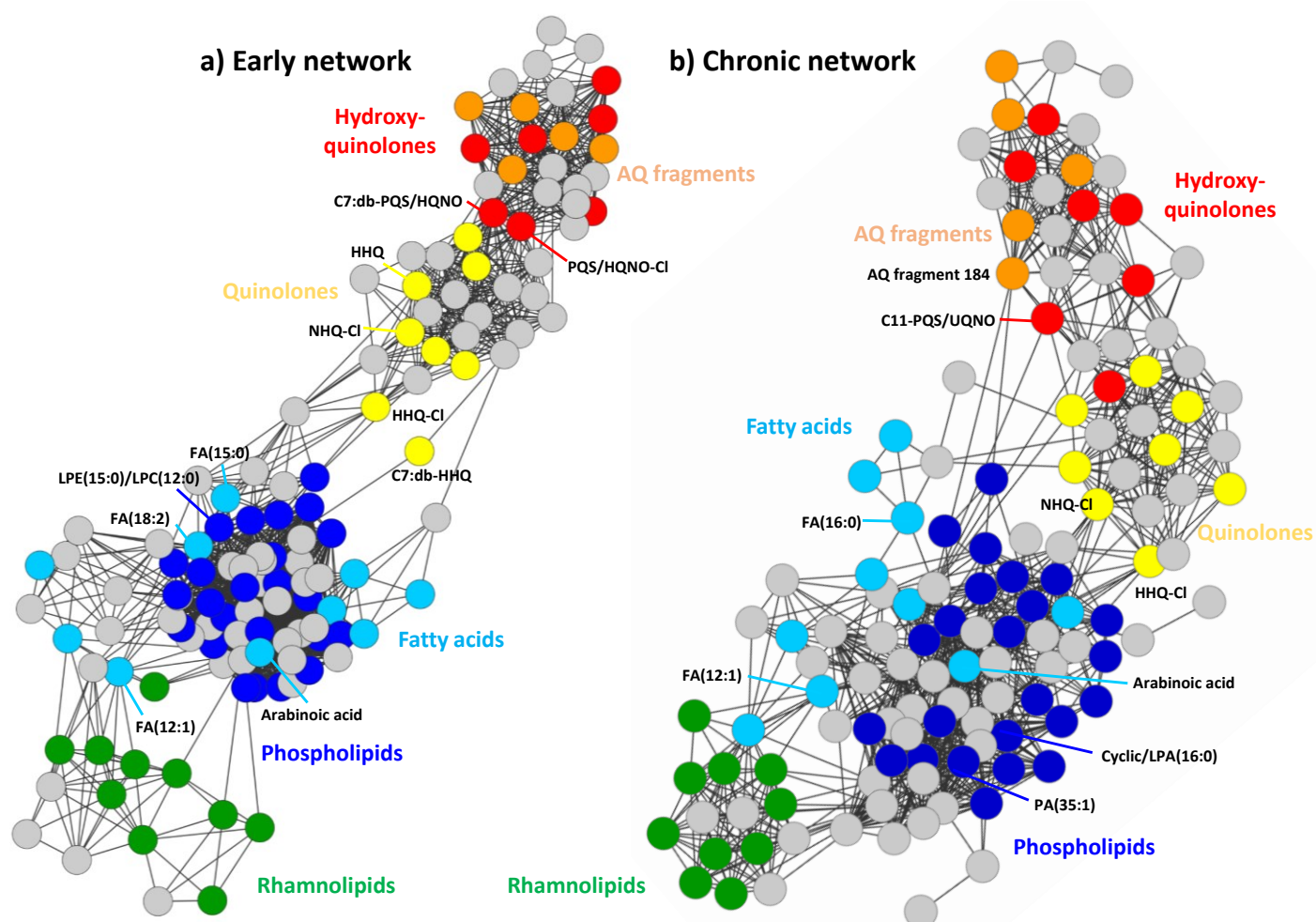


Figure 5.10. Correlation networks for isolates collected from (a) early infection and (b) chronic infection. The lines (edges) represent positive Spearman correlations greater than 0.5. The nodes correspond to the 128 highest spectral bins common to early and chronic isolates. They are coloured according to their metabolic family when known (or inferred): hydroxy-quinolones are in red, quinolones in yellow, AQ fragments in orange, rhamnolipids in green, fatty acids in light blue and phospholipids are in dark blue. Annotated nodes correspond to (tentatively) identified features amongst the 20 highest betweenness centrality for each network (built in Cytoscape).

On initial inspection the networks look similar, showing clusters of metabolites based on their chemical families: small organic QS molecules, lipids (including fatty acids and phospholipids), and rhamnolipids (glycolipids). Because the technique impacts similarly molecules of the same chemical families, these groupings may be the result of analytical effects. Very tight clusters, such as those observed in the early infection network, form due to a high number of connections within a subgroup of features. These are likely to result from technical correlations, for instance between metabolites, fragments (e.g. fatty acids clustering with phospholipids, AQ fragments with quinolones), ions with various adducts (e.g. Cl^- adducts of AQs), isotopes missed during pre-processing, etc. This phenomenon may be used for feature identification [226]. However, depending on the objective of the study, these underlying correlations may be considered “noise” or “background” correlations.

The detection of biological correlations requires substantial metabolic variability, which explains why they were rather limited in the early network. There were interesting connections between the rhamnolipids and the other lipids, whereas the tight clusters formed by phospholipids and by AQs may be due to analytical correlations. It was hypothesised that the increasing metabolic variability in the chronic isolates allowed more biological correlations to be revealed superimposed on technical correlations, which would explain why the network appeared looser. By contrast with the early network, the AQs were more connected to the lipids whilst the rhamnolipids formed a tighter cluster. Nevertheless, both networks showed the same global layout, with the lipid cluster positioned between the rhamnolipids and QS molecules. In addition, the latter seemed to divide into two subgroups: the quinolones directly linked to the lipids, and the hydroxy-quinolones further away. This disposition reflected the production pathways of rhamnolipids and AQs, which both require (amongst other precursors) similar hydroxy-fatty acid supplies [90]. However, here native fatty acids could not be distinguished from those produced in-source by phospholipid fragmentation. Furthermore, the quinolones are the direct precursors of hydroxy-quinolones [91].

Surprisingly, there were no connections between the rhamnolipids and the AQs, despite the fact that AQs are associated to the control of rhamnolipid production, which conversely contributes to the action of AQs by enhancing their solubility in aqueous media [108]. Altogether, these observations suggested that these networks are representative of synthetic pathways rather than metabolic functions.

In order to identify key metabolites in the networks, betweenness centrality was calculated for each of the 128 features (Appendix 12). Centrality is a measure of the importance of a variable in a network. In betweenness centrality, the importance of a variable is dependent on the number of times its node

stands in the shortest path between two other nodes. In other words, betweenness centrality quantifies the degree to which each feature acts as a bridge between other nodes, providing this feature with impact and control over the network. Metabolites with (tentative) identification amongst the 20 most central features are indicated on each network in Figure 5.10. The key metabolites were consistent at both infection stages; the early network particularly highlighted the PQS synthesis pathway, from fatty acids through quinolones (HHQ analogues). Although still visible in the chronic network, direct correlations with a significant centrality appeared between the lipids and the hydroxy-quinolones, leaving the quinolone precursors aside. This may reflect the synthesis pathway of isomer compounds which could not be differentiated from hydroxy-quinolones: quinolone-*N*-oxides. Indeed, these are generated via a similar pathway although, contrary to hydroxy-quinolones, they do not require the production of intermediary quinolones [227]. This may be another effect of the loss-of-function in *lasR* at chronic stage: due to the cessation of the conversion of quinolones to hydroxy-quinolones, the synthesis pathway of quinolone-*N*-oxides is revealed.

To further evaluate the disparities between the two networks, betweenness centralities of the 20 features with the highest absolute difference between early and chronic isolates are represented on Figure 5.11 in a comparative manner. (Tentatively) identified features included six AQs and three lipids.

Data in Figure 5.11 confirmed what was observed before on the networks. The PQS analogues and precursors played a central role in the early network, although C11-PQS/C11-UQNO was more important in the chronic network due to its direct connection with lipids. Following up on the previous discussion, this molecule may actually be a quinolone-*N*-oxide: 2-undecenyl-4(1H)-quinolone *N*-oxide (C11-UQNO). These observations supported again the hypothesis that there is a different AQ profile at different infection stages: the production of hydroxy-quinolones such as PQS seems important in the early isolates, whereas it is down-regulated in the chronic isolates, revealing the presence of other AQs with different metabolic functions, such as quinolones and quinolone-*N*-oxides.

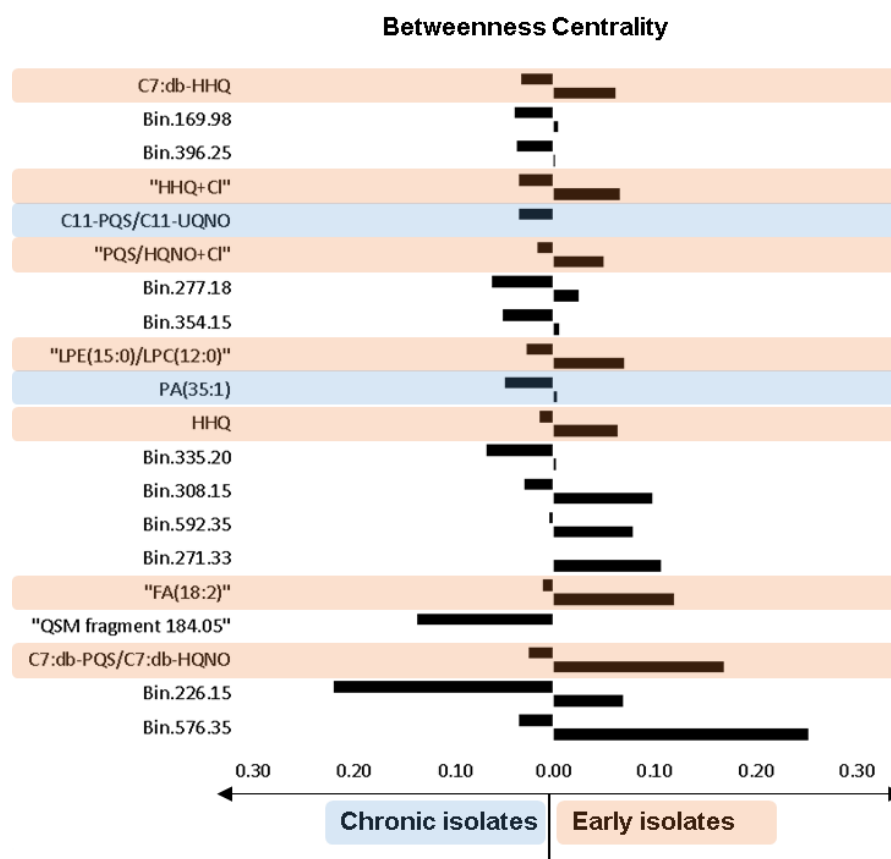


Figure 5.11. Betweenness centrality scores for the 20 features with the highest difference between the early and chronic networks.

Tentative identification are indicated in inverted commas; identified metabolites with higher betweenness centrality in the early network are highlighted in orange (notably PQS analogues and precursors), higher betweenness centrality in the chronic network are in blue.

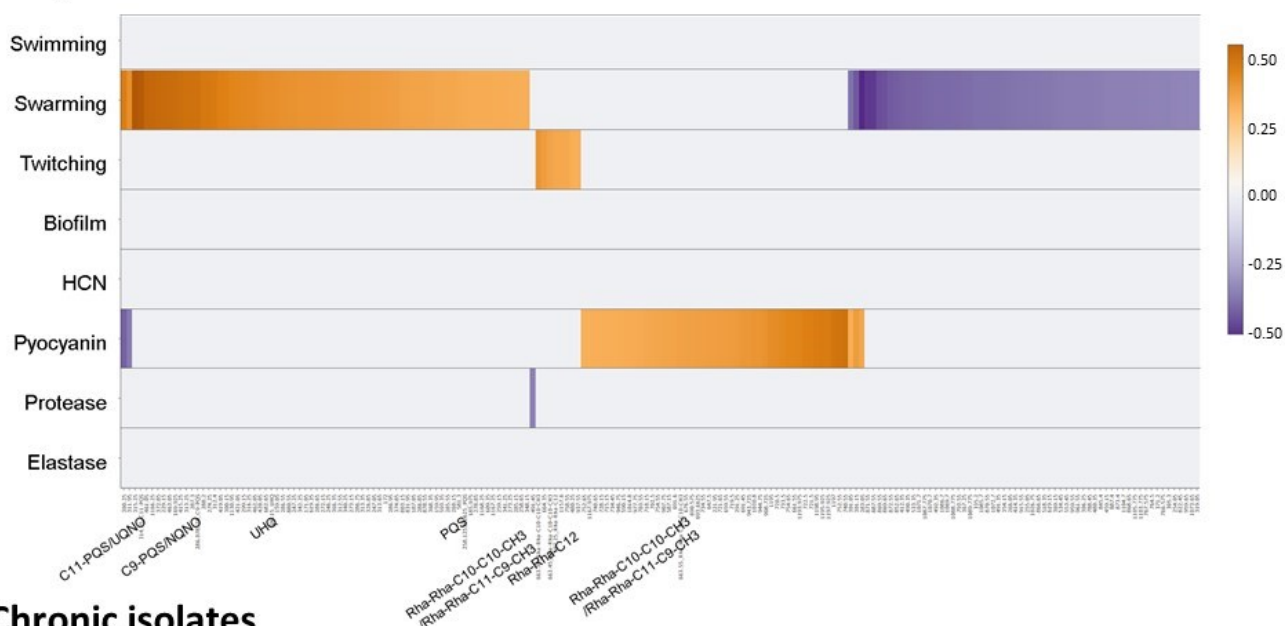
5.3.2.3.3. Correlations were found between REIMS metabolic profile and the other phenotypes

Correlations between the whole m/z range and eight measured phenotypes were explored. A technique called repeated measures correlation (rmcorr) was applied in order to take into account the confounding factor of patients providing multiple colonies [228]. The primary objective was to assess the utility of the detected metabolites to predict phenotypes. These were associated with bacterial virulence and included the three types of bacterial motility (swimming, swarming, twitching), biofilm formation, production of HCN, PYO and production of two enzymes (protease and elastase). In agreement with the literature, assays showed that, except for biofilm formation, these behaviours were observed more frequently in the early isolates, than in the chronic isolates as the bacteria adapt and settle in the host's airways. Correlations were calculated separately for early and chronic isolates, to unravel potential differences. This led to the detection of 190 and 319 correlated features within

the early and chronic isolates, respectively, 14 of which only could be identified. Significant correlations (p -value < 0.05 after Bonferroni correction; $r_{\text{rm}} > 0.5$) are displayed in heatmaps in Figure 5.12.

Detected correlations were different in early and chronic isolates, in terms of number, metabolites and phenotypes impacted. In the case of early infection, features correlated primarily with swarming motility, a key ability for colonising microorganisms, and PYO production. A higher number of correlations were observed in the chronic isolates, which was again consistent with their higher variability. Almost all the phenotypes displayed correlations, in particular swimming motility, and productions of HCN, PYO and protease. Unfortunately, the limited number of identified features renders difficult any metabolic interpretation. It can only be speculated that hydroxy-quinolones were positively associated to swarming motility in the early isolates, and rhamnolipids to PYO and to protease production in the chronic isolates. Surprisingly, rhamnolipids were poorly correlated to biofilm phenotypes and swarming, which they have been previously reported to be linked to [104, 229]. Furthermore, QS molecules, which are known to contribute to the regulation of motility, biofilm formation, enzyme and virulence factor production [230], showed very little correlations. One possible explanation for these missing correlations is the fact that the phenotypic assays were carried out on liquid cultures, whilst REIMS analysis required culturing on a solid matrix. It is therefore unlikely that such a versatile pathogen as *P. aeruginosa* would behave and express the same phenotypes or metabolomes in different growing conditions.

a) Early isolates



b) Chronic isolates

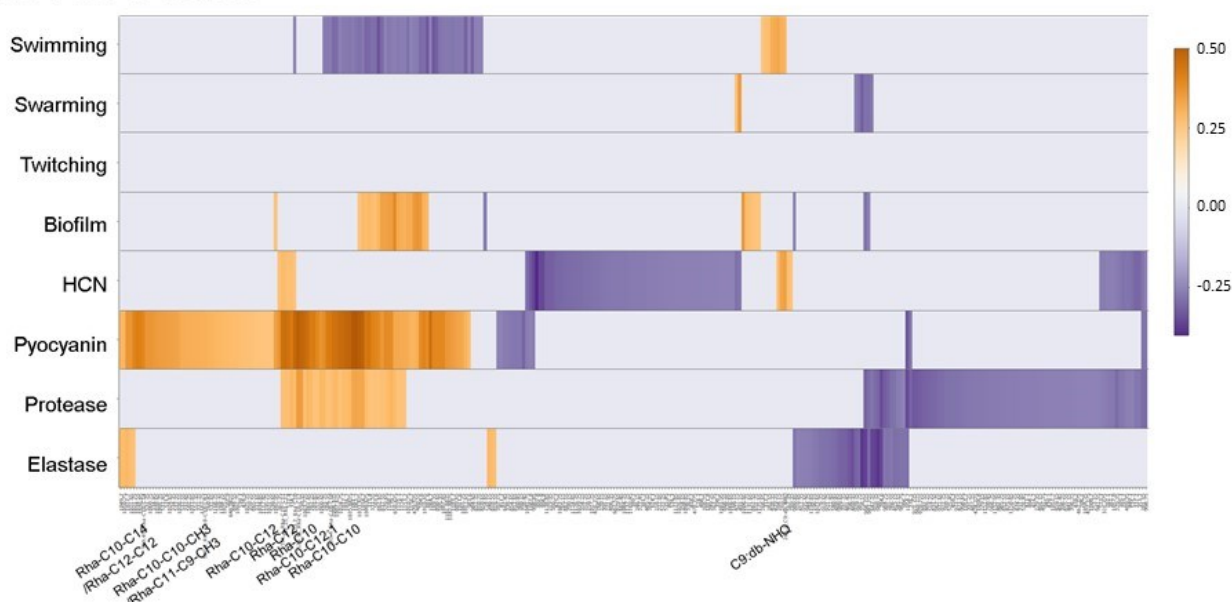


Figure 5.12. Heatmaps representing significant correlations between metabolic features and phenotypes for (a) early and (b) chronic isolates.

Identified metabolites are indicated at the bottom of each map. Positive correlations are in orange, negative correlations in purple; darkening colours show strengthening correlations.

5.4. Discussion

5.4.1. REIMS as a tool to explore and monitor *P. aeruginosa* metabolism and diversity

REIMS metabotyping revealed a high degree of diversity within the species *P. aeruginosa*. REIMS analyses were consistent with phenotypic assays and VNTR typing, in that they showed a metabolic variability higher between patients than within. Indeed, after repeated sampling, the colonies could be attributed to the patient they were collected from by comparing their metabolic profile. In concordance with VNTR typing, this confirmed that most of the patients were infected by the same isolate. However, REIMS offers further information on the diversity of a single isolate evolving within a patient, whilst this information is lost by typing methods focusing on slowly evolving DNA regions, such as MLVA and MLST. On the other hand, classification of isolates per patient was 95% accurate. This score was notably higher than in Chapter 4 (81%), underlying the importance of the batch effect due to the date of analysis. Furthermore, the reproducibility of REIMS allowed the detection of co-infection in patients, either by two types of *P. aeruginosa* or by different microbes. These results suggest that REIMS may be useful in the clinical monitoring of patients infected with *P. aeruginosa*, to detect co-infection, infection with a new strain, or transmission between patients, thereby allowing the adjustment of antibiotic treatment and of patient segregation. This represents a real potential since the current methods to differentiate *P. aeruginosa* strains rely on time-consuming DNA typing.

The correlations between the levels of metabolites detected by REIMS as observed by building networks were similar at both stages of infection. They seemed to represent the synthesis pathways of the metabolites rather than revealing metabolic activity. Generation of rhamnolipids and Aqs from lipids, and the conversion of quinolones to hydroxy-quinolones, could be observed. REIMS analysis thereby offers a global overview of these complex pathways. Whilst most studies usually target a limited number of metabolites such as PQS and its precursor HHQ, REIMS allows the rapid detection of the whole AQ profile. This provides a more comprehensive assessment of bacterial adaptation and pathogenicity. PQS analogues have previously demonstrated different biological activities according to their chemical structure [231], but the functions of many Aqs still remain to be understood. REIMS may contribute to further elucidating QS networks and to unravel new roles for Aqs.

5.4.2. *P. aeruginosa* produces different metabolic profiles at early and chronic stage of infection

The metabolic profile was more diversified in the isolates collected at a chronic stage of infection, both within and between patients. Although the patients generally seemed to harbour a single strain, chronicity favours the occurrence of mutations leading to the diversity of genomes, metabolomes and phenotypes within one strain. These observations on several strains support the findings of Mowat et al. obtained with LES [214], providing a more comprehensive understanding of *P. aeruginosa*. The higher metabolic variability observed in the respiratory isolates compared to other infection types can again be attributed to the fact that respiratory infections are more likely to be chronic. Chronicity implies lengthy adaptation of the organism to its specific infection site, which is impacted by individual host environment and treatment history. REIMS allows the detection of this metabolic variability, which is important in understanding bacterial behaviour within a niche, whereas partial DNA typing cannot provide this information.

According to the literature, phenotypes of early isolates are characterised by a high degree of motility and high production of enzymes, rhamnolipids, PYO, HCN and type III secretions. Chronic isolates on the other hand, are reported to produce lower levels of virulence factors, except for the type VI secretions and sometimes alginate, at the origin of the mucoid phenotype. They usually develop as a biofilm, with a low motility, improved nutrient utilisation and resistance to antibiotics. The profile of QS molecules, controlling most of these phenomena, changes as the infection progresses, playing a continuous role throughout the bacterial evolution. These modifications in the QS profile is likely to be subtle and progressive, which would explain why REIMS did not allow the diagnosis of the stage of infection. Nevertheless, a number of observations led to the formulation of the following theoretical metabolic patterns (see scheme in Figure 5.13). The hydroxy-quinolones such as PQS play a key role at the early stage of the infection. Swarming motility, which is important in acute infection, correlated positively with several hydroxy-quinolones in the early isolates. Due to mutations in the *lasR* gene, frequent in chronic isolates, PQS production (and hydroxy-quinolone analogues) was disrupted. Since it controls – through the *rhl* system – the generation of rhamnolipids, which contribute to establishing early infection, these were also decreased. Although no correlation was observed here, rhamnolipids are required for swarming motility [229]. The quinolones such as HHQ were not oxidised to form hydroxy-quinolones anymore and accumulated. HHQ has demonstrated QS activity on its own [91], and has notably been shown to repress swarming motility [232]. Furthermore, other Aqs such as quinolone-*N*-oxides seemed to be synthesised at least at a chronic stage, and may play a role – although not in QS – in the persistence of the infection. This switch in the QS profile was in agreement

with previous studies: Guina et al. found that PQS was produced by early isolates [99]; Jiricny et al. later showed that PQS was significantly reduced whilst HHQ significantly increased in chronic isolates [233]. Recent clinical studies also observed alterations in the systemic levels of AQs during pulmonary exacerbations [98, 234]. Contrary to the present observations, their results suggested that quinolones – and not hydroxy-quinolones – levels in plasma (e.g. HHQ, NHQ) may be markers for early infection. A larger cohort including higher numbers of controls and cases of new infection, together with the use of statistical methods to adjust for data sparsity [235] may strengthen their conclusions.

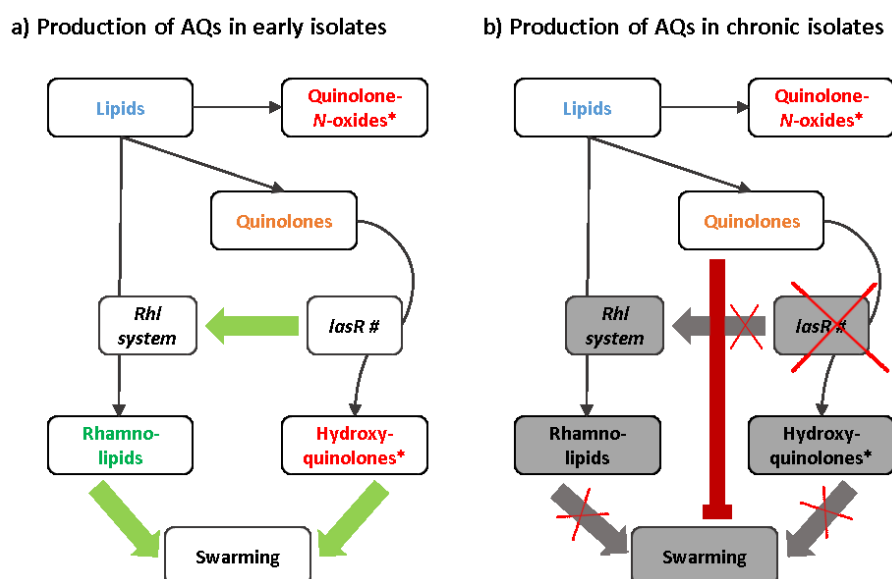


Figure 5.13. Simplistic and theoretical representation of metabolic pathways observed in (a) early and (b) chronic isolates.

*Fine black arrows symbolise chemical synthesis of the metabolites; green bold arrows show positive impact; red blunted arrow show negative impact; factors in grey are down-regulated in the chronic isolates. *Structural isomers, undistinguishable by classical MS; #Indirect action of the lasR gene.*

At the same time, some data supported the adjustment of the phospholipid composition of the bacterial membrane to its growing environment and during infection progression. On the other hand, some observations could not be explained. For instance, in the chronic isolates, PYO and protease secretion correlated with the level of some rhamnolipids. Since PYO and enzymes are water soluble, this phenomenon is unlikely to reflect a metabolic action of the rhamnolipids. Then, the most plausible cause is a confounding effect, since all of these virulence factors are typically upregulated at early stage of the infection.

5.4.3. Limitations and perspectives

The ability of REIMS to detect metabolic changes in agreement with genetic and phenotypic observations supports its use as a potentially clinically relevant tool. It may eventually serve to monitor the evolution of the bacterial population within patients' airways.

However, there were a number of limitations. Firstly, the diversity within patients was assessed looking at one sample only, when it is known that bacterial population differs in the lung regions of CF patients [56]. Moreover, it would be interesting to follow the genetic, phenotypic and metabolic evolution of these isolates over time, in a longitudinal study from an early to a chronic stage of infection, in order to get additional insight as to how *P. aeruginosa* is adapting and how it could be more efficiently combatted.

One difficulty was the inherent differences in the experimental conditions between REIMS analysis and the phenotypic assays. Indeed, the metabolic profile is impacted by culturing conditions [149]. It was also observed elsewhere that the intensity of the AQs and rhamnolipids signals was influenced by the population structure at the time of analysis, i.e. whether the bacteria are growing as individual colonies or as a lawn (preliminary data not shown). This is not surprising given the versatility of *P. aeruginosa* and further investigation into the conditions that impact metabolism would improve our understanding and interpretation of how it evolves and adapts.

One important shortcoming of this study was the small number of metabolites with confirmed molecular identification. A substantial amount of MS/MS data have been acquired on potentially interesting features, yet elucidation of this kind of data is a long process that is still on-going. However, the current experimental setup does not allow the differentiation of hydroxy-quinolones and quinolone-*N*-oxides. This would require the adjunct of a separating method, such as chromatography or ion mobility.

Finally, the biggest pitfall may still reside in the biological significance of *in vitro* results compared to the phenomena occurring *in vivo*. Indeed, previous studies have reported discrepancy in the metabolites detected in clinical sputum samples compared to associated *P. aeruginosa* cultures, e.g. in the AQ and rhamnolipid levels [226]. Possible experimental improvements could be done to better mimic the CF lungs, e.g. using artificial sputum media or a micro-aerobic atmosphere. Nevertheless, robust correlations between *in vitro* observations and clinical outcomes can still be useful.

In this chapter, the hypothesis stating that REIMS was able to detect metabolic differences between *P. aeruginosa* having adapted to different environments was verified, which allowed exploration of the metabolic diversity of *P. aeruginosa*. The study was successful in that it brought insight into *P. aeruginosa* metabolism and answers to the initial questions: REIMS revealed that chronic respiratory infection was associated with more diversified metabolic profiles. Although prediction of the infection stage was not achieved, substantial differences were observed between early and chronic isolates. In the future, such metabolomic studies may allow better understanding and prediction of virulence, antibiotic resistance or transmissible strains, providing crucial information for timely and appropriate treatment. Ideally, this information would be retrieved directly from clinical samples such as sputum, avoiding culture bias, or even better, from non-invasive samples, such as urine or skin secretions. The feasibility of these approaches will be tested in the following chapters.

Chapter 6

Direct analysis of sputum and urine samples using REIMS

6. Direct analysis of sputum and urine samples using REIMS

The previous chapters have shown that REIMS can detect specific *P. aeruginosa* metabolites such as AQs and rhamnolipids from organisms cultured on plates. Based on the initial hypothesis that infection impacts the metabolic composition of biofluids and that these changes can be detected by REIMS, this chapter focuses on testing REIMS as a tool to diagnose *P. aeruginosa* infection directly from clinical samples (Chapter 2, objective 3). At present, sputum is still the best sample to detect pulmonary infection in adults: it can be spontaneously produced by most CF patients (and induced in many more) and provides reliable microbiological results [57] whereas oropharyngeal swabs lack sensitivity [62, 63] and BAL is very invasive. However, as processing and culturing cause delays, REIMS was tested on raw sputum as part of a culture-free method. In addition, based on the fact that sputum is not available from young patients in whom detection is the most critical, urine samples, which are easily obtainable, were also tested.

6.1. Introduction

6.1.1. Sputum composition

Sputum is the expectorated mix of airway secretions, originating from the lower respiratory tract, the nose, the mouth and the pharynx [236]. It is not produced by healthy people – unless subjected to a sputum induction procedure – and is indicative of lung disease [237]. The main component of sputum is mucus, produced by specialised epithelial cells of the lower respiratory tract. The mucus layer lies on a layer of airway surface liquid (ASL). Motile airway cilia beat to clear dust, bacteria and other inhaled particles trapped in the mucus. This mechanism constitutes the mucociliary system, which is an important part of the innate immune defences [238]. Mucus is composed of approximately 95% water, 2% glycoproteins called mucins, 1% proteins, 1% lipids and 1% inorganic salts [239]. Mucus also contains DNA originating from necrotic cells, and antimicrobial peptides [237]. In CF patients, the mucus composition is altered, impairing the host defences and favouring the development of bacterial infection. Viscosity is greatly increased, which impedes mucociliary clearance and is primarily due to the airway dehydration and higher proportion of mucins [238]. Animal models suggest that the pH of ASL is decreased which inhibits antimicrobial host defences [3] and may add to mucus viscosity by favouring the formation of di-sulphide bonds within mucins. Furthermore, CF sputum has a higher

amino acid content, providing nutrients to potential infecting pathogens and favouring the development of auxotrophic strains [240].

Lipids constitute an important part of the dry mass of sputum, the majority being phospholipids [241]. Phosphatidylcholines are a primary component of lung tissues and fluids in vertebrates and have been accordingly found predominant in asthma and CF sputum [241]. Other lipids previously reported include PEs, PGs, phosphatidylinositols, triglycerides, cholesterol and sphingolipids, amongst which glyco-ceramides were characterised [241]. One of the metabolic peculiarities observed in CF patients is the consistently abnormal distribution of saturated and unsaturated fatty acids, due to the increased activity of fatty acid desaturases [242]. A decreased expression of CFTR has also been reported to lead to the synthesis of ceramides containing longer chain fatty acids, altering the cell membrane structure and functions [243]. Furthermore, they have been described as accumulating in the lungs of CF patients, contributing to many of the pathogenic processes of CF, including inflammation, susceptibility to bacterial infection, cell death and reduction in mucociliary clearance [244].

Thus, CF sputum contains information about a patient's lung condition and the lower airway microbiome, although there are regional differences in bacterial distribution within the lungs and airway fluids [55, 56, 97, 245].

6.1.2. Detection of metabolites characteristic of *P. aeruginosa* in biofluids

Characteristic metabolites of *P. aeruginosa* have long been tracked as markers of infection in biofluids. Rhamnolipids and above all QS molecules, have been targeted, initially by thin layer chromatography (TLC) and then by MS techniques coupled to chromatography. All methods required preliminary extraction of the metabolites using organic solvents. Rhamnolipids were detected in sputum several decades ago; high levels seemed to accompany acute exacerbations [109] although this phenomenon does not seem to have been much investigated subsequently. AQs were also detected in sputum, BAL and lung fluids, at concentrations that positively correlated with bacterial load [97]. The results confirmed that *P. aeruginosa* produced these virulence factors *in vivo*, and suggested that their level varied as infection progressed.

According to a recent profiling study, induced sputum and *P. aeruginosa* culture shared only 1.0% of metabolites, which included notably PE lipids [226]. The method applied LC-MS/MS analysis to organic extracts and allowed the detection of quinolones and quinolone-*N*-oxides in sputum, yet at very low levels compared to the rest of the detected metabolites. The most abundant AQ was HHQ,

representing 0.02% of the total signal, and surprisingly, hydroxy-quinolones such as PQS were not detected in sputum although they were present in the culture extracts. It was said to support the fact that *P. aeruginosa*'s secretions varied depending on the growth conditions. Similarly, only one rhamnolipid – namely Rha-Rha-C10-C10 – was detected out of the 12 produced by the *P. aeruginosa* cultures. More worryingly, clinical phenotypic assays, 16S rRNA sequencing and the detection of these metabolites were not consistent in diagnosing *P. aeruginosa* infection, which questions again the choice of a reference method. For instance, metabolites could be detected in sputum of patient without a diagnosed infection, and vice versa. Possible explanations may be the presence of non-cultivable bacteria, low bacterial density and poor sensitivity of the detection techniques. Other detected metabolites in CF sputum included xenobiotics, sphingolipids and glyco-ceramides, which did not seem to correlate with infection.

In other recent clinical studies, QS molecules were quantified in spontaneous sputum, plasma and urine samples [98, 234]. The method was targeted and involved metabolite extraction with an organic solvent and LC-MS/MS analysis. The molecules assayed included quinolones, quinolone-*N*-oxides, hydroxy-quinolones and homoserine lactones. Concentrations were higher in sputum than in plasma, where they were also higher than in urine, but correlated in the three types of fluids and with bacterial load. Notably, although below the limit of quantification, hydroxy-quinolones were detected contrary to another study [226], which seems to be due to the sample processing and dilution. Again, a few hydroxy-quinolones raised interest as potential markers for early infection or pulmonary exacerbations

6.1.3. Analysis of urine

Urine is certainly not the first sample that comes to mind when trying to detect pulmonary infections. Yet, urine as a sample possesses many advantages: it is readily available non-invasively, in great quantities, and may contain a great diversity of chemicals, including host and bacterial metabolites, drugs and pollutants. Besides the few studies mentioned above demonstrating that urine may carry *P. aeruginosa* metabolites, urine has been analysed previously mainly to investigate the presence of markers of inflammation and tissue destruction in CF. Indeed, biomarkers (e.g. glutathione sulphonamide, desmosine, cathepsin), have been detected during pulmonary exacerbations and correlated with lung inflammation [246], using commercial enzyme-linked immunosorbent assays (ELISA) [247-249] and LC-MS [249, 250]. However, due notably to the variable composition of urine, as well as large variations in pH and in dilution, it remains a challenging sample to analyse. In addition,

because of its high-water content, it is not the best candidate for ambient MS analysis which favours the detection of lipophilic compounds. Nevertheless, a few promising approaches have been successful in analysing urine deposited on various materials using ambient DESI-MS. Steroids for example could be detected with the use of reactive-DESI (involving the addition of a reactant to the solvent) [159]. Kauppila et al. also detected various drugs in urine samples collected from drug abusers deposited on a PTFE support, although the assay required preliminary extraction of the metabolites [251]. The utility of REIMS analysis on urine is currently under investigation in our laboratory. Similarly to DESI-MS, REIMS is not suitable for the analysis of liquid, thus samples are first deposited on adsorbing material or pelleted to retrieve solid biomass and eliminate water. Preliminary results for the detection of urinary infection have shown some potential (data not published, personal communication by Adam Burke, fellow PhD student).

6.1.4. Objectives

Biofluids such as sputum and urine are thought to contain valuable metabolic information regarding the host, including lipids and specific *P. aeruginosa* metabolites. Since these were readily detected from bacterial culture by REIMS, I sought to test sputum and urine directly. Samples were provided by CF patients with and without infection. REIMS-imaging and repetitions were implemented to evaluate heterogeneity of sputum and reproducibility of the REIMS analysis. The REIMS profile of sputum was further interrogated in order to identify potentially interesting metabolites. The collected data were empirically tested to determine whether the infection had a systematic impact on the sputum or urine metabolic profiles, thereby potentially providing a diagnostic test.

6.2. Material and methods

6.2.1. Clinical samples

All the samples were provided by different CF patients attending the Royal Brompton Hospital and frozen at -80°C after collection. The sputum samples were collected in March 2017; urine samples were retrieved from the CF biobank, from a previous collection, from different patients, dating back to 2006-2011. Infection status of the two groups of patients was assessed by the hospital microbiology laboratory with traditional culturing assays on selective media using a separate sputum sample taken

in parallel on the same visit (as described in Chapter 3, section 3.2. *Collection and culturing of microorganisms*).

The presence of *P. aeruginosa* cells in the sputum samples collected for this study was further checked through gene tracking. DNA was extracted as described in Chapter 3, section 3.3. *DNA extraction*. The genes *16S rRNA*, *exoA* and *oprL* were targeted to account for the presence of bacteria, through the 16S ribosomal subunit, and for *P. aeruginosa*, characterised by the exotoxin A and the outer membrane lipoprotein OprL. The gene loci were amplified using PCR and the amplicons detected through electrophoresis [252]. The specimens were considered positive for *P. aeruginosa* infection when the three loci were present; negative if the loci *exoA* and *oprL* were both absent; when only *exoA* or *oprL* was clearly detected the samples were considered “likely positive”; the other samples whose *exoA* and *oprL* signals were not clear were not classified (called “unsure”). Three samples were not congruent with the culture results: two sputa considered as not infected revealed the presence of *P. aeruginosa* DNA; one sample collected from an infected patient did not show either of the two gene markers. For the purposes of this study and since bacterial distribution in the lungs can be spatially heterogeneous, the molecular results were considered more reliable for categorising the sputum sample being analysed by REIMS. Final numbers of samples are detailed in Table 6.1, gene tracking results for individual samples are available in Appendix 13.

Table 6.1. Number of clinical samples collected from CF patients with infection status.

NB: sputum and urine were collected from two different groups of patients.

<i>P. aeruginosa</i> infection	Clinical samples	
	Sputum*	Urine
Negative	18	28
Positive	22	42
Likely positive	6	-
Unsure	5	-
Total	51	70

*Infection status as determined through the detection of the genes *16S rRNA*, *exoA* and *oprL*. Negative: *16S rRNA* present, *exoA* and *oprL* absent; Positive: *16S rRNA*, *exoA* and *oprL* present; Likely positive: *16S rRNA* present, *exoA* or *oprL* present; Unsure: *16S rRNA* present, signal unclear for *exoA* and/or *oprL*.

6.2.2. REIMS analysis of clinical samples

All the experiments were conducted using laser-assisted REIMS and the automated Tecan platform to acquire data in the $m/z = 50$ -2500 range, as described in Chapter 3, section 3.4.2.2. *Automated, laser-assisted REIMS*. Analysis and waiting times between each burn were 8 and 25 s, respectively, except for imaging analysis. Data were processed as described in Chapter 3, section 3.4.4. *Data analysis*.

6.2.2.1. Sputum

An approximate volume of 0.5 mL of the sputum samples were placed in 24-well plates, and analysed at a laser power of 4 W. In parallel, MS images of 4 cm² were acquired on a few sputum samples: one positive for infection (sample 4988), one likely positive (sample 4964) and one negative (sample 5005). The samples were spread on a CBA plate to avoid burning of the plastic support. Burns were applied every 1 mm, at a power of 2 W, for 2 s, with a waiting time of 5 s between burns (see example of a sputum sample after imaging in Figure 6.1). For the reproducibility study, analyses were repeated five times on 36 samples; the individual samples were mixed in each well with a plastic loop between the repeats and the burnt material was placed aside. All the analyses were carried out on a single day to avoid batch effect. TIC-normalisation was applied and the reduced $m/z = 50\text{--}1200$ range was used throughout the statistical analyses. ROC curve analysis was implemented and AUC were calculated on the Metaboanalyst platform after filtering of the 5000 most intense median bins, log-transform and auto-scaling. Classification was performed using the Random Forest algorithm. Images were built in R.

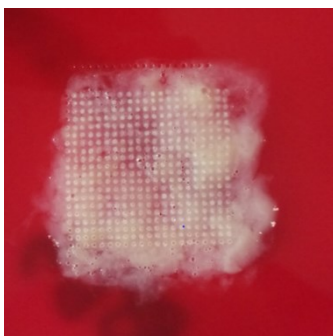


Figure 6.1. Picture of a sputum sample spread on a CBA plate, after laser-REIMS-imaging at a 1 mm spatial resolution.

6.2.2.2. Urine

Aliquots of 40 µL of the urine samples were adsorbed onto two different types of filter material, placed inside the wells of a 24-well-plate: WhatmanTM grade 595 qualitative filter paper (4-7 µm pore size, Whatman, Kent, UK); WhatmanTM nitrocellulose (NC) membrane filters (0.45 µm pore size, Whatman, Kent, UK). The filters and membranes were let to dry inside the biosafety cabinet before laser-REIMS analysis. The laser was operated in SuperPulse mode, at 4.5 W on filter paper, and 3.5 W on NC membranes. In addition, 1.8 mL of each of the raw urine samples were placed in a 96-Deepwell plate (Eppendorf, Stevenage, UK), sealed with an aluminium plate seal (VWR, Leicestershire, UK) before undergoing centrifugation at 3486 x g for 10 min. The supernatant was discarded, whilst the pelleted urine was analysed using a custom-made, far-focus laser head (Waters Research Centre, Budapest,

Hungary), fitting inside the wells of the 96-Deepwell plates. Laser power was 3.5 W, in SuperPulse mode. Classification was performed using the Random Forest algorithm on the reduced $m/z = 50$ -1200 range after TIC-normalisation.

6.3. Results

6.3.1. Detection of *P. aeruginosa* through sputum analysis

6.3.1.1. Sputum metabolic profile

The analysis of sputum samples with laser-assisted REIMS revealed a rich but variable metabolic profile, including in particular an intense signal of fatty acids ($m/z = 200$ -350), phospholipids ($m/z = 600$ -850) and some higher mass lipids ($m/z = 850$ -1100) (see two examples of mass spectra in Figure 6.2).

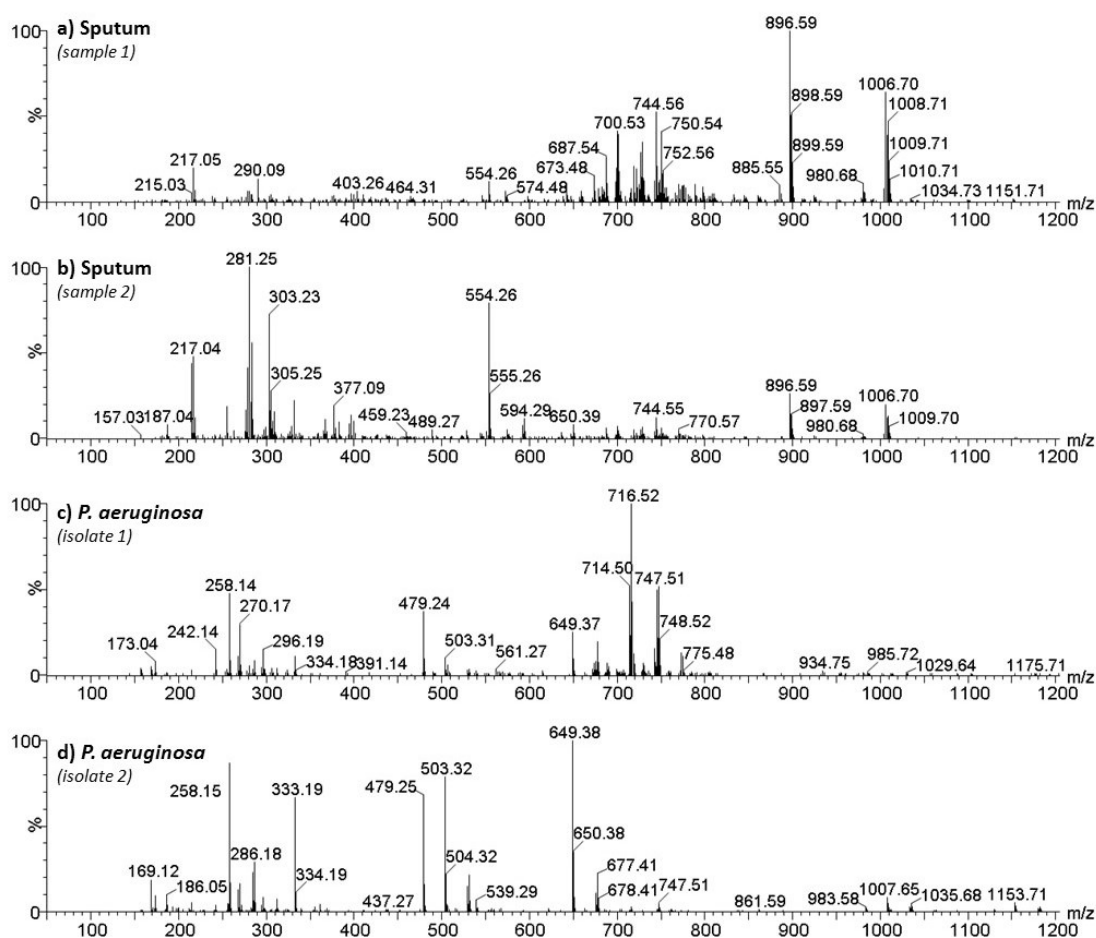


Figure 6.2. Compared mass spectra of (a & b) sputum samples infected by *P. aeruginosa* and (c & d) examples of *P. aeruginosa* colonies.

Acquired with laser-assisted REIMS; data were corrected for mass drift using the LeuEnk ion ($m/z = 554.26$), and background-subtracted.

Some of the highest peaks in the sputum spectra (Figure 6.2a and b) were tentatively identified by exact mass comparison. Thus, $m/z = 281.2575$ and $m/z = 303.2320$ are likely to be oleic acid (FA(18:1)), and arachidonic acid (FA(20:4)), respectively. Indeed, the production of arachidonate by the host is increased due to the higher activity of fatty acid desaturases in CF lungs [242]. Further oxidation of arachidonic acid leads to the formation of eicosanoids such as leukotrienes, previously reported in CF sputum and urine [253]. Eicosanoids are one mediator of inflammation, and their generation may be triggered by the development of *P. aeruginosa* infection [254]. These molecules have been previously detected in CF sputum [255] and further interrogation of the data may reveal their detection with REIMS.

The intense peaks in the higher mass range at $m/z = 896.5894$, $m/z = 980.6777$ and $m/z = 1006.6997$ had the typical isotopic distribution of chlorinated adducts and were putatively annotated as diglyco-ceramides: diglyco-Cer(34:1), diglyco-Cer(40:1) and diglyco-Cer(42:2). Ceramides are glycolipids belonging to the sphingolipid family which is little synthesised by bacteria. Their composition may vary in terms of fatty acid and sugar residues (according to the chemical structures in Figure 6.3, diglyco-ceramides are cerebroside with an additional sugar). They impact the physical properties and metabolic functions of the epithelial cell membranes and their accumulation may contribute to CF pathogenesis. Furthermore, *P. aeruginosa* has been reported to contribute to trigger ceramide release, which further favours the development of infection [256]. However, previous analyses of sputum did not observe correlations of sphingolipids and ceramides with bacterial infection [226, 257].

In agreement with previous observations [226], some phospholipid peaks were common to *P. aeruginosa* and sputum, although detected in different proportions (e.g. $m/z = 744.5563$ and $m/z = 770.5696$ which were tentatively annotated as PE(36:1) and PE(38:2)). Phosphatidylcholines are reported as being the predominant phospholipids, yet they are positively charged lipids and were not detected with the negative ionisation mode used here.

Preliminary examination of the data revealed neither QS molecules nor rhamnolipids in the mass spectra of sputum, whether it was infected or not. Nevertheless, the molecules detected with REIMS are metabolically meaningful and in agreement with the literature. Their direct detection in sputum may therefore be of clinical use, e.g. to evaluate a patient's status or risk of infection, to monitor inflammation and the efficiency or the impact of a treatment.

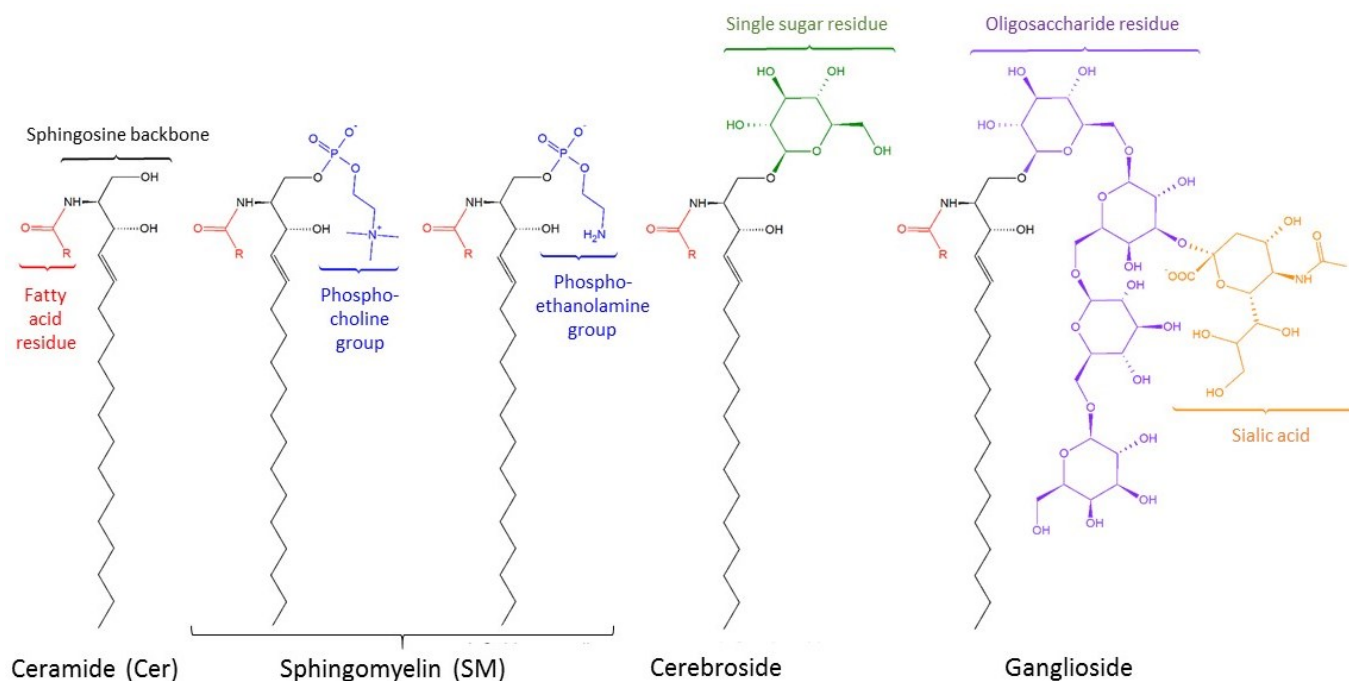


Figure 6.3. General chemical structure of sphingolipids.

(Reproduced under the Creative Commons Attribution-Share Alike 3.0 Unported license from Wikipedia user LHcheM).

6.3.1.2. Heterogeneity of the sputum samples: impact on the metabolic profile

Sputum samples are physically very heterogeneous, with variable content of water due to mixing with saliva, and the presence of solid plugs, leading to important changes in global viscosity. In Figure 6.4 are shown examples of sputum samples.

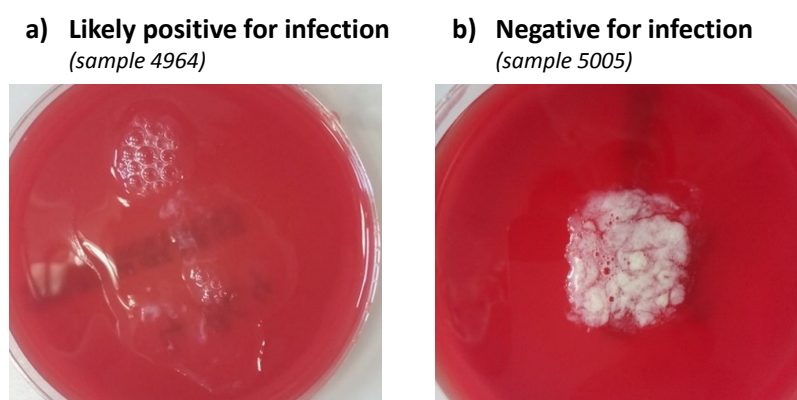


Figure 6.4. Pictures of sputum samples spread on a CBA plate.

Sputum collected from a CF patient (a) infected by *P. aeruginosa* and (b) not infected.

REIMS has been designed for direct rapid analysis and burning is usually applied on a small area ($< 1 \text{ mm}^2$) for a few seconds only. Thus, it was important to check how the location of the burning site would impact the detected metabolic profile. Therefore, three sputum samples – one positive, one

likely positive, one negative for infection – were screened using REIMS-imaging, in order to visualise their metabolic distribution, notably for a few molecules characteristic of *P. aeruginosa*, such as AQs and rhamnolipids. Notable ion images are presented thereafter, relative to the global metabolic distribution (Figure 6.5), the distribution of sputum metabolites (Figure 6.6) and the distribution of *P. aeruginosa* metabolites (Figure 6.7).

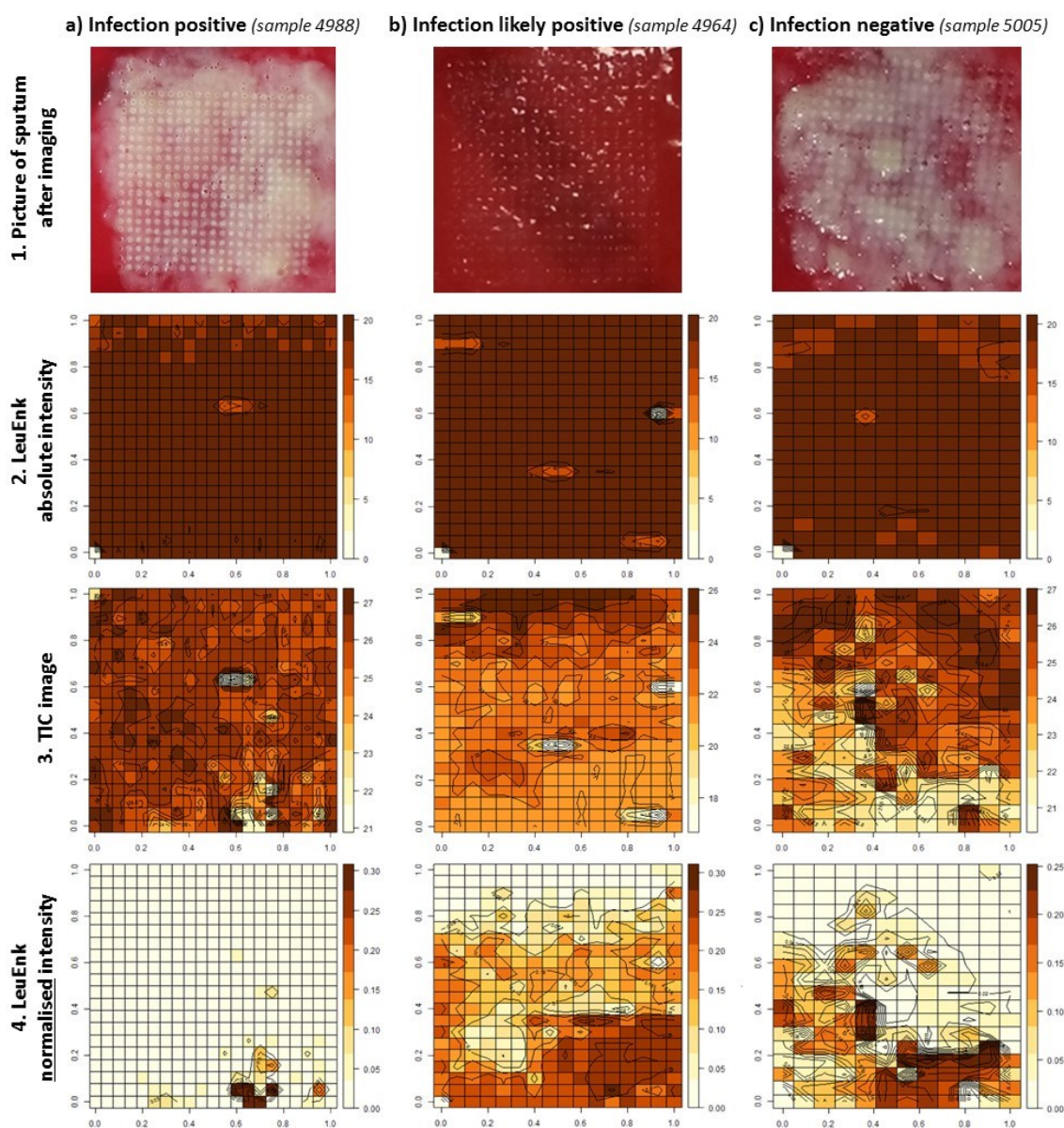


Figure 6.5. Pictures of sputum samples and ion images obtained with laser-REIMS showing global distribution of the signal.

Sample in column (a) is positive for *P. aeruginosa* infection, (b) likely positive, (c) negative; lines show (1) pictures of the samples after laser-REIMS imaging, and ion images: (2) absolute signal of the lock mass compound LeuEnk, (3) TIC signal, (4) LeuEnk signal normalised against the TIC. Intensities are shown as a darkening orange on a log2 scale for (2) and (3), relative scale for (4). Lines in ion images show the density level sets (built in R).

As a first approach, the total signal (TIC) and the LeuEnk reference signal shown in Figure 6.5 gave an overall picture of the global metabolic distribution and sample heterogeneity. The images in Figure

6.5, line (2) confirm that the LeuEnk signal was constant throughout the imaging of the three samples, meaning that the variations observed in signal intensity originated from the samples themselves and not from the analysis. The TIC images in Figure 6.5(3) represent the global intensity of the signal, i.e. the amount of metabolic information collected at each pixel: the higher the TIC, the more numerous and/or concentrated the metabolites. Samples (a) and (c) showed similar intensities with a maximum of $\log_2(\text{TIC}) = 27$ whereas sample (2) was lower with $\log_2(\text{TIC})_{\text{max}} = 26$, which represents a 2-fold variation in absolute intensity. This was consistent with the high level of water observed in this sample (see previous Figure 6.4 before imaging). The global metabolic distribution was rather homogeneous in sample (a) and heterogeneous in (b) and (c), with a gradient in (b) where concentration decreases from top to bottom. Since the LeuEnk signal was constant, its proportion increases as the total signal decreases and images in Figure 6.5(4), showing the normalised signal of LeuEnk, are the reverse of the TIC images.

The distribution of sputum metabolites was then examined to check whether they were consistent with the global signal. Ion images corresponding to the sputum peaks of FA(20:4), PE(36:1) and diglyco-Cer(42:2) are compared to the TIC distribution in Figure 6.6. PE(36:1) and diglyco-Cer(42:2) are constituents of cell membranes. This PE is present in *P. aeruginosa*, whereas sphingolipids such as diglyco-Cer(42:2) are rare in bacteria [150] and were not detected by REIMS in the previous analyses of *P. aeruginosa* cultures. Therefore, these may be produced by human cells, and their distribution globally followed that of the TIC as shown in Figure 6.6. Interestingly, FA(20:4) did not follow the same pattern. This suggests that this compound is not produced through the fragmentation of complex lipids due to the burning. Indeed, its distribution differs from that of the PE and ceramide compounds. Besides, since polyunsaturated fatty acids are very uncommon in bacteria [150, 258], it is likely to be from human origin. Thus, FA(20:4) may be released in the lung fluids by the epithelial cells rather than sticking to them.

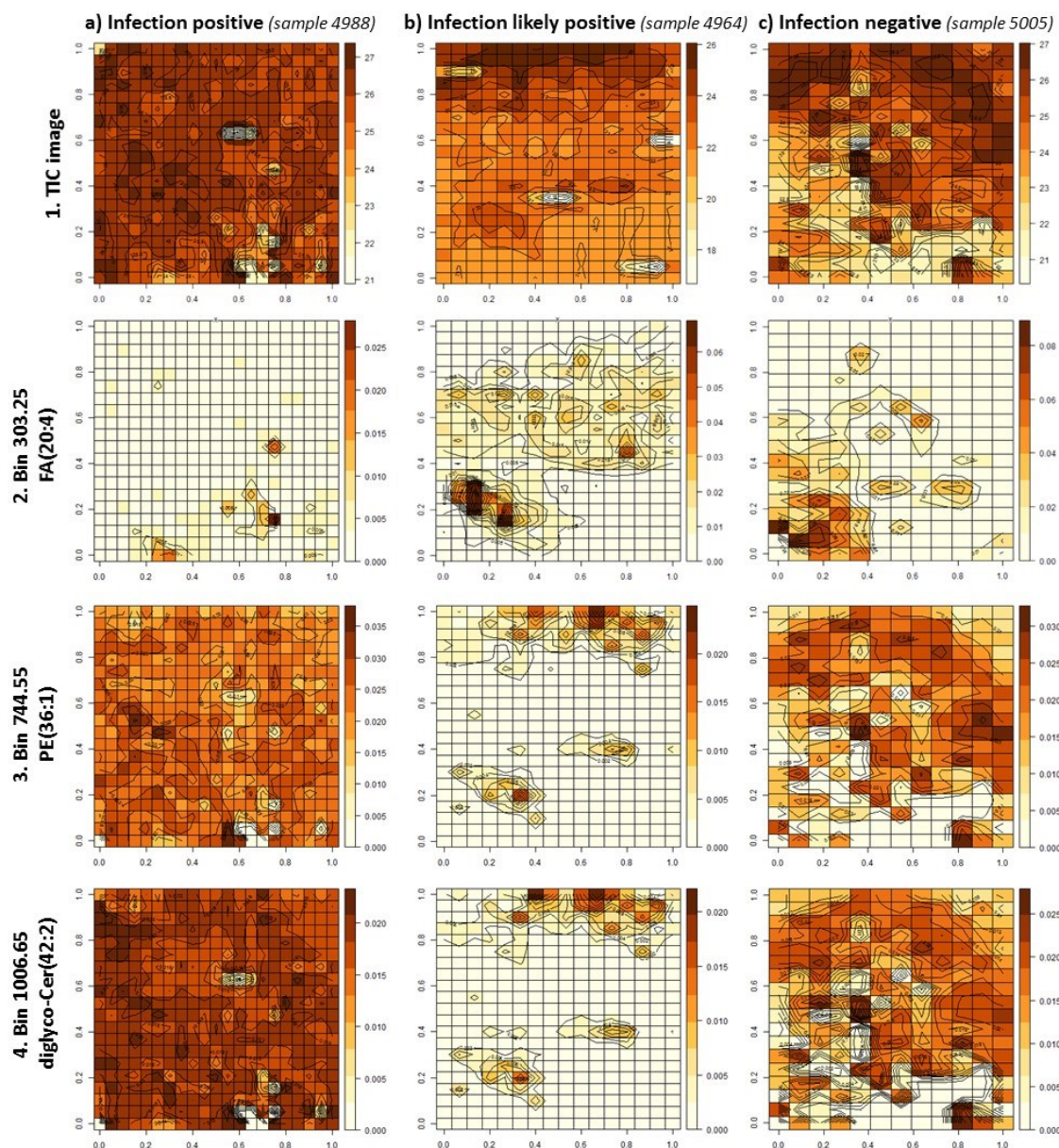


Figure 6.6. Ion distribution of sputum features obtained from sputum samples imaged with laser-REIMS.

Sample in column (a) is positive for *P. aeruginosa* infection, (b) likely positive, (c) negative; lines show (1) TIC signal, and relative distributions of sputum metabolites normalised against the TIC: (2) FA(20:4) (bin 303.25), (3) PE(36:1) (bin 744.55), (4) diglyco-Cer(42:2) (bin 1006.65). Intensities are shown as a darkening orange, on a log2 absolute scale for the TIC in (1), on a relative scale for (2), (3) and (4). Lines in ion images show the density level sets (built in R).

Finally, the distribution of characteristic metabolites of *P. aeruginosa* was sought and compared to that of the TIC and the sputum metabolites (Figure 6.7). The compounds HHQ, Rha-Rha-C10-C10 and PE(18:1/16:0) (i.e. PE(34:1)) were selected as representative of *P. aeruginosa* because they had been previously readily detected both with REIMS in *P. aeruginosa* and in previous studies in sputum [98, 226]. Their distribution is shown in Figure 6.7. They were compared to the TIC and sputum metabolic distribution in order to detect the possible presence of bacterial spots in the sputum, and thereby evaluate the risk of missing signal related to infection when analysing a restricted area. Indeed, a similar distribution for the three bacterial metabolites, but different from the sputum would suggest the presence of a bacterial spot. Although neither AQs nor rhamnolipids were visible at first examination of the sputum data, the instrument was set to record signal throughout the pre-defined m/z range, whether it belongs to the sample or to the background. Examination of the raw data for some of the pixels with the highest signal for the bins of HHQ and Rha-Rha-C10-C10 did not reveal any peak corresponding to their accurate mass. Indeed, their bins showed a very low signal (below 0.1% of the TIC) and followed the same distribution as the TIC, altogether suggesting that they belonged to the background. Therefore, it was not possible to conclude whether sputum contained bacterial spots or not. PE(18:1/16:0) displayed slightly higher intensities than the other *P. aeruginosa* markers (maximum 0.7% of the TIC). This PE is an important component of *P. aeruginosa* cells but may also be found in other cells such as the lung epithelium and its distribution was similar to that of the TIC.

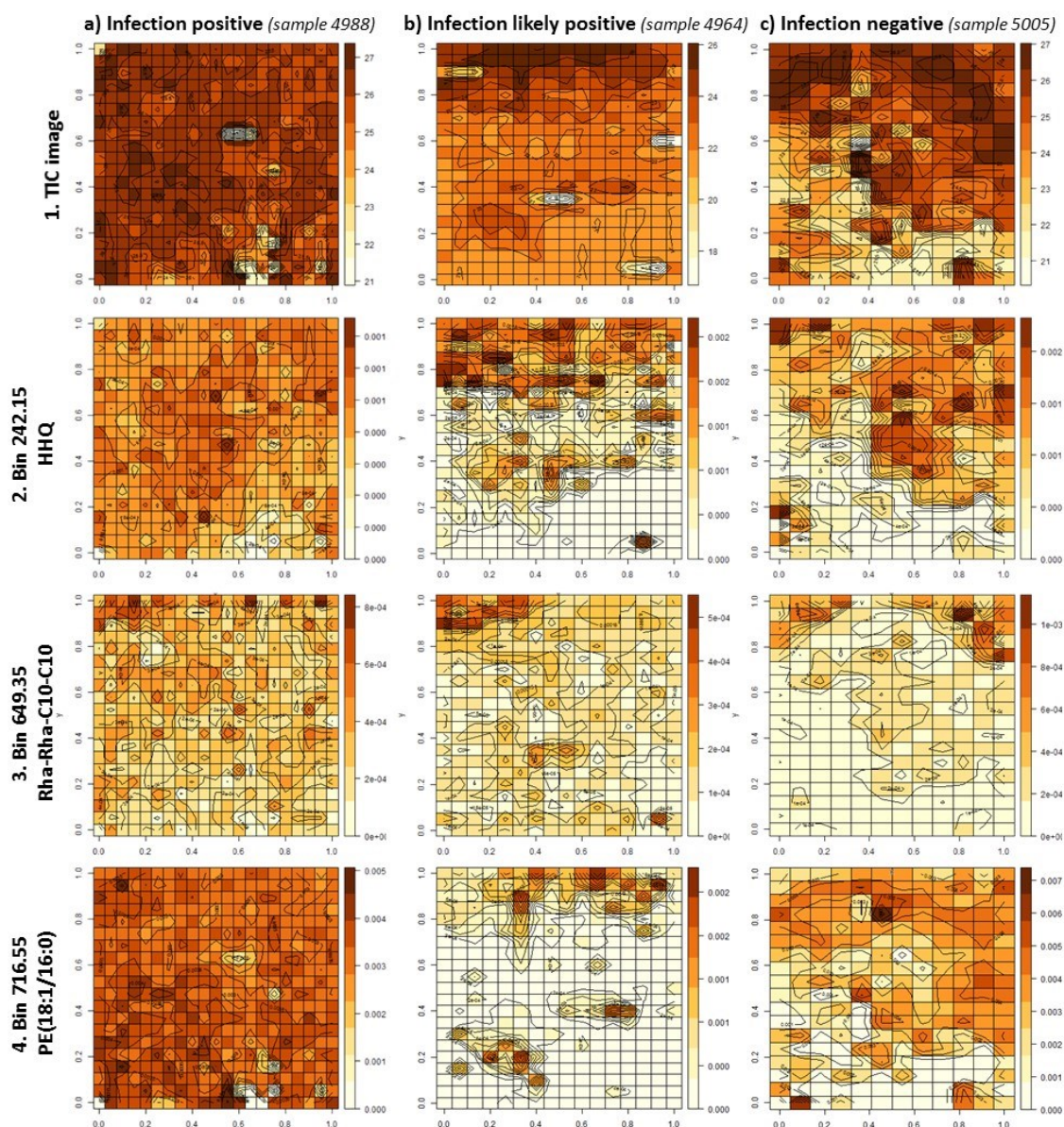


Figure 6.7. Ion distribution of bacterial features obtained from sputum samples imaged with laser-REIMS. Sample in column (a) is positive for *P. aeruginosa* infection, (b) likely positive, (c) negative; lines show (1) TIC signal, and relative distribution of *P. aeruginosa* metabolites normalised against the TIC: (2) HHQ (bin 242.15), (3) Rha-Rha-C10-C10 (bin 649.35), (3) PE(18:1/16:0). Intensities are shown as a darkening orange, on a log₂ scale for the TIC in (1), on a relative scale for (2), (3) and (4). Lines show the density level sets (built in R).

The distribution of the sputum metabolites confirmed that its metabolic profile could be spatially heterogeneous and may vary locally, quantitatively and qualitatively, which supports the acquisition of replicates on various portions of the sputum. However, bacterial metabolites were not detected and bacterial heterogeneity of sputum could not be assessed. This may require localised quantification of *P. aeruginosa* using molecular techniques, e.g. on fractionated sputum samples. Meanwhile, it also implies that any diagnostic value measured by REIMS would rely on modifications of the sputum metabolome due to the infection, rather than *P. aeruginosa* biomarkers.

6.3.1.3. Evaluation of the reproducibility of REIMS analysis on sputum samples

Results obtained in Chapter 5 suggested that REIMS reproducibility may allow the prediction of isolate identity with up to 95% accuracy. However, sputum samples were shown to be heterogeneous, therefore, the reproducibility of the analysis was further assessed, using 36 samples with five replicates. The 36 samples included 13 negative, 15 positive, six likely positive and two sputa with an uncertain infection status. The REIMS profiles detected between $m/z = 50$ -1200 were classified according to sample using the Random Forest algorithm in order to evaluate reproducibility. The cross-validation matrix is presented in Figure 6.8 and the detailed scores are available in Appendix 14.

The confusion matrix shows that, whilst some samples are well predicted (e.g. 5007, 5010, 5018, 4970 had F1-scores of 100%), others could not be classified (e.g. 4994, 4998, 5002, 5008, 5016, 4978, 4981, 4983 scored 0%), leading to a low overall accuracy of 57%. Since it has been established that REIMS reproducibility is high when repetitions are carried out within a short time, most of the variability observed can be attributed to the heterogeneity within samples. Another explanation may be the alteration of the samples caused by successive burning, although no consistent sequence effect was observed. The samples previously imaged, i.e. 4988, 4964 and 5005, had F1-scores of 80%, 89% and 44%, respectively. This confirmed that sample 5005 was the most heterogeneous amongst these, but further indicates that much more heterogeneous samples existed and that replicates are essential to reduce the risk of missing information regarding infection.

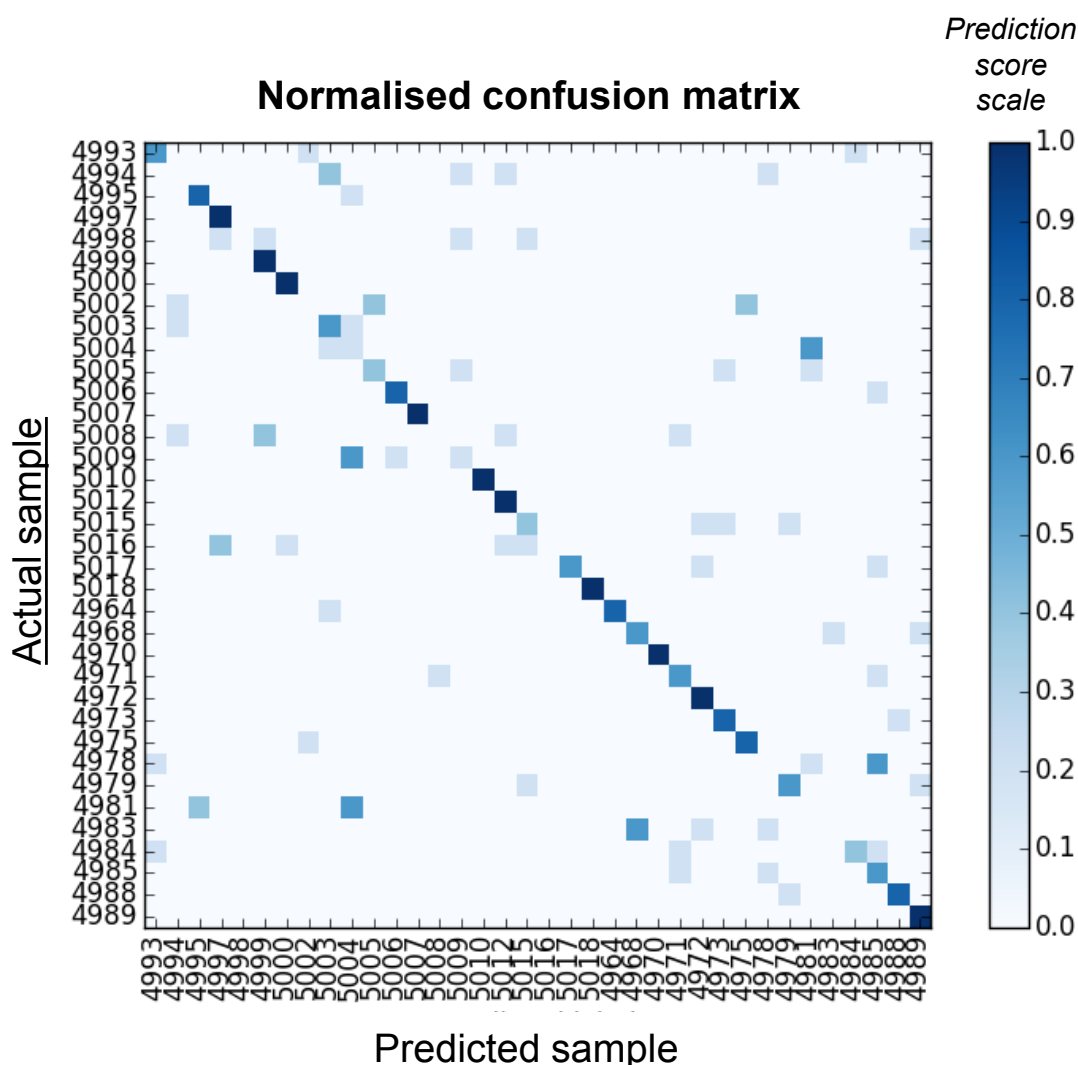


Figure 6.8. Confusion matrix for Random Forest classification of replicates obtained from the REIMS analysis of 36 CF sputum samples.

Leave-one-out cross-validation applied on a REIMS dataset including five replicates per sample. Real and predicted identities are displayed respectively on the left-hand and the bottom sides. Frequency of prediction for each isolate is represented in darkening blue. Overall prediction was 57% accurate.

6.3.1.4. Classification of infected sputum samples using REIMS

A total of 40 sputa, including the samples with definite typing results only – positive and negative for infection – were used to assess the utility of their REIMS profile to predict infection status. Between two and 13 replicates were acquired on the 18 negative and 22 positive samples for infection (see detailed numbers in Appendix 13). In order to account for the heterogeneity of the samples, a “maximum profile” was computed: for each variable – i.e. each m/z bin – of each sample, the maximum intensity amongst the replicates was kept. This was to avoid the dilution of discrete information – i.e. a specific peak showing in a limited number of replicates only – which would occur

if calculating the mean or the median. The data were also filtered and the 150 most intense features were selected within $m/z = 50$ -1200 for the Random Forest classification. This reduced the impact of the background, which contributes greatly to the variability of the detected profile. This data processing led to the best classification results, which are presented in Table 6.2.

Table 6.2. Random Forest cross-validation results for the classification of sputum samples according to infection status.

Leave-one-out cross-validation applied to the “maximum profiles” of 40 samples analysed with REIMS. Overall prediction was 63% accurate.

<i>P. aeruginosa</i> infection status	Precision	Recall	F1-score	Number of samples
Negative	0.62	0.44	0.52	18
Positive	0.63	0.77	0.69	22
Average / Total	0.62	0.62	0.61	40

The proportion of samples correctly identified – either as negative or positive for infection – was 63% accurate. Despite this low classification score, the features were further interrogated to unravel potential individual metabolic differences between the two groups of sputa. ROC curve analysis was performed and the AUC was calculated for the 150 most intense m/z bins (see list in Appendix 15). The maximum AUC was 0.77, a third of the features (47/150) had AUC above 0.60, and amongst them, five had been previously tentatively identified and discussed. Their distributions (of maximum intensities) throughout the samples are shown in boxplots in Figure 6.9.

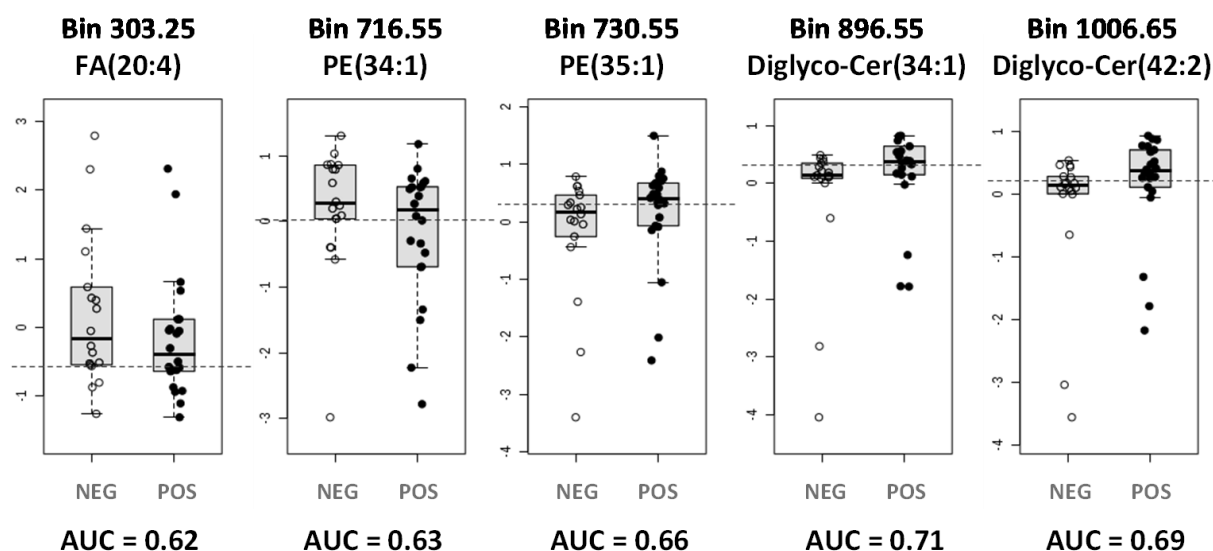


Figure 6.9. Boxplots showing the distribution of selected metabolites in sputum samples negative and positive for *P. aeruginosa* infection.

Negative samples (NEG) are plotted on the left-hand side of the boxes, positive samples (POS) are on the right-hand side. Bins, (putative) identification and AUCs are given above and below the boxplot of each features. The dashed lines show the optimal cutoff to separate the two classes for each compound. NB: Plots show the maximum intensity (relative to the TIC signal) detected by REIMS in one sample's replicates (generated on the Metaboanalyst platform).

Interestingly, the diglyco-ceramides are amongst the 10 most differentiating features (AUC around 0.7). They demonstrate higher levels in sputum samples infected by *P. aeruginosa*, which was in agreement with the fact that *P. aeruginosa* has been described to stimulate the conversion of sphingolipids to ceramides [256]. Notably, the two PEs showed distinctive trends. The bin of PE(34:1) may comprise PEs with varying fatty acid residues, including PE(18:1/16:0), which is abundant in sputum and *P. aeruginosa*, and seemed slightly higher in the negative samples. On the other hand, PE(35:1) contains a single fatty acid with an odd number of carbons, which is typical of bacteria and would be consistent with higher levels in the infected sputa. Finally, levels of arachidonic acid (FA(20:4)) appeared globally slightly lower in the infected sputa. This is again consistent with the literature which suggested that *P. aeruginosa* infection increased the synthesis of eicosanoids from arachidonic acid [254].

Beyond the heterogeneity of the samples, their variability between patients may be an important factor in the difficulty of diagnosing infection. Indeed, the sputum composition is very patient-dependant (and probably even varies between samples collected from the same patient). Since there were no detectable biomarkers for *P. aeruginosa*, the classification relied on metabolic variations occurring in the sputum, and it seems that the sputum metabolic profile – at least the one detected with REIMS – was more impacted by the host than the presence of *P. aeruginosa*. Nevertheless, some features showed trends differentiating between the infected and non-infected sputum samples. Although their significance was not robust enough to improve the classification when taken alone, all the observations were in concordance with the metabolic processes described in the literature.

6.3.2. Detection of *P. aeruginosa* through urine analysis

As part of this project exploring REIMS capacities for the detection of infection, urine samples collected in previous studies were subjected to REIMS analysis for screening. The detected metabolic profile was classified using Random Forest algorithm in order to assess its potential for infection diagnosis.

6.3.2.1. REIMS analysis of urine deposited on filter paper

Urine aliquots of 70 samples (28 negative and 42 positive for *P. aeruginosa* infection) were deposited onto two different adsorbing materials: filter paper and NC membranes. The raw samples had varying

appearances, both in colour and cloudiness. Similarly, the profiles detected were very variable with a high background signal as visible on the example spectra in Figure 6.10. The metabolic profile differed depending on the deposition material, which significantly contributed to the background signal in the case of filter paper. Some common peaks were found on both materials, particularly in the lower m/z range (e.g. $m/z = 145.06$, $m/z = 263.10$ or $m/z = 345.16$). Indeed, most of the signal was in the lower m/z range; exact mass comparison did not lead to identification of any reliable discriminant feature and no peaks characteristic of *P. aeruginosa* were detected.

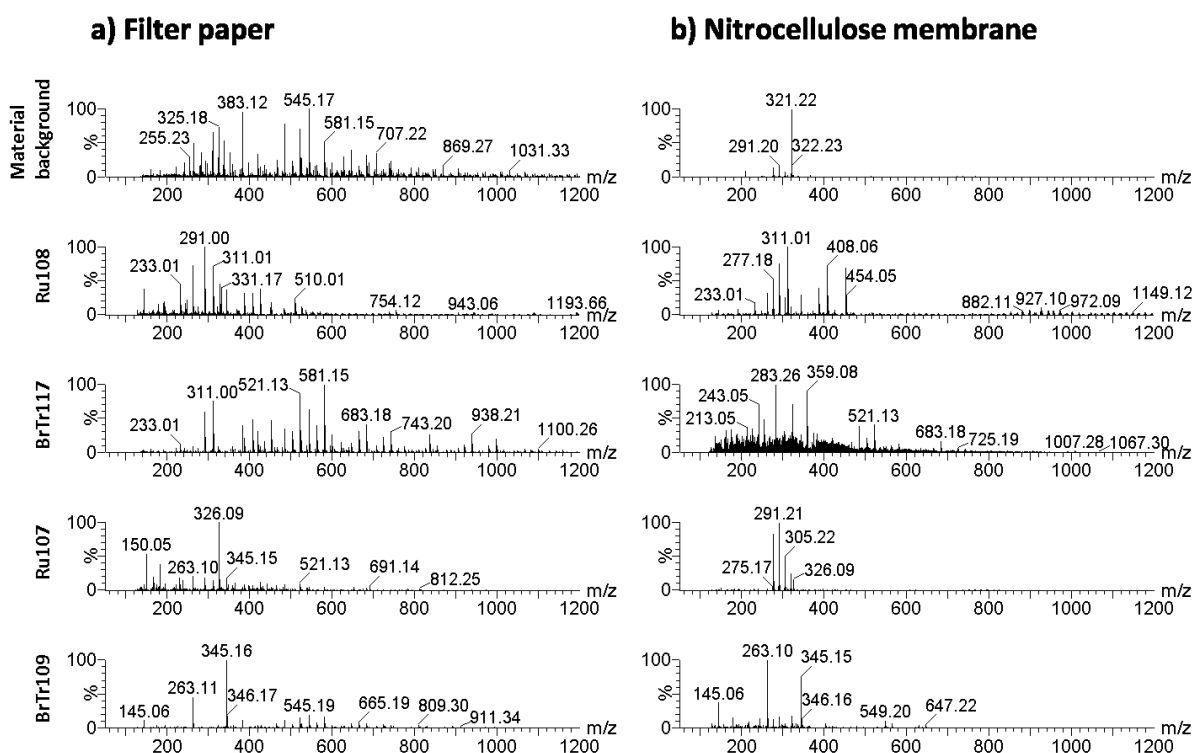


Figure 6.10. Comparative examples of mass spectra obtained with laser-REIMS analysis on urine deposited on (a) filter paper and (b) NC membranes.

The first spectra show the material background, followed by two spectra acquired on urine collected from non-infected patients (Ru108 and BrTr117), and two from patients infected with *P. aeruginosa* (Ru107 and BrTr109); sample numbers are given on the left-hand side. Data were corrected for mass drift using the LeuEnk ion ($m/z = 554.26$), and background subtracted.

All the bins within $m/z = 50$ -1200 were used for the Random Forest classification. Results of the cross-validation are presented in Table 6.3 and Table 6.4 for the analyses on filter paper and NC membranes, respectively.

Table 6.3. Random Forest cross-validation results for the classification of urine samples deposited on filter paper according to infection status.

Leave-one-out cross-validation applied on the REIMS profiles of 70 samples. Overall prediction was 54% accurate.

<i>P. aeruginosa</i> infection status	Precision	Recall	F1-score	Number of samples
Negative	0.33	0.14	0.20	28
Positive	0.59	0.81	0.68	42
Average / Total	0.49	0.54	0.49	70

Table 6.4. Random Forest cross-validation results for the classification of urine samples deposited on NC membranes according to infection status.

Leave-one-out cross-validation applied on the REIMS profiles of 61 samples. Overall prediction was 51% accurate.

<i>P. aeruginosa</i> infection status	Precision	Recall	F1-score	Number of samples
Negative	0.22	0.08	0.12	25
Positive	0.56	0.81	0.66	36
Average / Total	0.42	0.51	0.44	61

The metabolic profile detected on urine did not allow the prediction of the infection status on any of the deposition materials tested. Although *P. aeruginosa* metabolites and inflammation markers have been previously detected in urine, their low concentration, combined to the small volume analysed here (40 µL) and the important and intrinsic variability of urine samples, may preclude their detection with REIMS.

6.3.2.2. REIMS analysis of pelleted urine

In order to overcome the issue of dilution inherent to urine, and to concentrate the information-containing biomass, the samples underwent centrifugation and the resulting pellets were subjected to REIMS analysis. The aqueous phase, which increases background and interferes with metabolite detection, was discarded. Indeed, REIMS is more likely to detect water-insoluble metabolites. They may include bacterial metabolites, such as lipids and AQs which are poorly soluble in water [108], and host metabolites – e.g. fatty acids, triglycerides, cholesterol derivatives – whose levels have been previously reported to vary in the host plasma and lung tissue in case of infection [127, 128].

Pelleting of the urine samples resulted in various amounts of biomass of diverse colours, from white to pink; for some samples, no pellets were produced (see picture of the pellets in 96-well plate in Appendix 16). Examples of spectra acquired with REIMS are shown in Figure 6.11.

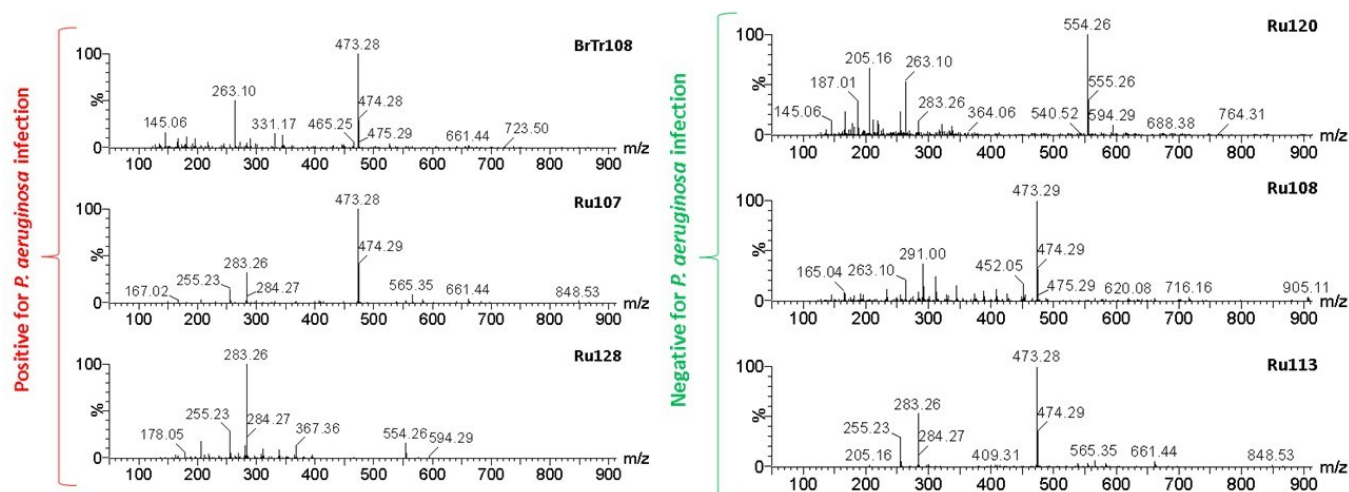


Figure 6.11. Examples of mass spectra obtained with laser-REIMS analysis on urine pellets.

Three spectra acquired on urine collected from non-infected patients (left-hand side), three from patients infected with *P. aeruginosa* (left-hand side); sample numbers are given on the top right-hand corner of the spectra. Data were corrected for mass drift using the LeuEnk ion ($m/z = 554.26$), and background subtracted.

A few peaks in the lower m/z range were consistent throughout the samples, including the saturated palmitic and stearic acids (FA(16:0) and FA(18:0) at $m/z = 255.23$ and 283.26 , respectively). The rest of the spectra was poor, except for a peak at $m/z = 473.28$, which could not be identified. Despite this sparse metabolic profile, Random Forest classification was tested on the range $m/z = 50$ -1200. Results of the cross-validation are presented in Table 6.5.

Table 6.5. Random Forest cross-validation results for the classification of urine pellets according to infection status.

Leave-one-out cross-validation applied on the REIMS profiles of 70 samples. Overall prediction was 54% accurate.

<i>P. aeruginosa</i> infection status	Precision	Recall	F1-score	Number of samples
Negative	0.25	0.07	0.11	28
Positive	0.58	0.86	0.69	42
Average / Total	0.45	0.54	0.46	70

The model built did not allow the prediction of the infection status of the patients. Although *P. aeruginosa* metabolites have been detected before in urine, the pellets produced here did not provide sufficient metabolic information. However, whether this is due to an insufficient amount of biomass or the inadequacy of REIMS for the analysis of urine metabolites is unclear.

6.4. Discussion

6.4.1. Limitations and perspectives for diagnosis

This first approach showed that our current REIMS setup and methodology are not suitable for the direct and untargeted analysis of unprocessed urine. The detected metabolic profiles were poor and did not allow the classification of infection status. Indeed, urine samples are very variable and diluted, and, although *P. aeruginosa* metabolites have been detected before using other methods [98, 234], it is likely that REIMS is not sensitive enough to allow the detection in urine of metabolites produced in the lungs. The methodology might benefit from further preliminary treatment of the samples, e.g. extraction or concentration; however, this would defeat one primary objective of this project, which was to develop a direct-from-sample diagnostic tool. Other ambient methods have been previously tested in other contexts and would be worth investigating, such as DESI-MS [251] and extractive-ESI [259], or even the analysis of the VOCs released by urine with SESI or electronic noses [260, 261]. However, the results may again greatly benefit if the dilution issue inherent to urine could be overcome.

The metabolic profile detected from sputum using the present methodology and data processing did not allow infection diagnosis. A challenging characteristic of sputum remains its heterogeneity and its variability between (and probably within) patients, which may mask the global metabolic impact of the infection. Whilst replicates are essential, they may not be sufficient in the case of a discrete analysis like REIMS, and it might be necessary to include a short pre-processing step to homogenise the sputum, for instance centrifugation or addition of Sputasol to denature the proteins. Previous studies have highlighted the low levels of *P. aeruginosa* metabolites in sputum [226, 234], which may explain the lack of discrimination here described. Preliminary assays conducted on sputum spiked with pure bacterial cells suggested that some metabolites could be detected with REIMS only if there was a high concentration of bacteria (data not shown). These experiments need to be repeated on a wide range of sputum types, spiked with different strains, in order to determine the bacterial load detection threshold. The profiling workflow followed here as a first approach is not adapted to the detection of traces of metabolites; other data (pre-)processing approaches exist for biomarker discovery that may be more successful. Moreover, a more sensitive targeted technique, detecting a reduced number of specific compounds using tandem MS, would be worth testing, by focusing on bacterial metabolites, or on a selection of host metabolites which differentiate patients with and without *P. aeruginosa* infection. Other interesting work on sputum include the detection of another family of QS molecules produced by *P. aeruginosa*, namely AHLs, by cell-free biosensors [262]. Others have investigated the

utility of analysing VOCs released by sputum to diagnose *P. aeruginosa* infection using electronic noses [179] or GC-MS [162].

Finally, analysis in positive ionisation mode of urine and sputum may bring interesting complementary information, for instance with the potential detection of phosphatidylcholines and PYO.

6.4.2. The potential of REIMS for the direct analysis of biofluids

REIMS profiling analysis on sputum detected a rich metabolic profile consisting, at first glance, primarily of lipids. However, it did not reveal the presence of *P. aeruginosa* biomarkers, and a prediction model based on the whole detected profile did not allow the accurate classification of patients. Nevertheless, interesting host metabolites have been detected. The observations made here on a small number of features were in agreement with the literature. Further interrogation of the spectral data in future work may reveal the detection of many more meaningful metabolites. The direct detection in sputum of host-produced compounds may be useful to microbiologists and clinicians in understanding biological phenomena occurring during infection or treatment. Molecules such as ceramides and eicosanoids provide information about lung health, and may define targets for therapeutic intervention. Indeed, treatments inhibiting the action of acid sphingomyelinase early on in the ceramide pathway, showed promising improvements in FEV₁ [263]. Besides, preliminary results suggested that polyunsaturated fatty acids such as arachidonic acid varied during a pulmonary exacerbation and may be useful in predicting relapse [264]. In addition, the visualisation of the distribution of the different chemical families within the sputum itself might as well lead to the discovery of important information, e.g. about metabolic function of host cells, distribution of drug compounds.

This experimental plan did not confirm the initial hypothesis and REIMS was unsuccessful in diagnosing *P. aeruginosa* infection or identifying specific biomarkers directly from clinical samples. Nevertheless, REIMS is a promising technique for sputum analysis, and the profile revealed potentially relevant information about the host physiology. Further testing and adaptation of the method are needed to determine the value of REIMS as a diagnostic tool. In the following chapter, another non-invasive combination of [sample + technique] will be tested for the diagnosis *P. aeruginosa* infection.

Chapter 7

Analysis of skin secretions using DESI- MS: a feasibility study

7. Analysis of skin secretions using DESI-MS: a feasibility study

Following up on Chapter 6, and the attempt to use REIMS to diagnose *P. aeruginosa* infection from sputum and urine, this study aimed to explore another non-invasive, culture-free alternative approach. *P. aeruginosa* infection was hypothesised to have an impact on all body fluids, thus skin secretions were considered as a possibly informative, easy-to-collect sample. DESI-MS is well suited for the analysis of lipid-rich surfaces and has been applied before on fingermarks [265]. Here, I intended to develop a protocol for the collection and the analysis of skin secretions using DESI-MS and to assess the feasibility of this method for the diagnosis of *P. aeruginosa* infection in CF patients using pilot data (Chapter 2, objective 4).

7.1. Introduction

7.1.1. Skin secretions, lipids and analysis

Compounds on the skin surface are either produced and excreted by the skin itself, or external substances that might have come in contact with the skin, leading to a mix of endogenous and exogenous species. Endogenous compounds may be inorganic, such as metal ions, or organic, such as lipids, proteins, vitamins, or amino acids. Exogenous compounds refer to external substances. They may have been deposited on skin, or ingested and subsequently excreted, metabolised or not, such as drugs or food preservatives. Water-soluble species are usually released in sweat. Most lipids found on skin are synthesised by the skin to constitute a protective barrier against environmental microbial aggressions as part of the innate immune system [266]. Long, straight-chain, saturated fatty acids, mostly C20 to C28, constitute the principal lipids on the skin surface together with ceramides and cholesterol [266, 267]. The skin surface differs from other human tissues in that it also contains polyterpenoid squalenes [267].

Several ambient ionisation techniques are now safe to be performed directly on skin. DESI-MS studies on skin have previously been conducted and have demonstrated potential for the analysis of drugs or excreted fatty acids [144, 268]. SESI-MS has been used to characterise skin vapours. Fatty acids and small metabolites were observed [173] and showed that different skin areas of the body produce distinctive patterns [269]. Native and environmental information about the donors can thus be

retrieved from the chemical composition of skin secretions, e.g. fingerprints, which is notably of great interest to forensic scientists.

7.1.2. Fingerprints

The composition of fingerprints has been widely investigated and reviewed using numerous analytical techniques, and free fatty acids have been identified by many as the most abundant lipids [270]. Mass spectrometry, in particular, is a powerful tool with great potential to study the chemical composition of fingerprints, especially since the development of MSI, by MALDI-MS or DESI-MS, allows joint chemical analysis and ridge pattern visualisation [137]. For instance, MALDI-MSI has been used to investigate the spatial chemical composition of fingerprints and localisation of sweat pores has been achieved. Sex was predicted with a mitigated accuracy (85%) by comparing protein and peptide profiles [271]. Caffeine, fatty acid and amino acid detection has also been reported [137, 271]. DESI-MS has been used to detect spiked exogenous drugs or explosives in fingerprints in addition to endogenous lipids [158].

When analysing fingerprints, aging of the samples is to be considered, as this can result in dehydration and oxidation of lipids. The fingerprint composition is also influenced by the deposition conditions, the substrate nature and the donor's characteristics, such as age, ethnic origin, medication or diet, as well as use of or contact with exogenous substances [270]. Fingerprint composition reflects a donor's condition and has huge potential for disease diagnosis [272], particularly in babies and infants who cannot talk but can easily provide them.

7.1.3. Objectives

A method for the collection and DESI-MS analysis of skin secretion samples and subsequent data processing was developed using a trial and error approach. The impact of certain collection parameters was assessed, including the sampling device, sampling site and storage. Measures were implemented to overcome a number of obstacles, such as sampling reproducibility and environmental contamination. The method was evaluated by testing its performance for donor classification. Furthermore, clinical samples were collected from CF patients, with and without *P. aeruginosa* infection, for an early pilot study. The data acquired with DESI-MS was interrogated using multivariate

statistical tools to determine whether it may provide valuable clinical information, notably for the detection of *P. aeruginosa* infections.

7.2. Method development

7.2.1. Sample collection for DESI-MS analysis and processing

Initially, fingerprints were collected by gently pressing the donor's finger onto microscope glass slides as a support material. Slides were analysed as described in Chapter 3, section 3.5.1. *DESI-MS profiling*. During method development, data were recorded in positive and negative ionisation modes within the range $m/z = 150$ -2000. However, the fingerprints were not always clearly visible on the slides, and one of the first difficulties was to locate the sample on the support material in order to position the DESI-MS set-up accordingly. To overcome this, lines were drawn on the support using a black marker (Sharpie permanent black marker, Newell Rubbermaid, Atlanta, GA, USA) to demarcate the sample. Lines were drawn either before collection, separated by 2 cm so that the donor would apply his/her finger in between, or after deposition, on both sides of the mark (see Figure 7.1a).

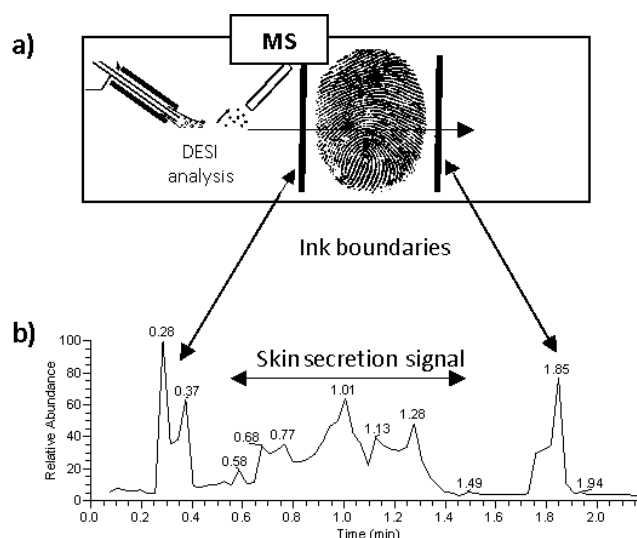


Figure 7.1. DESI-MS analysis of a skin secretion sample.

(a) Schematic representation of the DESI sprayer and the MS inlet progressing through a fingerprint deposited on a microscope glass slide flanked by two ink lines for profiling analysis. (b) Example of a chromatogram (i.e. TIC signal) obtained with DESI-MS profiling of the fingerprint in (a). The highest peaks on the left and right-hand sides show the signal of the ink, whereas the signal in the middle corresponds to the detection of the secretions deposited in between.

The ink is characterised by specific ions (namely $m/z = 372.2429$ in positive ionisation mode; $m/z = 311.1687$; 635.4094 ; 649.1251 in negative ionisation mode) that produce an intense signal during analysis (see TIC signal in Figure 7.1b). Thereby, the lines also served to delimit the spectral

zone of interest for data processing. For that purpose, the in-house program used to handle the DESI-MS data was amended to detect the specific ink ions (m/z tolerance set to 10 ppm); any signal outside of these boundaries was considered background and subtracted, together with the ink peaks, from the fingerprint zone in between (see the signal breakdown in Figure 7.2).

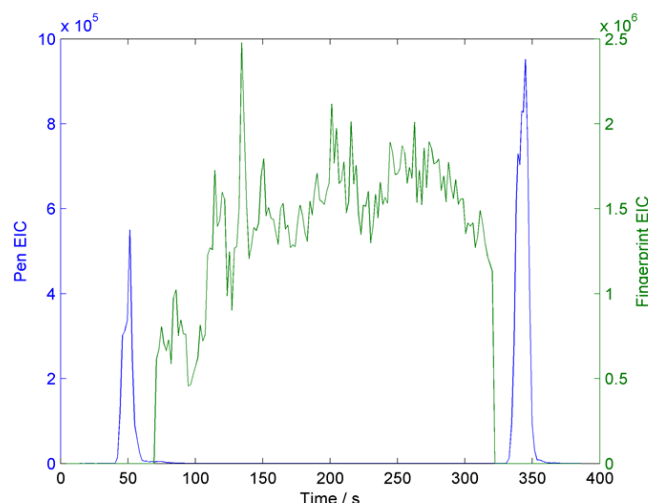


Figure 7.2. Example of the pre-processing of skin secretion data acquired by DESI-MS profiling. Shows the extracted ion count (EIC) for the ink signal (in blue) and for the fingerprint signal (in green) against time (corresponding to spatial position on the sample; signal intensity in arbitrary units).

7.2.2. Choice of the collection material

Various supports were tested to collect the fingerprints:

- Superfrost Plus glass slides (Fisher Scientific Ltd, Loughborough, UK),
- Polysine glass slides (Fisher Scientific Ltd, Loughborough, UK),
- Frosted glass slides (X144 full frosted slide, Fisher Scientific Ltd, Loughborough, UK),
- Plexiglass, laser-cut pieces (polymethylmethacrylate sheet, Sigma-Aldrich, St Louis, MO, USA),
- Polyethylene, hand-cut pieces (white container lids, low density polyethylene, Fisher Scientific Ltd, Loughborough, UK),
- Collection cards (903 protein saver cards, Whatman, Kent, UK).

Other materials were considered, such as double-sided adhesive tape and TLC plates with cellulose, polyamide 6 or C18-reversed phase on silica gel. These were dismissed early on because of the detection of an intense background signal due to the presence of polymers (data not shown).

Examples of total signal and spectral data recorded from fingerprints deposited on Polysine and Plexiglass are given in Figure 7.3 for both ionisation modes; the results for the six materials tested are available in Appendix 17. These data were acquired on fingerprints collected and analysed on the same

day, from a single volunteer healthy donor, to avoid ageing, instrumental or donor confounding effects. Figure 7.3 shows the TIC recorded on each sample and the mean mass spectra of the fingerprint and of the material background. The ink signal, well visible on all materials and in both ionisation modes, delimits the fingerprint (in between) and the material background (outside). Negative ionisation revealed the presence of low mass lipids, such as fatty acids ($m/z = 200-600$), in the finger secretions as well as in the material background, whereas in positive ionisation, higher mass complex lipids were predominant in the fingerprints and notably absent from the background.

In order to assess the suitability of the various materials for fingerprint collection, signal detected from the fingerprints was compared to that of the materials' background. Peak intensities were summed across the whole m/z range (equivalent to the TIC) for the background and for the fingerprint, and across reduced lipid-specific m/z regions of the fingerprint signal. The maximum summed intensities are summarised in Table 7.1 and Table 7.2 for the six materials in both ionisation modes.

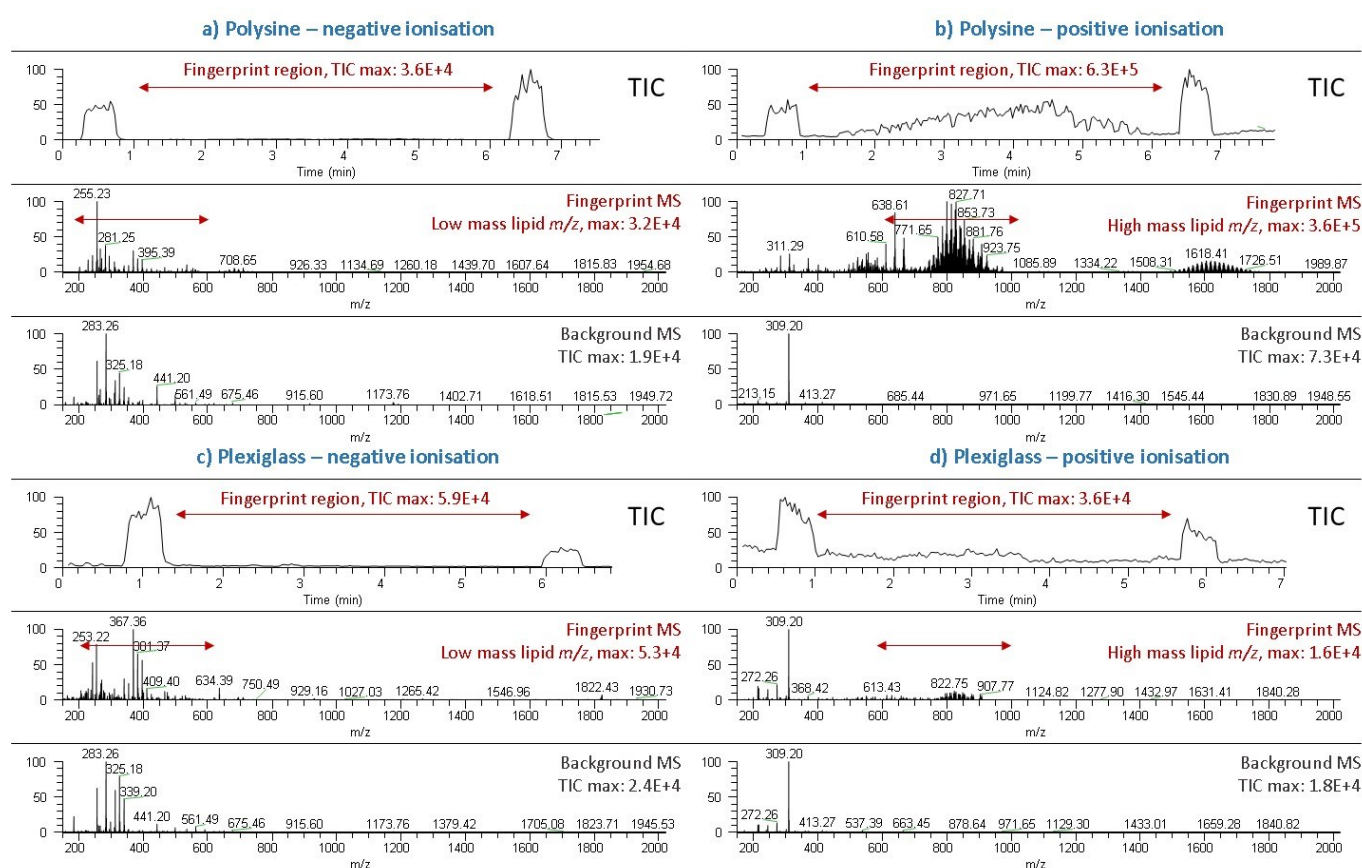


Figure 7.3. Examples of data resulting from DESI-MS profiling of fingerprints deposited on (a & b) Polysine glass and (c & d) Plexiglass.

Analysis in negative (a & c) and positive (b & d) ionisation modes, showing for each: TIC with the maximum intensity (absolute unit) detected on the fingerprint area; averaged and background-subtracted mass spectrum obtained on the fingerprint area with the maximum summed intensity detected within $m/z = 200-600$ in negative ionisation mode, and within $m/z = 600-1000$ in positive ionisation mode; the averaged mass spectrum of the material's background with the maximum detected TIC.

In parallel, samples were collected repeatedly from several volunteers using the different materials (see total numbers in Table 7.1 and Table 7.2). The objective was to build one model per material and to challenge it for donor classification in order to evaluate the sampling reproducibility and sample homogeneity provided by a given material. A subset of five (or 10) spectra were randomly selected between the ink boundaries and used as analytical replicates. Processing of the data is described in Chapter 3, section 3.5.2. *Data analysis* and included mean-fold change normalisation, log-transform, PCA and supervised PCA-LDA. The accuracy of each model for donor prediction is reported in Table 7.1 and Table 7.2

Table 7.1. Summary of the intensities recorded by DESI-MS in positive ionisation mode on fingerprints deposited on six materials.

Maximum intensity of the TIC on the fingerprint (Fgp), maximum summed intensity on the fingerprint in the high mass lipid range (% of fingerprint TIC), maximum TIC on the raw material (background, % of fingerprint TIC). Outstanding values are highlighted in green or red. Total numbers of samples collected on each material and associated classification accuracy for donor prediction using a subset of five spectra per sample, PCA-LDA model, 10-fold cross-validation and 1-nn classifier.

Positive ionisation					
Material	Fingerprint max TIC (Fgp)	High mass lipid region max signal (% Fgp)	Background max signal (% Fgp)	Number of samples	Classification accuracy
Superfrost glass	6.3E+5	3.7E+5 (59%)	6.1E+4 (10%)	16	78%
Polysine glass	6.3E+5	3.6E+5 (57%)	7.3E+4 (12%)	12	79%
Frosted glass	1.7E+5	8.1E+4 (48%)	6.2E+4 (36%)	11	80%
Plexiglass	3.6E+4	1.6E+4 (45%)	1.8E+4 (50%)	8	73%
Polyethylene	3.2E+5	1.8E+5 (58%)	8.7E+4 (27%)	9	82%
Collection card	2.0E+4	1.5E+3 (7%)	1.6E+4 (80%)	7	71%

Table 7.2. Summary of the intensities recorded by DESI-MS in negative ionisation mode on fingerprints deposited on six materials.

Maximum intensity of the TIC on the fingerprint (Fgp), maximum summed intensity on the fingerprint in the low mass lipid range (% of fingerprint TIC), maximum TIC on the raw material (background, % of fingerprint TIC). Outstanding values are highlighted in green or red. Total numbers of samples collected on each material and associated classification accuracy for donor prediction using a subset of five spectra per sample, PCA-LDA model, 10-fold cross-validation and 1-nn classifier.

Negative ionisation					
Material	Fingerprint max TIC (Fgp)	Low mass lipid region max signal (% Fgp)	Background max signal (% Fgp)	Number of samples	Classification accuracy
Superfrost glass	2.4E+4	2.2E+4 (90%)	1.5E+4 (63%)	24	78%
Polysine glass	3.6E+4	3.2E+4 (88%)	1.9E+4 (53%)	5*	78%
Frosted glass	5.6E+4	5.2E+4 (92%)	7.1E+4 (127%)	16	81%
Plexiglass	5.9E+4	5.3E+4 (89%)	2.4E+4 (41%)	15	77%
Polyethylene	2.5E+5	2.2E+5 (86%)	1.7E+5 (68%)	9	80%
Collection card	1.1E+4	1.0E+4 (88%)	3.0E+3 (27%)	11	83%

* A subset of 10 spectra per sample was used for statistical processing due to the low number of samples.

Overall, most of the materials provided comparable results. In positive ionisation mode, the complex lipids represented around 50% of the TIC of the fingermarks and the background to fingermark ratio varied from 10 to 50%, except for the collection card, which displayed a particularly low signal for the

fingerprint and a high background. Superfrost Plus and Polysine glasses showed the lowest background. Prediction accuracies were around 80%, with lower results for Plexiglass and the collection card, likely due to a lower fingerprint signal. In negative ionisation mode, the lower mass range including the fatty acids, made up around 90% of the fingerprint signal. Frosted glass displayed a particularly high background, and the signal on the card was again very low compared to the other materials. Similarly to the positive mode, overall prediction accuracies were around 80%.

These observations suggested that frosted glass and the tested collection card were not appropriate materials for sampling fingerprints, due to a very high background and/or a low signal for the fingerprint. The four other materials gave equivalent results. The Superfrost Plus and Polysine glass gave a notably lower background in positive ionisation mode, and since microscope glass slides constituted the easiest sampling device, they were retained to collect the fingerprints.

7.2.3. Choice of the sampling site

Although first tests showed that the metabolites secreted by the fingers of a single donor were highly similar (data not shown), one of the main obstacles was the poor reproducibility of the sampling. Hands and fingers are particularly exposed to environmental contamination, e.g. plastic, food, soap or moisturising creams. To eliminate these sources of interference, the skin was cleaned using pure IPA (Chromasolv for HPLC, 99.9%, Scientific Laboratory Supplies Ltd, Nottingham, UK) and left to secrete for a few minutes before imprinting. However, whether cleaned with IPA or not, secretions from the skin of the fingers were very variable, even for one subject throughout the same day, and in some cases, secretions were so faint they barely generated signal with DESI-MS. Wolstenholme et al. previously encountered this issue whilst imaging latent fingerprints and managed to enhance the print by rubbing fingers on the sebum-rich forehead, nose and chin, thereby supplying “groomed” fingerprints [273]. Facial skin secretes abundant amounts of lipids due to a higher concentration of sebaceous glands. The lipidome of the forehead surface has previously been analysed, and its composition was observed to vary between individuals while remaining constant in the same individual over time [274]. Therefore, the face was considered as an alternative sampling site and the bridge of the nose was chosen as it presented easy access for sampling with a microscope glass slide. First assays showed that cosmetic products may generate a significant signal in DESI-MS and that this was readily removed by cleaning the skin with IPA beforehand, which also presented the advantage of eliminating potential contaminating microbes (e.g. when blowing one’s nose). Contrary to fingers, secretion from the face skin was quick enough to allow sampling after a few minutes. Samples were

collected after IPA cleaning from fingers and from the bridge of the nose using Superfrost Plus microscope glass slides and analysed in both ionisation modes for comparison. Examples of spectra obtained from both sites of the same donor are shown in Figure 7.4.

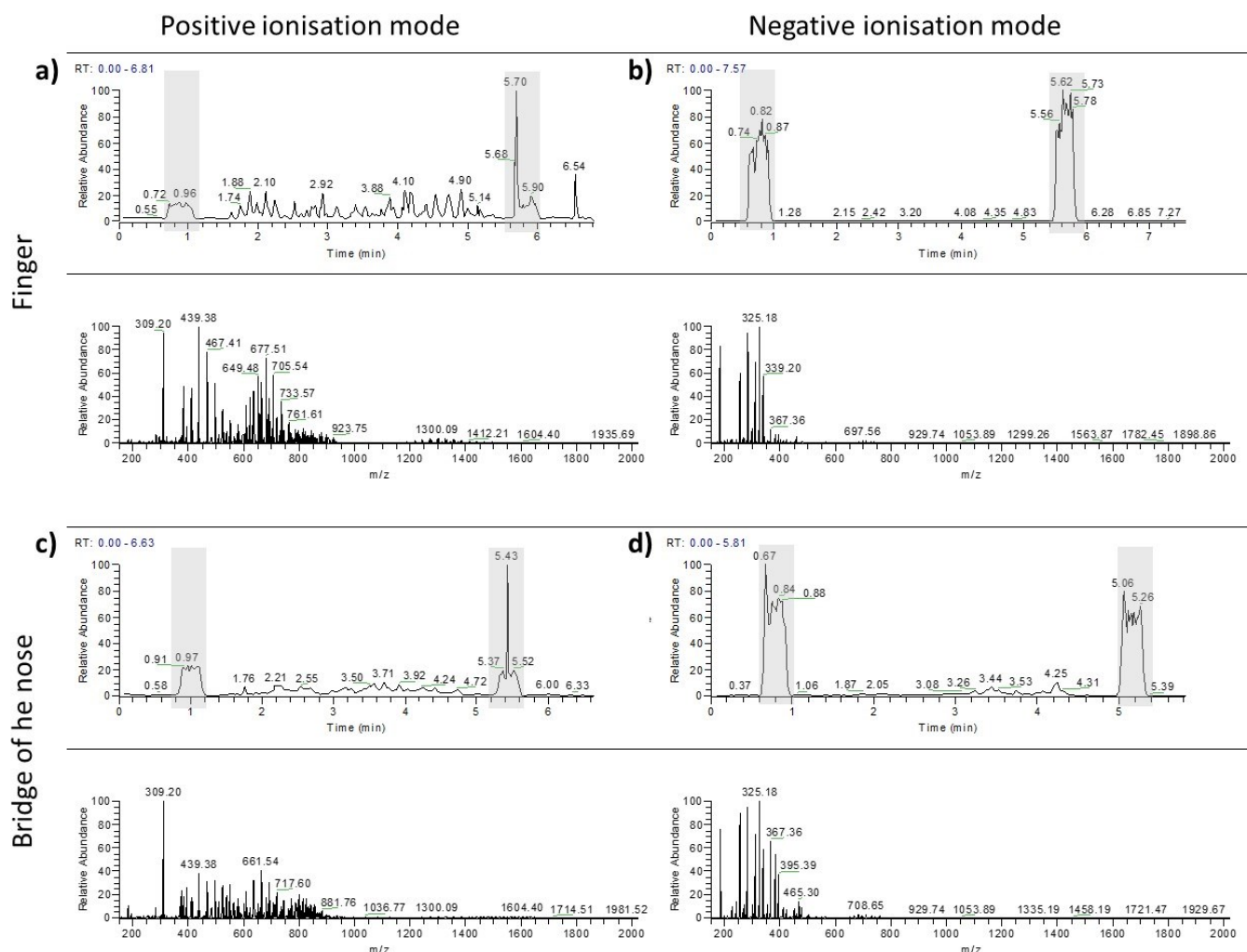


Figure 7.4. DESI-MS profiling analysis of skin secretions collected from the finger (a & b) and the bridge of the nose (c & d).

Samples collected from a single donor on Superfrost microscope glass slides; data acquired in positive (a & c) and negative (b & d) ionisation modes. Shows TIC and background-subtracted mean mass spectrum. The TIC signal of the delimitating ink is highlighted in grey.

Figure 7.4 shows that both sites generated signal in similar m/z ranges in the respective ionisation modes, although intensity distribution differed throughout the profiles. In order to assess the extent of this profile divergence, repeated samples from the same donor were collected and analysed within one day: 12 finger samples and seven nose samples in positive ionisation mode; six finger samples and six nose samples in negative ionisation mode. A subset of five spectra per sample was used to perform PCA (Figure 7.5) and t -tests (FDR = 0.05) (Figure 7.6) after median-fold change normalisation and log-

transform. Given the very high number of discriminant features, q -values were calculated to reduce the number of false-positive.

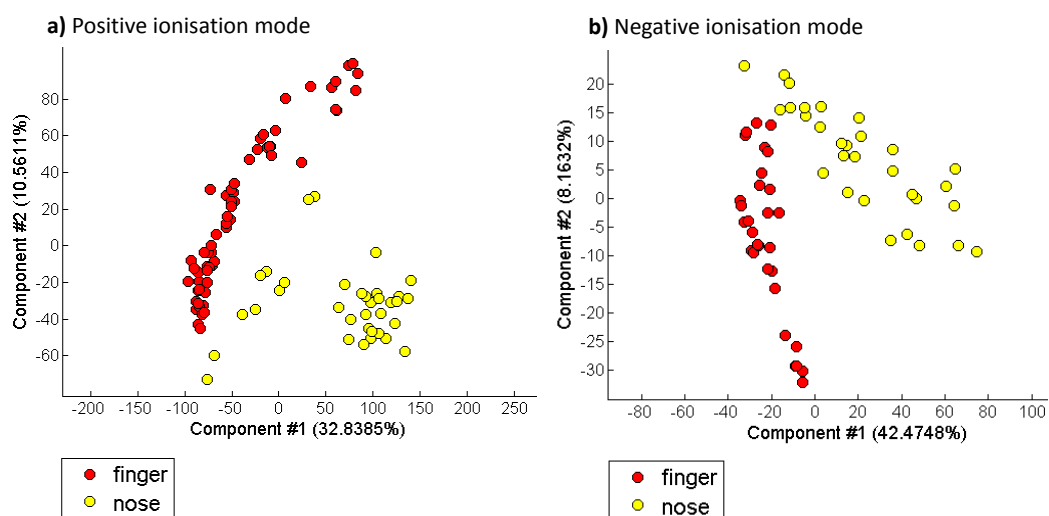


Figure 7.5. PCA score plots using DESI-MS profiling data acquired on skin secretions collected from the finger and from the nose.

Data acquired in positive (a) and negative (b) ionisation modes. Shows the two first principal components; uses a subset of five mass spectra for each sample (built on Matlab).

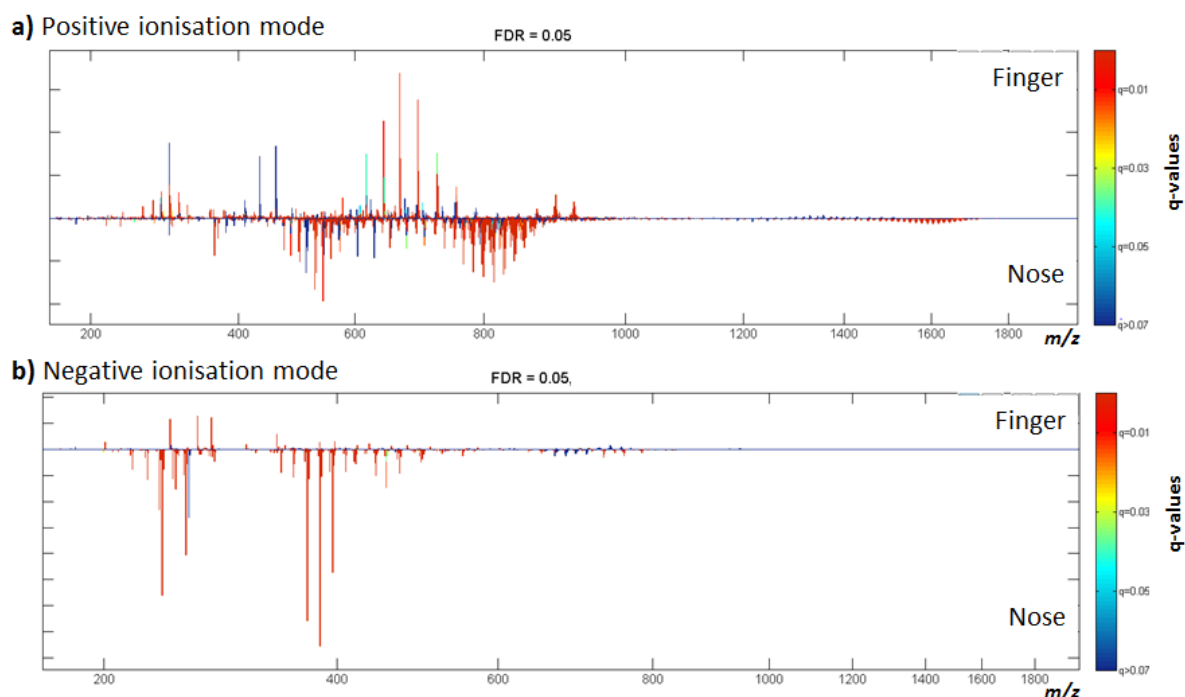


Figure 7.6. t -tests showing differentiating features between finger and nose skin secretions.

Represented as mass spectra in positive (a) and negative (b) ionisation modes, with features pointing up or down for characteristic peaks of finger or nose skin secretions, respectively; features are coloured on a scale from blue to red as the q -value decreases and significance increases (FDR = 0.05) (built on Matlab).

In both ionisation modes, PCA showed a separation according to the two first PCs and the t -tests highlighted the great number of features discriminating between finger and nose. This suggests that

the secreted metabolites differ depending on the skin area. Although the experiment was inconclusive regarding which site presented the best reproducibility (no clear clusters on PCAs in Figure 7.5), The bridge of the nose provided an easy access to more abundant secretions and was thus selected as a sampling site.

7.2.4. Sample collection – Impact of IPA cleaning

As mentioned in the previous section, environmental contamination can have a significant impact on data quality in DESI-MS analysis of skin secretions. One example is given in Figure 7.7a, showing the detection of a polymer in a skin secretion sample. It was found IPA could be used to clean the skin before sampling to eliminate these interferences (Figure 7.7b).

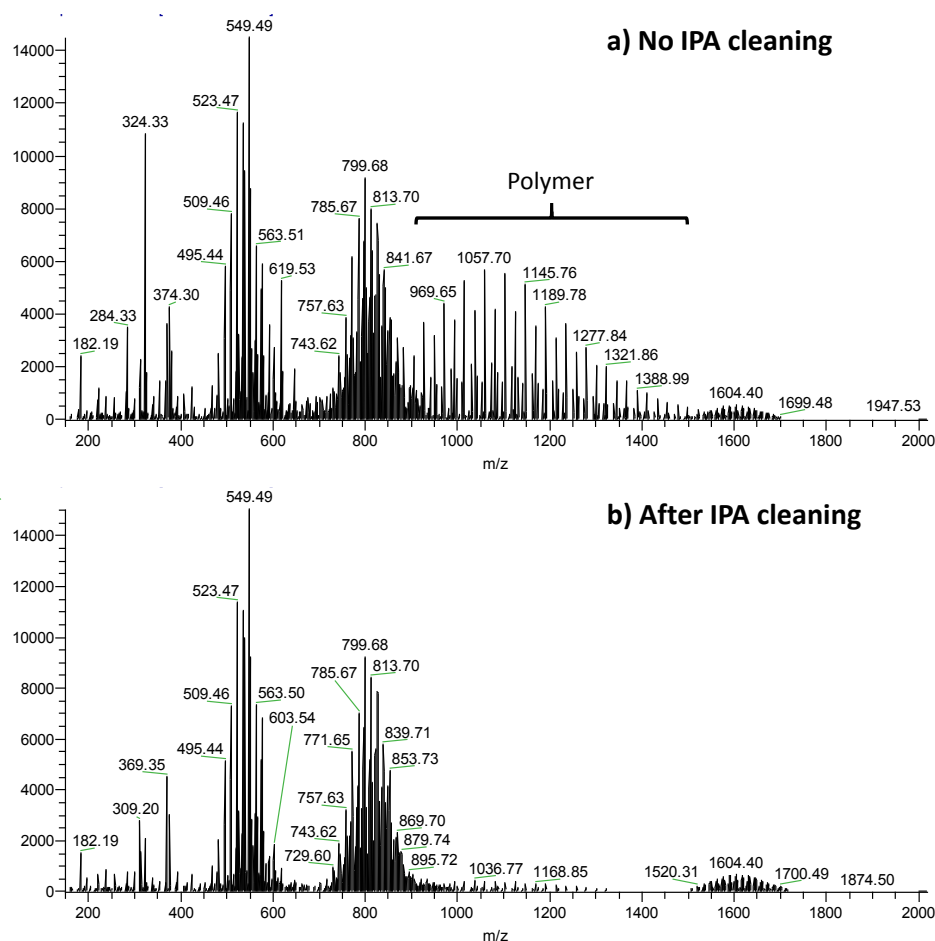


Figure 7.7. Examples of mass spectra acquired by DESI-MS profiling on skin secretions before cleaning the skin (a) and after IPA cleaning (b).

Samples were collected from the bridge of the nose of one donor; spectra were acquired in positive ionisation mode, averaged and background-subtracted. The first spectrum (a) reveals the presence of a polymer within $m/z = 700-1500$.

In order to determine whether IPA cleaning had a systematic impact on the detected profile, samples collected without IPA cleaning (22 samples) were compared to samples collected after IPA cleaning (20 samples). Skin secretions were sampled from the bridge of the nose from a total of eight voluntary donors. PCA and *t*-tests (FDR = 0.05, *p*-values) were performed to detect differentiating trends between the median spectra, after median-fold change normalisation and log-transform (Figure 7.8 and Figure 7.9).

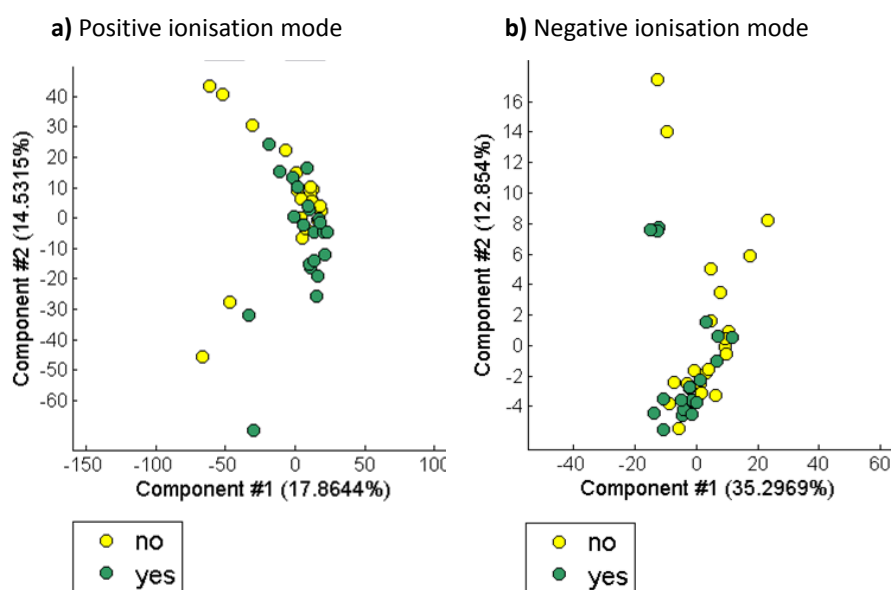


Figure 7.8. PCA score plots using DESI-MS profiling data acquired on secretions collected without cleaning the skin and after IPA cleaning.

Data acquired in positive (a) and negative (b) ionisation modes; samples collected from the bridge of the nose without cleaning the skin (class “no”) and after IPA cleaning (class “yes”). Shows the two first principal components; uses median spectra for each sample (built on Matlab).

The PCAs showed no separation between the two groups of samples. However, the *t*-tests suggested some features tended to be decreased after IPA cleaning, notably ranges $m/z = 800-900$ in positive ionisation and $m/z = 200-500$ in negative ionisation mode. This was attributed to the removal of sebum which is known to be secreted slowly [275].

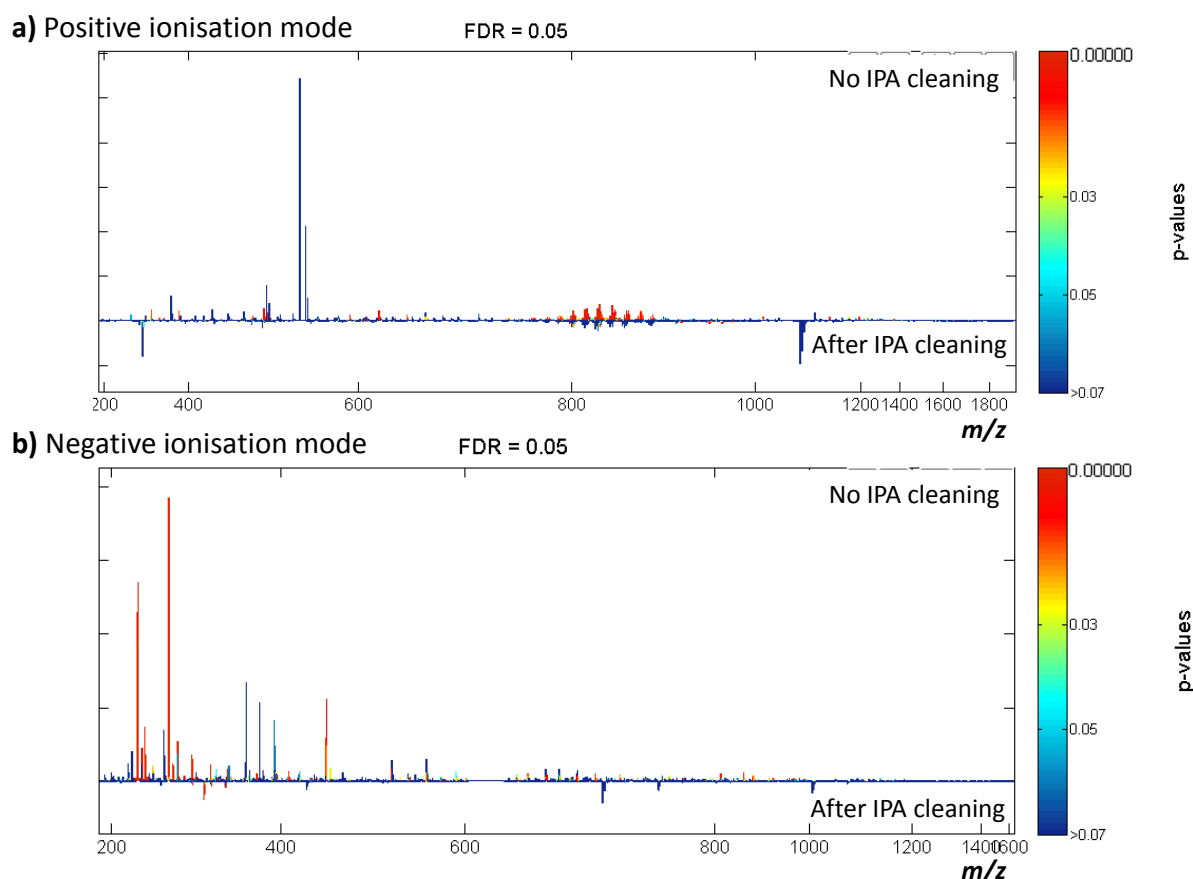


Figure 7.9. t-tests showing differentiating features between secretions collected without cleaning the skin and after IPA cleaning.

Represented as mass spectra in positive (a) and negative (b) ionisation modes, with features pointing up or down when characteristic peaks of secretions collected from skin without or after IPA cleaning, respectively; features are coloured on a scale from blue to red as the p-value decreases and significance increases (FDR = 0.05) (built on Matlab).

IPA cleaning was efficient in eliminating environmental contamination that could be detrimental to sampling reproducibility and data interpretation. Its impact on detection of the secreted profile was considered to be negligible. Additionally, IPA was thought to help eliminate potential microbial contamination from the skin microflora or from the environment. Thus, a preliminary cleaning step was systematically included in the collection protocol, allowing the skin to secrete for 5 min before sampling. IPA-containing medical swabs, suitable for clinical use, were tested. These were carefully selected to contain only pure diluted IPA (70%) since additional antiseptic compounds may engender a high signal during DESI-MS analysis (see spectra acquired when using such swabs containing chlorhexidine gluconate in Appendix 18).

7.2.5. Sample storage

Since ageing was known to alter the metabolic composition of fingerprints [270], samples needed to be stored under protective conditions before being subjected to analysis. Freezing was hypothesised to best conserve the metabolic profile of skin secretions.

7.2.5.1. Impact of freezing

In order to detect a potential systematic effect of the freezing, the metabolic profile detected on fresh samples was compared to that of samples that had been stored in a -80°C freezer for one to 122 days. Secretions were collected from four donors, from the skin on the bridge of the nose after IPA cleaning. PCA and t -tests (FDR = 0.05, p -values) were performed to detect differentiating trends between the median spectra after median-fold change normalisation and log-transform (Figure 7.10 and Figure 7.11).

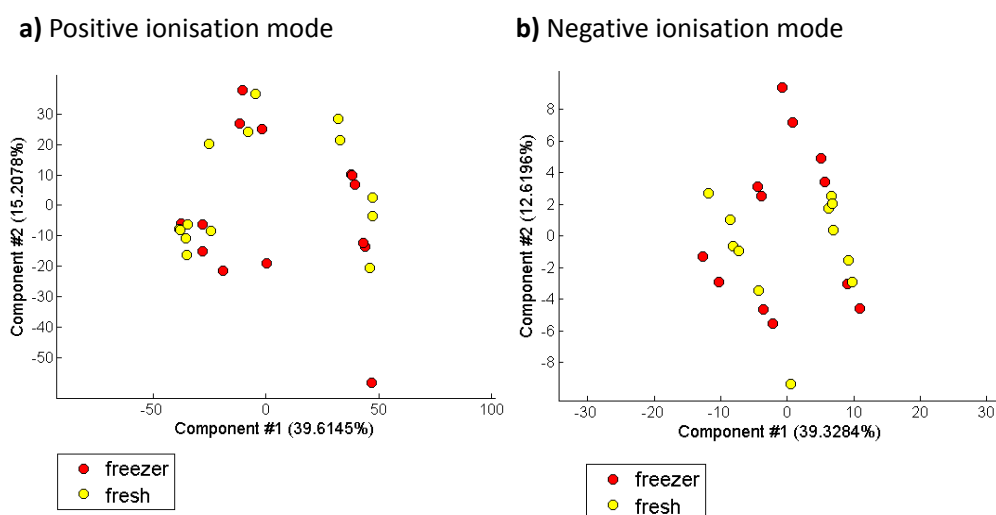


Figure 7.10. PCA score plots using DESI-MS profiling data acquired on fresh and frozen skin secretion samples. Data acquired in positive (a) and negative (b) ionisation modes; samples collected from the cleaned skin of the bridge of the nose and analysed fresh, or after storage in a -80°C freezer. Shows the two first principal components; uses median spectra for each sample (built on Matlab).

The PCAs showed no separating trends between the two classes. The t -tests revealed very few discriminant features; most of the features with significant p -values seemed to decrease after freezing (e.g. some complex lipids detected in positive ionisation).

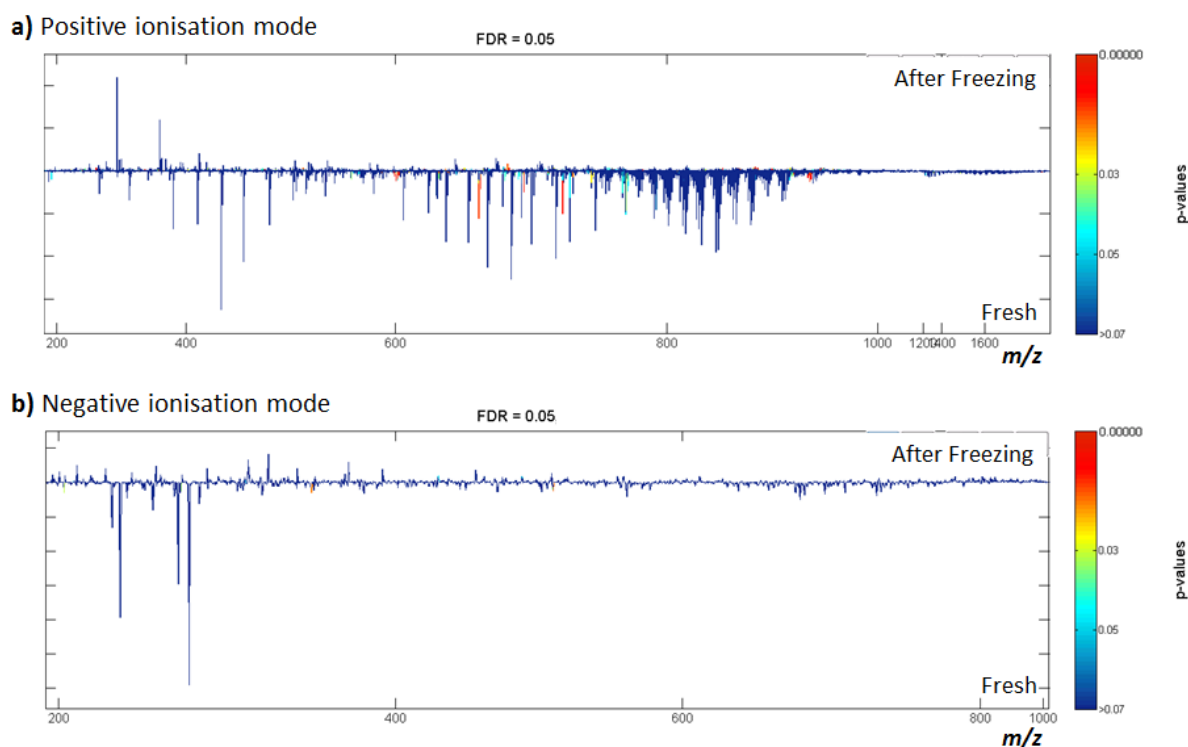


Figure 7.11. t-tests showing differentiating features between fresh samples and after freezing. Represented as mass spectra in positive (a) and negative (b) ionisation modes, with features pointing up or down for characteristic peaks of secretions analysed after storage in a -80°C freezer or fresh, respectively; features are coloured on a scale from blue to red as the p-value decreases and significance increases (FDR = 0.05) (built on Matlab).

7.2.5.2. Impact of freeze/thaw cycles

The impact of the thawing process was further assessed, using fingerprint samples formerly collected from four donors. The four samples were frozen and thawed 10 times, and analysed after one, two, five and 10 freeze/thaw cycles. PCA and ANOVA (FDR = 0.05; q -values were calculated to reduce the number of significant features) were performed on the median spectra, after median-fold change normalisation and log-transform (Figure 7.12 and Figure 7.13).

The analyses suggested that multiple freeze/thaw cycles impacted the skin secretions samples and their detected profile. The PCA showed a separation which is particularly clear in positive ionisation mode for the samples subjected to 10 cycles (Figure 7.12a). The ANOVA further confirmed that the positive profile was affected throughout the entire m/z range, and in particular in the higher m/z range (Figure 7.13a).

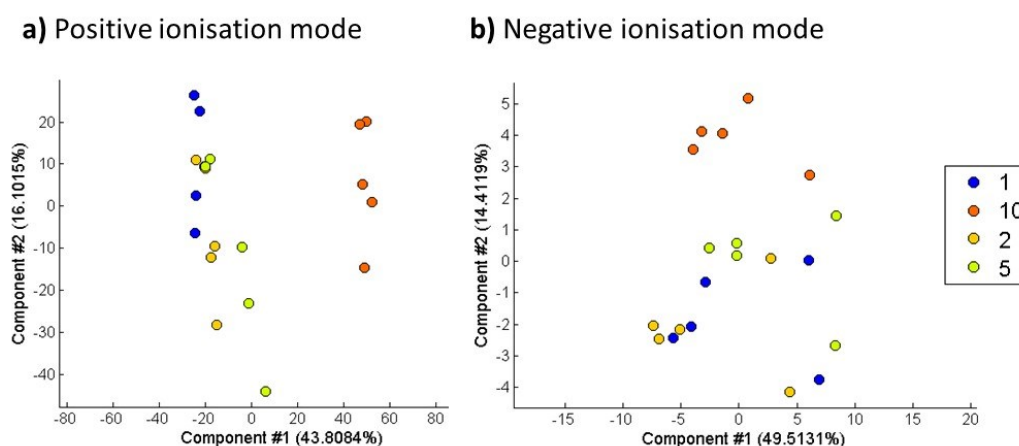


Figure 7.12. PCA score plots using DESI-MS profiling data acquired on skin secretion samples after several freeze/thaw cycles.

Data acquired in positive (a) and negative (b) ionisation modes, after one, two, five and 10 freeze/thaw cycles, on fingermarks without IPA cleaning. Shows the two first principal components; uses median spectra for each sample (built on Matlab).

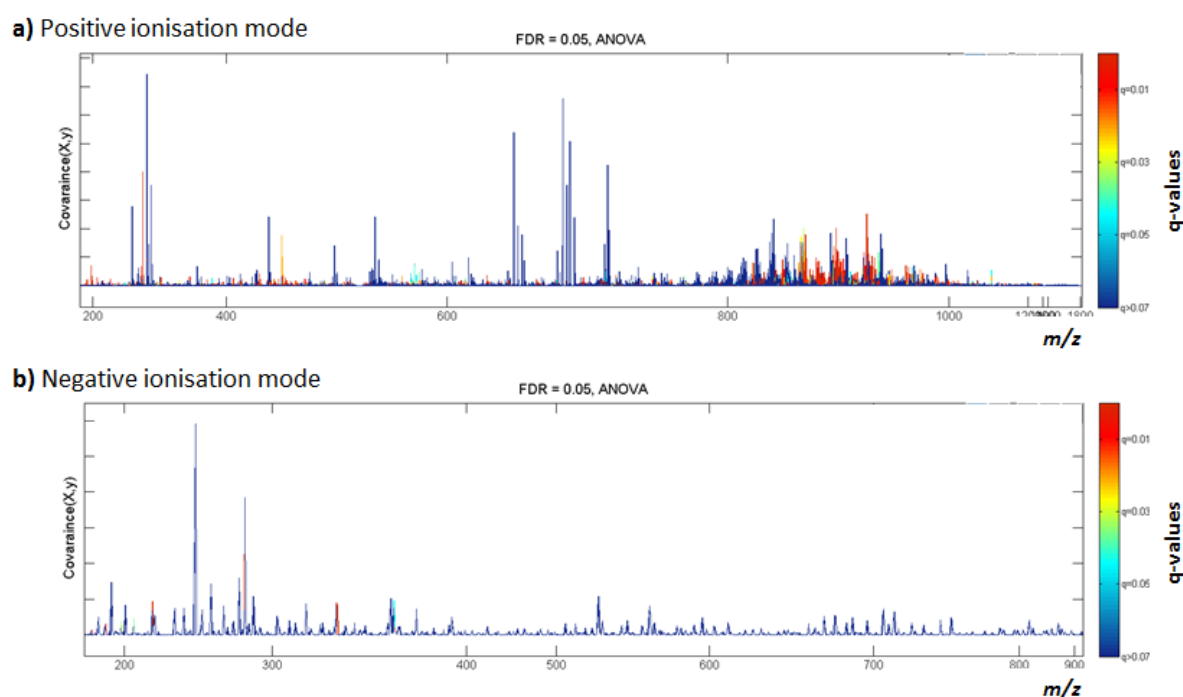


Figure 7.13. ANOVA showing differentiating features depending on the number of freeze/thaw cycles.

Represented as mass spectra in positive (a) and negative (b) ionisation modes; features discriminating between samples subjected to a varying number of freeze/thaw cycles are coloured on a scale from blue to red as q-value decreases and significance increases (FDR = 0.05) (built on Matlab).

Overall, one-off freezing does not appear detrimental to the DESI-MS profiling analysis. However, repeated thawing may significantly alter the metabolic profile, notably in positive ionisation mode. Samples were thus systematically frozen after collection, and freeze/thaw cycles kept to a minimum.

7.3. Evaluation of the method

7.3.1. Selection of the sampling parameters

Based on the observations described above, the following parameters were set for the rest of the study in order to minimise possible confounding effects:

- Skin was cleaned with IPA before sampling;
- Superfrost Plus microscope glass slides were used for sampling;
- Ink lines were drawn to delimit the sample deposition and process the data;
- Secretions were collected from the skin on the bridge of the nose;
- Slides were stored at -80°C until analysis.

The samples were analysed and the data processed as described in Chapter 3, sections 3.5.1. *DESI-MS profiling* and 3.5.2. *Data analysis*.

7.3.2. Metabolic profile of nose skin secretions

Typical mass spectra obtained using the above conditions are presented in Figure 7.14; peaks were tentatively identified by exact mass comparison, chemical families are indicated on the spectra.

In accordance with the literature, ceramides, cholesterol and above all fatty acids were detected in the negative ionisation mode [266, 267]. Glycerides (monoglyceride (MG), diglyceride (DG) and triglyceride (TG)) were largely predominant in positive ionisation mode and numerous fatty acids esters were detected.

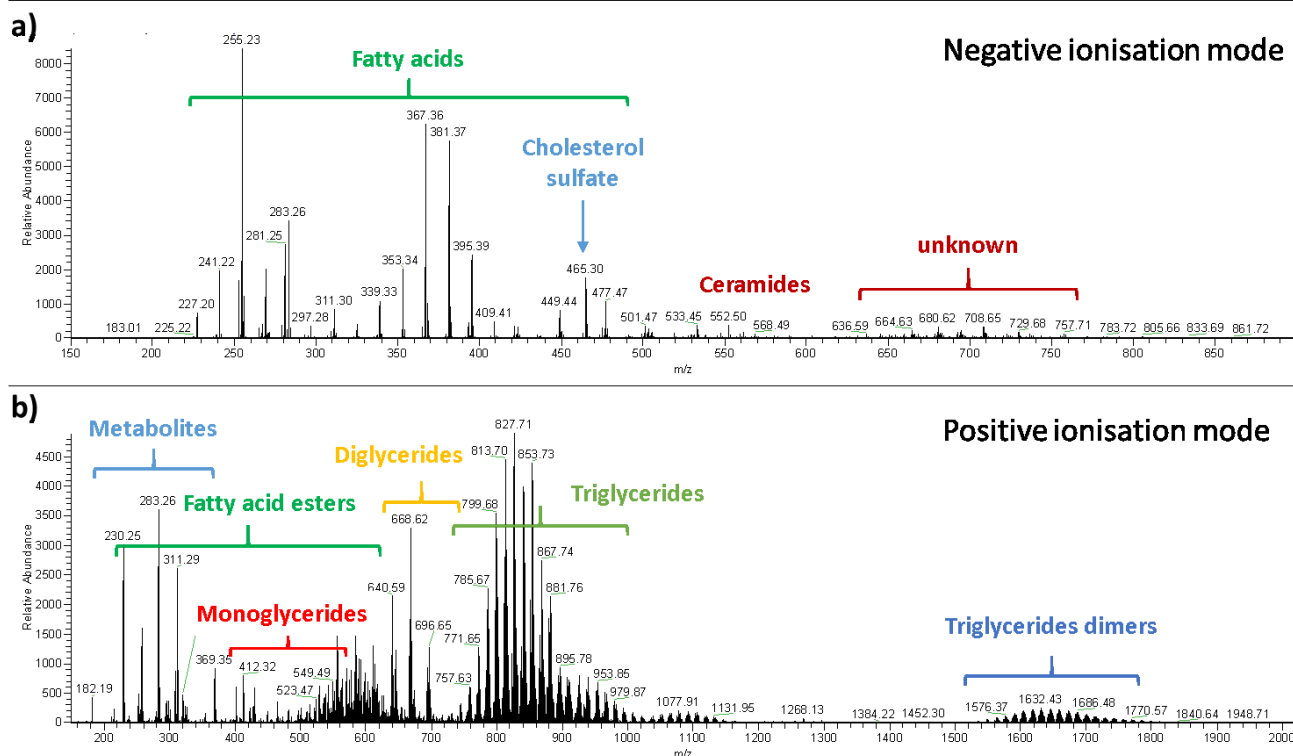


Figure 7.14. Examples of typical mass spectra acquired in (a) negative and (b) positive ionisation modes, from fresh skin secretions collected on Superfrost Plus glass slides from the bridge of the nose after cleaning the skin with IPA.

Spectra were averaged throughout the sample signal area and background-subtracted; m/z regions with tentatively identified spectral features are indicated.

7.3.3. Evaluation of reproducibility through classification of donors

A total of 14 healthy subjects provided between six and 90 samples; collection and analysis were carried out over a time period of several months. The metabolic profiles detected in both ionisation modes by DESI-MS profiling were averaged per sample and examined using unsupervised (PCA) and supervised (RMMC) analyses. Figure 7.15 shows the distribution of the 14 donors according to the profiling data recorded in positive ionisation mode, after median-fold change normalisation and log-transform.

The first two PCs (Figure 7.15a) did not show any separation between the subjects but groupings per donors were made visible when looking at the following PCs (Figure 7.15b) and enhanced in the RMMC plots (Figure 7.15c). Negative ionisation profiling led to a similar distribution (Appendix 19).

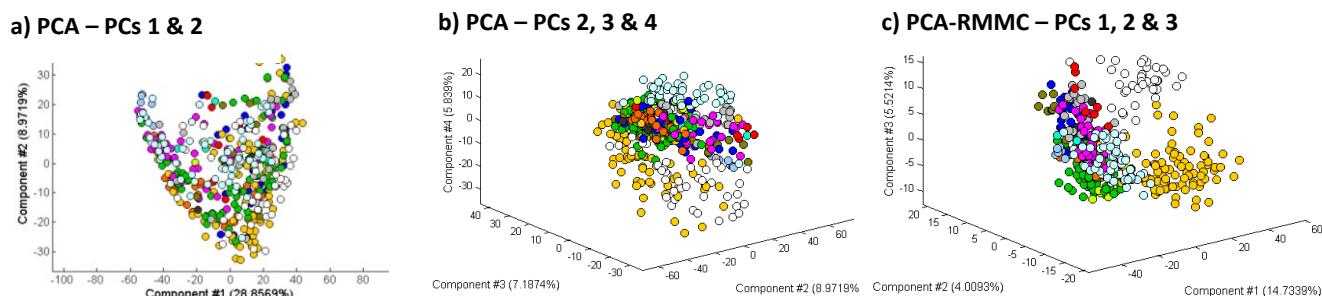


Figure 7.15. Score plots showing unsupervised and supervised analysis of DESI-MS profiling data acquired in positive ionisation mode on skin secretions collected from 14 subjects.

(a) Two-dimensional PCA with the two first PCs 1 & 2; (b) Three-dimensional PCA with PCs 2, 3 & 4; (c) Supervised PCA-RMMC with the three first PCs; uses median spectra for each sample; colouring corresponds to the donors' identity (built on Matlab).

The RMMC model was tested for subject classification; its degree of accuracy was an indicator of the reproducibility of the method for sampling, analysis and data processing. The cross-validation obtained with the positive ionisation model is shown in Figure 7.16, with the frequency of prediction for each subject; overall prediction was 90% accurate. Unsurprisingly, the lowest classification accuracies were obtained for the subjects with smallest numbers of repetitions, which happens when dealing with unbalanced sample sets. Nevertheless, models using reduced data sets in order to homogenise group numbers led to similar overall accuracy. An equivalent data workflow in negative ionisation mode led to 78% accuracy for donor classification (confusion matrix in Appendix 20).

ANOVA implemented in both modes revealed that most of the features contributed significantly to discriminating the donors (FDR = 0.05, q -values; data not shown).

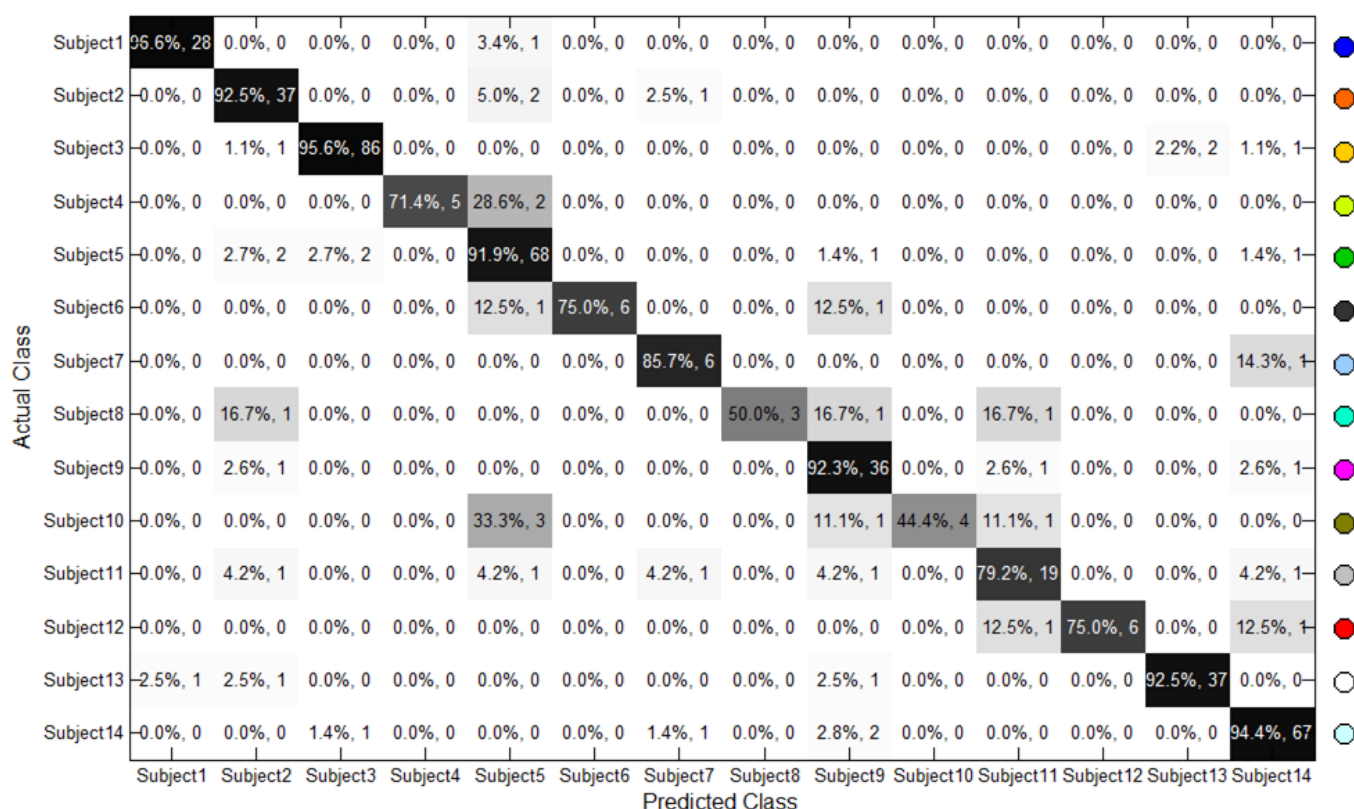


Figure 7.16. Confusion matrix for donor classification using DESI-MS profiling data acquired on skin secretions. Samples analysed in positive ionisation mode; prediction uses an RMMC supervised model, 10-fold cross-validation, 5-nn classifier. Actual and predicted classes are displayed respectively on the left-hand and the bottom sides. Each line refers back to a subject in Figure 7.15 (see dot colours). Frequency of prediction for each subject is represented in darkening grey with the proportion and number of samples indicated in each box. Overall classification was 90% accurate (built on Matlab).

The lipidome detected by DESI-MS profiling from skin secretions was rich and complex. The results obtained for classification indicated that individuals secreted characteristic lipidomes, involving the whole m/z range, and that this method was reproducible enough to differentiate between them. This unique lipid profile may contain useful clinical information about a subject's condition. Thus, its utility for the diagnosis of *P. aeruginosa* infection in CF patients was tested.

7.4. Analysis of skin secretions from CF patients

7.4.1. Collection and analysis of clinical samples

Samples of skin secretions were collected from CF patients attending routine visits at the Royal Brompton Hospital using the optimal conditions determined previously (section 7.3.1. *Selection of the sampling parameters*) with patients' written consent (see Chapter 3, section 3.1. *Health, safety and ethical considerations*). The skin on the bridge of the donors' nose was cleaned using clinical wipes

containing pure diluted IPA (70%, FastAid pre-injection swabs, Robinson Healthcare Ltd, Worksop, UK) and left for 5 min to secrete again. Microscope glass slides (Superfrost Plus, Fisher Scientific Ltd, Loughborough, UK) were used to collect skin secretion samples. Silicone tubing (2 mm ID; 4 mm OD, VWR International Ltd, Lutterworth, UK) was cut lengthwise and adjusted along the edges of the glass slides in order to protect the donor's face during sampling (see an example of a protected glass slide in Figure 7.17). The glass slides were wiped before collection using low lint tissue (VWR International Ltd, Lutterworth, UK) to eliminate any residual deposition or dust. The protected glass slide was then put in contact with the skin of the bridge of the nose of the donor for 2 s. After collection, ink lines were drawn on each side on the deposited sample to allow future analysis and data processing as explained above (section 7.2.1. *Sample collection for DESI-MS analysis and processing*). Samples were kept on dry ice during the time of the clinic visits (< 4 hours) before being transferred to a -80°C freezer. Nitrile gloves were worn throughout the procedure to avoid contamination when manipulating the glass slides.

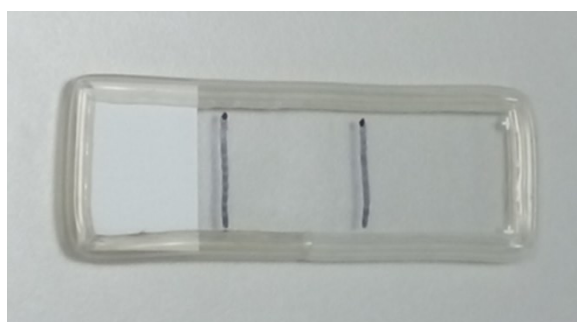


Figure 7.17. Example of a glass slide with edges covered by protective silicon tubing.

Donors included 16 healthy volunteers, 22 patients positive for *P. aeruginosa* infection and 20 non-infected patients (see Table 7.3). Infection status was assessed using sputum samples taken in parallel by the hospital microbiology laboratory (see Chapter 3, section 3.2. *Collection and culturing of microorganisms*).

Table 7.3. Numbers of donors recruited for the pilot clinical study using skin secretions analysed by DESI-MS profiling.

Donors	Healthy subjects	CF patients	
<i>P. aeruginosa</i> infection	N/A	Negative	Positive
Males	6	14	17
Females	12	7	5
Total	18	20	22

DESI-MS profiling was carried out in both ionisation modes within $m/z = 150$ -1000 following the protocols described in this chapter, section 7.2.1. *Sample collection for DESI-MS analysis and processing* and in Chapter 3, section 3.5. *DESI-MS analysis*. Data processing included averaging mass

spectra per sample, TIC-normalisation, log-transform, PCA, RMMC (leave-one-out cross-validation procedure), and *t*-tests (FDR = 0.05, *p*-values).

The method was first evaluated by checking if it would detect differences between categories of subjects that are clearly distinct, such as males and females, healthy donors and CF patients. Then, the data was challenged to predict CF patients with *P. aeruginosa* infection.

7.4.2. Classification of gender

Differentiating features between males and females were sought using *t*-tests. The results obtained in both ionisation modes are shown in Figure 7.18 in the form of mass spectra. A significant number of discriminating features (*p*-values < 0.05), characteristic of males or females, was scattered throughout the mass spectra. They were more numerous in the data detected with positive ionisation – covering various chemical families – than with negative ionisation –mostly detecting fatty acids.

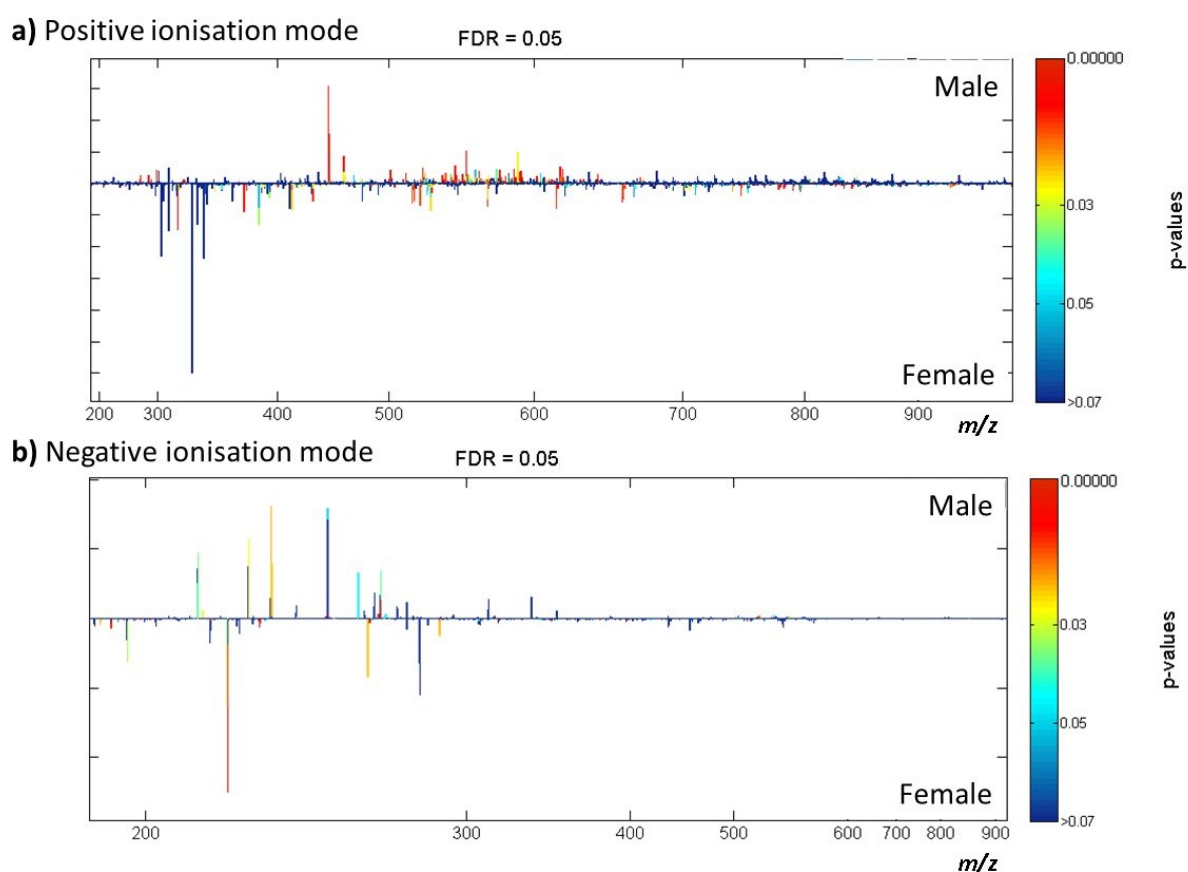


Figure 7.18. *t*-tests showing differentiating features between the skin secretions of males and females. Represented as mass spectra in positive (a) and negative (b) ionisation modes, with features pointing up or down when characteristic peaks of male or female, respectively; features are coloured on a scale from blue to red as *p*-value decreases and significance increases (built on Matlab).

An RMMC analysis was implemented and various classification parameters were applied in an attempt to classify males and females. The models were built using the entire data matrix or a reduced m/z range in an attempt to target the discriminating features, such as in the case of the negative ionisation where most of the information was detected within $m/z = 200-300$ (see Figure 7.18b). The models were tested with various classifiers (k -nn and mahalanobis) and the combinations of parameters that resulted in the best overall accuracy for gender classification are reported in Table 7.4.

Table 7.4. Gender prediction using skin secretion profiles.

Samples analysed by DESI-MS profiling, classified with an RMMC model, leave-one-out cross-validation, using various selections of features and classifier.

Ionisation mode	Positive	Negative	
m/z range	150-1000	150-1000	200-300
Classifier	Mahalanobis	Mahalanobis	Mahalanobis
Overall accuracy	72%	61%	67%

Classification of gender was higher in positive ionisation mode, scoring 72%, vs. 67% in negative ionisation. Data reduction enhanced the classification accuracy in negative ionisation modes, which may be the result of lessening noise signal.

7.4.3. Classification of healthy subjects and CF patients

The metabolic profile of healthy donors and CF patients was compared using t -tests; mass spectra highlighting the discriminant features (p -values < 0.05) are displayed in Figure 7.19. The most striking difference showed in positive ionisation mode, in the $m/z = 600-900$ range. This m/z region covers complex lipids; most of the signal was attributed to triglycerides by exact mass comparison and appeared more intense in the secretions of CF patients. Almost no significant features were detected in negative ionisation mode.

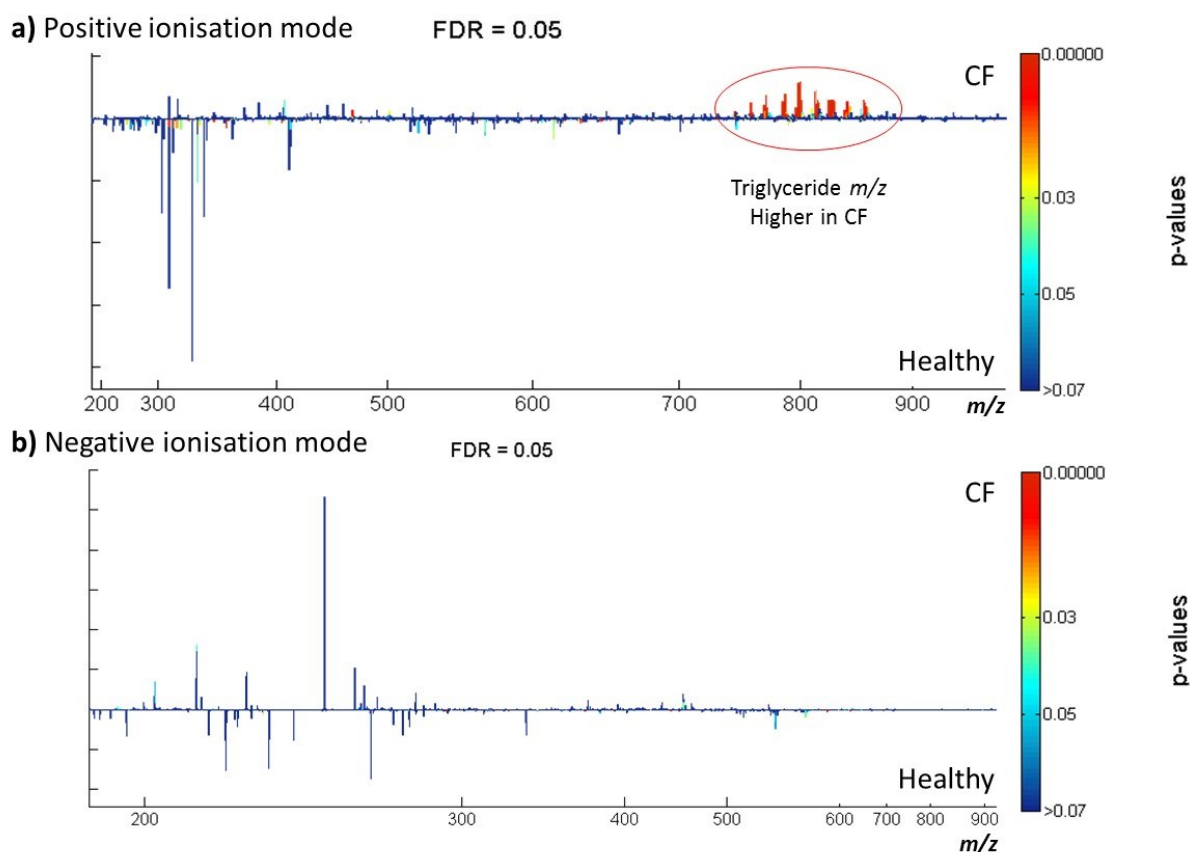


Figure 7.19. t-tests showing differentiating features between the skin secretions of healthy subjects and CF patients.

Represented as mass spectra in positive (a) and negative (b) ionisation modes, with features pointing up or down when characteristic peaks of CF or healthy subjects, respectively; features are coloured on a scale from blue to red as p-value decreases and significance increases (built on Matlab).

Again, various classification models were tested for the separation of healthy subjects from CF patients; parameters that allowed the best classifications are reported in Table 7.5.

Table 7.5. Prediction of the CF condition using skin secretion profiles.

Samples analysed by DESI-MS profiling, classified with an RMMC model, leave-one-out cross-validation, using various selections of features and classifier.

Ionisation mode	Positive		Negative	
m/z range	150-1000	600-900	150-1000	200-300
Classifier	Mahalanobis	Mahalanobis	1-nn	Mahalanobis
Overall accuracy	65%	70%	68%	70%

Although little discriminant information was detected in negative ionisation mode, both ionisation modes provided similar classification accuracy, i.e. 70%. In both cases, reduction of the m/z range slightly enhanced the classification accuracy.

7.4.4. Detection of *P. aeruginosa* infection through the analysis of skin secretions

Finally, *P. aeruginosa* positive and negative CF patients were compared. PCA was carried out as a first approach to visualise the data (Figure 7.20); however, the two first principal components did not reveal any separation between the samples.

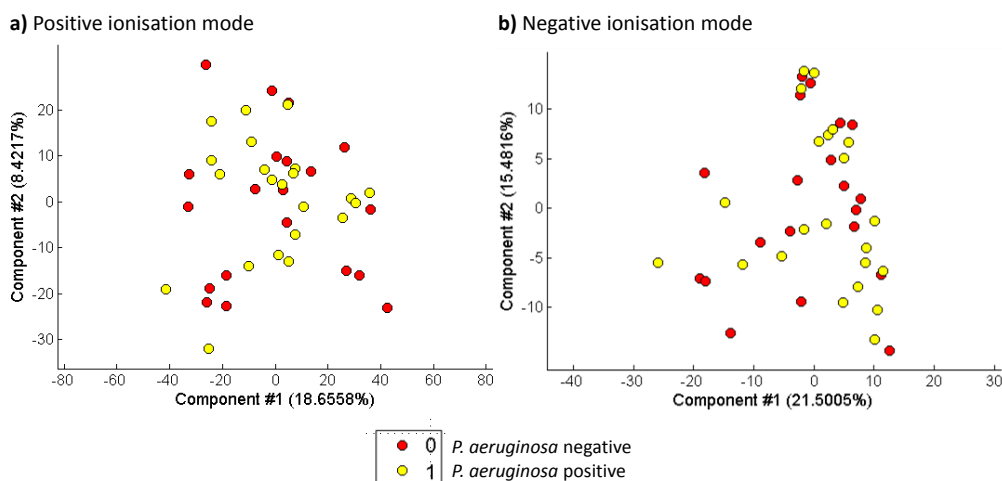


Figure 7.20. PCA score plots using DESI-MS profiling data acquired on skin secretions collected from CF patients negative and positive for *P. aeruginosa* infection.

Data acquired in positive (a) and negative (b) ionisation modes. Showing the two first principal components (built on Matlab).

Seeking differentiating features between infected and non-infected patients, a *t*-test was applied to each feature in the data sets. The results, represented as mass spectra in Figure 7.21 for positive and in Figure 7.22 for negative ionisation mode, showed that a small number of features had significant *p*-values (*p*-value < 0.05); a few of them were tentatively identified by exact mass comparison in positive ionisation mode (Figure 7.21). Fatty acid esters (two TGs, one alkyl ester) and one cholesterol derivate were globally lower in infected patients, whilst one phospholipid was higher. The latter was tentatively identified as PG(34:2), which has been notably detected before by REIMS in *P. aeruginosa* (Chapter 4). Another metabolite significantly higher in infected patients was tentatively identified as hydrocortisone. This compound may originate from the host metabolic response or from medical treatment (although unlikely in those patients), which in both case could explain a higher level in infected patients. Very few features were discriminant in negative ionisation mode and none could be reliably identified (Figure 7.22).

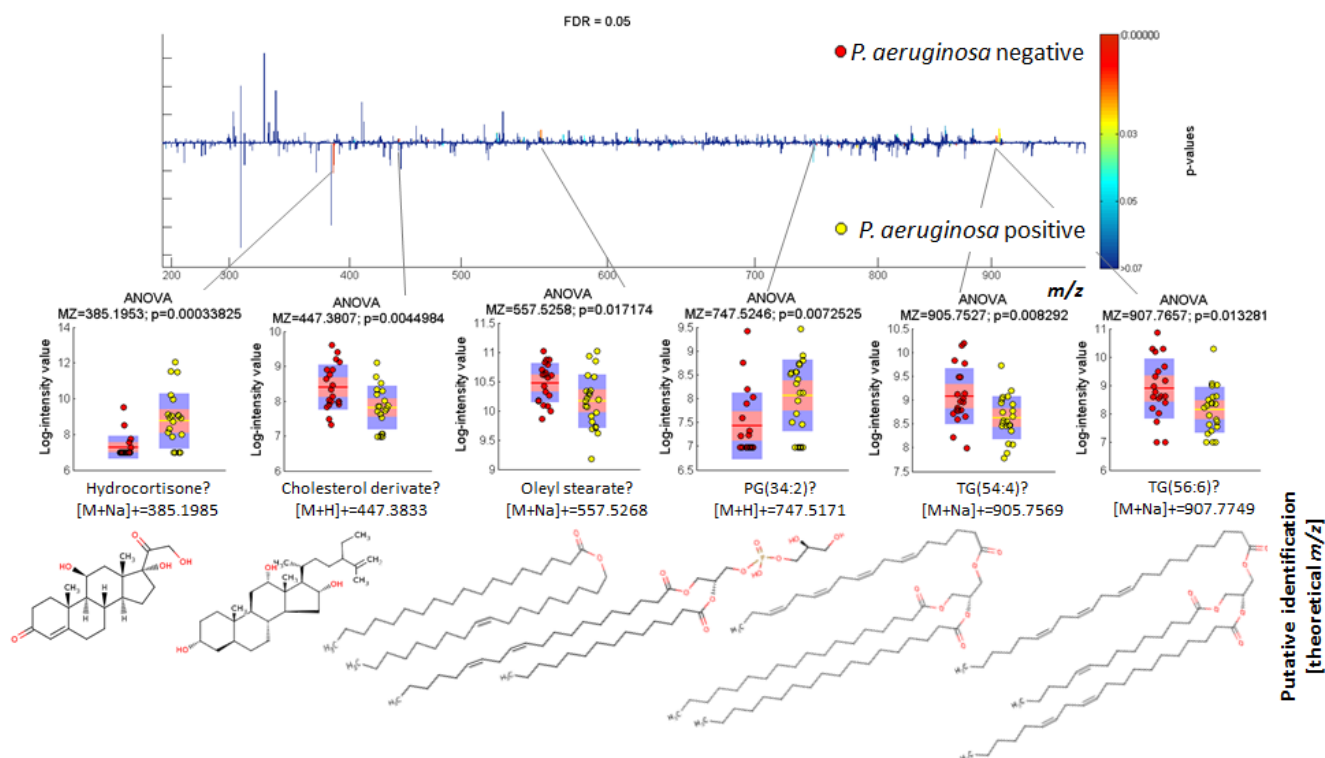


Figure 7.21. t-tests showing differentiating features, acquired in positive ionisation mode, between CF patients negative and positive for *P. aeruginosa* infection.

Represented as a mass spectrum, with features pointing up or down when characteristic peaks of negative or positive infection, respectively; features are coloured on a scale from blue to red as p-value decreases and significance increases (FDR = 0.05). Includes boxplots showing the distribution of selected metabolites with p-values < 0.05 within non-infected (red dots) and infected patients (yellow dots) (built on Matlab).

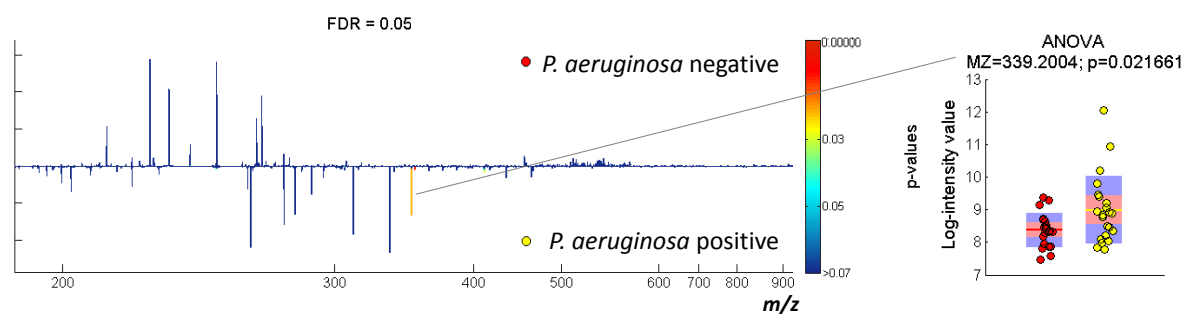


Figure 7.22. t-tests showing differentiating features, acquired in negative ionisation mode, between CF patients negative and positive for *P. aeruginosa* infection.

Represented as a mass spectrum, with features pointing up or down when characteristic peaks of negative or positive infection, respectively; features are coloured on a scale from blue to red as p-value decreases and significance increases (FDR = 0.05). Includes a boxplot showing the distribution of one metabolite with a p-value < 0.05 within non-infected (red dots) and infected patients (yellow dots) (built on Matlab).

Several classification models were examined in order to differentiate the two groups of samples, including PCA-LDA, RMMC and the Random Forest algorithm, and various methods of cross-validation and classifiers. Since the discriminant features were scattered throughout the spectra, reduced m/z ranges were not tested for modelling. All the methods led to similarly low classification accuracy; the

best results were obtained in positive ionisation mode with an RMMC model with an overall accuracy of 64% (leave-one-out cross-validation, Mahalanobis classifier). Overall accuracy using the negative ionisation data and the same cross-validation parameters was 55%.

7.5. Discussion

7.5.1. Evaluation of the method

The metabolic profile detected from skin secretions with DESI-MS was very rich in lipids. Although no definite identification was carried out, fatty acids and triglycerides seemed largely predominant, in agreement with the literature.

Amongst the studied factors, the choice of the skin site to be sampled showed a significant impact on the detected profile. This suggests that various skin area secrete different metabolites, and that the site to be sampled must be chosen in accordance with the objective of the experiment. Target molecules may be secreted by a specific area of skin; for instance, compounds excreted through sweat may be better detected from the skin of the palm than from that of the face. Analysis of skin vapours previously reported distinct profiles depending on the body area to be sampled [269]; however, a recent study comparing the lipidomes of the forehead and fingers found that the difference was not significant [276]. Their experimental plan included samples from various donors, looking for systematic metabolic differences, whilst the present study compared the profiles of a single donor acquired on one day, which may explain diverging conclusions.

IPA cleaning and freezing steps caused minor alterations of the metabolic profile, although fatty acids and triglycerides notably decreased after IPA cleaning. However, multiple freeze/thaw cycles had a significant impact, notably on the triglyceride m/z range. The quick and repeated change of atmospheric conditions may cause lipids to oxidise or hydrolyse.

It has been observed before that the skin of the forehead secretes lipid profiles unique to individuals [274]. Accordingly, the present protocol was able to predict a donor's identity according to his/her metabolic profile with 90% accuracy. Given the samples were collected and analysed over several months, it also implies that the method was reproducible and that differences between individuals were more important than potential secretion variations within donors over time. The lipid profiles detected in the present study classified gender with a maximum of 72% accuracy. Some features detected displayed levels significantly different in males and females, most of them likely to be

glycerides. Gender classification was previously achieved with 85% accuracy using peptidic profiles detected on fingerprints with MALDI-MS [271]. More recently, the DESI-MS analysis of lipids collected from the forehead achieved 89% accuracy for gender classification [276]. Their dataset included 194 samples (vs. 60 in the present study) and classification was performed using a machine learning algorithm called gradient boosting tree ensemble, possibly resulting in a stronger model.

Indeed, the data processing workflow employed here was basic and must be enhanced. Data reduction sometimes improved the prediction model and it is likely that selection of the specific discriminating features (p -value < 0.05) for modelling will enable a stronger classification. This requires the implementation of statistical functions that were not available at the time of the project but will be tested in the near future.

7.5.2. Clinical relevance

The lipidome detected with this method did not allow the diagnosis of *P. aeruginosa* infection. Nevertheless, a few interesting features were tentatively identified as differentiating patients with and without infection.

In previous studies, analysis of blood plasma revealed higher levels of some cholesterol derivatives and triglycerides in CF patients with chronic infection whilst a few phosphatidylcholines were reported at lower levels [128]. The analysis of lung tissue also showed an increase of cholesterol sulphate, and a decrease of fatty acids in infected patients [127]. The observations made in the present study were not congruent with these findings. Notably, tentatively identified molecules included two triglycerides and one cholesterol derivative that were significantly decreased in infected patients. Possible reasons for this divergence in lipid distribution may lie in the different metabolism and lipid utilisation of skin compared to other organs or fluids.

A defect in *CFTR* is known to impact the metabolism of lipids, resulting in abnormal profiles of fatty acids, cholesterol and sphingolipids [277]. Higher levels of triglycerides have also been reported in blood [278]. In this study, the dominance of triglycerides in CF patients compared to healthy individuals was clearly observed. However, although the selection of the triglyceride m/z range improved the model, accuracy for CF patient prediction was not higher than 78%, suggesting that other factors may impact triglycerides level in healthy subjects, or that CF patients do not demonstrate systematic higher levels of triglycerides.

This pilot study has limitations in its profiling approach. The detected metabolic profile was very rich, and although it showed patterns unique to each individual, human beings still share a common base profile that may conceal discrete information. The selection of reduced m/z ranges or significant features may improve the statistical models, and specific medications such as hydrocortisone may be targeted for therapeutic drug monitoring or forensic applications. Nevertheless, this study shed some light on the glyceride chemical family which seemed to be associated with various metabolic processes. However, these compounds are often overlooked, possibly due to the high diversity of their fatty acid substitution, which renders exact structure elucidation and metabolic interpretation laborious.

According to the present study, the lipidome secreted by the skin was not significantly impacted by the presence of bacterial infection in the respiratory tract; the initial hypothesis was thus not verified. Nevertheless, a simple, non-invasive procedure for the analysis of skin secretions was developed. The method led to the detection of a rich lipidome with a high reproducibility. Although it did not allow the prediction of *P. aeruginosa* infection, future targeted analyses on specific lipids, metabolites or drug compounds may be of clinical interest.

Chapter 8

General discussion

8. General discussion

The methods tested in this thesis revealed a great potential for bacterial phenotyping and for metabolomic studies with clinical applications. Unfortunately, the current protocols were not successful in diagnosing *P. aeruginosa* infections in CF patients; yet, there is scope for improvement and many other potential approaches exist.

8.1. Summary

8.1.1. Achievements

The initial hypotheses were that novel ambient mass spectrometric technologies, such as REIMS and DESI-MS, would be able to detect, by direct-from-sample analysis, theoretical modifications occurring in (a) the metabolome produced by the evolving pathogen; and (b) in the body fluids of CF patients, due to *P. aeruginosa* infection. Based on this, I intended to (a) provide insight into the mechanisms of *P. aeruginosa* infection and (b) develop a non-invasive culture-free diagnosis method.

In Chapters 4 and 5, aim (a) was addressed. REIMS was successful in identifying *P. aeruginosa* amongst common CF pathogens, and in characterising its metabolic profile directly from cultures in a high-throughput fashion. A wide range of clinical isolates were analysed and important virulence-associated metabolites were identified. These were shown to vary between strains, infection type (acute vs. respiratory) and stage (early vs. chronic in CF); isolates collected from patients with chronic respiratory infection in particular produced more diversified metabolic profiles. These studies confirmed the first hypothesis and demonstrated the potential of REIMS to provide insight on *P. aeruginosa*'s metabolism, adaptation and community mechanisms that may support, eventually, a change in clinical management and treatment strategies.

In Chapters 6 and 7, solutions were tested to achieve aim (b). Experiments were carried out to examine the value of REIMS and DESI-MS as diagnostic tools on clinical samples, seeking previously identified biomarkers for *P. aeruginosa* in order to discriminate infected patients. The methods employed did not allow the detection of significant differences in sputum, urine or skin secretions related to *P. aeruginosa* infection, disproving the second hypothesis for these specific clinical samples and methods. Other tools or protocols, including additional preparatory steps or other data processing workflows, might demonstrate that *P. aeruginosa* infection actually impacts the metabolome of those clinical samples, as previously shown using LC-MS on sputum and urine. Nevertheless, REIMS detected

a novel metabolic profile from sputum with potential clinical value. In addition, the protocol developed for the DESI-MS analysis of skin secretions revealed a very rich lipidome, which could bring information relevant to the study of CF and indeed other diseases, studies or fields of application.

8.1.2. Strengths of the studies and techniques

The real strength of novel mass spectrometric tools is their capacity to carry out real-time analysis directly from samples, in ambient conditions, on a wide range of sample types. The metabolites observed in this thesis typically require preliminary processing of the samples so that they can be subjected to LC-MS, which itself is time-consuming. Pre-treatment of the samples results in a method that is highly dependent on experimental conditions and prone to experimental errors, confounding factors and batch effects. Thanks to ambient methods, such as REIMS and DESI-MS, many variables are removed and we can get closer to understanding *in vivo* phenomena.

As far as possible, a large number of samples were included in each study. This led to the secure identification of virulence-associated molecules and demonstrated the high metabolic diversity of *P. aeruginosa*. In addition, it proved REIMS and DESI-MS to be reproducible and high-throughput, both in terms of number of samples and amount of information, bestowing a great potential for metabolomic studies on these techniques.

The laser-assisted REIMS platform is a powerful instrument which could soon be an alternative to MALDI-ToF systems for the speciation of microorganisms in clinical microbiological laboratories. Additionally, it provides metabolic information on bacteria cultured *in vitro* that is consistent over repeated cultures and analysis. REIMS was sensitive enough to classify to the isolate level, which is useful in the detection of transmission between patients. Furthermore, it was able to detect metabolic variations even below the strain classification level, i.e. between isolates belonging to the same strain, giving access to information about quickly evolving bacterial features, including virulence factors. Thus, REIMS provides a high resolution on real-time pathogenicity that may be even more clinically relevant than expensive and time-consuming DNA typing methods.

The REIMS technique and set-up allow analytic flexibility, not only regarding the diversity of samples that can be tested (e.g. bacterial cultures on plate, freshly expectorated sputum, pellets from urine), but also in terms of spatial applications. Multiple analyses may be carried out on various areas of a single sample, either for imaging purposes (e.g. as was done to evaluate sputum homogeneity), or for distinct targeted experiments, for instance the detection of contaminations in a cultured isolate, the

analysis of selected areas of heterogeneous samples (e.g. sputum), or the study of metabolic distribution within bacterial communities, etc.

8.1.3. Limitations

One of the main weaknesses of such flexible ambient methods is the inherent impact of the variable environmental and instrumental conditions at the time of analysis, leading to frequent and potentially important batch effects. The signal detected with DESI-MS in particular was very sensitive to the set-up geometry and ambient conditions. Although the use of a commercial DEFFI sprayer improved the robustness of the method, factors impacting the signal detected by DESI-MS are still under investigation, and studies are being carried out to reduce this variability, notably in collaboration with MS instrument manufacturers. Nevertheless, the noise introduced by extraneous factors can generally be attenuated by data processing (e.g. background subtraction, normalisation) or by using reference standards (external or internal (re-)calibration to adjust for intensity or m/z drifts). In this thesis, although batch effects, notably related to the date of analysis, were commonly observed, the richness of the detected metabolic profiles, together with mathematical corrections of the data, often allowed separation of the studied groups.

Although efforts have been made when planning the studies, experiments may lead to unexpected findings and sample numbers and availability sometimes limited the conclusions. This was true for instance when replicates were used instead of independent isolates to evaluate the typing-power of REIMS for MLST classification (Chapter 4). Again in Chapter 5, the great number of isolates analysed included multiple replicates per sputum samples, which rendered them dependant, and raises question around the differences observed between early and chronic infection. Likewise, higher numbers and more balanced sample sets of clinical skin secretions in Chapter 7 may have strengthened the prediction models, if not for infection diagnosis at least to confirm classifications achieved elsewhere [271, 276] for individual characteristics, e.g. gender, ethnicity, diet, allergies, etc.

Other difficulties encountered are intrinsic to the nature of the samples, e.g. urine samples may be highly and variably diluted and sputum samples are variable and heterogeneous; some of these issues may be overcome in the future with preliminary steps such as sample concentration or extraction.

Although most of the studies unravelled rich and diversified metabolic profiles, little feature identification could be carried out due to time constraints. This thwarted further metabolic investigation which could have led to novel insights into the biological processes relevant to

Chapters 5, 6, and 7 in particular, regarding bacterial phenotyping, sputum and skin secretion metabolism, respectively. Skin secretions were particularly rich in lipids, yet profiling analysis did not lead to conclusive classification models. A higher number of identified molecules may benefit these models and promote the utility of these data by focusing on relevant metabolites or chemical families.

Regarding the specific application of REIMS to cultured microorganisms, a number of limitations must be acknowledged. Speciation is still a recent application of REIMS [148]; libraries have not yet reached the number of species encompassed by the MALDI-ToF clinical platforms though databases are being continuously expanded. Despite its high degree of sensitivity, classification in accordance with defined genetic types, such as VNTR or MLST, could not be achieved, preventing any link with phylogenetics. Furthermore, the significance of bacterial expression *in vitro* with regard to the *in vivo* behaviour of bacteria infecting a patient's lungs has not been assessed and discrepancy has been reported previously [226]. Therefore, the clinical utility of the REIMS data relies, for now, on indirect correlations between *in vitro* cultures and clinical observations. This bias is introduced by the necessity to culture the bacteria in order to perform REIMS analysis and would have been overcome by the detection of the pathogen directly from sputum. Further drawbacks of the culturing constraints include the difficulty to compare the REIMS findings with assays employing different culturing methods, e.g. in the case of phenotypic studies.

8.2. Future research

8.2.1. Follow-up work

Next steps will be first to confirm the identities of the metabolic features, to expand their number in bacteria, sputum and skin secretions, and to explore the positive ionisation data that can now be acquired with REIMS for potential interesting metabolites in bacteria as well as in sputum samples, such as PYO, pyoverdine, AHLs, lipid A, phospholipids, etc.

With regard to the study of microorganisms using REIMS, studies involving more numerous patients/samples/isolates are needed to identify further correlations between the detected metabolic profile (or specific metabolites) and bacterial virulence, infection stage, high-risk clones, antibiotic resistance or impact of other treatments, and genetic data. Longitudinal studies are required to (1) characterise the bacteria adapting to its environment, (2) monitor the diversification of the bacterial population within the lungs of one patient as the infection progresses, (3) understand and control strain transmission between patients within a hospital. *In vitro* microbiological and molecular

experiments will unravel the biological functions of poorly characterised metabolites and confirm and/or elucidate metabolic synthesis pathways.

REIMS-imaging also has a great potential for the spatial study of bacterial communities and their interaction and competition with other microbes. This is currently a popular subject of research since significant proportions of the CF population are co-infected with distinct species, e.g. with *P. aeruginosa* and *S. aureus* [279] or *Aspergillus fumigatus* [280], which have been reported to worsen prognosis. Traditional microbiological assays have unravelled cooperative behaviours between these pathogens [281, 282], but imaging mass spectrometry offers a more comprehensive overview of these processes. MALDI-ToF [283, 284] and nano-DESI imaging techniques [285] have been successfully applied on co-cultured microorganisms revealing biotransformation events and spatial metabolite distributions (reviewed by Stasulli et al. [286]). The REIMS platform is now operational for imaging applications and offers additional insight thanks to its specific attributes (different metabolic targets, quick and direct-from-sample analysis). It is currently being tested on co-cultures of *P. aeruginosa* and *A. fumigatus*, showing interesting spatial distributions of AQs and rhamnolipids that may reveal collaborative or antagonist behaviours.

The analysis of sputum with REIMS also showed promising applications. Regarding its utility to diagnose infection, further testing using *in vitro* spiked sputum will determine the lowest level of bacterial load that can be expected to show signal and identify appropriate biomarkers (a few AQs and rhamnolipids were promising in preliminary assays). These specific biomarkers could be then sought using tools that are more suitable for targeted analysis, such as triple-quadrupole mass spectrometers which have higher sensitivity and wider linear dynamic range [287]. In addition, even if diagnosis is not achievable, sputum samples may contain highly valuable information about the host's metabolism and signal pathways that would be helpful in monitoring treatment effects or to predict exacerbations. A number of interesting metabolites are likely to be detected with REIMS, emphasising the importance of identifying more features.

8.2.2. Perspectives for *P. aeruginosa* infection diagnostics

The development of a method encompasses the elaboration of a protocol defining all the steps, from sample collection and potential pre-treatment to analysis and data processing, in order to answer the initial question. The diagnosis – allowing prompt treatment – of *P. aeruginosa* infection is an ongoing issue, upon which the health (and indeed, life) of most CF patients depends. Since the approaches

tested in this thesis were not successful in detecting infection, other combinations of samples and techniques, mass spectrometric or not, must be considered.

The adjunct of ion mobility to mass spectrometric systems allows the further separation of ions with close m/z ratios. It is notably relevant when dealing with samples showing very rich and dense profiles, such as in the case of skin secretions in positive ionisation mode. The CF Trust SRC for *P. aeruginosa* is now collaborating with Owlstone Ltd. (Cambridge, UK), which manufactures a device employing the FAIMS technology that could be implemented within the DESI-MS set-up. Declustering the signal of the ions may bring some clarity, uncover separating trends and facilitate interpretation and feature identification.

The detection of *P. aeruginosa* biomarkers directly in sputum is still topical, and cell-free biosensors have demonstrated potential as a quick and low-cost assay to detect QS molecules [262]. Although still relying on the expectoration of sputum – which is not possible for the youngest patients – the objective is to accelerate and enhance diagnosis as well as to provide metabolic information representative of the *in vivo* processes by eliminating the biasing culturing step.

P. aeruginosa QS molecules have also been detected in urine using LC-MS [98, 234] but other techniques and/or protocols are needed to enable a reliable diagnosis that could be applied routinely in the clinic. DESI-MS has been applied before on organic fractions of urine deposited on a surface for the tracking of drugs [251]. In order to eliminate any preliminary step, Chen et al. have combined two sprayers – one nebulising the sample and a second one spraying the extracting and ionising solvent – in a set-up called extractive ESI [259]. The set-up allowed the detection of signal in the low m/z range from complex biological matrices, including urine. These approaches may reveal other interesting metabolites in urine.

Beyond the samples themselves, others have focused on studying the vapours released by skin [269], sputum [162, 179] and urine [260, 261] using GC-MS, SESI-MS or electronic noses. Breath belongs to the same class of gaseous samples containing non-volatile compounds – namely aerosols; numerous techniques have been inconclusively tested in order to detect bacterial signal in breath as reviewed in Chapter 1. Ambient SESI-MS is a promising solution and had been considered in the course of this thesis. Preliminary data confirmed the detection of volatile and non-volatile molecules in breath and nebulised solutions (data not shown). Further optimisation of the set-up was required to analyse bacterial headspace before targeting biomarkers in breath. Other current hopeful strategies for the analysis of breath rely on electronic noses and are being developed by expert collaborators at the University of Amsterdam and at the Owlstone Company (Cambridge, UK). The Spironose, elaborated

to diagnose respiratory diseases and lung cancer [181], and Owlstone FAIMS technology, developed originally for the diagnosis of gastro-intestinal diseases [182-184], will be tested at our clinic in the near future. Another empirical approach adopted within the SRC for *P. aeruginosa* consists in training dogs to recognise the scent emitted by *P. aeruginosa*. The well-known superior olfactory abilities of dogs is already employed in forensics and public-safety applications. Their utility in the medical field is being increasingly documented, notably for the detection of cancers [288-292] and pathogens [293-295]. Experience showed that dogs can be successfully trained to recognise the *P. aeruginosa* volatile profile amongst other bacterial cultures (abstracts [296, 297]). The next stage must challenge the dogs with clinical samples, including sputum, exhaled breath condensates and breath, collected from infected and non-infected CF patients and appropriate controls.

The final step of a method is data processing. Metabolomic studies lead to the generation of large amounts of data, which entails the use of powerful data management and analysis strategies. In this thesis, a few statistical tools were tested, including classical approaches as well as more recent machine learning algorithms. But many others exist and keep flourishing with the advances in machine learning and big data analysis. For instance, the gradient boosting tree ensemble methods, which gave good predictions using skin secretion data elsewhere [276], offer novel implementations, e.g. extreme gradient boosting [298], that have demonstrated higher performance than the Random Forest algorithm and may unravel new possibilities in clinical applications.

To conclude, this thesis aimed to explore the potential of novel mass spectrometry methods to answer some key issues in CF. REIMS showed great promises for the metabolic phenotyping of bacterial isolates, providing unique information that is complementary to what can be obtained with other current techniques. Although metabolic profiling did not allow the direct diagnosis of infection here, the data generated remains largely unexplored. Further interrogation of the data based on refined research targets may unveil metabolic processes, and eventually enable infection detection. Experiments demonstrated that REIMS and DESI-MS were powerful tools for the collection of metabolomics data directly from sample; they have generated hypotheses which may be of benefit to a broader medical field.

List of communications

Oral presentations

- Novel strategies for *P. aeruginosa* infection diagnostics.

Bardin, E., Bolt, F., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; Cystic Fibrosis Conference (UKCFC); Manchester; September 2016; talk for lay/patients audience.

- Strain typing and metabolite monitoring of *P. aeruginosa* using rapid evaporative ionisation mass spectrometry.

Bardin, E., Cameron, S. J. S., Perdones-Montero, A., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; American Society of Microbiology (ASM); New Orleans (LA); June 2017; travel grant.

- Metabolic phenotyping and strain characterisation of *P. aeruginosa* isolates from cystic fibrosis patients using rapid evaporative ionisation mass spectrometry.

Bardin, E., Cameron, S. J. S., Perdones-Montero, A., Hardiman, K., Bolt, F., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; UK microbiology consortium; Liverpool; November 2017.

- Investigating *P. aeruginosa* using rapid evaporative ionisation mass spectrometry: metabolic characterisation, strain typing and clinical applications.

Bardin, E., Cameron, S. J. S., Perdones-Montero, A., Bolt, F., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; Mass Spectrometry Applications in Clinical Laboratories (MSACL) US; Palm Springs (CA); January 2018; travel grant.

Poster presentations

- Discriminatory ability of ambient ionisation mass spectrometry in analysing skin secretions: Towards a non-invasive diagnostic tool?

Bardin, E., Bolt, F., McKenzie, J., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; American Society of Mass Spectrometry (ASMS); San Antonio (TX); June 2016.

- Application of desorption electrospray ionisation mass spectrometry in analysing skin secretions: a non-invasive diagnostic tool?

Bardin, E., Bolt, F., McKenzie, J., Cameron, S. J. S., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; Mass Spectrometry Applications in Clinical Laboratories (MSACL) Europe; Salzburg; September 2016; travel grant.

- Metabolomic characterisation of *P. aeruginosa* isolates from cystic fibrosis patients using rapid evaporative ionisation mass spectrometry.

Bardin, E., Bolt, F., Cameron, S. J. S., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; North America Cystic Fibrosis Conference (NACFC); Orlando (FL); October 2016.

- Metabolomic characterisation of *P. aeruginosa* isolates from cystic fibrosis patients using rapid evaporative ionisation mass spectrometry.

Bardin, E., Bolt, F., Cameron, S. J. S., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; UK microbiology consortium; Liverpool; November 2016.

- Metabolite monitoring of *P. aeruginosa* using rapid evaporative ionisation mass spectrometry and applications for strain typing and infection diagnosis.

Bardin, E., Cameron, S. J. S., Perdones-Montero, A., Hardiman, K., Bolt, F., Alton, E. W., Bush, A., Davies, J. C., Takats, Z.; American Society of Mass Spectrometry (ASMS); Indianapolis (IN); June 2017.

- Looking for novel strategies to diagnose *P. aeruginosa* infection.

Bardin, E., Cameron, S. J. S., Perdones-Montero, A., Hardiman, K., Bolt, F., Alton, E. W., Bush, A., Davies, J. C., Takats, Z., UK Cystic Fibrosis Conference (UKCFC); Nottingham; September 2017.

Publications

- *Pseudomonas aeruginosa* infection in cystic fibrosis: pathophysiological mechanisms and therapeutic approaches.

Lund-Palau, H., Turnbull, A. R., Bush, A., Bardin, E., Cameron, L., Soren, O., Wierre-Gore, N., Alton, E. W., Bundy, J. G., Connett, G., Faust, S. N., Filloux, A., Freemont, P., Jones, A., Khoo, V., Morales, S., Murphy, R., Pabary, R., Simbo, A., Schelenz, S., Takats, Z., Webb, J., Williams, H. D., Davies, J. C., Expert Rev Respir Med, 2016. **10**(6): p. 685-97.

- *Pseudomonas aeruginosa* infection in patients with cystic fibrosis.

Smith, W.D., Bardin, E., Cameron, L., Edmondson, C. L., Farrant, K. V., Martin, I., Murphy, R. A., Soren, O., Turnbull, A. R., Wierre-Gore, N., Alton, E. W., Bundy, J. G., Bush, A., Connett, G. J., Faust, S. N., Filloux, A., Freemont, P. S., Jones, A. L., Takats, Z., Webb, J. S., Williams, H. D., Davies, J. C., Current and future therapies for FEMS Microbiol Lett, 2017. **364**(14).

- Metabolic Phenotyping and Strain Characterisation of *Pseudomonas aeruginosa* Isolates from Cystic Fibrosis Patients using Rapid Evaporative Ionisation Mass Spectrometry.

Bardin, E., Cameron, S. J. S., Perdones-Montero, A., Hardiman, K., Bolt, F., Alton, E. W., Bush, A., Davies, J. C., Takats, Z., Scientific Report, 2018. **7**.

- Analysis of Fingerprints by Desorption Electrospray Ionisation Mass Spectrometry Imaging.

Bardin, E., Claude, E., Takats, Z., Waters Technology Brief, April 2017.

References

1. UK CF Trust, *2017 Registry Annual Data Report*. 2017.
2. Reddy, M.M. and M.J. Stutts, *Status of fluid and electrolyte absorption in cystic fibrosis*. Cold Spring Harb Perspect Med, 2013. **3**(1): p. a009555.
3. Pezzulo, A.A., et al., *Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung*. Nature, 2012. **487**(7405): p. 109-13.
4. Konstan, M.W. and M. Berger, *Current understanding of the inflammatory process in cystic fibrosis: onset and etiology*. Pediatr Pulmonol, 1997. **24**: p. 137-142.
5. Cantin, A.M., et al., *Inflammation in cystic fibrosis lung disease: Pathogenesis and therapy*. J Cyst Fibros, 2015. **14**(4): p. 419-30.
6. Bruscia, E.M. and T.L. Bonfield, *Innate and Adaptive Immunity in Cystic Fibrosis*. Clin Chest Med, 2016. **37**(1): p. 17-29.
7. Goldstein, W. and G. Doring, *Lysosomal enzymes from polymorphonuclear leukocytes and proteinase inhibitors in patients with cystic fibrosis*. American Review of Respiratory Disease, 1986. **134**(1): p. 49-56.
8. Su, X., et al., *Role of CFTR expressed by neutrophils in modulating acute lung inflammation and injury in mice*. Inflamm Res, 2011. **60**(7): p. 619-32.
9. Painter, R.G., et al., *The role of chloride anion and CFTR in killing of Pseudomonas aeruginosa by normal and CF neutrophils*. J Leukoc Biol, 2008. **83**(6): p. 1345-53.
10. Kerem, E., et al., *Factors associated with FEV1 decline in cystic fibrosis: analysis of the ECFS patient registry*. Eur Respir J, 2014. **43**(1): p. 125-33.
11. Kosorok, M.R., et al., *Acceleration of lung disease in children with cystic fibrosis after Pseudomonas aeruginosa acquisition*. Pediatric Pulmonology, 2001. **32**(277-287).
12. Jarad, N.A., et al., *Factors associated with reduced FEV1 in adult patients with cystic fibrosis in a relatively affluent area*. Chronic Respiratory Disease, 2005. **2**: p. 133-137.
13. Sadikot, R.T., et al., *Pathogen-host interactions in Pseudomonas aeruginosa pneumonia*. Am J Respir Crit Care Med, 2005. **171**(11): p. 1209-23.
14. Smith, W.D., et al., *Current and future therapies for Pseudomonas aeruginosa infection in patients with cystic fibrosis*. FEMS Microbiol Lett, 2017. **364**(14).
15. Stover, C.K., et al., *Complete genome sequence of pseudomonas aeruginosa PA01, an opportunistic pathogen*. Nature, 2000. **406**: p. 959-964.
16. Schuster, M. and E.P. Greenberg, *A network of networks: quorum-sensing gene regulation in Pseudomonas aeruginosa*. Int J Med Microbiol, 2006. **296**(2-3): p. 73-81.
17. Kidd, T.J., et al., *Pseudomonas aeruginosa genotypes acquired by children with cystic fibrosis by age 5-years*. J Cyst Fibros, 2015. **14**(3): p. 361-9.
18. Burns, J.L., et al., *Longitudinal Assessment of Pseudomonas aeruginosa in Young Children with Cystic Fibrosis*. The Journal of Infectious Diseases, 2001. **183**(444-452).
19. Nixon, G.M., et al., *Clinical outcome after early Pseudomonas aeruginosa infection in cystic fibrosis*. J Pediatr, 2001. **138**(5): p. 699-704.
20. Hansen, C.R., T. Pressler, and N. Hoiby, *Early aggressive eradication therapy for intermittent Pseudomonas aeruginosa airway colonization in cystic fibrosis patients: 15 years experience*. J Cyst Fibros, 2008. **7**(6): p. 523-30.
21. Lee, T.W.R., *Eradication of early Pseudomonas infection in cystic fibrosis*. Chronic Respiratory Disease, 2009. **6**: p. 99-107.
22. UK CF Trust, *2016 Registry Annual Data Report*. 2016.
23. Lee, T.W.R., et al., *Evaluation of a new definition for chronic Pseudomonas aeruginosa infection in cystic fibrosis patients*. Journal of Cystic Fibrosis, 2003. **2**(1): p. 29-34.
24. Aaron, S.D., et al., *Infection With Transmissible Strains of Pseudomonas aeruginosa and Clinical Outcomes in Adults With Cystic Fibrosis*. 2010.

25. Armstrong, D.S., et al., *Detection of a widespread clone of Pseudomonas aeruginosa in a pediatric cystic fibrosis clinic*. Am J Respir Crit Care Med, 2002. **166**(7): p. 983-7.
26. Kidd, T.J., et al., *The social network of cystic fibrosis centre care and shared Pseudomonas aeruginosa strain infection: a cross-sectional analysis*. Lancet Respir Med, 2015. **3**: p. 640-650.
27. Fothergill, J.L., M.J. Walshaw, and C. Winstanley, *Transmissible strains of Pseudomonas aeruginosa in cystic fibrosis lung infections*. Eur Respir J, 2012. **40**(1): p. 227-38.
28. Emerson, J., et al., *Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis*. Pediatr Pulmonol, 2002. **34**(2): p. 91-100.
29. Hoiby, N., et al., *Pseudomonas aeruginosa and the in vitro and in vivo biofilm mode of growth*. Microbes and Infection, 2001. **3**: p. 23-35.
30. Davies, D.G., et al., *The Involvement of Cell-to-Cell Signals in the Development of a Bacterial Biofilm*. Science Translational Medicine, 1998. **280**: p. 295-298.
31. Lee, J. and L. Zhang, *The hierarchy quorum sensing network in Pseudomonas aeruginosa*. Protein Cell, 2015. **6**(1): p. 26-41.
32. Jensen, P.O., et al., *The immune system vs. Pseudomonas aeruginosa biofilms*. FEMS Immunol Med Microbiol, 2010. **59**(3): p. 292-305.
33. Hoiby, N., O. Ciofu, and T. Bjarnsholt, *Pseudomonas aeruginosa biofilms in cystic fibrosis*. Future Microbiol, 2010. **5**(11): p. 1663-74.
34. Jensen, P.O., et al., *Rapid necrotic killing of polymorphonuclear leukocytes is caused by quorum-sensing-controlled production of rhamnolipid by Pseudomonas aeruginosa*. Microbiology, 2007. **153**(Pt 5): p. 1329-38.
35. Mahenthiralingam, E., M.E. Campbell, and D.P. Speert, *Nonmotility and Phagocytic Resistance of Pseudomonas aeruginosa Isolates from Chronically Colonized Patients with Cystic Fibrosis*. Infection and Immunity, 1994. **62**(2): p. 596-605.
36. Faure, E., K. Kwong, and D. Nguyen, *Pseudomonas aeruginosa in Chronic Lung Infections: How to Adapt Within the Host?* Front Immunol, 2018. **9**: p. 2416.
37. Ernst, R.K., et al., *Pseudomonas aeruginosa lipid A diversity and its recognition by Toll-like receptor 4*. J Endotoxin Res, 2003. **9**(6): p. 395-400.
38. Nikaido, H., *Molecular Basis of Bacterial Outer Membrane Permeability Revisited*. Microbiology and Molecular Biology Reviews, 2003. **67**(4): p. 593-656.
39. Hancock, R.E., et al., *Pseudomonas aeruginosa isolates from patients with cystic fibrosis: A class of serum-sensitive, nontypable strains deficient in lipopolysaccharide O side chains*. Infect Immun, 1983. **42**(1): p. 170-177.
40. Kharazmi, A., *Mechanisms involved in the evasion of the host defence by Pseudomonas aeruginosa*. Immunology Letters, 1991. **30**: p. 201-206.
41. Wagner, V.E. and B.H. Iglewski, *P. aeruginosa Biofilms in CF Infection*. Clin Rev Allergy Immunol, 2008. **35**(3): p. 124-34.
42. Chiang, W.C., et al., *Extracellular DNA shields against aminoglycosides in Pseudomonas aeruginosa biofilms*. Antimicrob Agents Chemother, 2013. **57**(5): p. 2352-61.
43. Licht, T.R., et al., *Plasmid transfer in the animal intestine and other dynamic bacterial populations: the role of community structure and environment*. Microbiology, 1999. **145**: p. 2615-2622.
44. Fux, C.A., et al., *Survival strategies of infectious biofilms*. Trends Microbiol, 2005. **13**(1): p. 34-40.
45. Stewart, P.S. and J.W. Costerton, *Antibiotic resistance of bacteria in biofilms*. The Lancet, 2001. **358**: p. 135-38.
46. Winstanley, C., S. O'Brien, and M.A. Brockhurst, *Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections*. Trends Microbiol, 2016. **24**(5): p. 327-37.
47. Mulcahy, L.R., V.M. Isabella, and K. Lewis, *Pseudomonas aeruginosa biofilms in disease*. Microb Ecol, 2014. **68**(1): p. 1-12.

48. Smith, E.E., et al., *Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients*. Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8487-92.
49. Aires, J.R., et al., *Involvement of an Active Efflux System in the Natural Resistance of Pseudomonas aeruginosa to Aminoglycosides*. Antimicrobial Agents and Chemotherapy, 1999. **43**(11): p. 2624-2628.
50. Giwercman, B., et al., *High-Level 3-Lactamase Activity in Sputum Samples from Cystic Fibrosis Patients during Antipseudomonal Treatment*. Antimicrob Agents Chemother, 1992. **31**(1): p. 71-76.
51. Daniels, T.W., et al., *Impact of antibiotic treatment for pulmonary exacerbations on bacterial diversity in cystic fibrosis*. J Cyst Fibros, 2013. **12**(1): p. 22-8.
52. Rodriguez-Rojas, A., et al., *Antibiotics and antibiotic resistance: a bitter fight against evolution*. Int J Med Microbiol, 2013. **303**(6-7): p. 293-7.
53. Bobay, L.M., C.C. Traverse, and H. Ochman, *Impermanence of bacterial clones*. Proc Natl Acad Sci U S A, 2015. **112**(29): p. 8893-900.
54. Tramper-Stranders, G.A., C.K. van der Ent, and T.F. Wolfs, *Detection of Pseudomonas aeruginosa in patients with cystic fibrosis*. J Cyst Fibros, 2005. **4 Suppl 2**: p. 37-43.
55. Gilchrist, F.J., et al., *Bronchoalveolar lavage in children with cystic fibrosis: how many lobes should be sampled?* Arch Dis Child, 2011. **96**(3): p. 215-7.
56. Ronchetti, K., et al., *The CF-Sputum Induction Trial (CF-SpIT) to assess lower airway bacterial sampling in young children with cystic fibrosis: a prospective internally controlled interventional trial*. Lancet Respir Med, 2018. **6**(6): p. 461-471.
57. Aaron, S.D., et al., *Sputum versus bronchoscopy for diagnosis of Pseudomonas aeruginosa biofilms in cystic fibrosis*. Eur Respir J, 2004. **24**(4): p. 631-7.
58. Forton, J., *Induced sputum in young healthy children with cystic fibrosis*. Paediatr Respir Rev, 2015. **16 Suppl 1**: p. 6-8.
59. Blau, H., et al., *Induced sputum compared to bronchoalveolar lavage in young, non-expectorating cystic fibrosis children*. J Cyst Fibros, 2014. **13**(1): p. 106-10.
60. Ferreira, A.C.M., et al., *Hypertonic Saline as a Useful Tool for Sputum Induction and Pathogen Detection in Cystic Fibrosis*. Lung, 2017. **195**(4): p. 431-439.
61. D'Sylva, P., et al., *Induced sputum to detect lung pathogens in young children with cystic fibrosis*. Pediatr Pulmonol, 2017. **52**(2): p. 182-189.
62. Rosenfeld, M., et al., *Diagnostic Accuracy of Oropharyngeal Cultures in Infants and Young Children With Cystic Fibrosis*. Pediatr Pulmonol, 1999. **28**(321-328).
63. Zemanick, E.T., et al., *Assessment of airway microbiota and inflammation in cystic fibrosis using multiple sampling methods*. Ann Am Thorac Soc, 2015. **12**(2): p. 221-9.
64. Hery-Arnaud, G., et al., *Evaluation of quantitative PCR for early diagnosis of Pseudomonas aeruginosa infection in cystic fibrosis: a prospective cohort study*. Clin Microbiol Infect, 2017. **23**(3): p. 203-207.
65. Mellmann, A., et al., *Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria*. J Clin Microbiol, 2008. **46**(6): p. 1946-54.
66. Patel, J.B., *16S rRNA Gene Sequencing for Bacterial Pathogen Identification in the Clinical Laboratory*. Molecular Diagnosis, 2001. **6**(4).
67. Janda, J.M. and S.L. Abbott, *16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls*. J Clin Microbiol, 2007. **45**(9): p. 2761-4.
68. Yarza, P., et al., *Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences*. Nat Rev Microbiol, 2014. **12**(9): p. 635-45.
69. de Vos, D., et al., *Direct Detection and Identification of Pseudomonas aeruginosa in Clinical Samples Such as Skin Biopsy Specimens and Expectorations by Multiplex PCR Based on Two Outer Membrane Lipoprotein Genes, oprI and oprL*. Journal of Clinical Microbiology, 1997. **35**(6): p. 1295-1299.

70. Khan, A.A. and C.E. Cerniglia, *Detection of Pseudomonas aeruginosa from Clinical and Environmental Samples by Amplification of the Exotoxin A Gene Using PCR*. Applied and Environmental Microbiology, 1994. **60**(10): p. 3739-3745.
71. Salman, M., A. Ali, and A. Haque, *A novel multiplex PCR for detection of Pseudomonas aeruginosa: A major cause of wound infections*. Pakistan Journal of Medical Sciences, 2013. **29**(4).
72. Ullah, W., et al., *Beta-lactamase-producing Pseudomonas aeruginosa: Phenotypic characteristics and molecular identification of virulence genes*. J Chin Med Assoc, 2017. **80**(3): p. 173-177.
73. Porte, L., et al., *Head-to-head comparison of Microflex LT and Vitek MS systems for routine identification of microorganisms by MALDI-TOF mass spectrometry in Chile*. PLoS One, 2017. **12**(5): p. e0177929.
74. Pinkart, H.C. and D.C. White, *Lipids of Pseudomonas*, ed. M.T.C.e.P.B. Handbooks. Vol. 10. 1998: Springer, Boston, MA.
75. Wilkinson, S.G., *Gram-negative bacteria*. Microbial Lipids, 1988. **1**: p. 333-348.
76. Sohlenkamp, C. and O. Geiger, *Bacterial membrane lipids: diversity in structures and pathways*. FEMS Microbiol Rev, 2016. **40**(1): p. 133-59.
77. Benamara, H., et al., *Impact of the biofilm mode of growth on the inner membrane phospholipid composition and lipid domains in Pseudomonas aeruginosa*. Biochim Biophys Acta, 2011. **1808**(1): p. 98-105.
78. Mykytczuk, N.C., et al., *Fluorescence polarization in studies of bacterial cytoplasmic membrane fluidity under environmental stress*. Prog Biophys Mol Biol, 2007. **95**(1-3): p. 60-82.
79. De Kievit, T.R., et al., *Quorum-sensing genes in Pseudomonas aeruginosa biofilms: their role and expression patterns*. Appl Environ Microbiol, 2001. **67**(4): p. 1865-73.
80. Lee, K. and S.S. Yoon, *Pseudomonas aeruginosa Biofilm, a Programmed Bacterial Life for Fitness*. J Microbiol Biotechnol, 2017. **27**(6): p. 1053-1064.
81. D'Argenio, D.A., et al., *Growth phenotypes of Pseudomonas aeruginosa lasR mutants adapted to the airways of cystic fibrosis patients*. Mol Microbiol, 2007. **64**(2): p. 512-33.
82. Bjarnsholt, T., et al., *Pseudomonas aeruginosa tolerance to tobramycin, hydrogen peroxide and polymorphonuclear leukocytes is quorum-sensing dependent*. Microbiology, 2005. **151**(Pt 2): p. 373-83.
83. Srivastava, D. and C.M. Waters, *A tangled web: regulatory connections between quorum sensing and cyclic Di-GMP*. J Bacteriol, 2012. **194**(17): p. 4485-93.
84. Lee, J., et al., *A cell-cell communication signal integrates quorum sensing and stress response*. Nat Chem Biol, 2013. **9**(5): p. 339-43.
85. Toder, D.S., et al., *lasA and lasB Genes of Pseudomonas aeruginosa: Analysis of Transcription and Gene Product Activity*. Am Soc Mic, 1994. **62**(4): p. 1320-1327.
86. Brint, J.M. and D.E. Ohman, *Synthesis of Multiple Exoproducts in Pseudomonas aeruginosa Is under the Control of RhIR-RhII, Another Set of Regulators in Strain PAO1 with Homology to the Autoinducer-Responsive LuxR-LuxI Family*. Journal of Bacteriology, 1995. **177**(24): p. 7155-7163.
87. Winzer, K., et al., *The Pseudomonas aeruginosa Lectins PA-IL and PA-III Are Controlled by Quorum Sensing and by RpoS*. Journal of Bacteriology, 2000. **182**(22): p. 6401-6411.
88. Cooley, M., S.R. Chhabra, and P. Williams, *N-Acylhomoserine lactone-mediated quorum sensing: a twist in the tail and a blow for host immunity*. Chem Biol, 2008. **15**(11): p. 1141-7.
89. Cornforth, J.W. and A.T. James, *Structure of a naturally occurring antagonist of dihydrostreptomycin*. Biochem J, 1956. **63**(1): p. 124-130.
90. Bredenbruch, F., et al., *Biosynthetic pathway of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines*. J Bacteriol, 2005. **187**(11): p. 3630-5.

91. Deziel, E., et al., *Analysis of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication*. Proc Natl Acad Sci U S A, 2004. **101**(5): p. 1339-44.
92. Diggle, S.P., et al., *The Pseudomonas aeruginosa 4-quinolone signal molecules HHQ and PQS play multifunctional roles in quorum sensing and iron entrapment*. Chem Biol, 2007. **14**(1): p. 87-96.
93. Mashburn, L.M. and M. Whiteley, *Membrane vesicles traffic signals and facilitate group activities in a prokaryote*. Nature, 2005. **437**(7057): p. 422-5.
94. Bredenbruch, F., et al., *The Pseudomonas aeruginosa quinolone signal (PQS) has an iron-chelating activity*. Environ Microbiol, 2006. **8**(8): p. 1318-29.
95. Toyofuku, M., et al., *Influence of the Pseudomonas quinolone signal on denitrification in Pseudomonas aeruginosa*. J Bacteriol, 2008. **190**(24): p. 7947-56.
96. D'Argenio, D.A., et al., *Autolysis and Autoaggregation in Pseudomonas aeruginosa Colony Morphology Mutants*. Journal of Bacteriology, 2002. **184**(23): p. 6481-6489.
97. Collier, D.N., et al., *A bacterial cell to cell signal in the lungs of cystic fibrosis patients*. FEMS Microbiol Lett, 2002. **215**: p. 41-46.
98. Barr, H.L., et al., *Pseudomonas aeruginosa quorum sensing molecules correlate with clinical status in cystic fibrosis*. Eur Respir J, 2015. **46**(4): p. 1046-54.
99. Guina, T., et al., *Quantitative proteomic analysis indicates increased synthesis of a quinolone by Pseudomonas aeruginosa isolates from cystic fibrosis airways*. Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2771-6.
100. Lépine, F., et al., *A stable isotope dilution assay for the quantification of the Pseudomonas quinolone signal in Pseudomonas aeruginosa cultures*. Biochimica et Biophysica Acta (BBA) - General Subjects, 2003. **1622**(1): p. 36-41.
101. Abdel-Mawgoud, A.M., F. Lepine, and E. Deziel, *Rhamnolipids: diversity of structures, microbial origins and roles*. Appl Microbiol Biotechnol, 2010. **86**(5): p. 1323-36.
102. Zgola-Grzeskowiak, A. and E. Kaczorek, *Isolation, preconcentration and determination of rhamnolipids in aqueous samples by dispersive liquid-liquid microextraction and liquid chromatography with tandem mass spectrometry*. Talanta, 2011. **83**(3): p. 744-50.
103. Van Gennip, M., et al., *Inactivation of the rhIA gene in Pseudomonas aeruginosa prevents rhamnolipid production, disabling the protection against polymorphonuclear leukocytes*. APMIS, 2009. **117**(7): p. 537-46.
104. Boles, B.R., M. Thoendel, and P.K. Singh, *Rhamnolipids mediate detachment of Pseudomonas aeruginosa from biofilms*. Mol Microbiol, 2005. **57**(5): p. 1210-23.
105. Itoh, S.H., H., F. Tomita, and T. Suzuki, *Rhamnolipids produced by Pseudomonas aeruginosa grown on n-paraffin*. The Journal of Antibiotics, 1971. **24**(12): p. 855-859.
106. Lang, S., E. Katsiwela, and F. Wagner, *Antimicrobial Effects of Biosurfactants*. Lipid / Fett, 1989. **91**(9): p. 363-366.
107. Read, R.C., et al., *Effect of Pseudomonas aeruginosa rhamnolipids on mucociliary transport and ciliary beating*. The American Physiological Society, 1992. **72**(6): p. 2271-2277.
108. Calfee, M.W., et al., *Solubility and Bioactivity of the Pseudomonas Quinolone Signal Are Increased by a Pseudomonas aeruginosa-Produced Surfactant*. Infection and Immunity, 2005. **73**(2): p. 878-882.
109. Kownatzki, R., B. Tummier, and G. Doring, *Rhamnolipid of Pseudomonas aeruginosa in Sputum of Cystic Fibrosis Patients*. The Lancet, 1987. **329**(8540): p. 1026-1027.
110. Pollack, P., *The Role of Exotoxin A in Pseudomonas Disease and Immunity*. Reviews of Infectious Diseases, 1983. **5**: p. S979-S984.
111. Michalska, M. and P. Wolf, *Pseudomonas Exotoxin A: optimized by evolution for effective killing*. Front Microbiol, 2015. **6**: p. 963.
112. Chemani, C., et al., *Role of LecA and LecB lectins in Pseudomonas aeruginosa-induced lung injury and effect of carbohydrate ligands*. Infect Immun, 2009. **77**(5): p. 2065-75.

113. Banin, E., M.L. Vasil, and E.P. Greenberg, *Iron and Pseudomonas aeruginosa biofilm formation*. Proc Natl Acad Sci U S A, 2005. **102**(31): p. 11076-81.
114. Lamont, I.L., et al., *Siderophore-mediated signaling regulates virulence factor production in Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A, 2002. **99**(10): p. 7072-7.
115. Muller, M., *Pyocyanin induces oxidative stress in human endothelial cells and modulates the glutathione redox cycle*. Free radical Biology and Medicine, 2002. **33**(11): p. 1527-1533.
116. Usher, L.R., et al., *Induction of Neutrophil Apoptosis by the Pseudomonas aeruginosa Exotoxin Pyocyanin: A Potential Mechanism of Persistent Infection*. The Journal of Immunology, 2002. **168**(4): p. 1861-1868.
117. Mahajan-Miklos, S., et al., <Mahajan-Miklos 1999.pdf>. cell, 1999. **96**(47-56).
118. Wilson, R., et al., *Pyocyanin and 1-hydroxyphenazine produced by Pseudomonas aeruginosa inhibit the beating of human respiratory cilia in vitro*. J Clin Invest, 1987. **79**(1): p. 221-9.
119. Nair, C., et al., *Cyanide levels found in infected cystic fibrosis sputum inhibit airway ciliary function*. Eur Respir J, 2014. **44**(5): p. 1253-61.
120. Wilson, R., et al., *Measurement of Pseudomonas aeruginosa Phenazine Pigments in Sputum and Assessment of Their Contribution to Sputum Sol Toxicity for Respiratory Epithelium*. Infection and Immunity, 1988. **56**(9): p. 2515-2517.
121. Pastells, C., et al., *Immunochemical Determination of Pyocyanin and 1-Hydroxyphenazine as Potential Biomarkers of Pseudomonas aeruginosa Infections*. Anal Chem, 2016. **88**(3): p. 1631-8.
122. Williams, H.D., J.E.A. Zlosnik, and B. Ryall, *Oxygen, Cyanide and Energy Generation in the Cystic Fibrosis Pathogen Pseudomonas aeruginosa*. 2006. **52**: p. 1-71.
123. Gallagher, L.A. and C. Manoil, *Pseudomonas aeruginosa PAO1 kills Caenorhabditis elegans by cyanide poisoning*. J Bacteriol, 2001. **183**(21): p. 6207-14.
124. Carterson, A.J., et al., *The transcriptional regulator AlgR controls cyanide production in Pseudomonas aeruginosa*. J Bacteriol, 2004. **186**(20): p. 6837-44.
125. Ryall, B., et al., *Pseudomonas aeruginosa, cyanide accumulation and lung function in CF and non-CF bronchiectasis patients*. Eur Respir J, 2008. **32**(3): p. 740-7.
126. Enderby, B., et al., *Hydrogen cyanide as a biomarker for Pseudomonas aeruginosa in the breath of children with cystic fibrosis*. Pediatr Pulmonol, 2009. **44**(2): p. 142-7.
127. Desbenoit, N., et al., *Localized lipidomics in cystic fibrosis: TOF-SIMS imaging of lungs from Pseudomonas aeruginosa-infected mice*. Int J Biochem Cell Biol, 2014. **52**: p. 77-82.
128. Ollero, M., et al., *Plasma lipidomics reveals potential prognostic signatures within a cohort of cystic fibrosis patients*. J Lipid Res, 2011. **52**(5): p. 1011-22.
129. Becker, K.A., et al., *Lipids in cystic fibrosis*. Expert Rev Respir Med, 2011. **5**(4): p. 527-35.
130. Hoerr, V., et al., *Gram-negative and Gram-positive bacterial infections give rise to a different metabolic response in a mouse model*. J Proteome Res, 2012. **11**(6): p. 3231-45.
131. Robroeks, C.M., et al., *Metabolomics of volatile organic compounds in cystic fibrosis patients and controls*. Pediatr Res, 2010. **68**(1): p. 75-80.
132. Robroeks, C.M., et al., *Structural lung changes, lung function, and non-invasive inflammatory markers in cystic fibrosis*. Pediatr Allergy Immunol, 2010. **21**(3): p. 493-500.
133. Palmer, A., et al., *FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry*. Nat Methods, 2017. **14**(1): p. 57-60.
134. Whitehouse, C.M., et al., *Electrospray interface for liquid chromatographs and mass spectrometers*. Anal Chem, 1985. **57**(3): p. 675-9.
135. Banerjee, S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte*. Int J Anal Chem, 2012. **2012**: p. 282574.
136. Karas, M., D. Bachmann, and F. Hillenkamp, *Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules*. Analytical Chemistry, 1985. **57**(14): p. 2935-2939.

137. Francese, S., et al., *Beyond the ridge pattern: multi-informative analysis of latent fingerprints by MALDI mass spectrometry*. Analyst, 2013. **138**(15): p. 4215-28.
138. Laiko, V.V., M.A. Baldwin, and A.L. Burlingame, *Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry*. Anal Chem, 2000. **72**(4): p. 652-7.
139. Nicholson, J.K., et al., *Metabolic phenotyping in clinical and surgical environments*. Nature, 2012. **491**(7424): p. 384-92.
140. Ryzhov, V. and C. Fenselau, *Characterisation of the Protein Subset Desorbed by MALDI from Whole Bacterial Cells*. Anal. Chem., 2001. **73**: p. 746-750.
141. Kim, Y.W., et al., *MALDI-MS-based quantitative analysis for ketone containing homoserine lactones in Pseudomonas aeruginosa*. Anal Chem, 2015. **87**(2): p. 858-63.
142. Price, N.P., et al., *MALDI-TOF mass spectrometry of naturally occurring mixtures of monorhamnolipids and dirhamnolipids*. Carbohydr Res, 2009. **344**(2): p. 204-9.
143. Zavalin, A., et al., *Tissue protein imaging at 1 µm laser spot diameter for high spatial resolution and high imaging speed using transmission geometry MALDI TOF MS*. Anal Bioanal Chem, 2015. **407**(8): p. 2337-42.
144. Takats, Z., et al., *Mass spectrometry sampling under ambient conditions with desorption electrospray ionization*. Science, 2004. **306**(5695): p. 471-3.
145. Schafer, K.C., et al., *In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry*. Angew Chem Int Ed Engl, 2009. **48**(44): p. 8240-2.
146. Abbassi-Ghadi, N., et al., *Repeatability and reproducibility of desorption electrospray ionization-mass spectrometry (DESI-MS) for the imaging analysis of human cancer tissue: a gateway for clinical applications*. Anal. Methods, 2015. **7**(1): p. 71-80.
147. Balog, J., et al., *Identification of Biological Tissues by Rapid Evaporative Ionization Mass Spectrometry*. Anal Chem, 2010. **82**: p. 7343-7350.
148. Strittmatter, N., et al., *Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry*. Chem Commun (Camb), 2013. **49**(55): p. 6188-90.
149. Strittmatter, N., et al., *Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry*. Anal Chem, 2014. **86**(13): p. 6555-62.
150. Lechevalier, M.P. and C.W. Moss, *Lipids in Bacterial Taxonomy - A Taxonomist's View*. CRC Critical Reviews in Microbiology. Vol. 5. 1977. 109-210.
151. Bolt, F., et al., *Automated High-Throughput Identification and Characterization of Clinically Important Bacteria and Fungi using Rapid Evaporative Ionization Mass Spectrometry*. Analytical Chemistry, 2016. **88**(19): p. 9419-9426.
152. Bolt, F., et al., *Automated High-Throughput Identification and Characterization of Clinically Important Bacteria and Fungi using Rapid Evaporative Ionization Mass Spectrometry*. Anal Chem, 2016.
153. Takats, Z., J.M. Wiseman, and R.G. Cooks, *Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology*. J Mass Spectrom, 2005. **40**(10): p. 1261-75.
154. Tillner, J., et al., *Faster, More Reproducible DESI-MS for Biological Tissue Imaging*. J. Am. Soc. Mass Spectrom., 2017. **28**: p. 2090-2098.
155. Abbassi-Ghadi, N., et al., *Imaging of Esophageal Lymph Node Metastases by Desorption Electrospray Ionization Mass Spectrometry*. Cancer Res, 2016. **76**(19): p. 5647-5656.
156. Guenther, S., et al., *Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry*. Cancer Res, 2015. **75**(9): p. 1828-37.
157. Veselkov, K., et al., *BASIS: High-performance bioinformatics platform for processing of large-scale mass spectrometry imaging data in chemically augmented histology*. Scientific Report, 2018. **8**.
158. Ifa, D.R., et al., *Forensic applications of ambient ionization mass spectrometry*. Anal Bioanal Chem, 2009. **394**(8): p. 1995-2008.

159. Huang, G., et al., *Rapid screening of anabolic steroids in urine by reactive desorption electrospray ionization*. Anal Chem, 2007. **79**(21): p. 8327-32.
160. Cotte-Rodriguez, I. and R.G. Cooks, *Non-proximate detection of explosives and chemical warfare agent simulants by desorption electrospray ionization mass spectrometry*. Chem Commun (Camb), 2006(28): p. 2968-70.
161. Carroll, W., et al., *Detection of volatile compounds emitted by Pseudomonas aeruginosa using selected ion flow tube mass spectrometry*. Pediatr Pulmonol, 2005. **39**(5): p. 452-6.
162. Goeminne, P.C., et al., *Detection of Pseudomonas aeruginosa in sputum headspace through volatile organic compound analysis*. Respiratory Research, 2012. **13**(87).
163. Scott-Thomas, A.J., et al., *2-Aminoacetophenone as a potential breath biomarker for Pseudomonas aeruginosa in the cystic fibrosis lung*. BMC Pulm Med, 2010. **10**: p. 56.
164. Chambers, S.T., A. Scott-Thomas, and M. Epton, *Developments in novel breath tests for bacterial and fungal pulmonary infection*. Curr Opin Pulm Med, 2012. **18**(3): p. 228-32.
165. Gilchrist, F.J., et al., *Exhaled breath hydrogen cyanide as a marker of early Pseudomonas aeruginosa infection in children with cystic fibrosis*. ERJ Open Research, 2015. **1**(2): p. 00044-2015-44-2015.
166. Chen, W., et al., *Detection of hydrogen cyanide from oral anaerobes by cavity ring down spectroscopy*. Sci Rep, 2016. **6**: p. 22577.
167. Spanel, P., et al., *Do linear logistic model analyses of volatile biomarkers in exhaled breath of cystic fibrosis patients reliably indicate Pseudomonas aeruginosa infection?* J Breath Res, 2016. **10**(3): p. 036013.
168. Pabary, R., et al., *Does mass spectrometric breath analysis detect Pseudomonas aeruginosa in cystic fibrosis?* Eur Respir J, 2016.
169. Chen, H., et al., *Rapid in vivo fingerprinting of nonvolatile compounds in breath by extractive electrospray ionization quadrupole time-of-flight mass spectrometry*. Angew Chem Int Ed Engl, 2007. **46**(4): p. 580-3.
170. Martinez-Lozano, P., et al., *Secondary electrospray ionization (SESI) of ambient vapors for explosive detection at concentrations below parts per trillion*. J Am Soc Mass Spectrom, 2008. **20**(2): p. 287-94.
171. Huang, M.Z., et al., *Ambient ionization mass spectrometry: a tutorial*. Anal Chim Acta, 2011. **702**(1): p. 1-15.
172. Berchtold, C., et al., *Real-time monitoring of exhaled drugs by mass spectrometry*. Mass Spectrom Rev, 2014. **33**(5): p. 394-413.
173. Martinez-Lozano, P. and J.F. de la Mora, *On-line detection of human skin vapors*. J Am Soc Mass Spectrom, 2009. **20**(6): p. 1060-3.
174. Gu, H., N. Xu, and H. Chen, *Direct analysis of biological samples using extractive electrospray ionization mass spectrometry (EESI-MS)*. Anal Bioanal Chem, 2012. **403**(8): p. 2145-53.
175. Bean, H.D., J. Zhu, and J.E. Hill, *Characterizing bacterial volatiles using secondary electrospray ionization mass spectrometry (SESI-MS)*. J Vis Exp, 2011(52).
176. Zhu, J., et al., *Secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting of multiple bacterial lung pathogens, a mouse model study*. J Appl Physiol (1985), 2013. **114**(11): p. 1544-9.
177. Zhu, J., et al., *Robust detection of P. aeruginosa and S. aureus acute lung infections by secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting: from initial infection to clearance*. J Breath Res, 2013. **7**(3): p. 037106.
178. Pereira, J., et al., *Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview*. Metabolites, 2015. **5**(1): p. 3-55.
179. Fend, R., et al., *Prospects for clinical application of electronic-nose technology to early detection of Mycobacterium tuberculosis in culture and sputum*. J Clin Microbiol, 2006. **44**(6): p. 2039-45.

180. Scarlata, S., et al., *Exhaled breath analysis by electronic nose in respiratory diseases*. Expert Rev Mol Diagn, 2015. **15**(7): p. 933-56.
181. de Vries, R., et al., *Integration of electronic nose technology with spirometry: validation of a new approach for clinical breath analysis*. Eur Respir J, 2014. **44**(58).
182. Arasaradnam, R.P., et al., *Review article: next generation diagnostic modalities in gastroenterology--gas phase volatile compound biomarker detection*. Aliment Pharmacol Ther, 2014. **39**(8): p. 780-9.
183. Arasaradnam, R.P., et al., *Differentiating coeliac disease from irritable bowel syndrome by urinary volatile organic compound analysis--a pilot study*. PLoS One, 2014. **9**(10): p. e107312.
184. Covington, J.A., et al., *The detection of patients at risk of gastrointestinal toxicity during pelvic radiotherapy by electronic nose and FAIMS: a pilot study*. Sensors (Basel), 2012. **12**(10): p. 13002-18.
185. Montuschi, P., et al., *Metabolomic Analysis by Nuclear Magnetic Resonance Spectroscopy as a New Approach to Understanding Inflammation and Monitoring of Pharmacological Therapy in Children and Young Adults With Cystic Fibrosis*. Front Pharmacol, 2018. **9**: p. 595.
186. Montuschi, P., et al., *Liquid chromatography-mass spectrometry measurement of leukotrienes in asthma and other respiratory diseases*. J Chromatogr B Analyt Technol Biomed Life Sci, 2014. **964**: p. 12-25.
187. Effros, R.M., *Exhaled Breath Condensate, Delusion or Dilution?* Chest, 2010. **138**(3): p. 471-72.
188. Veselkov, K., et al., *Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer*. 2013.
189. Veselkov, K.A., et al., *Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery*. Anal Chem, 2011. **83**(15): p. 5864-72.
190. Breiman, L., *Random Forests*. 2001.
191. Wolpert, D.H. and W.G. Mcready, *No Free Lunch Theorems for Optimization*. IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, 1997. **1**(1).
192. Xia, J., et al., *MetaboAnalyst 3.0--making metabolomics more meaningful*. Nucleic Acids Res, 2015. **43**(W1): p. W251-7.
193. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.
194. Strittmatter, N., R.A. During, and Z. Takats, *Analysis of wastewater samples by direct combination of thin-film microextraction and desorption electrospray ionization mass spectrometry*. Analyst, 2012. **137**(17): p. 4037-44.
195. Forbes, T.P., T.M. Brewer, and G. Gillen, *Desorption electro-flow focusing ionization of explosives and narcotics for ambient pressure mass spectrometry*. Analyst, 2013. **138**(19): p. 5665-73.
196. Urwin, R. and M.C.J. Maiden, *Multi-locus sequence typing: a tool for global epidemiology*. Trends in Microbiology, 2003. **11**(10): p. 479-487.
197. Livermore, D.M., *Multiple Mechanisms of Antimicrobial Resistance in Pseudomonas aeruginosa: Our Worst Nightmare?* Antimicrobial Resistance, 2002. **34**: p. 634-640.
198. Woodford, N., J.F. Turton, and D.M. Livermore, *Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance*. FEMS Microbiol Rev, 2011. **35**(5): p. 736-55.
199. Kidd, T.J., et al., *Comparison of three molecular techniques for typing Pseudomonas aeruginosa isolates in sputum samples from patients with cystic fibrosis*. J Clin Microbiol, 2011. **49**(1): p. 263-8.
200. Li, W., D. Raoult, and P.E. Fournier, *Bacterial strain typing in the genomic era*. FEMS Microbiol Rev, 2009. **33**(5): p. 892-916.
201. Vos, P., et al., *AFLP: a new technique for DNA fingerprinting*. Nucleic Acids Res, 1995. **23**(21): p. 4407-4414.

202. Turton, J.F., et al., *Evaluation of a nine-locus variable-number tandem-repeat scheme for typing of Pseudomonas aeruginosa*. Clin Microbiol Infect, 2010. **16**(8): p. 1111-6.
203. Curran, B., et al., *Development of a multilocus sequence typing scheme for the opportunistic pathogen Pseudomonas aeruginosa*. J Clin Microbiol, 2004. **42**(12): p. 5644-9.
204. Mencacci, A., et al., *Typing of nosocomial outbreaks of Acinetobacter baumannii by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry*. J Clin Microbiol, 2013. **51**(2): p. 603-6.
205. Josten, M., et al., *Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of Staphylococcus aureus identifies mutations that allow differentiation of the main clonal lineages*. J Clin Microbiol, 2013. **51**(6): p. 1809-17.
206. Cabrolie, N., et al., *Matrix-assisted laser desorption ionization-time of flight mass spectrometry identifies Pseudomonas aeruginosa high-risk clones*. J Clin Microbiol, 2015. **53**(4): p. 1395-8.
207. Fleurbaey, F., et al., *Typing Pseudomonas aeruginosa Isolates with Ultrahigh Resolution MALDI-FTICR Mass Spectrometry*. Anal Chem, 2016. **88**(11): p. 5996-6003.
208. Jolley, K.A.M., M. C. J., *BIGSdb: Scalable analysis of bacterial genome variation at the population level*. BMC Bioinformatics, 2010. **11**(595): p. 1-11.
209. Bardin, E., et al., *Metabolic Phenotyping and Strain Characterisation of Pseudomonas aeruginosa Isolates from Cystic Fibrosis Patients using Rapid Evaporative Ionisation Mass Spectrometry*. Scientific Report, 2018. **7**.
210. Wright, L.L., et al., *Dominance of international 'high-risk clones' among metallo-beta-lactamase-producing Pseudomonas aeruginosa in the UK*. J Antimicrob Chemother, 2015. **70**(1): p. 103-10.
211. De Soya, A., et al., *Developing an international Pseudomonas aeruginosa reference panel*. Microbiologyopen, 2013. **2**(6): p. 1010-23.
212. Chugani, S., et al., *Strain-dependent diversity in the Pseudomonas aeruginosa quorum-sensing regulon*. Proc Natl Acad Sci U S A, 2012. **109**(41): p. E2823-31.
213. Foweraker, J.E., et al., *Phenotypic variability of Pseudomonas aeruginosa in sputa from patients with acute infective exacerbation of cystic fibrosis and its impact on the validity of antimicrobial susceptibility testing*. J Antimicrob Chemother, 2005. **55**(6): p. 921-7.
214. Mowat, E., et al., *Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections*. Am J Respir Crit Care Med, 2011. **183**(12): p. 1674-9.
215. Ha, D.-G., S.L. Kuchma, and G.A. O'Toole, *Plate-Based Assay for Swimming Motility in Pseudomonas aeruginosa*, in *Pseudomonas Methods and Protocols*, A. Filloux and J.-L. Ramos, Editors. 2014, Springer New York: New York, NY. p. 59-65.
216. Coffey, B.M. and G.G. Anderson, *Biofilm Formation in the 96-Well Microtiter Plate*, in *Pseudomonas Methods and Protocols*, A. Filloux and J.-L. Ramos, Editors. 2014, Springer New York: New York, NY. p. 631-641.
217. Nair, C.G., B. Ryall, and H.D. Williams, *Cyanide Measurements in Bacterial Culture and Sputum*, in *Pseudomonas Methods and Protocols*, A. Filloux and J.-L. Ramos, Editors. 2014, Springer New York: New York, NY. p. 325-336.
218. Essar, D.W., et al., *Identification and Characterization of Genes for a Second Anthranilate Synthase in Pseudomonas aeruginosa: Interchangeability of the Two Anthranilate Synthases and Evolutionary Implications*. Journal of Bacteriology, 1990. **172**(2): p. 884-890.
219. Kessler, E. and M. Safran, *Elastinolytic and Proteolytic Enzymes*, in *Pseudomonas Methods and Protocols*, A. Filloux and J.-L. Ramos, Editors. 2014, Springer New York: New York, NY. p. 135-169.
220. Shannon, P., et al., *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks*. Genome Res, 2002. **13**: p. 2498-2504.
221. Woo, T.E., et al., *Virulence Adaptations of Pseudomonas aeruginosa Isolated from Patients with Non-Cystic Fibrosis Bronchiectasis*. Microbiology, 2016.

222. Puchelle, E., O. Bajolet, and M. Abely, *Airway mucus in cystic fibrosis*. Paediatr Respir Rev, 2002. **3**: p. 115-119.
223. McKnight, S.L., *The Pseudomonas Quinolone Signal Regulates rhl Quorum Sensing in Pseudomonas aeruginosa*. 2000.
224. Rampioni, G., et al., *Unravelling the Genome-Wide Contributions of Specific 2-Alkyl-4-Quinolones and PqsE to Quorum Sensing in Pseudomonas aeruginosa*. PLoS Pathog, 2016. **12**(11): p. e1006029.
225. Benamara, H., et al., *Characterization of membrane lipidome changes in Pseudomonas aeruginosa during biofilm growth on glass wool*. PLoS One, 2014. **9**(9): p. e108478.
226. Quinn, R.A., et al., *Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome*. ISME J, 2015.
227. Lepine, F., et al., *Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (HAQs) produced by Pseudomonas aeruginosa*. J Am Soc Mass Spectrom, 2004. **15**(6): p. 862-9.
228. Bakdash, J.Z. and L.R. Marusich, *Repeated Measures Correlation*. Front Psychol, 2017. **8**: p. 456.
229. Kohler, T., et al., *Swarming of Pseudomonas aeruginosa Is Dependent on Cell-to-Cell Signaling and Requires Flagella and Pili*. Journal of Bacteriology, 2000. **182**(21): p. 5990-5996.
230. Dubern, J.F. and S.P. Diggle, *Quorum sensing by 2-alkyl-4-quinolones in Pseudomonas aeruginosa and other bacterial species*. Molecular BioSystems, 2008. **4**(9): p. 882.
231. Hodgkinson, J., et al., *Structure-activity analysis of the Pseudomonas quinolone signal molecule*. J Bacteriol, 2010. **192**(14): p. 3833-7.
232. Ha, D.G., et al., *2-Heptyl-4-quinolone, a precursor of the Pseudomonas quinolone signal molecule, modulates swarming motility in Pseudomonas aeruginosa*. J Bacteriol, 2011. **193**(23): p. 6770-80.
233. Jiricny, N., et al., *Loss of social behaviours in populations of Pseudomonas aeruginosa infecting lungs of patients with cystic fibrosis*. PLoS One, 2014. **9**(1): p. e83124.
234. Barr, H.L., et al., *Diagnostic and prognostic significance of systemic alkyl quinolones for P. aeruginosa in cystic fibrosis: A longitudinal study*. J Cyst Fibros, 2017. **16**(2): p. 230-238.
235. Ayubi, E., J. Hallajzadeh, and S. Safiri, *Comment on data sparsity - Diagnostic and prognostic significance of systemic alkyl quinolones for P. aeruginosa in cystic fibrosis: A longitudinal study*. J Cyst Fibros, 2017. **16**(6): p. e19-e20.
236. Rubin, B.K., *Physiology of airway mucus clearance*. Respiratory Care, 2002. **47**(7): p. 761-768.
237. Richardson, M., *The physiology of mucus and sputum production in the respiratory system*. Respiratory care, 2003. **99**(23): p. 63-64.
238. Bhagirath, A.Y., et al., *Cystic fibrosis lung environment and Pseudomonas aeruginosa infection*. BMC Pulm Med, 2016. **16**(174).
239. Samet, J.M. and P.-W. Chen, *The Role of Airway Mucus in Pulmonary Toxicology*. Environmental Health Perspectives, 1994. **102**: p. 89-103.
240. Barth and T. Pitt, *The high amino-acid content of sputum from cystic fibrosis patients promotes growth of autotrophic pseudomonas aeruginosa*. J Med Microbiol, 1996. **45**: p. 110-119.
241. Sahu, S. and W.S. Lynn, *Lipid composition of sputum from patients with asthma and patients with cystic fibrosis*. Inflammation, 1978. **3**(1): p. 27-36.
242. Seegmiller, A.C., *Abnormal unsaturated fatty acid metabolism in cystic fibrosis: biochemical mechanisms and clinical implications*. Int J Mol Sci, 2014. **15**(9): p. 16083-99.
243. Hamai, H., et al., *Defective CFTR increases synthesis and mass of sphingolipids that modulate membrane composition and lipid signaling*. J Lipid Res, 2009. **50**(6): p. 1101-8.
244. Ziobro, R.M., et al., *Ceramide in cystic fibrosis*. Clinical Lipidology, 2013. **8**(6): p. 681-692.
245. Willner, D., et al., *Spatial distribution of microbial communities in the cystic fibrosis lung*. ISME J, 2012. **6**(2): p. 471-4.

246. Viglio, S., et al., *MEKC of desmosine and isodesmosine in urine of chronic destructive lung disease patients*. Eur Respir J, 2000. **15**(6): p. 1039-1045.
247. Martin, S.L., et al., *Association of airway cathepsin B and S with inflammation in cystic fibrosis*. Pediatr Pulmonol, 2010. **45**(9): p. 860-8.
248. Parekh, G., et al., *Inflammatory biomarkers in urine of cystic fibrosis patients*. European Respiratory Journal, 2014. **44**(Suppl 58).
249. Laguna, T.A., et al., *Biomarkers of inflammation in infants with cystic fibrosis*. Respiratory Research, 2018. **19**(6).
250. Dickerhof, N., et al., *Oxidized glutathione and uric acid as biomarkers of early cystic fibrosis lung disease*. J Cyst Fibros, 2017. **16**(2): p. 214-221.
251. Kauppila, T.J., et al., *Rapid analysis of metabolites and drugs of abuse from urine samples by desorption electrospray ionization-mass spectrometry*. Analyst, 2007. **132**(9): p. 868-75.
252. Xu, J., et al., *Early detection of Pseudomonas aeruginosa--comparison of conventional versus molecular (PCR) detection directly from adult patients with cystic fibrosis (CF)*. Ann Clin Microbiol Antimicrob, 2004. **3**: p. 21.
253. Sampson, A.P., et al., *Leukotrienes in the sputum and urine of cystic fibrosis children*. Br J Clin Pharmacol, 1990. **30**: p. 861-869.
254. Saliba, A.M., et al., *Eicosanoid-mediated proinflammatory activity of Pseudomonas aeruginosa ExoU*. Cell Microbiol, 2005. **7**(12): p. 1811-22.
255. Zakrzewski, J.T., et al., *Detection of sputum eicosanoids in cystic fibrosis and in normal saliva by bioassay and radioimmunoassay*. Br J Clin Pharmacol, 1987. **23**: p. 19-27.
256. Grassme, H., et al., *Host defense against Pseudomonas aeruginosa requires ceramide-rich membrane rafts*. Nat Med, 2003. **9**(3): p. 322-30.
257. Karandashova, S., et al., *Neutrophil elastase correlates with increased sphingolipid content in cystic fibrosis sputum*. Pediatr Pulmonol, 2018. **53**(7): p. 872-880.
258. Yoshida, K., et al., *Bacterial Long-Chain Polyunsaturated Fatty Acids: Their Biosynthetic Genes, Functions, and Practical Use*. Mar Drugs, 2016. **14**(5).
259. Chen, H., A. Venter, and R.G. Cooks, *Extractive electrospray ionization for direct analysis of undiluted urine, milk and other complex mixtures without sample preparation*. Chem Commun (Camb), 2006(19): p. 2042-4.
260. Capelli, L., et al., *Application and Uses of Electronic Noses for Clinical Diagnosis on Urine Samples: A Review*. Sensors (Basel), 2016. **16**(10).
261. Esfahani, S., et al., *Variation in Gas and Volatile Compound Emissions from Human Urine as It Ages, Measured by an Electronic Nose*. Biosensors (Basel), 2016. **6**(1).
262. Wen, K.Y., et al., *A Cell-Free Biosensor for Detecting Quorum Sensing Molecules in P. aeruginosa-Infected Respiratory Samples*. ACS Synth Biol, 2017. **6**(12): p. 2293-2301.
263. Riethmuller, J., et al., *Therapeutic Efficacy and Safety of Amitriptyline in Patients with Cystic Fibrosis*. Cell Physiol Biochem, 2009. **24**: p. 65-72.
264. Wojewodka, G., et al., *Candidate markers associated with the probability of future pulmonary exacerbations in cystic fibrosis patients*. PLoS One, 2014. **9**(2): p. e88567.
265. Ifa, D.R., et al., *Latent Fingerprint Chemical Imaging by Mass Spectrometry*. 2008.
266. Drake, D.R., et al., *Thematic review series: skin lipids. Antimicrobial lipids at the skin surface*. J Lipid Res, 2008. **49**(1): p. 4-11.
267. De Luca, C. and G. Valacchi, *Surface lipids as multifunctional mediators of skin responses to environmental stimuli*. Mediators Inflamm, 2010. **2010**: p. 321494.
268. Katona, M., et al., *Intact skin analysis by desorption electrospray ionization mass spectrometry*. Analyst, 2011. **136**(4): p. 835-40.
269. Chen, H., A. Wortmann, and R. Zenobi, *Neutral desorption sampling coupled to extractive electrospray ionization mass spectrometry for rapid differentiation of biosamples by metabolomic fingerprinting*. J Mass Spectrom, 2007. **42**(9): p. 1123-35.

270. Girod, A., R. Ramotowski, and C. Weyermann, *Composition of fingermark residue: a qualitative and quantitative review*. Forensic Sci Int, 2012. **223**(1-3): p. 10-24.
271. Ferguson, L.S., et al., *Direct detection of peptides and small proteins in fingermarks and determination of sex by MALDI mass spectrometry profiling*. Analyst, 2012. **137**(20): p. 4686-92.
272. Dutkiewicz, E.P., et al., *Hydrogel micropatches for sampling and profiling skin metabolites*. Anal Chem, 2014. **86**(5): p. 2337-44.
273. Wolstenholme, R., et al., *Study of latent fingermarks by matrix-assisted laser desorption/ionisation mass spectrometry imaging of endogenous lipids*. Rapid Commun Mass Spectrom, 2009. **23**(19): p. 3031-9.
274. Downing, D.T., J.S. Strauss, and P.E. Pochi, *Variability in the chemical composition of human skin surface lipids*. The journal of investigative dermatology, 1969. **53**(5): p. 322-327.
275. Cunliffe, W.J. and S. Shuster, *The Rate of sebum excretion in man*. Br J Derm, 1969. **81**(697).
276. Zhou, Z. and R.N. Zare, *Personal Information from Latent Fingerprints Using Desorption Electrospray Ionization Mass Spectrometry and Machine Learning*. Anal Chem, 2017. **89**(2): p. 1369-1372.
277. Worgall, T.S., *Lipid metabolism in cystic fibrosis*. Current Opinion in Clinical Nutrition and Metabolic Care, 2009. **12**(2): p. 105-109.
278. Figueroa, V., et al., *Abnormal lipid concentrations in cystic fibrosis*. Am J Clin Nutr, 2002. **75**(6): p. 1005-11.
279. Maliniak, M.L., A.A. Stecenko, and N.A. McCarty, *A longitudinal analysis of chronic MRSA and Pseudomonas aeruginosa co-infection in cystic fibrosis: A single-center study*. J Cyst Fibros, 2016. **15**(3): p. 350-6.
280. Reece, E., et al., *Co-colonisation with Aspergillus fumigatus and Pseudomonas aeruginosa is associated with poorer health in cystic fibrosis patients: an Irish registry analysis*. BMC Pulmonary Medicine, 2017. **17**(1): p. 70.
281. Frydenlund Michelsen, C., et al., *Staphylococcus aureus Alters Growth Activity, Autolysis, and Antibiotic Tolerance in a Human Host-Adapted Pseudomonas aeruginosa Lineage*. J Bacteriol, 2014. **196**(22): p. 3903-3911.
282. Reen, F.J., et al., *Harnessing Bacterial Signals for Suppression of Biofilm Formation in the Nosocomial Fungal Pathogen Aspergillus fumigatus*. Front Microbiol, 2016. **7**: p. 2074.
283. Moree, W.J., et al., *Interkingdom metabolic transformations captured by microbial imaging mass spectrometry*. Proc Natl Acad Sci U S A, 2012. **109**(34): p. 13811-6.
284. Frydenlund Michelsen, C., et al., *Evolution of metabolic divergence in Pseudomonas aeruginosa during long-term infection facilitates a proto-cooperative interspecies interaction*. ISME J, 2016. **10**(6): p. 1323-36.
285. Watrous, J., et al., *Metabolic profiling directly from the Petri dish using nanospray desorption electrospray ionization imaging mass spectrometry*. Anal Chem, 2013. **85**(21): p. 10385-91.
286. Stasulli, N.M. and E.A. Shank, *Profiling the metabolic signals involved in chemical communication between microbes using imaging mass spectrometry*. FEMS Microbiol Rev, 2016. **40**(6): p. 807-813.
287. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole-time-of-flight mass spectrometry*. J Mass Spectrom, 2001. **36**(8): p. 849-65.
288. McCulloch, M., et al., *Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers*. Integr Cancer Ther, 2006. **5**(1): p. 30-9.
289. Sonoda, H., et al., *Colorectal cancer screening with odour material by canine scent detection*. Gut, 2011. **60**(6): p. 814-9.
290. Rudnicka, J., et al., *Is It Possible to Detect Lung Cancer by Trained Dogs? Health Problems of Civilization*, 2015. **2**: p. 19-26.
291. Hackner, K., et al., *Canine scent detection for the diagnosis of lung cancer in a screening-like situation*. J Breath Res, 2016. **10**(4): p. 046003.

292. Guerrero-Flores, H., et al., *A non-invasive tool for detecting cervical cancer odor by trained scent dogs*. BMC Cancer, 2017. **17**(1): p. 79.
293. Bomers, M.K., et al., *A detection dog to identify patients with Clostridium difficile infection during a hospital outbreak*. J Infect, 2014. **69**(5): p. 456-61.
294. Angle, C., et al., *Canine Detection of the Volatilome: A Review of Implications for Pathogen and Disease Detection*. Front Vet Sci, 2016. **3**: p. 47.
295. Koivusalo, M., et al., *Canine scent detection as a tool to distinguish meticillin-resistant Staphylococcus aureus*. J Hosp Infect, 2017. **96**(1): p. 93-95.
296. Johnston, N.V., et al., *120 Canine detection of *Pseudomonas aeruginosa* (PA) volatile organic compounds (VOC's)*. Journal of Cystic Fibrosis, 2012. **11**: p. S87.
297. Davies, J.C., et al., *"Dr Dog"; will see you now? First steps in assessing the utility of trained sniffer dogs to detect Pseudomonas aeruginosa (*Pa*) airway infection in non-expectorating CF patients*. Journal of Cystic Fibrosis, 2018. **17**: p. S72-S73.
298. Chen, T. and C. Guestrin, *XGBoost*. 2016: p. 785-794.

Appendices

Appendix 1. Culturing conditions for the isolates analysed with REIMS in Chapter 4.

Genus and species	Temperature (°C)	Atmosphere	Medium	Period (hours)
<i>Achromobacter xylosoxidans</i>	37	Aerobic	CBA	24
<i>Burkholderia cepacia</i>	37	5% CO ₂	CBA	24
<i>Burkholderia metallica</i>	37	5% CO ₂	CBA	24
<i>Pseudomonas aeruginosa</i>	37	Aerobic	CBA/LB	24
<i>Pseudomonas alcaligenes</i>	37	Aerobic	CBA	24
<i>Pseudomonas koreensis</i>	37	Aerobic	CBA	24
<i>Pseudomonas monteilii</i>	37	Aerobic	CBA	24
<i>Pseudomonas oryzae</i>	37	Aerobic	CBA	24
<i>Pseudomonas putida</i>	37	Aerobic	CBA	24
<i>Pseudomonas stutzeri</i>	37	Aerobic	CBA	24
<i>Serratia marcescens</i>	37	5% CO ₂	CBA	24
<i>Staphylococcus aureus</i>	37	5% CO ₂	CBA	24
<i>Stenotrophomonas maltophilia</i>	37	Aerobic	CBA	24

Appendix 2. Origins of *Pseudomonas* isolates other than *aeruginosa* species analysed in Chapter 4. Obtained from Charing Cross Hospital microbiological bank (London, UK).

Isolate number	Genus and Species	Isolated from:
P-01	<i>Pseudomonas stutzeri</i>	NOSE
P-02	<i>Pseudomonas monteilii</i>	URINE
P-03	<i>Pseudomonas monteilii</i>	URINE
P-04	<i>Pseudomonas monteilii</i>	BLOOD CULTURE
P-05	<i>Pseudomonas monteilii</i>	HIGH VAGINAL
P-06	<i>Pseudomonas koreensis</i>	RECTAL
P-07	<i>Pseudomonas oryzae</i>	WOUND
P-08	<i>Pseudomonas alcaligenes</i>	DISABLED TOILET
P-09	<i>Pseudomonas alcaligenes</i>	BIG YELLOW
P-10	<i>Pseudomonas oryzae</i>	BLOOD CULTURE
P-11	<i>Pseudomonas stutzeri</i>	BLOOD CULTURE
P-12	<i>Pseudomonas putida</i>	BLOOD CULTURE
P-13	<i>Pseudomonas mosselii</i>	BLOOD CULTURE
P-14	<i>Pseudomonas putida</i> Group	BLOOD CULTURE
P-15	<i>Pseudomonas oryzae</i>	BLOOD CULTURE
P-16	<i>Pseudomonas putida</i>	RESP - CF

Appendix 3. Identification of significant molecules based on MS/MS spectral data acquired with direct REIMS analysis on *P. aeruginosa* colonies.

The method did not include preliminary chromatographic separation; parent ions may therefore represent a mix of molecules. The four most intense fragment ions are given together with their parent ion exact mass. Intensities are given related to the most intense fragment ion. Unattributed fragment ions were included in the table; isotope ions were excluded (reproduced under the Creative Commons Attribution license from Ref. [209]).

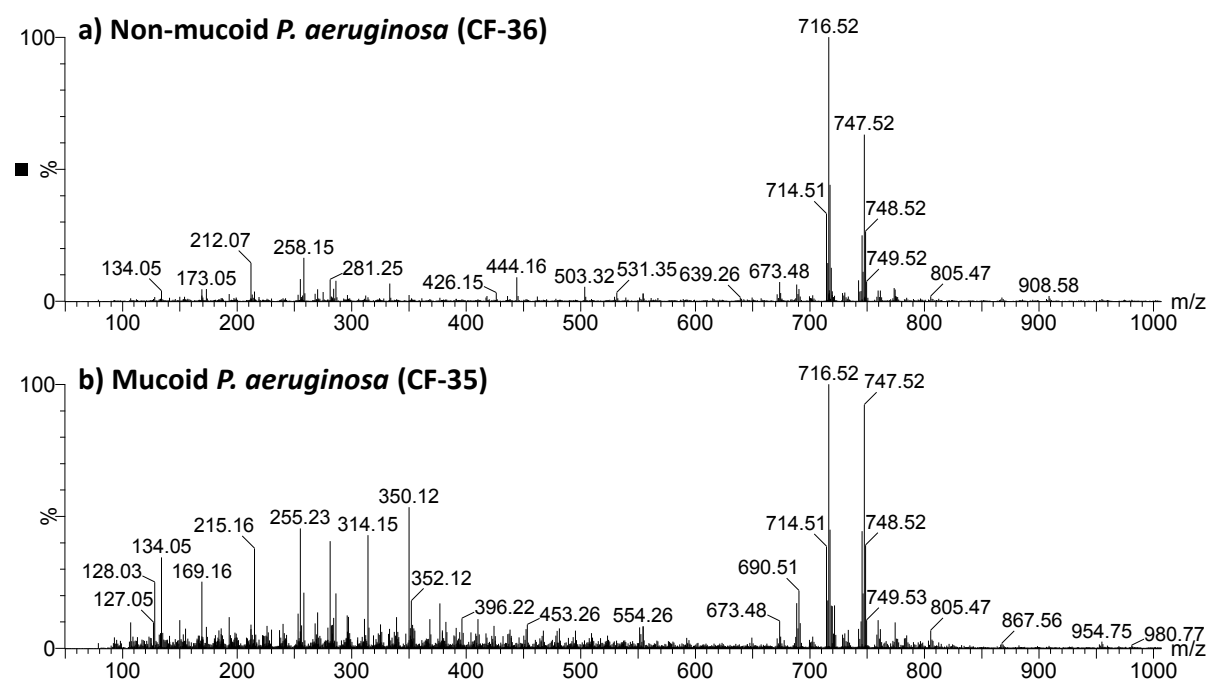
Alkyl-quinolones	Theoretical <i>m/z</i> [M-H] ⁻	CID	Resolution	Four main fragments <i>m/z</i> (intensity relative to the most abundant fragment)				Parent ion [M-H] ⁻	Mass deviation (ppm)
C7:db-HHQ	240.1394	26	12	157.1 (100%)	170.1 (62%)	182.1 (38%)	143.0 (33%)	240.1357 (37%)	15
HHQ	242.1550	28	12	157.1 (100%)	170.1 (30%)	143.0 (12%)	115.0 (7%)	242.1521 (30%)	12
C7:db-PQS C7:db-HQNO	256.1343	29	12	157.1 (100%)	170.1 (45%)	186.1 (41%)	184.0 (39%)	256.1312 (32%)	12
PQS HQNO	258.1500	29	12	173.0 (100%)	170.1 (45%)	186.1 (28%)	159.0 (15%)	258.1472 (34%)	11
C9:db-NHQ	268.1707	30	18	157.1 (100%)	182.1 (89%)	196.1 (52%)	143.0 (19%)	268.1678 (33%)	11
NHQ	270.1863	30	18	157.1 (100%)	170.1 (44%)	143.0 (16%)	184.1 (11%)	270.1834 (35%)	11
C9:db-PQS C9:db-NQNO	284.1656	25	18	196.1 (100%)	144.0 (37%)	173.0 (29%)	158.1 (26%)	284.1628 (21%)	10
C9-PQS C9-NQNO	286.1813	25	18	170.1 (100%)	173.0 (47%)	144.0 (29%)	186.1 (15%)	286.1784 (35%)	10
C11:db-UHQ	296.2020	29	18	158.1 (100%)	157.1 (95%)	170.1 (43%)	258.2 (27%)	296.1992 (26%)	9
UHQ	298.2176	32	18	157.1 (100%)	170.1 (56%)	158.1 (16%)	143.0 (16%)	298.2136 (37%)	14
C11:db-PQS C11:db-UQNO	312.1969	23	18	170.1 (100%)	144.0 (33%)	173.0 (23%)	294.2 (18%)	312.1941 (40%)	9
C11-PQS C11-UQNO	314.2126	28	18	170.1 (100%)	144.0 (35%)	173.0 (26%)	296.0 (15%)	314.2092 (38%)	11

Rhamnolipids	Theoretical <i>m/z</i> [M-H] ⁻	CID	Resolution	Four main fragments <i>m/z</i> (intensity relative to the most abundant fragment)				Parent ion [M-H] ⁻	Mass deviation (ppm)
Rha-C10	333.1919	13	18	169.1 (100%)	297.2 (7%)	163.1 (6%)	101.0 (6%)	333.1886 (28%)	10
Rha-C12	361.2232	14	18	197.2 (100%)	325.2 (21%)	343.3 (8%)	151.1 (7%)	361.2212 (29%)	5
Rha-Rha-C10	479.2498	22	15-18	205.1 (100%)	143.1 (55%)	169.1 (50%)	282.3 (36%)	479.2483 (29%)	3
Rha-C10-C10	503.3226	14	18	333.2 (100%)	169.1 (94%)	339.3 (17%)	163.1 (7%)	503.3209 (35%)	3
Rha-Rha-C12	507.2811	22	15-18	205.1 (100%)	197.2 (63%)	143.1 (54%)	131.1 (36%)	507.2790 (46%)	4
Rha-C10-C10-CH3 Rha-C11-C9-CH3	517.3382	16	18	169.1 (100%)	333.2 (89%)	183.1 (66%)	347.2 (48%)	517.3353 (39%)	6
Rha-C10-C12:1	529.3382	15	18	333.2 (100%)	169.1 (79%)	151.2 (15%)	365.3 (9%)	529.3360 (31%)	4
Rha-C10-C12	531.3539	15	18	169.1 (100%)	333.2 (99%)	197.2 (48%)	361.2 (37%)	531.3518 (37%)	4
Rha-C10-C14 Rha-C12-C12	559.3852	16	18	197.2 (100%)	361.2 (68%)	523.4 (21%)	225.2 (12%)	559.3815 (31%)	7
Rha-Rha-C10-C10	649.3805	19	15-18	479.3 (100%)	169.1 (34%)	205.1 (20%)	393.4 (15%)	649.3790 (33%)	2
Rha-Rha-C10-C10-CH3 Rha-Rha-C11-C9-CH3	663.3961	22	15-18	479.3 (100%)	493.4 (54%)	205.1 (46%)	169.1 (45%)	663.3970 (78%)	-1
Rha-Rha-C10-C12:1	675.3961	20	15-18	479.3 (100%)	393.4 (36%)	169.1 (26%)	205.1 (21%)	675.3952 (37%)	1
Rha-Rha-C10-C12	677.4118	22	15-18	479.3 (100%)	169.1 (41%)	205.1 (34%)	507.4 (33%)	677.4097 (20%)	3
Rha-Rha-C12-C12	705.4431	20	15-18	507.4 (4%)	-	-	-	705.4482 (100%)	-7

Phospholipids	Theoretical m/z [M-H] ⁻	CID	Resolution	Four main fragments m/z (intensity relative to the most abundant fragment)				Parent ion [M-H] ⁻	Mass deviation (ppm)
LPG(18:1)	509.2885	20	18	281.3 (100%)	153.0 (8%)	227.1 (4%)	-	509.2871 (34%)	3
PA(16:1/18:1)	671.4657	22	18	281.3 (100%)	253.2 (41%)	417.3 (40%)	435.3 (21%)	671.4642 (37%)	2
PA(16:0/18:1)	673.4814	22	18	281.3 (100%)	255.2 (62%)	417.3 (37%)	153.0 (22%)	673.4791 (34%)	3
PE(16:1/16:1) PE(18:1/14:1)	686.4766	22	18	253.2 (100%)	281.3 (22%)	255.2 (17%)	225.2 (13%)	686.4785 (29%)	3
PA(16:0/19:1) PA(18:1/17:0)	687.497	22	18	255.2 (100%)	281.3 (59%)	295.3 (38%)	253.2 (38%)	687.4922 (44%)	7
PE(16:1/16:0) PE(18:1/14:0)	688.4923	22	18	253.2 (100%)	255.2 (49%)	227.2 (27%)	281.3 (23%)	688.4911 (43%)	2
PE(16:0/16:0)	690.5079	22	18	255.2 (100%)	-	-	-	690.5034 (11%)	7
PE(18:1/16:0) -NH ₃	699.4970	24	18	255.2 (100%)	281.3 (63%)	-	-	699.4946 (44%)	3
PE(18:1/16:1)	714.5079	24	18	253.3 (100%)	281.3 (42%)	478.4 (9%)	460.4 (2%)	714.5057 (34%)	3
PE(18:1/16:0)	716.5236	24	10-18	255.3 (100%)	281.3 (68%)	478.4 (8%)	-	716.5201 (47%)	5
PG(18:1/14:0) PG(16:1/16:0)	719.4868	28	10-18	253.3 (100%)	255.3 (56%)	281.3 (21%)	227.2 (17%)	719.4871 (48%)	0
PE(17:1/18:1) PE(16:1/19:1)	728.5236	25	10-18	267.3 (100%)	281.3 (43%)	478.4 (7%)	253.3 (7%)	728.5198 (39%)	5
PE(16:0/19:1) PE(17:0/18:1) PE(17:1/18:0)	730.5392	26	10-18	295.3 (100%)	255.3 (36%)	281.3 (23%)	267.3 (23%)	730.5308 (44%)	11
PE(18:1/18:1) PE(19:1/17:1)	742.5392	25	10-18	281.3 (100%)	267.3 (12%)	295.3 (8%)	478.4 (7%)	742.5331 (36%)	8
PG(18:1/16:1)	745.5025	28	10-18	253.3 (100%)	281.3 (45%)	509.4 (7%)	153.0 (4%)	745.4998 (49%)	4
PG(18:1/16:0)	747.5181	30	10-18	255.3 (100%)	281.3 (76%)	253.3 (10%)	153.0 (6%)	747.5148 (42%)	4
PG(18:1/17:1) PG(19:1/16:1)	759.5181	35	15-17	253.2 (100%)	267.2 (98%)	281.3 (49%)	295.3 (43%)	759.5161 (53%)	3
PG(19:1/16:0) PG(18:1/17:0) PG(18:0/17:1)	761.5338	35	15-17	295.3 (100%)	255.2 (88%)	281.3 (7%)	267.2 (6%)	761.5291 (50%)	6
PG(19:0 / 16:0)	763.5283	38	15-17	255.2 (100%)	295.3 (77%)	297.3 (46%)	257.3 (32%)	763.5297 (37%)	2
PG(18:1/18:1) PG(19:1/17:1)	773.5338	36	15-17	281.3 (100%)	267.2 (59%)	295.3 (27%)	153.0 (5%)	773.5305 (40%)	4
PC(16:0/18:1) +Cl	794.5342	33	15-17	779.7 (100%)	281.3 (51%)	255.3 (46%)	253.2 (39%)	794.5327 (30%)	2
PC(16:0/18-OH) +Cl	812.5447	37	15-17	281.3 (100%)	255.3 (84%)	253.2 (38%)	417.3 (31%)	812.549 (43%)	5
NAPE(52:2)	980.7689	43	15-17	255.3 (100%)	281.3 (67%)	295.3 (35%)	267.3 (31%)	980.7604 (43%)	9

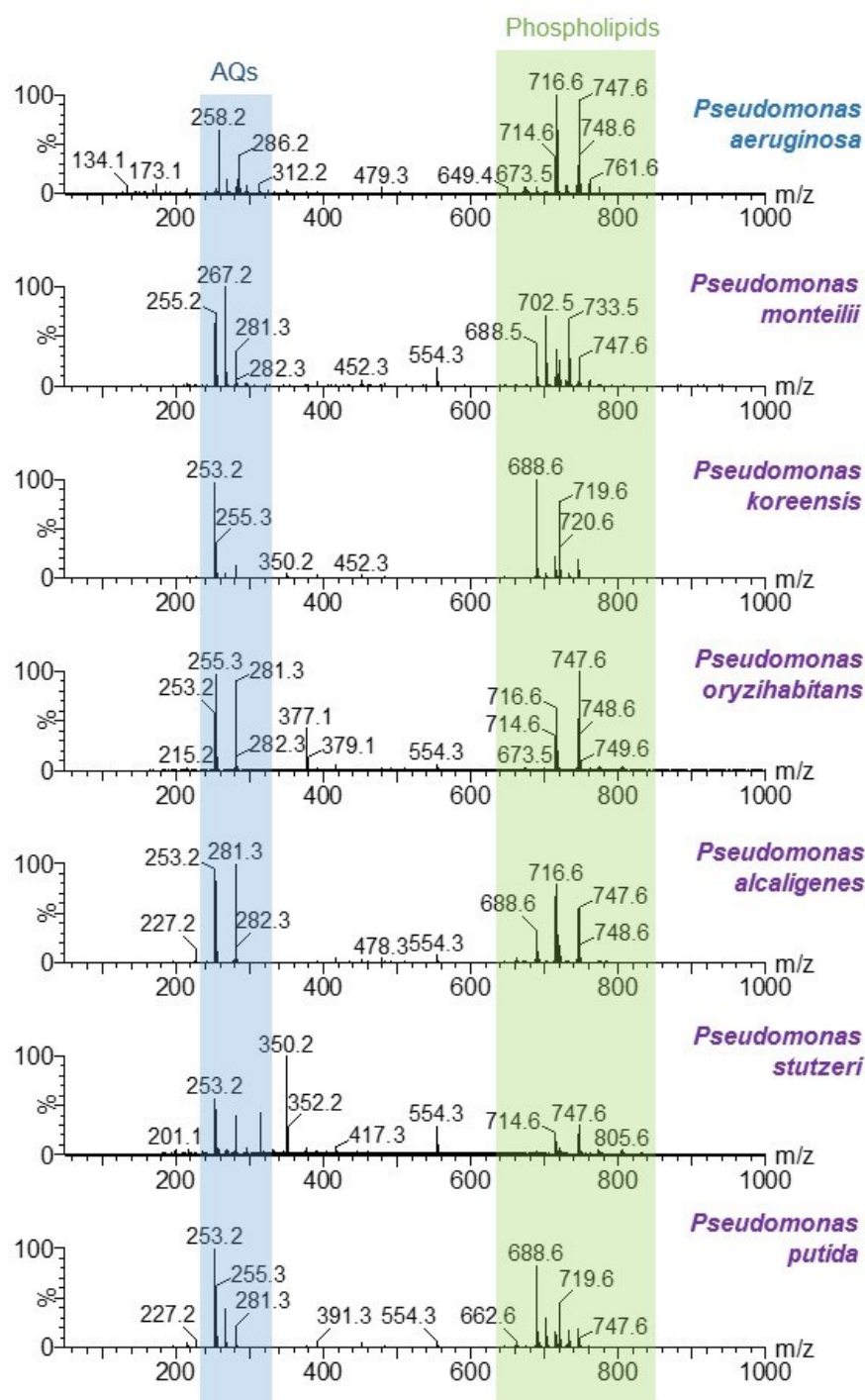
Appendix 4. Comparison of mass spectra obtained using REIMS in negative ionisation mode on (a) non-mucoid and (b) mucoid isolates of *P. aeruginosa*.

Spectra are shown after background removal and mass drift correction using the LeuEnk ion ($m/z = 554.26$).



Appendix 5. Examples of mass spectra obtained from various *Pseudomonas* species using REIMS in negative ionisation mode.

Spectra are shown after background removal and mass drift correction using the LeuEnk ion ($m/z = 554.26$). Phospholipid and AQ spectral regions are highlighted in green and blue respectively.



Appendix 6. Analysis dates for the 10 replicates of the 45 *P. aeruginosa* isolated from CF lungs analysed in Chapter 4.

Isolates which grew mucoid more than once are indicated (reproduced under the Creative Commons Attribution license from Ref. [209]).

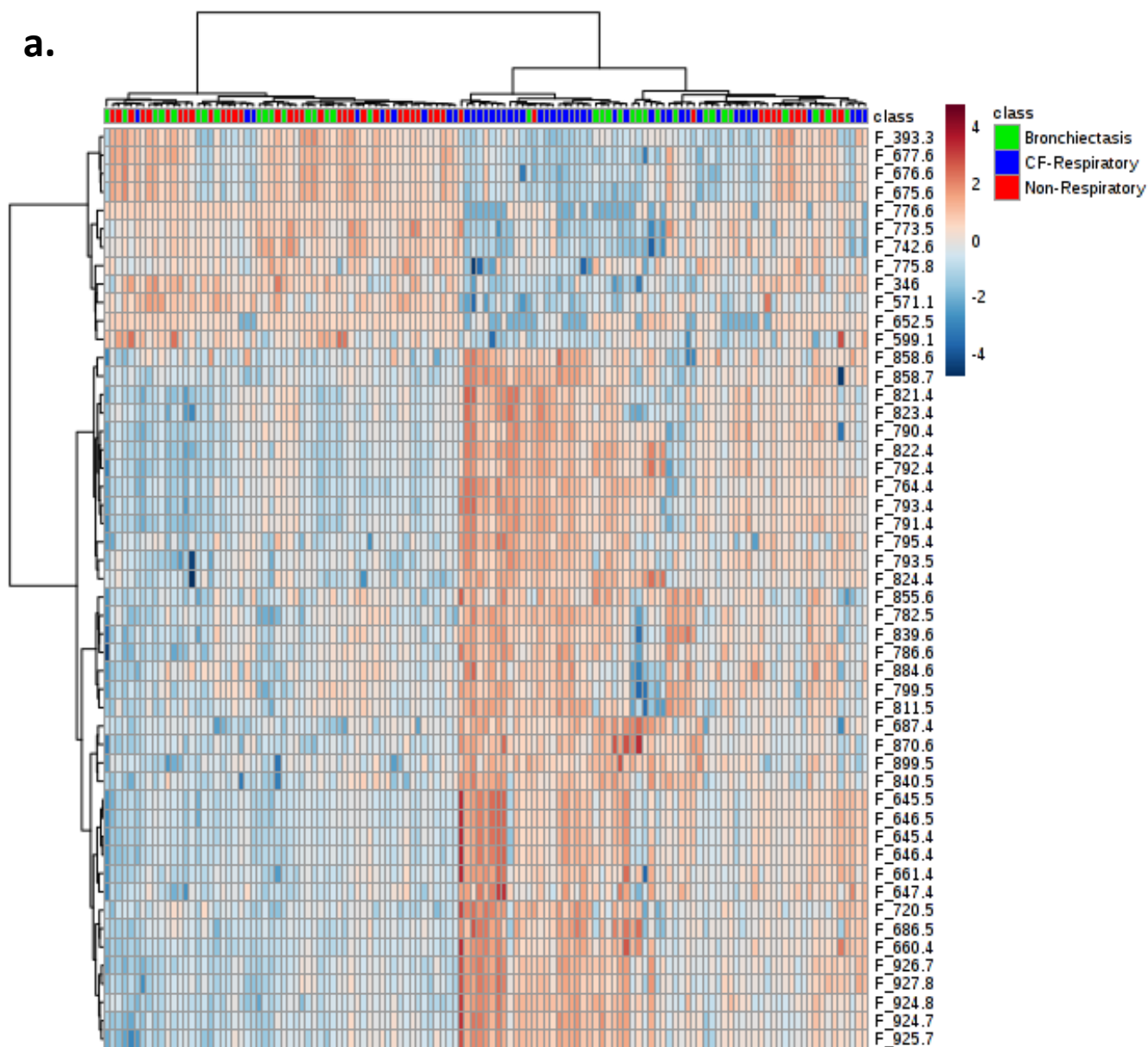
Isolates	Mucoid phenotype	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	TOTAL
CF-01		-	-	1	-	2	1	1	2	2	1	10
CF-02	X	-	-	1	-	2	1	1	2	2	1	10
CF-03		-	-	1	-	2	1	1	2	2	1	10
CF-04		-	-	1	-	2	1	1	2	2	1	10
CF-05		-	-	1	-	2	1	1	2	2	1	10
CF-06		-	-	1	-	2	1	1	2	2	1	10
CF-07		-	-	1	-	2	1	1	2	2	1	10
CF-08	X	-	-	1	-	2	1	1	2	2	1	10
CF-09		-	-	1	-	2	1	1	2	2	1	10
CF-10		-	-	1	-	2	1	1	2	2	1	10
CF-11		-	-	1	-	2	1	1	2	2	1	10
CF-12	X	-	-	1	-	2	1	1	2	2	1	10
CF-13		-	-	1	-	2	1	1	2	2	1	10
CF-14		-	-	1	-	2	1	1	2	2	1	10
CF-15		-	-	1	-	2	1	1	2	2	1	10
CF-16		-	-	1	-	2	1	1	2	2	1	10
CF-17		-	-	1	-	2	1	1	2	2	1	10
CF-18		-	-	1	-	2	1	1	2	2	1	10
CF-19		-	-	1	-	2	1	1	2	2	1	10
CF-20		-	-	1	-	2	1	1	2	2	1	10
CF-21		-	-	1	-	2	1	1	2	2	1	10
CF-22		-	-	1	-	2	1	1	2	2	1	10
CF-23		-	-	1	-	2	1	1	2	2	1	10
CF-24		-	-	1	-	2	1	1	2	2	1	10
CF-25		-	-	1	-	2	1	1	2	2	1	10
CF-26		-	-	1	-	2	1	1	2	2	1	10
CF-27		-	-	1	-	2	1	1	2	2	1	10
CF-28	X	-	-	1	-	2	1	1	2	2	1	10
CF-29		-	-	1	-	2	1	1	2	2	1	10
CF-30		-	-	1	-	2	1	1	2	2	1	10
CF-31		-	-	1	-	2	1	1	2	2	1	10
CF-32		-	-	1	-	2	1	1	2	2	1	10
CF-33		3	2	-	-	-	1	1	1	1	1	10
CF-34		3	2	-	-	-	1	1	1	1	1	10
CF-35	X	-	3	-	-	-	1	1	2	2	1	10
CF-36		-	3	-	-	-	1	1	2	2	1	10
CF-37		-	-	1	-	2	1	1	2	2	1	10
CF-38		-	-	1	-	1	2	1	2	2	1	10
CF-39	X	-	-	1	-	2	1	1	2	2	1	10
CF-40		-	-	1	-	2	1	1	2	2	1	10
CF-41		-	-	1	-	2	1	1	2	2	1	10
CF-42		-	-	1	-	2	1	1	2	2	1	10
CF-43		-	-	1	-	2	1	1	2	2	1	10
CF-44		-	-	1	-	2	1	1	2	2	1	10
CF-45		-	-	-	1	2	1	1	2	2	1	10

Appendix 7. List and characteristics of colonies analysed in Chapter 5.

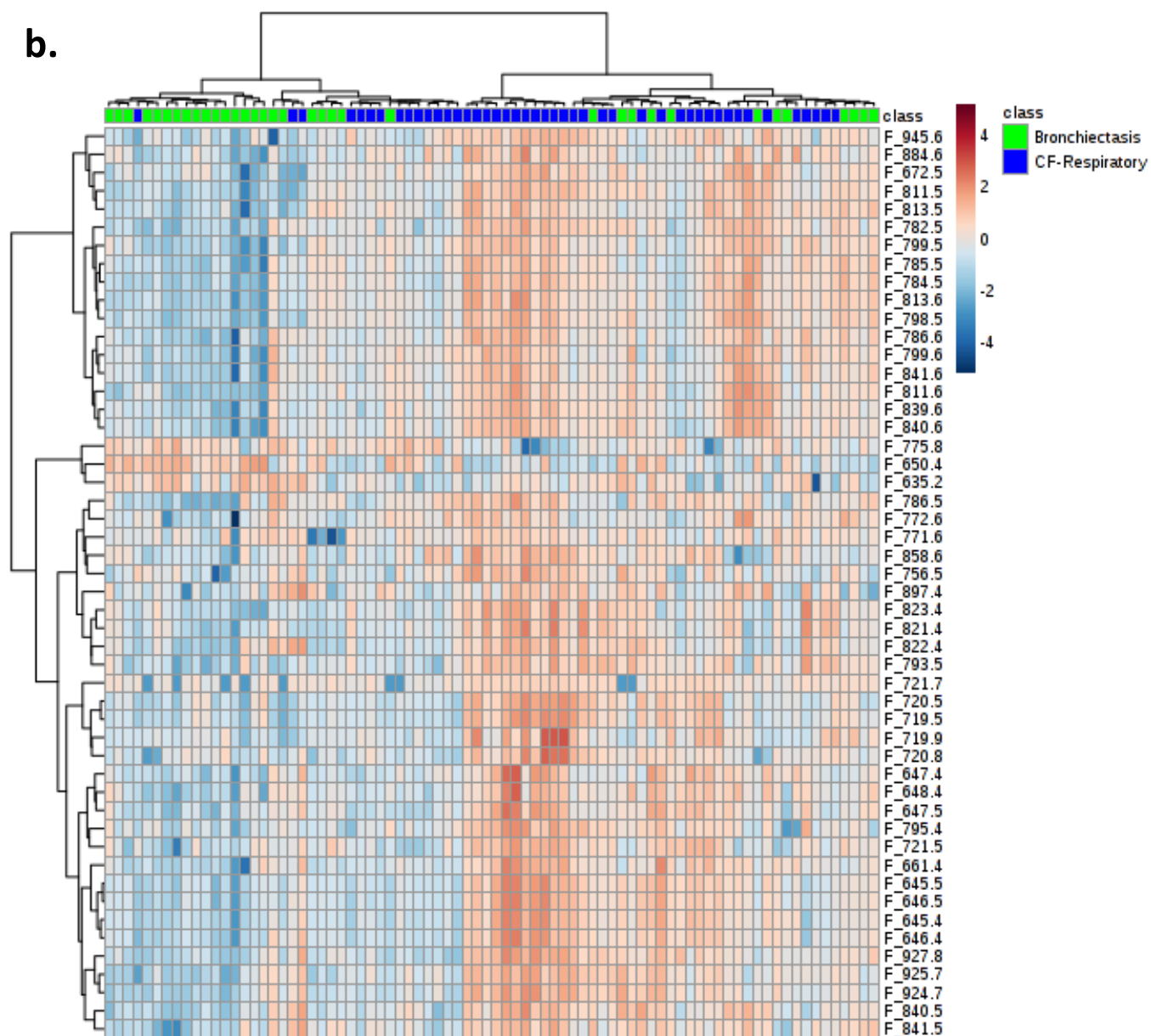
Patient	Infection stage	Number of colonies	VNTR-typed (12 colonies)
E01	Early	26	
E02	Early	13	
E03	Early	36	
E04	Early	24	
E05	Early	60	
E06	Early	36	
E07	Early	24	
C01	Chronic	24	
C02	Chronic	12	X
C03	Chronic	24	
C04	Chronic	24	
C05	Chronic	24	
C06	Chronic	24	X
C07	Chronic	24	X
C08	Chronic	12	2 types (2 vs 10 colonies)
C09	Chronic	24	X
C10	Chronic	12	X
C11	Chronic	24	
C12	Chronic	24	
C13	Chronic	24	X (LES)
C14	Chronic	24	
C15	Chronic	24	X
C16	Chronic	24	X
C17	Chronic	24	
C18	Chronic	24	
PA14	Control	32	
PAQ1	Control	32	
LB	Control	32	

Appendix 8. HCA and associated heatmaps representing (a) non-respiratory, bronchiectasis and CF isolates and (b) bronchiectasis and CF isolates only; using data collected in Chapter 5.

HCA were built using the normalised intensities of the top 50 most differentiating mass spectral bins, based on ANOVA tests, within the 50 to 1200 m/z range (reproduced under the Creative Commons Attribution license from Ref. [209]).



b.



Appendix 9. Random Forest cross-validation results for the classification of *P. aeruginosa* colonies collected from 17 chronic patients (Cxx), seven early infections (Exx) and LB, PAO1 and PA14 as controls, using REIMS spectral data collected in Chapter 5.

Leave-one-colony-out cross-validation; the model was built using all the bins within $m/z = 50-1200$. Overall classification per patient was 93% accurate.

Patient	Precision	Recall	F1-score	Number of colonies
C01	1	0.83	0.91	24
C02	0.80	1	0.89	12
C03	0.92	1	0.96	24
C04	1	0.88	0.93	24
C05	1	0.92	0.96	24
C06	0.96	1	0.98	24
C07	0.83	0.83	0.83	24
C08	0.75	0.25	0.38	12
C09	1	0.96	0.98	24
C10	1	1	1	12
C11	1	1	1	24
C12	0.64	0.96	0.77	24
C13	1	1	1	24
C14	1	0.88	0.93	24
C15	1	0.96	0.98	24
C16	0.88	0.92	0.90	24
C18	0.89	1	0.94	24
E01	0.92	0.88	0.90	26
E02	1	1	1	13
E03	0.97	1	0.99	36
E04	0.95	0.88	0.91	24
E05	1	1	1	60
E06	1	0.97	0.99	36
E07	0.92	1	0.96	24
LB	0.84	1	0.91	32
PA14	0.93	0.88	0.90	32
PAO1	0.97	0.88	0.92	32
Average / Total	0.94	0.93	0.93	687

Appendix 10. Random Forest cross-validation results for the classification of colonies collected from seven early infected and 17 chronic patients in Chapter 5.

The model used all the bins within $m/z = 50-1200$ and a leave-one-patient-out cross-validation. Overall classification was 65% accurate.

Infection stage	Precision	Recall	F1-score	Number of colonies
Chronic	0.66	0.91	0.76	372
Early	0.57	0.20	0.29	219
Average / Total	0.62	0.65	0.59	591

Appendix 11. Identified features amongst the 150 most intense m/z bins, with ROC curve univariate analysis results for the differentiation of early and chronic infection in Chapter 5.

Features with AUC > 65% presented as boxplots in the main document are highlighted in yellow.

m/z bin	Metabolite	ROC curve analysis	Higher levels in:	
		AUC	Early	Chronic
240.15	C7:db-HHQ	57.7		+
242.15	HHQ	71.3		+
253.25	FA(16:1)	60.3		+
255.25	FA(16)	60.3		+
256.15	C7:db-PQS	61.8		+
258.28	PQS	51.3	+	
268.20	C9:db-NHQ	57.0		+
270.30	NHQ	71.8		+
284.15	C9:db-PQS	71.6	+	
296.20	C11:db-UHQ	61.6		+
297.25	FA(19)	63.0		+
298.33	UHQ	69.7		+
311.25	FA(20)	57.2		+
312.25	C11:db-PQS	54.9	+	
314.15	C11-PQS	69.7	+	
333.20	Rha-C10	63.3	+	
479.30	Rha-Rha-C10	72.8	+	
503.25	Rha-C10-C10	66.5	+	
507.25	Rha-Rha-C12	71.3	+	
529.35	Rha-C10-C12:1	65.3	+	
649.69	Rha-Rha-C10-C10	74.9	+	
675.40	Rha-Rha-C10-C12:1	74.2	+	
677.48	Rha-Rha-C10-C12	68.9	+	
714.58	PE(34:2)	65.5	+	
716.58	PE(34:1)	64.8	+	
745.58	PG(34:2)	70.7	+	
761.58	PG(35:1)	52.9	+	

Appendix 12. Compiled list of the features used to build both the early and chronic networks in Chapter 5.

Includes identification when known (tentative identification are in inverted comas) and associated betweenness centralities.

Bin.	Tentative identification	Early	Chronic	Bin.	Tentative identification	Early	Chronic
157.050	Bin.157.05	1.03E-03	3.47E-03	308.150	Bin.308.15	9.80E-02	2.77E-02
158.050	Bin.158.05	4.24E-03	1.58E-02	310.150	Bin.310.15	3.48E-03	1.14E-03
159.050	Bin.159.05	5.35E-03	2.83E-02	310.250	Bin.310.25	1.79E-02	3.40E-03
165.100	"arabinonic acid"	4.16E-02	6.43E-02	311.250	Bin.311.25	6.10E-03	5.80E-04
169.200	"FA(10:1)"	1.53E-02	2.77E-02	312.250	C11:db-PQS/C11:db-UQNO	6.24E-03	1.45E-03
169.975	Bin.169.98	5.35E-03	3.80E-02	314.150	Bin.314.15	2.01E-03	0.00E+00
173.050	"QSM fragment 173.05"	4.62E-04	2.38E-03	314.250	C11-PQS/C11-UQNO	6.59E-04	3.39E-02
184.050	"QSM fragment 184.05"	3.37E-04	1.34E-01	322.150	"C9-PQS/C9-NQNO+CI"	4.41E-03	3.54E-03
185.975	"QSM fragment 185.98"	4.62E-04	4.24E-03	322.250	Bin.322.25	7.72E-03	1.63E-03
192.975	Bin.192.98	1.24E-03	1.58E-02	323.150	Bin.323.15	1.46E-04	1.56E-02
197.150	"FA(12:1)"	5.86E-02	5.10E-02	324.250	Bin.324.25	8.46E-03	1.64E-02
199.050	"C9:db-PQS/NQNO fragment"	1.56E-03	0.00E+00	333.200	Rha-C10	6.43E-03	1.53E-02
200.050	Bin.200.05	4.62E-04	2.38E-03	335.200	Bin.335.20	3.10E-03	6.55E-02
212.050	"QSM fragment 212.05"	1.98E-02	1.93E-02	340.150	Bin.340.15	1.70E-04	0.00E+00
214.050	Bin.214.05	7.59E-03	3.73E-04	350.150	Bin.350.15	0.00E+00	1.56E-02
215.338	"hydroxy lauric acid"	5.76E-04	3.33E-03	352.300	Bin.352.30	1.53E-03	2.13E-02
216.200	Bin.216.20	5.76E-04	3.33E-03	353.250	Bin.353.25	5.39E-03	8.18E-03
226.150	Bin.226.15	6.90E-02	2.17E-01	354.150	Bin.354.15	6.56E-03	4.94E-02
227.100	Bin.227.10	2.56E-02	4.33E-02	354.250	Bin.354.25	7.41E-03	2.36E-02
230.050	Bin.230.05	3.25E-02	3.55E-02	355.200	Bin.355.20	8.48E-03	6.96E-03
240.150	C7:db-HHQ	6.19E-02	3.17E-02	361.250	Rha-C12	2.82E-03	6.41E-03
241.150	"FA(15:0)"	3.54E-02	2.44E-02	368.150	Bin.368.15	2.51E-03	3.98E-03
242.150	HHQ	6.42E-02	1.34E-02	368.250	Bin.368.25	4.92E-03	1.21E-02
253.250	"FA(16:1)"	1.50E-02	7.85E-03	377.100	Bin.377.10	1.38E-02	0.00E+00
255.250	"FA(16:0)"	1.13E-02	3.33E-02	382.200	Bin.382.20	4.92E-03	1.16E-02
256.150	C7:db-PQS/C7:db-HQNO	1.69E-01	2.42E-02	389.200	Bin.389.20	9.29E-04	6.72E-03
257.150	Bin.257.15	2.76E-04	1.82E-02	391.250	"cyclic/LPA(16:0)"	4.26E-03	3.32E-02
259.150	Bin.259.15	2.30E-03	5.69E-03	394.200	Bin.394.20	4.00E-03	2.97E-02
268.200	C9:db-NHQ	1.12E-02	1.61E-03	396.250	Bin.396.25	2.26E-03	3.53E-02
270.300	NHQ	9.21E-03	3.47E-03	410.250	"LPE(13:0)/LPC(10:0)"	1.85E-03	2.51E-02
271.050	Bin.271.05	3.33E-03	0.00E+00	417.250	"cyclic/LPA(18:1)"	2.93E-03	8.04E-03
271.150	Bin.271.15	9.21E-03	1.01E-03	424.300	"LPE(14:0)/LPC(11:0)"	5.05E-03	3.24E-02
271.325	Bin.271.33	1.06E-01	0.00E+00	425.200	Bin.425.20	0.00E+00	3.64E-04
272.150	Bin.272.15	2.71E-03	2.38E-03	436.250	"LPE(15:1)"	4.91E-03	1.12E-02
277.175	Bin.277.18	2.53E-02	6.04E-02	437.250	"LPA(18:0)"	2.72E-02	2.42E-02
278.150	"HHQ+CI"	6.63E-02	3.32E-02	438.250	"LPE(15:0)/LPC(12:0)"	6.99E-02	2.65E-02
279.050	"FA(18:2)"	1.19E-01	1.02E-02	439.200	Bin.439.20	5.08E-03	1.60E-02
281.325	"FA(18:1)"	0.00E+00	0.00E+00	439.350	Bin.439.35	4.73E-03	2.56E-03
282.150	Bin.282.15	4.36E-03	0.00E+00	450.300	"LPE(16:1)"	3.19E-03	5.59E-03
282.250	Bin.282.25	4.04E-02	2.07E-02	452.300	"LPE(16:0)"	2.94E-03	1.37E-02
283.150	Bin.283.15	1.35E-02	1.56E-02	467.300	"LPG(P-16:0)"	1.16E-03	1.39E-02
284.150	C9:db-PQS/C9:db-NQNO	4.41E-03	2.79E-02	479.300	Rha-Rha-C10	1.41E-02	6.24E-04
284.250	Bin.284.25	1.11E-03	2.35E-02	481.250	Bin.481.25	7.90E-02	7.45E-02
286.156	C9-PQS/C9-NQNO	4.41E-03	1.54E-03	495.225	Bin.495.23	2.94E-03	3.30E-02
286.400	Bin.286.40	7.45E-06	0.00E+00	503.250	Rha-C10-C10	1.97E-03	0.00E+00
287.300	Bin.287.30	4.41E-03	1.54E-03	503.500	Bin.503.50	2.35E-03	1.08E-03
294.150	"PQS/HQNO+CI"	4.94E-02	1.55E-02	504.350	Bin.504.35	6.65E-03	1.08E-03
295.300	"FA(19:1)"	1.34E-02	0.00E+00	507.250	Rha-Rha-C12	1.48E-02	6.48E-03
296.200	C11:db-UHQ	4.24E-03	9.88E-03	509.300	LPG(18:1)	2.28E-03	1.48E-03
298.325	UHQ	9.27E-03	1.84E-02	529.350	Rha-C10-C12:1	8.69E-03	3.95E-04
306.150	"NHQ+CI"	6.54E-02	5.38E-02	531.450	Rha-C10-C12	8.69E-03	6.41E-03

Bin.	Tentative identification	Early	Chronic
539.300	Bin.539.30	3.32E-02	6.94E-03
556.350	Bin.556.35	4.73E-03	7.39E-04
574.350	Bin.574.35	5.33E-03	1.37E-02
576.350	Bin.576.35	2.52E-01	3.37E-02
577.350	Bin.577.35	2.35E-02	8.76E-03
592.350	Bin.592.35	7.89E-02	3.25E-03
649.694	Rha-Rha-C10-C10	0.00E+00	6.24E-04
650.400	Bin.650.40	0.00E+00	6.24E-04
673.500	PA(34:1)	2.62E-03	2.47E-03
675.400	Rha-Rha-C10-C12:1	1.41E-02	1.08E-03
677.475	Rha-Rha-C10-C12	7.17E-03	1.08E-03
687.500	PA(35:1)	3.39E-03	4.76E-02
688.500	PE(32:1)	3.16E-03	7.17E-04
690.575	PE(32:0)	8.04E-05	3.10E-04
714.575	PE(34:2)	8.42E-04	5.61E-03
716.575	PE(34:1)	1.86E-03	1.21E-03
719.500	PG(32:1)	9.19E-04	1.42E-03
730.575	PE(35:1)	3.64E-03	0.00E+00
745.575	PG(34:2)	2.84E-03	9.07E-03
746.500	Bin.746.50	2.84E-03	1.04E-02
747.575	PG(34:1)	4.39E-03	1.49E-03
759.575	PG(35:2)	4.97E-03	3.14E-02
761.575	PG(35:1)	3.55E-03	3.09E-02
765.500	"PI(30:1)"	5.81E-03	1.03E-02
766.550	Bin.766.55	2.11E-03	3.93E-03
780.600	Bin.780.60	3.28E-03	1.93E-03

Appendix 13. List of sputum samples and number of replicates acquired with REIMS in Chapter 6, with infection status as verified by the detection of the genes 16S rRNA, *exoA* and *oprL*.

Sputum sample ID	<i>P. aeruginosa</i> infection status*	Number of REIMS replicates
4961	unsure	3
4962	unsure	3
4964	likely positive	7
4965	negative	3
4966	positive	6
4967	negative	3
4968	negative	13
4970	positive	10
4971	negative	7
4972	positive	7
4973	likely positive	10
4974	positive	5
4975	positive	10
4977	positive	3
4978	positive	11
4979	positive	10
4980	unsure	5
4981	positive	7
4982	positive	3
4983	positive	7
4984	likely positive	7
4985	positive	7
4986	positive	2
4987	negative	3
4988	positive	5
4989	positive	7
4991	negative	3
4992	negative	5
4993	likely positive	7
4994	negative	7
4995	positive	9
4996	positive	2
4997	unsure	7
4998	likely positive	5
4999	positive	7
5000	positive	7
5001	positive	2
5002	positive	7
5003	negative	7
5004	negative	7
5005	negative	7
5006	negative	7
5007	negative	9
5008	likely positive	7
5009	positive	7
5010	negative	7
5012	unsure	7
5015	negative	7
5016	negative	7
5017	negative	7
5018	negative	7

*Negative: 16S rRNA present, *exoA* and *oprL* absent; Positive: 16S rRNA, *exoA* and *oprL* present; Likely positive: 16S rRNA present, *exoA* or *oprL* present; Unsure: 16S rRNA present, signal unclear for *exoA* and/or *oprL*.

Appendix 14. Random Forest cross-validation results for the classification of replicates obtained from the REIMS analysis of 36 CF sputum samples in Chapter 6.

Leave-one-out cross-validation applied on a REIMS dataset including five replicates per sample. Overall prediction was 57% accurate.

Sputum sample	Precision	Recall	F1-score	Number of REIMS replicates	<i>P. aeruginosa</i> infection status*
4964	1	0.8	0.89	5	likely positive
4968	0.5	0.6	0.55	5	negative
4970	1	1	1	5	positive
4971	0.5	0.6	0.55	5	negative
4972	0.62	1	0.77	5	positive
4973	0.67	0.8	0.73	5	likely positive
4975	0.67	0.8	0.73	5	positive
4978	0	0	0	5	positive
4979	0.6	0.6	0.6	5	positive
4981	0	0	0	5	positive
4983	0	0	0	5	positive
4984	0.67	0.4	0.5	5	likely positive
4985	0.3	0.6	0.4	5	positive
4988	0.8	0.8	0.8	5	positive
4989	0.62	1	0.77	5	positive
4993	0.6	0.6	0.6	5	likely positive
4994	0	0	0	5	negative
4995	0.67	0.8	0.73	5	positive
4997	0.62	1	0.77	5	unsure
4998	0	0	0	5	likely positive
4999	0.62	1	0.77	5	positive
5000	0.83	1	0.91	5	positive
5002	0	0	0	5	positive
5003	0.43	0.6	0.5	5	negative
5004	0.11	0.2	0.14	5	negative
5005	0.5	0.4	0.44	5	negative
5006	0.8	0.8	0.8	5	negative
5007	1	1	1	5	negative
5008	0	0	0	5	likely positive
5009	0.25	0.2	0.22	5	positive
5010	1	1	1	5	negative
5012	0.62	1	0.77	5	unsure
5015	0.4	0.4	0.4	5	negative
5016	0	0	0	5	negative
5017	1	0.6	0.75	5	negative
5018	1	1	1	5	negative
Average / Total	0.51	0.57	0.53	180	

*Negative: 16S rRNA present, *exoA* and *oprL* absent; Positive: 16S rRNA, *exoA* and *oprL* present; Likely positive: 16S rRNA present, *exoA* or *oprL* present; Unsure: 16S rRNA present, signal unclear for *exoA* and/or *oprL*.

Appendix 15. List of the 150 most intense bins detected in sputum samples by REIMS in Chapter 6, with AUC values obtained from ROC curve univariate analysis associated to the separation of patients infected and non-infected by *P. aeruginosa*.

Name ↕	AUC ↕	Name ↕	AUC ↕
Bin.184.05	0.77146	Bin.1008.75	0.66919
Bin.240.05	0.74495	Bin.400.15	0.66162
Bin.215.05	0.7298	Bin.339.15	0.66162
Bin.898.55	0.70833	Bin.217.05	0.65783
Bin.896.55	0.70707	Bin.402.15	0.65657
Bin.151.05	0.70076	Bin.730.55	0.64773
Bin.897.55	0.69949	Bin.1009.75	0.64773
Bin.1006.65	0.69444	Bin.899.55	0.64646
Bin.166.05	0.69318	Bin.1010.75	0.64394
Bin.384.15	0.69192	Bin.290.05	0.64141
Bin.546.15	0.68434	Bin.576.25	0.64141
Bin.238.05	0.68182	Bin.1007.75	0.64141
Bin.718.55	0.67929	Bin.220.05	0.63889
Bin.199.05	0.67424	Bin.577.25	0.63636
Bin.719.55	0.67298	Bin.313.25	0.63131
Name ↕	AUC ↕	Name ↕	AUC ↕
Bin.717.55	0.62879	Bin.321.15	0.6048
Bin.353.25	0.62753	Bin.325.25	0.59975
Bin.716.55	0.62753	Bin.197.15	0.59848
Bin.293.15	0.61616	Bin.846.65	0.59722
Bin.303.25	0.61364	Bin.187.05	0.59596
Bin.297.15	0.61237	Bin.203.05	0.59596
Bin.165.05	0.61111	Bin.878.55	0.5947
Bin.163.05	0.60985	Bin.728.55	0.59343
Bin.337.15	0.60985	Bin.279.25	0.59217
Bin.1006.75	0.60985	Bin.826.65	0.59091
Bin.243.15	0.60859	Bin.167.05	0.58838
Bin.272.05	0.60859	Bin.181.05	0.58838
Bin.280.25	0.60732	Bin.529.45	0.58838
Bin.319.15	0.60732	Bin.221.15	0.5846
Bin.240.15	0.60606	Bin.276.15	0.58333

Name ↕	AUC ↕
Bin.750.55	0.58333
Bin.687.55	0.58081
Bin.365.35	0.58081
Bin.828.65	0.57828
Bin.241.15	0.57576
Bin.592.25	0.57449
Bin.847.65	0.57449
Bin.848.65	0.57323
Bin.282.25	0.57323
Bin.556.25	0.56944
Bin.844.65	0.56944
Bin.279.15	0.56818
Bin.657.55	0.56692
Bin.684.55	0.56692
Bin.685.55	0.56692

Name ↕	AUC ↕
Bin.297.25	0.55303
Bin.700.55	0.55303
Bin.636.25	0.55177
Bin.396.25	0.55177
Bin.222.15	0.55051
Bin.311.25	0.55051
Bin.164.05	0.54924
Bin.683.55	0.54545
Bin.263.15	0.54419
Bin.199.15	0.54293
Bin.701.55	0.54293
Bin.279.05	0.5404
Bin.227.15	0.53662
Bin.329.05	0.53662
Bin.744.55	0.53662

Name ↕	AUC ↕
Bin.845.65	0.56566
Bin.686.65	0.56439
Bin.298.25	0.56313
Bin.594.25	0.56313
Bin.554.35	0.56187
Bin.682.55	0.56187
Bin.898.65	0.56187
Bin.281.25	0.56061
Bin.305.25	0.56061
Bin.355.15	0.56061
Bin.247.15	0.55934
Bin.283.15	0.55808
Bin.595.25	0.55808
Bin.375.15	0.55682
Bin.253.25	0.55429

Name ↕	AUC ↕
Bin.436.25	0.53535
Bin.333.15	0.53409
Bin.573.45	0.53409
Bin.291.05	0.53283
Bin.259.05	0.5303
Bin.410.25	0.5303
Bin.735.55	0.52904
Bin.242.15	0.52778
Bin.271.25	0.52652
Bin.572.45	0.52652
Bin.734.55	0.52652
Bin.269.25	0.52525
Bin.277.05	0.52525
Bin.574.45	0.52525
Bin.736.55	0.52146

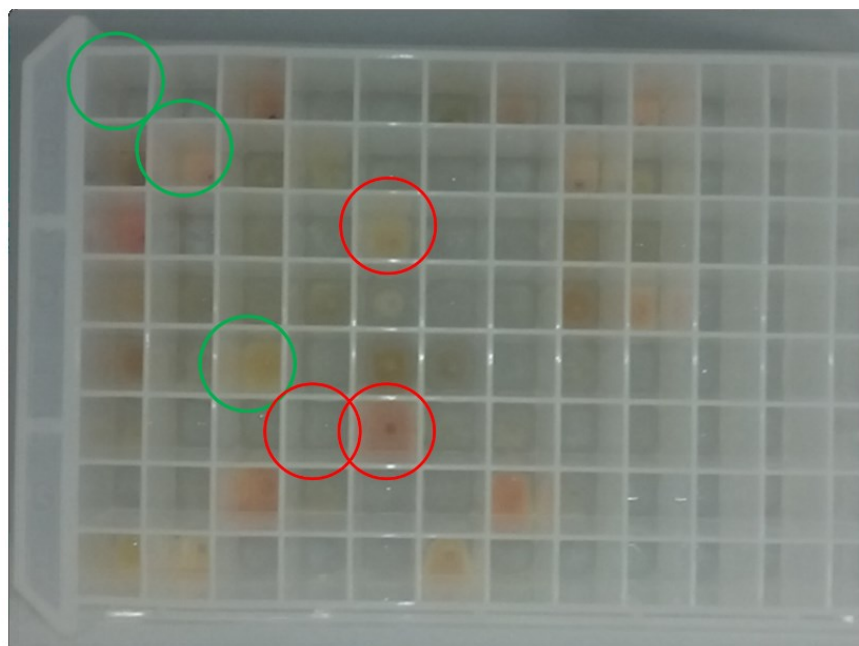
Name ↕	AUC ↕
Bin.842.65	0.52146
Bin.897.65	0.52146
Bin.325.15	0.5202
Bin.339.25	0.5202
Bin.896.65	0.5202
Bin.737.55	0.51894
Bin.307.25	0.51768
Bin.284.25	0.51515
Bin.205.05	0.51389
Bin.283.25	0.51136
Bin.312.15	0.51136
Bin.424.25	0.51136
Bin.439.25	0.5101
Bin.745.55	0.5101
Bin.170.05	0.50758

Name ↕	AUC ↕
Bin.411.25	0.50758
Bin.299.25	0.50631
Bin.311.15	0.50631
Bin.446.25	0.50631
Bin.255.25	0.50379
Bin.256.25	0.50379
Bin.340.15	0.5

Appendix 16. Picture of pelleted urine in a 96-deep well plate (a), with associated sample identification in Table (b), analysed by REIMS in Chapter 6.

Mass spectra of the circled samples are available in the main manuscript; samples positive and negative for *P. aeruginosa* infection are in red and green, respectively.

a)

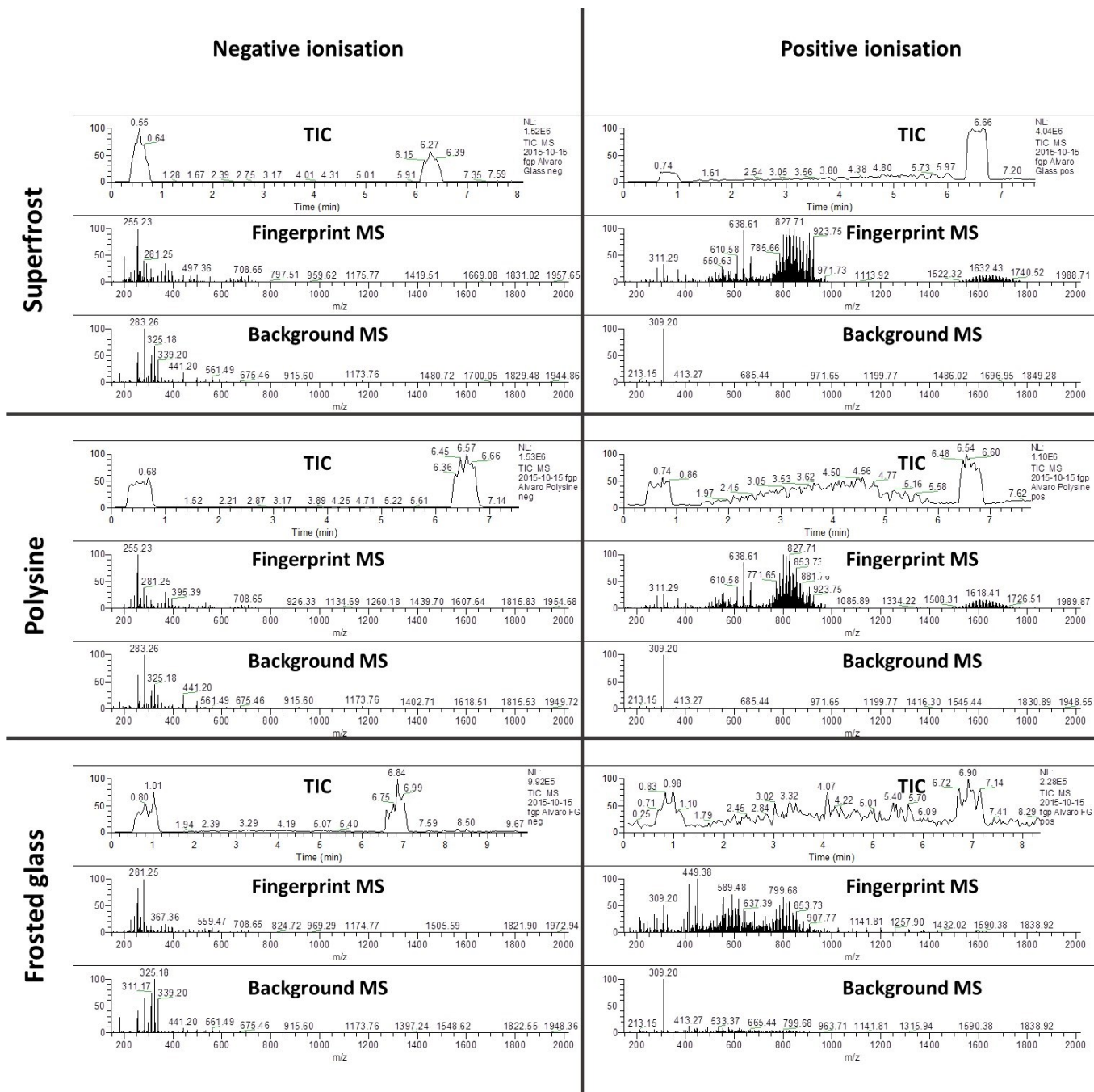


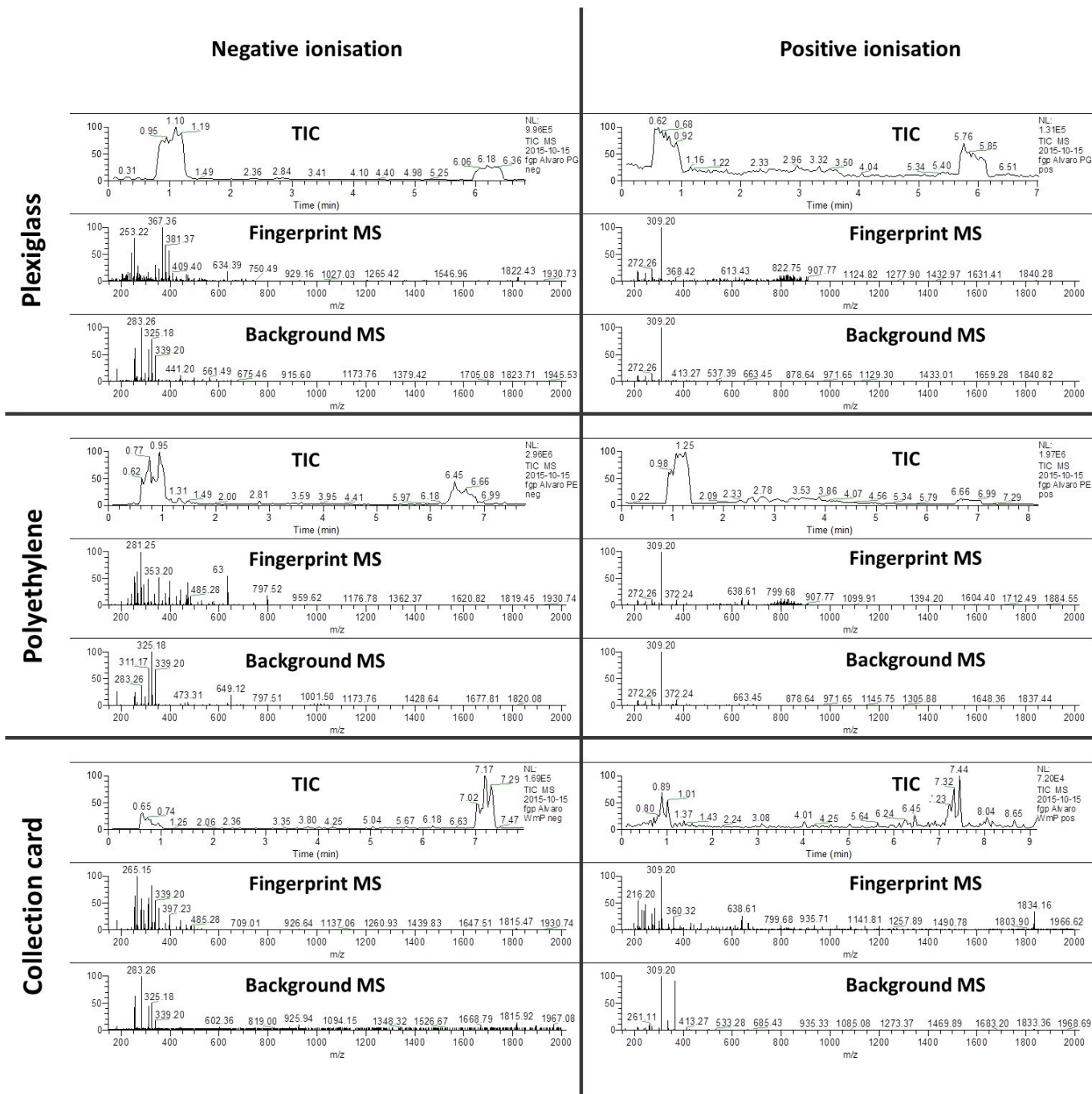
b)

	01	02	03	04	05	06	07	08	09	10	11	12
A	Ru113	BrTr103	BrTr113	Ru104	BrTr121	BrTr106	BrTr109	BrTr122	BrTr110			
B	BrTr107	Ru108	Ru118	Ru112	Ru132	BrTr124	Ru123	Ru126	Ru138			
C	BrTr115	BrTr114	Ru216	BrTr123	Ru107	Ru111	Ru103	Ru136	BrTr101			
D	BrTr112	BrTr119	Ru146	Ru144	Ru129	Ru141	BrTr100	Ru101	Ru131			
E	Ru191	BrTr105	Ru120	Ru140	Ru122	BrTr120	Ru102	BrTr111	Ru110			
F	Ru149	Ru109	Ru137	Ru128	BrTr108	Ru124	BrTr104	BrTr117	Ru145			
G	BrTr118	Ru121	Ru114	Ru143	Ru115	BrTr102	BrTr116	Ru130				
H	Ru178	Ru135	Ru127	Ru134	Ru116	Ru206	Ru133	Ru105				

Appendix 17. Total ion count (TIC) and mass spectra (MS) for the fingerprint and the background obtained with DESI-MS profiling in both ionisation modes on the six materials tested in Chapter 7.

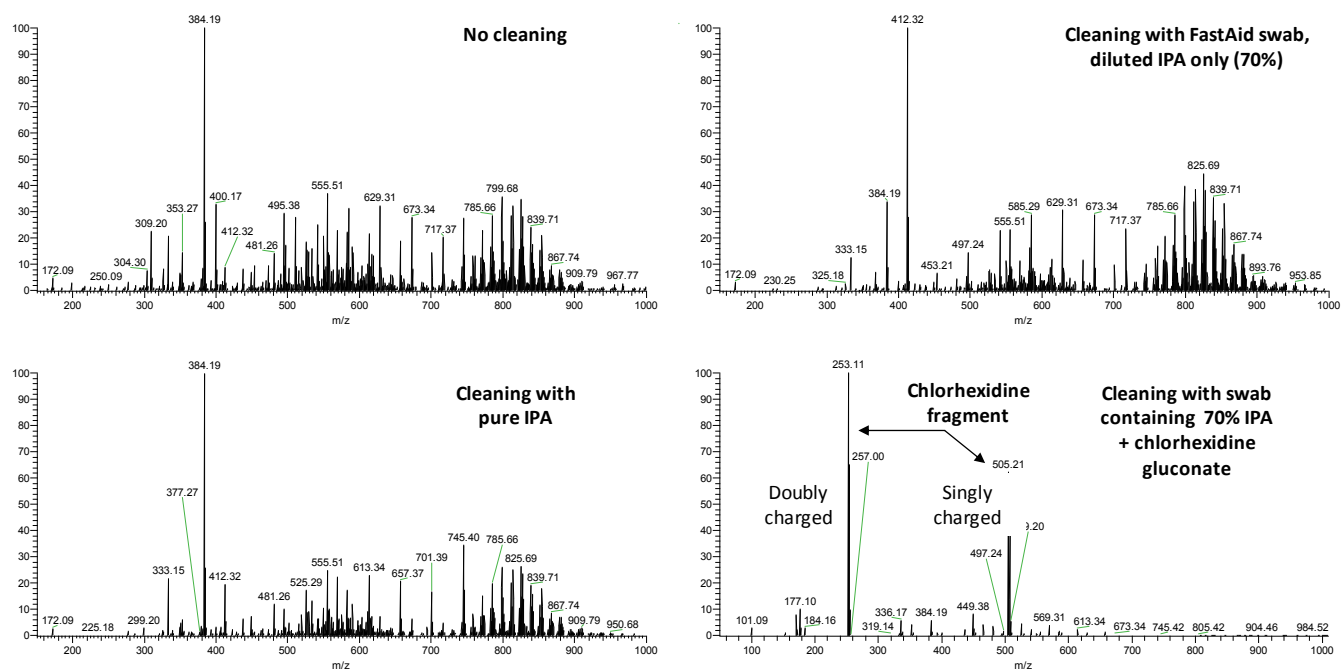
From first material to last: Superfrost Plus glass, Polysine glass, frosted glass, plexiglass, polyethylene, collection card; negative and ionisation data are shown on the left- and right-hand side, respectively. Spectra are averaged and background subtracted for the fingerprint; averaged for the background.



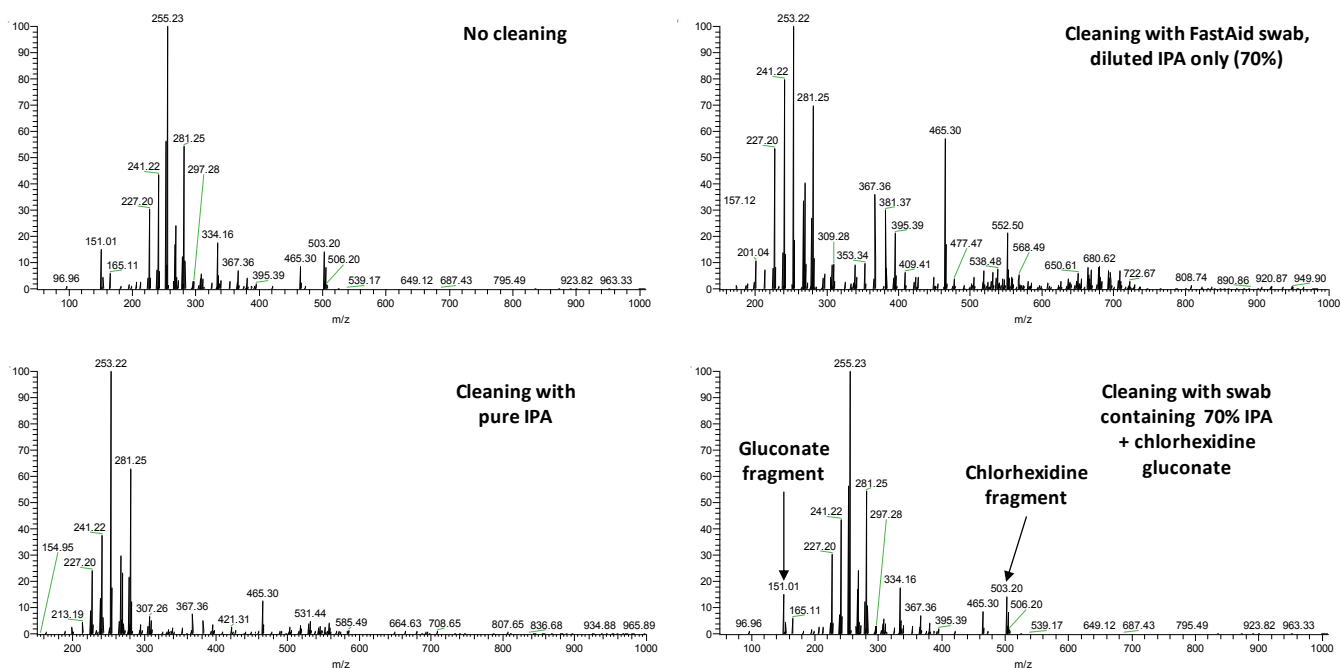


Appendix 18. Examples of mass spectra acquired by DESI-MS profiling in (a) positive and (b) negative ionisation mode, on skin secretions collected from the bridge of the nose of one donor.
Averaged, background subtracted spectra, showing signal from skin with and without IPA cleaning, and cleaning using IPA swabs with or without chlorhexidine gluconate.

a) Positive ionisation mode

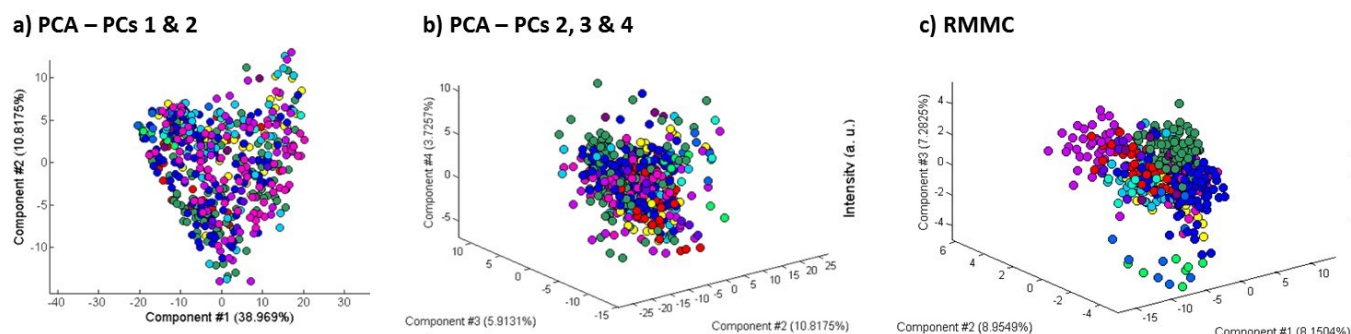


b) Negative ionisation mode



Appendix 19. Score plots showing unsupervised and supervised analysis of DESI-MS profiling data acquired in negative ionisation mode in Chapter 7 on skin secretions collected from 14 subjects.

(a) Two-dimensional PCA with the two first PCs 1 & 2; (b) Three-dimensional PCA with PCs 2, 3 & 4; (c) Supervised RMMC with the three first PCs; uses median spectra for each sample; colouring corresponds to the 14 donors' identity (built on Matlab).



Appendix 20. Confusion matrix for donor classification using DESI-MS profiling data acquired in negative ionisation mode on skin secretions in Chapter 7.

Metabolic profiles detected from skin secretions collected from the bridge of the nose; prediction uses an RMMC supervised model, 10-fold cross-validation, 5-nn classifier. Actual and predicted classes are displayed respectively on the left-hand and the bottom sides. Each line refers back to a subject in Appendix 19 (see dot colours). Frequency of prediction for each subject is represented in darkening grey with the proportion and number of samples indicated in each box. Overall classification was 78% accurate (built on Matlab).

●	72.4%, 21	0.0%, 0	3.4%, 1	0.0%, 0	3.4%, 1	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	3.4%, 1	0.0%, 0	3.4%, 1	6.9%, 2	6.9%, 2	0.0%, 0
●	0.0%, 0	70.0%, 28	7.5%, 3	0.0%, 0	12.5%, 5	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	7.5%, 3	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	2.5%, 1
●	0.9%, 1	0.0%, 0	85.8%, 91	1.9%, 2	7.5%, 8	0.0%, 0	0.0%, 0	1.9%, 2	0.9%, 1	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.9%, 1
●	0.0%, 0	0.0%, 0	0.0%, 0	42.9%, 3	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	14.3%, 1	14.3%, 1	28.6%, 2
●	2.3%, 2	1.1%, 1	2.3%, 2	0.0%, 0	81.8%, 72	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	10.2%, 9	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	2.3%, 2
●	0.0%, 0	0.0%, 0	14.3%, 1	14.3%, 1	42.9%, 3	14.3%, 1	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	14.3%, 1	0.0%, 0	0.0%, 0
●	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	100.0%, 7	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0
●	46.7%, 1	0.0%, 0	16.7%, 1	0.0%, 0	16.7%, 1	0.0%, 0	0.0%, 0	50.0%, 3	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0
●	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	5.1%, 2	0.0%, 0	0.0%, 0	0.0%, 0	92.3%, 36	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	2.6%, 1
●	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	22.2%, 2	22.2%, 2	11.1%, 1	0.0%, 0	0.0%, 0	44.4%, 4	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0
●	3.8%, 1	0.0%, 0	7.7%, 2	0.0%, 0	15.4%, 4	0.0%, 0	0.0%, 0	0.0%, 0	3.8%, 1	0.0%, 0	61.5%, 16	0.0%, 0	0.0%, 0	0.0%, 0	7.7%, 2
●	40.0%, 4	0.0%, 0	10.0%, 1	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	10.0%, 1	0.0%, 0	0.0%, 0	0.0%, 0	40.0%, 4	0.0%, 0	0.0%, 0	0.0%, 0
●	5.0%, 2	0.0%, 0	5.0%, 2	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	0.0%, 0	2.5%, 1	0.0%, 0	0.0%, 0	87.5%, 35	0.0%, 0	0.0%, 0
●	0.0%, 0	1.4%, 1	1.4%, 1	0.0%, 0	5.8%, 4	0.0%, 0	0.0%, 0	0.0%, 0	4.3%, 3	0.0%, 0	5.8%, 4	0.0%, 0	0.0%, 0	81.2%, 58	0.0%, 0

Permissions to reprint

Permissions to reprint were obtained for references [11, 22, 133, 137, 138, 183] and are attached to this document.

11. Jarad, N.A., et al., *Factors associated with reduced FEV1 in adult patients with cystic fibrosis in a relatively affluent area*. Chronic Respiratory Disease, 2005. 2: p. 133-137.



RightsLink®

Home

Create Account

Help



Title: Factors associated with reduced FEV1 in adult patients with cystic fibrosis in a relatively affluent area
Author: N A Jarad, S Higgs, T Jeffcote, et al
Publication: Chronic Respiratory Disease
Publisher: SAGE Publications
Date: 07/01/2005
Copyright © 2005, © SAGE Publications

LOGIN

If you're a **copyright.com** user, you can login to RightsLink using your copyright.com credentials.
Already a **RightsLink** user or want to learn more?

Gratis Reuse

Permission is granted at no cost for use of content in a Master's Thesis and/or Doctoral Dissertation. If you intend to distribute or sell your Master's Thesis/Doctoral Dissertation to the general public through print or website publication, please return to the previous page and select 'Republish in a Book/Journal' or 'Post on intranet/password-protected website' to complete your request.

BACK

CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

22. UK CF Trust, 2016 Registry Annual Data Report. 2016.

Bardin, Emmanuelle

From: Registry <registry@cysticfibrosis.org.uk>
Sent: 16 October 2018 15:37
To: Bardin, Emmanuelle; Registry
Subject: RE: Permission to reproduce

Dear Emmanuelle,

We hereby grant your request to re-use the figure from the UK CF Registry Annual Data Report 2016.

Please can you cite it as the 'UK CF Registry Annual Data Report 2016' and kindly let us know when you publish any document with the graph present, as we like to keep a record of the impact of our registry data.

Best wishes,

Kieran

Kieran Earlam
Registry Administrator
UK CF Registry

cysticfibrosis.org.uk
Cystic Fibrosis Trust, One Aldgate, London EC3N 1RE

N.B. Never send patient details by email to the UK CF Registry Team. If you need to reference specific data, please use the Registry CaseID.

From: Bardin, Emmanuelle <e.bardin15@imperial.ac.uk>
Sent: 15 October 2018 21:48
To: Registry <registry@cysticfibrosis.org.uk>
Subject: RE: Permission to reproduce

Dear Sir or Madam,

I am a PhD student funded by the CF Trust (SRC001), and I am writing to ask for the permission to reproduce a figure in my thesis: "1.15 Lung infections in 2016" published in the registry annual data report from 2016, page 23.

Thank you very much for your feedback,
Best regards,

Emmanuelle Bardin
PhD student
Department of Surgery & Cancer
Imperial College London

We're fighting for Life Unlimited by cystic fibrosis for everyone affected by the condition.



The contents and any attachments of this email are confidential and intended solely for the person to whom it is addressed. If you are not the intended recipient, please notify the sender and delete this email from your system immediately. Do not disclose, use, store or copy the contents in any way. Views expressed in this email are the individual sender's except where specifically stated to be the views of the Cystic Fibrosis Trust. The Cystic Fibrosis Trust hereby excludes any warranty or liability as to the quality or accuracy of the contents of this email and any attached transmitted files. Copyright in this email belongs to the Cystic Fibrosis Trust. When communicating with anyone in the Cystic

133. Takats, Z., et al., *Mass spectrometry sampling under ambient conditions with desorption electrospray ionization*. Science, 2004. **306**(5695): p. 471-3.

10/15/2018

RightLink Printable License

**THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE
TERMS AND CONDITIONS**

Oct 15, 2018

This Agreement between Emmanuelle Bardin ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	4450380754214
License date	Oct 15, 2018
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization
Licensed Content Author	Zoltan Takats, Justin M. Wiseman, Bogdan Gologan, R. Graham Cooks
Licensed Content Date	Oct 15, 2004
Licensed Content Volume	306
Licensed Content Issue	5695
Volume number	306
Issue number	5695
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1
Order reference number	
Title of your thesis / dissertation	Exploring Ambient Mass Spectrometry Capabilities for Rapid Detection and Phenotyping of <i>Pseudomonas aeruginosa</i> Infection in Cystic Fibrosis Patients
Expected completion date	May 2019
Estimated size(pages)	1
Requestor Location	Emmanuelle Bardin Imperial College London, South Kensington campus Sir Alexander Fleming Building, room 660 London, Choose a state SW7 2AZ United Kingdom Attn:
Billing Type	Invoice
Billing Address	Emmanuelle Bardin Imperial College London, South Kensington campus Sir Alexander Fleming Building, room 660 London, United Kingdom SW7 2AZ Attn: Emmanuelle Bardin
Total	0.00 GBP

Terms and Conditions

American Association for the Advancement of Science TERMS AND CONDITIONS

Regarding your request, we are pleased to grant you non-exclusive, non-transferable permission, to republish the AAAS material identified above in your work identified above, subject to the terms and conditions herein. We must be contacted for permission for any uses other than those specifically identified in your request above.

The following credit line must be printed along with the AAAS material: "From [Full Reference Citation]. Reprinted with permission from AAAS."

All required credit lines and notices must be visible any time a user accesses any part of the AAAS material and must appear on any printed copies and authorized user might make.

This permission does not apply to figures / photos / artwork or any other content or materials included in your work that are credited to non-AAAS sources. If the requested material is sourced to or references non-AAAS sources, you must obtain authorization from that source as well before using that material. You agree to hold harmless and indemnify AAAS against any claims arising from your use of any content in your work that is credited to non-AAAS sources.

If the AAAS material covered by this permission was published in Science during the years 1974 - 1994, you must also obtain permission from the author, who may grant or withhold permission, and who may or may not charge a fee if permission is granted. See original article for author's address. This condition does not apply to news articles.

The AAAS material may not be modified or altered except that figures and tables may be modified with permission from the author. Author permission for any such changes must be secured prior to your use.

Whenever possible, we ask that electronic uses of the AAAS material permitted herein include a hyperlink to the original work on AAAS's website (hyperlink may be embedded in the reference citation).

AAAS material reproduced in your work identified herein must not account for more than 30% of the total contents of that work.

AAAS must publish the full paper prior to use of any text.

AAAS material must not imply any endorsement by the American Association for the Advancement of Science.

This permission is not valid for the use of the AAAS and/or Science logos.

AAAS makes no representations or warranties as to the accuracy of any information contained in the AAAS material covered by this permission, including any warranties of merchantability or fitness for a particular purpose.

If permission fees for this use are waived, please note that AAAS reserves the right to charge for reproduction of this material in the future.

Permission is not valid unless payment is received within sixty (60) days of the issuance of this permission. If payment is not received within this time period then all rights granted herein shall be revoked and this permission will be considered null and void.

In the event of breach of any of the terms and conditions herein or any of CCC's Billing and Payment terms and conditions, all rights granted herein shall be revoked and this permission will be considered null and void.

AAAS reserves the right to terminate this permission and all rights granted herein at its discretion, for any purpose, at any time. In the event that AAAS elects to terminate this permission, you will have no further right to publish, publicly perform, publicly display, distribute or otherwise use any matter in which the AAAS content had been included, and all fees paid hereunder shall be fully refunded to you. Notification of termination will be sent to the contact information as supplied by you during the request process and termination shall be immediate upon sending the notice. Neither AAAS nor CCC shall be liable for any costs, expenses, or damages you may incur as a result of the termination of this permission, beyond the refund noted above.

This Permission may not be amended except by written document signed by both parties.

The terms above are applicable to all permissions granted for the use of AAAS material. Below you will find additional conditions that apply to your particular type of use.

FOR A THESIS OR DISSERTATION

If you are using figure(s)/table(s), permission is granted for use in print and electronic versions of your dissertation or thesis. A full text article may be used in print versions only of a dissertation or thesis.

Permission covers the distribution of your dissertation or thesis on demand by ProQuest / UMI, provided the AAAS material covered by this permission remains in situ.

If you are an Original Author on the AAAS article being reproduced, please refer to your License to Publish for rules on reproducing your paper in a dissertation or thesis.

137. Strittmatter, N., et al., *Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry*. Chem Commun (Camb), 2013. **49**(55): p. 6188-90.



Copyright
Clearance
Center

Notes: Copyright.com supplies permissions but not the copyrighted content itself.

1
PAYMENT

2
REVIEW

3
CONFIRMATION

Step 3: Order Confirmation

Thank you for your order! A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us 24 hrs/day, M-F at +1.855.239.3415 Toll Free, or write to us at info@copyright.com. This is not an invoice.

Confirmation Number: 11756646
Order Date: 10/15/2018

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

Emmanuelle Bardin
e.bardin15@imperial.ac.uk
+44 7419991197
Payment Method: n/a

Order Details

Chemical communications

Order detail ID: 71610252
Order License Id: 4450380446546
ISSN: 1364-548X
Publication Type: e-Journal
Volume:
Issue:
Start page:
Publisher: ROYAL SOCIETY OF CHEMISTRY
Author/Editor: Royal Society of Chemistry (Great Britain)

Permission Status: **Granted**

Permission type: Republish or display content
Type of use: Thesis/Dissertation

Requestor type: Academic institution

Format: Print, Electronic

Portion: image/photo

Number of images/photos requested: 1

The requesting person/organization: E. Bardin / Imperial College London

Title or numeric reference of the portion(s): Figure 1

Title of the article or chapter the portion is from: Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry

Editor of portion(s): N/A

Author of portion(s): N/A

Volume of serial or monograph: N/A

Issue, if republishing an article from a serial	55
Page range of portion	6189
Publication date of portion	2013
Rights for	Main product
Duration of use	Current edition and up to 5 years
Creation of copies for the disabled	no
With minor editing privileges	yes
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	Exploring Ambient Mass Spectrometry Capacities for Rapid Detection and Phenotyping of Pseudomonas aeruginosa Infection in Cystic Fibrosis Patients
Institution name	Imperial College London
Expected presentation date	May 2019

Note: This item will be invoiced or charged separately through CCC's RightsLink service. More info

\$ 0.00

Total order items: 1

This is not an invoice.

Order Total: 0.00 USD

138. Strittmatter, N., et al., *Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry*. Anal Chem, 2014. **86**(13): p. 6555-62.



RightsLink®

Home

Account
Info

Help



ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Characterization and Identification of Clinically Relevant Microorganisms Using Rapid Evaporative Ionization Mass Spectrometry

Author: Nicole Strittmatter, Monica Rebec, Emrys A. Jones, et al

Publication: Analytical Chemistry

Publisher: American Chemical Society

Date: Jul 1, 2014

Copyright © 2014, American Chemical Society

Logged in as:
Emmanuelle Bardin
Account #: 3001350686

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK

CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).

Comments? We would like to hear from you. E-mail us at customer@copyright.com

183. Strittmatter, N., R.A. During, and Z. Takats, *Analysis of wastewater samples by direct combination of thin-film microextraction and desorption electrospray ionization mass spectrometry*. *Analyst*, 2012. **137**(17): p. 4037-44.

11/21/2018

Copyright Clearance Center



Note: Copyright.com supplies permissions but not the copyrighted content itself.

1
PAYMENT

2
REVIEW

3
CONFIRMATION

Step 3: Order Confirmation

Thank you for your order! A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us 24 hrs/day, M-F at +1.855.239.3415 Toll Free, or write to us at info@copyright.com. This is not an invoice.

Confirmation Number: 11767101
Order Date: 11/21/2018

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

Emmanuelle Bardin
e.bardin15@imperial.ac.uk
+44 7419991197
Payment Method: n/a

Order Details

The analyst online

Order detail ID: 71672765
Order License Id: 4473570655193

ISSN: 1364-5528
Publication Type: e-Journal

Volume:
Issue:

Start page:

Publisher: Royal Society of Chemistry

Author/Editor: Society of Public Analysts (Great Britain) ; et al

Permission Status: **Granted**

Permission type: Republish or display content

Type of use: Thesis/Dissertation

Requestor type: Academic institution

Format: Electronic

Portion: image/photo

Number of images/photos requested: 1

The requesting person/organization: E. Bardin / Imperial College London

Title or numeric reference of the portion(s): Figure 1. Home-built DESI source mounted on an Orbitrap Discovery hybrid instrument

Title of the article or chapter the portion is from: Analysis of wastewater samples by direct combination of thin-film microextraction and desorption electrospray ionization mass spectrometry

Editor of portion(s): N/A

Author of portion(s): N. Strittmatter

Volume of serial or monograph	N/A
Issue, if republishing an article from a serial	17
Page range of portion	4039
Publication date of portion	2012
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	yes
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	Exploring Ambient Mass Spectrometry Capacities for Rapid Detection and Phenotyping of <i>Pseudomonas aeruginosa</i> Infection in Cystic Fibrosis Patients
Institution name	Imperial College London
Expected presentation date	May 2019

Note: This item will be invoiced or charged separately through CCC's RightsLink service. More info

\$ 0.00

Total order items: 1

This is not an invoice.

Order Total: 0.00 USD