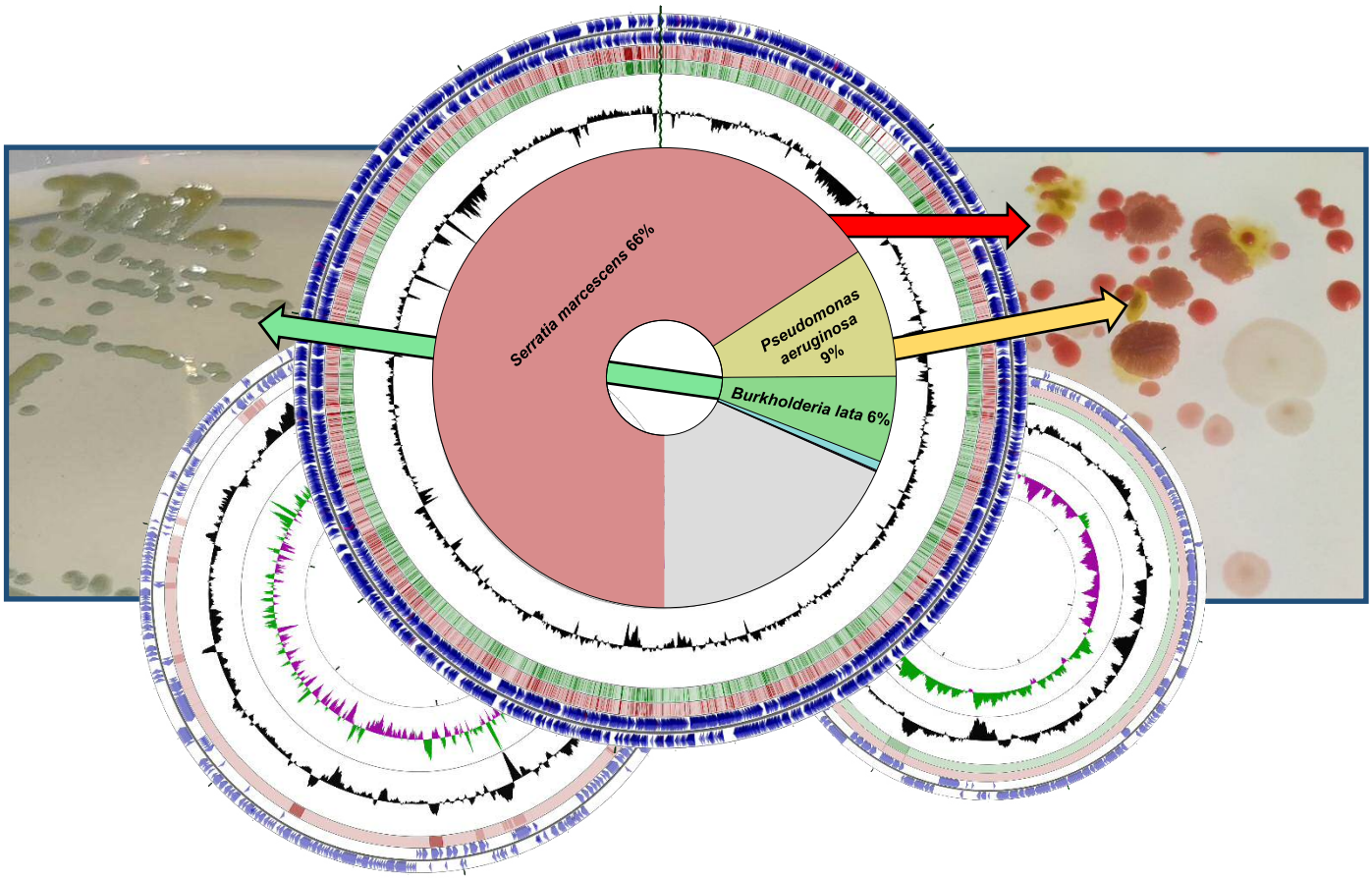
The Genomic Basis of Preservative Resistance



Thesis presented by

Edward Cunningham-Oakes

BSc (Hons) Pharmacology with Extra Mural Year

In candidature for the degree of Doctor of Philosophy

Microbiomes, Microbes, and Informatics Group

School of Biosciences

Cardiff University

April 2020

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Unilever

BBSRC Doctoral Training Partnerships

SWBio DTP

“Why scratch with the hens, when you can fly with the eagles?”

Presentation-related encouragement from a Belgian professor, Dublin, 2018

Acknowledgements

The journey throughout my doctoral studies has been an adventure, rife with new experiences, hardships, and unforgettable experiences. Whilst I do not consider myself a creature of sentiment, I can wholeheartedly say that I would not have made it through these studies, nor developed as a researcher, and individual without several people.

Firstly, I thank my primary supervisor, Eshwar Mahenthiralingam for constant advice, scientific expertise, and both scientific and emotional support. I am proud to have been a part of the Mahenthiralingam lab, and no longer consider this man to be only a mentor, but also a dear friend.

I would also like to thank my secondary supervisor Thomas Connor for his informatic expertise and support in a field I was very much inexperienced in, and unfamiliar with.

I thank my industrial supervisors, Tom Pointon and Barry Murphy, for their constant mentorship in the art of navigating industrial microbiology. Their expertise, humour and friendship helped me to develop from a clueless first year student, to prize winner amongst my cohort of industrial PhD students.

I thank the entire Preservation & Applied Microbiology team at Unilever Research & Development, Port Sunlight, for assisting the transition period into a placement and area I was unfamiliar with. I especially thank Stuart Campbell-Lee for his constant support, and for conducting the biofilm methodologies described in my *Pseudomonas* chapter. I also thank Sam Ridler for his support in developing and granting access for my factory ecology methodologies.

Thank you to the entirety of the Microbiomes, Microbes and Informatics group at Cardiff for making the PhD experience an unforgettable one. I especially thank Laura “lab mum” Rushton, “Uncle” Gordon Webster, and the one and only Alex Mullins, who have been the bedrock of my day-to-day lab life, which has been full of laughter and discovery. An additional special thanks to Lauren Wiltshire for generating the phenotypic data for the role of megaplasmids in my *Pseudomonas* chapter. I also thank the Cardiff Genome Research Hub, with special thanks to Angela Marchbank and Georgina Smethurst, for their friendship, tolerance of my endless sequencing projects, and terrible humour.

I thank my family, in particular my mother Hannah, who has supported my endeavours from day one, and my little brother Reuben, who has made my PhD all the more memorable by coming into my life as I came to Cardiff. I would also like to thank my newborn daughter Eva, for prompting me to write this thesis in a month and a half, and my partner Katie for her love and support throughout quite intense periods of writing and life. I look forward to spending many more years with you both.

Last but not least, I owe a special thanks to my dearest friends, all of whom are godparents to Eva, the numbers of which probably constitute a small private militia. I have known all of you for many years and consider you all to be family. To credit you all by name alone at the end of an acknowledgements page sells your worth somewhat short, but I could not have done this without you. In no particular order, a special thanks to Derek Tsiang, Jonathan Callow, Angie Halsey, Ollie Cornell, James Price, Bala Muthiah, James O’Donnell, Samuel Pargeter, Sîon Edwards, Joel Southgate and Jordan Cuff.

Scientific Awards and Conferences

AWARDED FUNDING

Date	Description	Amount	Contribution
Jan 2019	Wellcome Trust Institutional Strategic Support Fund Public engagement proof of concept: "Microbiology: small worlds, big opportunities"	**£5,361**	Co-Investigator

SELECTED PRESENTATIONS

Date	Appointed	Event	Location
Jul 2019	Accepted Speaker	17th International Conference on *Pseudomonas*	Kuala Lumpur, Malaysia
Jun 2019	Invited Speaker	Unilever-BBSRC Student Conference (Presentation winner)	Liverpool, UK
Jun 2018	Accepted Speaker	American Society of Microbiology (ASM) Microbe	Atlanta, GA, USA
May 2018	Invited Speaker	21st International *Burkholderia cepacia* Working Group Conference	Dublin, Ireland

Publications

Beaton A, Lood C, **Cunningham-Oakes E**, MacFadyen A, Mullins AJ, Bestawy W El, *et al*. Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas baetica* a390T sequenced by Ion semiconductor and Nanopore technologies. FEMS Microbiol Lett. 2018.

Weiser R, Green AE, Bull MJ, **Cunningham-Oakes E**, Jolley KA, Maiden MCJ, *et al*. Not all *Pseudomonas aeruginosa* are equal: strains from industrial sources possess uniquely large multireplicon genomes. Microb Genomics. 2019.

Webster G, Mullins AJ, Bettridge AS, Jones C, **Cunningham-Oakes E**, Connor TR, *et al*. The Genome Sequences of Three *Paraburkholderia* sp. Strains Isolated from Wood-Decay Fungi Reveal Them as Novel Species with Antimicrobial Biosynthetic Potential. Microbiol Resour Announc. 2019.

Webster G, Mullins AJ, Watkins AJ, **Cunningham-Oakes E**, Weightman AJ, Mahenthiralingam E, *et al*. Genome Sequences of Two Choline-Utilizing Methanogenic Archaea, *Methanococcoides* spp., Isolated from Marine Sediments. Microbiol Resour Announc. 2019.

Cunningham-Oakes E, Weiser R, Pointon T, Mahenthiralingam E. Understanding the challenges of non-food industrial product contamination. FEMS Microbiol Lett. 2020.

Cunningham-Oakes E, Pointon T, Murphy B, Connor TR, Mahenthiralingam E. Genome Sequence of *Pluralibacter gergoviae* ECO77, a Multireplicon Isolate of Industrial Origin. Stedman KM, editor. Microbiol Resour Announc. 2020

Summary

Antimicrobial resistance (AMR) is of increasing global concern, especially in the face of a stagnant antibiotic pipeline. As such, accurate global microbiological surveillance is more important than ever. This global crisis has been extensively highlighted in the nosocomial setting, where ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa and Enterobacter* spp.) are responsible for persistent infections, which bolster patient morbidity and mortality. In these settings, vigilant tracking and identification has been invaluable in limiting further outbreaks and infection. However, intrinsically resistant microorganisms are also problematic in manufacturing settings, where they can overcome antimicrobial preservatives present in food and cosmetics and enrich due to the presence of proteins and formulation additives. Therefore, the accurate identification of these organisms and understanding the mechanisms underpinning preservative tolerant phenotypes is essential for both the billion-dollar revenue of the HPCP industry, and consumer health. In this study, we first highlight that 49% of organisms responsible for non-food product recalls recorded in public recall databases remain unidentified, whilst the predominant identified organisms for recall are the Gram-negative bacteria, *P. aeruginosa*, *Pseudomonas* spp. and *Enterobacteriaeceae*. We then taxonomically characterise bacteria and fungi in the factory environment and demonstrate that the predominant identified microorganisms do not mirror those responsible for non-food product recalls. Utilising genomics and whole genome sequencing, we then illustrate that the use of historical taxonomic classification techniques such as multi locus sequence typing is insufficient for the accurate species-level identification of intrinsically preservative-tolerant *Burkholderia cepacia complex* spp. Finally, we use genomics to unveil the unique genome organisation of the problematic organisms *P. aeruginosa* and *Pluralibacter gergoviae,* revealing the presence of unique large extrachromosomal elements known as megaplasmids in bacteria from the industrial environment, before characterising these megaplasmids in the context of related megaplasmids from a variety of other stressful environments.

Contents

1. Introduction

This chapter was published in part as a mini-review (see *Cunningham-Oakes et al.* 2020, FEMS Microbiology Letters)

The globalisation of consumer markets has resulted in the expansion of a number of worldwide corporations. Whilst globalisation is beneficial for market growth, it means that the prevention of microbial spoilage is a key issue for several industrial sectors, as consumer goods can unintentionally serve as vectors for microbiological contamination (Saker *et al.* 2004; Centers for Disease Control and Prevention 2011). The home and personal care product (HPCP) industry is an example of where tight microbiological control is required to prevent contamination (Stewart *et al.* 2016). Personal care products are used to beautify, improve cosmetic appearance and maintain personal hygiene, and are often applied directly to the body (i.e. skin, or hair) (Sutton 2006). Home care products are cleaning products such as general purpose cleaners, dish wash liquid and laundry detergent (Goodyear *et al.* 2015), which may be purely chemical, or rely on biological factors such as enzymes in their antimicrobial mechanisms of action (Gupta *et al.* 2019). Many HPCPs are largely based on water and other components such as oils and proteins (Smart and Spooner 1972). In contrast to the pharmaceutical industry, products are not required to be sterile, but are required to be safe for consumer use (Orus and Leranoz 2005). The legislation concerning acceptable levels of microorganisms in cosmetics and HPCPs varies globally, (as detailed in Table 1), but all state that pathogenic organisms such as *Pseudomonas aeruginosa* (*P. aeruginosa*) and *Staphylococcus aureus* (*S. aureus*) must be absent from product (Halla *et al.* 2018).

As such, microorganisms may be present at low levels during HPCP manufacture, and proliferate with the correct physiochemical conditions such as redox potential, pH, water activity, water content available to microbial utility (A_W) (Lundov and Zachariae 2008), and the absence of adequate preservation. When microorganisms surmount preservation barriers and grow in a formulation, each component of the formulation will influence how they grow (Orth *et al.* 2006). For instance, when considering A_W in isolation, minimum A_W required for growth is subject to natural variance between organisms, and this will dictate which organisms grow in a product (Cundell 2015). A_W is however, standardised to 25°C (Cundell 2015), meaning that organisms will have a different minimum A_W in products shipped to a climate where the average temperature differs from this standard; this needs to be taken into account when considering the formulation

of the product. Once additional factors such as pH and nutrients are considered, HPCPs become incredibly complex microbiological niches, where each physicochemical property adds an additional layer of complexity to ensure acceptable levels of sterility (Orth *et al.* 2006).

Table 1.1: Global regulations for the microbiological safety of non-sterile products				
Region	**Advisory body**	**Advised product levels of microorganisms (CFU/ml or CFU/g)**		**Relevant documentation**
		Prioritised product type	**Other**	
European Union (EU)*	EU Council Directive 76/768/EEC.	**Eye/Infant** <1.0 x 10^2 (aerobic mesophilic bacteria)	<1.0 x 10^3	ISO 17516:2014 Cosmetics-Microbiology-Microbiological limits (1976)/(2015)
United States	FDA**/Personal Care Products Council (PCPC)	**Eye/Infant** <5.0 x 10^2 (No specified microorganism)	<1.0 x 10^3	United States Pharmacopeia (2016)
Japan***	Ministry of Health, Labor, and Welfare (MHLW)	**Oromucosal/gingival/cutaneous/nasal** <1.0 x 10^2 (aerobic mesophilic bacteria) <1.0 x 10^1 (yeasts and moulds)	N/A	Japan Pharmacopeia 17th Edition (2016)

Key:
* The EU additionally consider *Candida albicans* a primary potential pathogen alongside *P. aeruginosa* and *S. aureus*
** Unlike the PCPC, the FDA do not advise any particular levels of bioburden for microorganisms, but stipulate pathogens should not be present.
*** Cosmetics in Japan are regulated under the Pharmaceutical affairs law and divided into quasi-drugs and cosmetics.

1.1. WHY IS IT IMPORTANT TO PREVENT CONTAMINATION OF PRODUCTS?

As of 2017, the accumulative revenue from the sale of personal care products amounted to €77.6 billion in Europe alone, whilst the home care industry achieved a market value of €28.6 billion (statistics available via The Personal Care Association web page: https://cosmeticseurope.eu/news-events/socio-economic-contribution). Therefore, given the considerable revenue generated by the sale and production of HPCPs, the damages caused by contamination can range from moderate to catastrophic. In the case of early detection, basic costs include the decontamination of factory equipment, and discard of contaminated batches of raw material (Jimenez 2004). If contamination is detected at later stages in production, it can result in the removal of product from an entire supply chain, or in the worst-case scenario, a public recall, which results in damages to company profit margins and reputation, and, most importantly, to consumer health (Orth *et al.* 2006). It is therefore imperative that HPCPs are adequately preserved against contamination, to both minimise their risk to public health, and prevent product recall (Cunningham-Oakes, Weiser, *et al.* 2020).

In recent years, whilst HPCPs have a good track record with regard to microbiological safety, the risk of product contamination due to the presence of undesirable (or objectionable) organisms (see section 1.4) and subsequent infection is still possible. For example, in 2016, genome sequence analysis and single nucleotide polymorphism (SNP) mapping of a *Burkholderia lata* strain responsible for outbreaks in intensive care units of two tertiary hospitals in Australia was carried out (Leo*ng et al.* 2018). This analysis revealed a mouthwash preserved with 0.2 mg/mL chlorhexidine as a common source (Leong *et al.* 2018). The isolated strain was cultivable from patients, opened, and unopened bottles of mouthwash, suggesting contamination during the manufacturing process as opposed to subsequent use (Leong *et al.* 2018). In the same year, a contaminated piercing after-care solution resulted in a national outbreak of *P. aeruginosa* in England. Variable Number Tandem Repeat analysis was used to confirm that 29 infections were due to use of the solution (Evans *et al.* 2018). The aforementioned outbreaks illustrate the very real risk that intrinsically antimicrobial-resistant organisms pose to these products, and their consumers.

1.2. THE ROLE OF PRESERVATIVES IN PREVENTING PRODUCT CONTAMINATION

The manufacturers of HPCPs use synthetic (see Table 1.2) and natural (see Table 1.3) chemicals known as preservatives to prevent the spoilage of product by microbial

contamination, preserve the shelf life of the product, and protect the consumer (Halla *et al.* 2018). Microbial contamination can occur at many points during the life cycle of a product, from initial manufacture, through to the consumer level use, where HPCPs are unavoidably inoculated with the microbiota and flora of the skin (Orth *et al.* 2006; Sutton 2006). When considering the ideal preservative system for a product, multiple preservation strategies can be employed as follows:

1.2.1. Primary preservation

Adherence to good manufacturing practices (GMPs) during the manufacture of HPCPs and cosmetics by employing aseptic practices such as equipment sterilisation, water treatment and microbiological control of raw materials (Moore 2009). GMPs have been near unanimously accepted by regulatory bodies under ISO 22716:200, but are not enforced by the FDA (Corby-Edwards 2013). These practices are therefore largely down to the manufacturer to implement and enforce for high quality, low contamination risk production and use of products.

1.2.2. Secondary preservation

Physical: Preservation provided by physical packaging, such as narrow openings on shampoo and shower gel bottles, which limit contamination from bath/shower water by minimising the risk of microorganism ingress (Halla *et al.* 2018). The correct type of packaging is also critical to prevent the loss of preservative activity via interactions such as adsorption (Briasco *et al.* 2016).

Chemical: Certain chemical preservatives and multifunctional ingredients (described herein) can be obtained either from a natural and synthetic source. For example, organic acids can be either synthesised (Nicolaou 2014) or obtained naturally by means of fermentation (Couto and Sanromán 2006), but others, such as parabens, can only be obtained by means of synthesis. Overall, the current shift towards the utilisation of natural antimicrobials will result in new industrial portfolios of antimicrobial actives with no current track record in preservation, thus requiring rigorous testing (Russell 2003) (see 2.2.2) to ensure that they perform as required.

Synthetic preservatives: Chemical compounds with the primary purpose of preventing the growth of microorganisms by exerting antimicrobial activity (see Table 1.2). Synthetic preservatives are selected based on being non-toxic

to humans, and their antimicrobial efficacy and compatibility with other components of a formulation (Thiemann and Jänichen 2014). They are subject to Annex V of EU regulation (Commission Regulation (EU) No 1004/ 2014).

Natural preservatives: Compounds from a natural source, which have some antimicrobial, antioxidant and anti-inflammatory properties (Saxena and Nidhi 2015) (see Table 1.3). They are typically added as a free component, or using nanoparticles or microcapsules as a vector (Mokarizadeh *et al.* 2017). They are however usually discouraged in cosmetics, due to their loss of antimicrobial efficacy in dilutions, potential adverse impact on odour and lipophilic properties, and pH-dependent effects (Halla *et al.* 2018).

Multifunctional ingredients: Compounds where preservation effect is secondary to a primary function, such as fragrance (Schmaus *et al.* 2014) (see Table 1.3). These include, but are not limited to chelating agents, humectants, phenolic compounds and surfactants (Thiemann and Jänichen 2014). Multifunctional ingredients, which possess antimicrobial efficacy, but are not listed in Annex V cannot be marketed as preservatives (Halla *et al.* 2018). Commonly used multifunctionals include ethylenediaminetetraacetic acid (EDTA) a chelating agent which increases membrane permeability to antimicrobial compounds, and interferes with iron metabolism (Sutton 2006), and caprylyl glycol, an ampihilic surfactant which exerts antimicrobial effects via disruption of the bacterial membrane (Toliver and Narasimhan 2018).

1.2.3. Physicochemical

The A_W, pH and emulsion of a product, and the temperature at which the product is stored can also dictate the growth of microorganisms (Orth *et al.* 2006). Table 4 provides information as to the optimum physicochemical conditions for specific microorganisms which are considered "objectionable" organisms of interest by the HPCP industry (see section 1.4):

A_W: The majority of cosmetics are based on water, which alongside oils and proteins present in formulations can facilitate the growth of microorganisms (Smart and Spooner 1972). For this reason, formulations should be kept to a A_W of 0.80 or below (Berthele *et al.* 2014) via use of salts (e.g. sodium chloride), and other constituents such as polyols,

hydrocolloids, and amino acids (Erkmen and Bozoglu 2016). The constituents added to maintain a low A_W will vary based upon product type (Thiemann and Jänichen 2014). The choice of packaging such as the use of vapour-resistant bottles can also reduce A_W (Cundell 2015).

pH: The relative acidity or alkalinity of the formulation directly influences the ability of a microorganism to grow. Microorganisms capable of infection generally prefer to grow in a neutral pH, and do not survive well in an environmental pH of less than 4, or greater than 10 (Berthele *et al.* 2014). It is for this reason that the pH of liquid hand soaps (pH 9.5) (Gfatter *et al.* 1997) exert an antimicrobial effect, via ionisation of fatty acids and subsequent membrane destabilisation (Kabara 1997).

Emulsions: Creating an emulsion that prevents microbial growth can be a challenge, as reducing the oil to water content (i.e. creating a water-in-oil emulsion) (Varvaresou *et al.* 2009) means fewer lipids are available for microorganisms to metabolise, but can in turn destabilise a formulation. Additionally, certain compounds such as phenols exert their antimicrobial effects after forming a layer of oil (Buranasuksombat *et al.* 2011; Terjung *et al.* 2012).

The HPCP industry is gradually moving towards greener, minimal formulations in accordance with new regulations to reduce their global environmental impact (Bom *et al.* 2019). Consumers are also demanding preservative free, low preservative or natural preservative based products. Novel preservation strategies that will prevent the establishment of niches for microorganisms in HPCPs are therefore desirable. This will require complete identification and rigorous investigation of organisms isolated from HPCP contamination to understand their impact on the changing industry practices.

Table 1.2: Synthetic preservatives in use in the European Union			
Preservative class	**Examples**	**Permitted concentrations/restrictions**	**Mechanism of action**
Parabens	Methylparaben Butylparaben Propylparaben	Methylparaben: 0.4% Butyl/Propylparaben: 0.14%* Mixtures: 0.8%* *Sum of propyl and butyl cannot exceed 0.14% n.b. Cannot be used in baby products intended for nappy area	Inhibition of DNA/RNA synthesis (Nes and Eklund 1983) Inhibition of protein synthesis via reaction with amino acids (preferentially glutamate and aspartate) (Garner *et al.* 2014) Inhibition of nutrient transport (Elkund, 1980) Inhibition of mitochondrial oxygen consumption (fungi) (Ito *et al.* 2015)
Isothiazolinones	Methylisothiazolinone (MIT) Chloromethylisothiazolinone Methylisothiazolinone (CITMIT)	MIT: 0.0015% (rinse-off products only) CITMIT: 0.0015% (3:1 ratio of CIT to MIT)	Oxidisation of proteins (preferential action towards thiol groups and cysteine residues) (Lambert 2013b)
Formaldehyde releasing agents	Formadehyde Paraformaldehyde DMDM Hydantoin	Formaldehyde/paraformaldehyde: 0.1% (oral), 0.2% (other products) DMDM Hydantoin: 0.6%	Release of formaldehyde following hydrolysis in the presence of water. (De Groot and Veenstra 2010) React with amino acids to modify their nature, activity, and form disulphide bridges, thereby altering the structure of bacterial proteins. (Chen *et al.* 2016)
Biguanides	Chlorhexidine Polyaminopropyl biguanide	Chlorhexidine: 0.3% Polyaminopropyl biguanide: 0.1%	Chlorhexidine: • Positively charged, allowing it to bind to the negatively charged bacterial membrane: ▪ Low concentration – loss of metabolic capacity and osmosis regulation (D.A. *et al.* 2017) ▪ High concentration – Cytoplasmic coagulation and membrane disruption (D.A. *et al.* 2017) Polyaminopropyl biguanide: ▪ Alterations in membrane permeability via membrane integration (Kamaruzzaman *et al.* 2017) ▪ Interaction with, and damage to, genomic DNA (Chindera *et al.* 2016)

Table 1.3: Natural preservatives and multifunctional ingredients in use in the European Union			
Preservative class	**Examples**	**EU permitted concentrations/restrictions**	**Mechanism of action**
Alcohols & Phenols	Benzyl Alcohol Phenoxyethanol	Benzyl alcohol: 1.0% Phenoxyethanol: 1.0%	Phenoxyethanol: Concentration dependent effects including: ▪ Leakage of intracellular content via membrane interference (Halla *et al.* 2018) ▪ Dissipation of proton-motor force (Lucchini *et al.* 1990) Benzyl alcohol: • Solubilisation of the lipid bilayer (Scheinpflug *et al.* 2017)
Chelating agents	Ethylenediaminetetraacetic acid (EDTA) Glutamic acid (GLDA)	N/A	Inhibition of iron uptake for metabolism (Siegert 2014) Increased permeability of bacterial cell membrane, leading to increased preservative efficacy (Siegert 2014)
Humectants	Glycerin	N/A	Provide formulation stability by reducing A_w (Barel *et al.* 2009)
Organic acids	Anisic acid Benzoic acid and its sodium salt Dehydroacetic acid Other Benzoic acid salts Propionic acid Salicylic acid Sorbic acid	Propionic acid: 2% Sorbic acid: 0.6% Benzoic acid (and its sodium salt): 2.5% (rinse-off), 1.7% (oral), 0.5% (leave on) Other Benzoic acid salts: 0.5% Salicyclic acid: 0.5% Dehydroacetic acid	Acidification of the cytoplasm/external environment (Ricke 2003) Changing membrane fluidity (e.g. long-chain organic acids) (Alexandre *et al.* 1996) Chelation of metal ions and nutrients (e.g. anionic acids) (Stratford and Eklund 2003) Inhibition of metabolic enzymes
Surfactants	Glyceryl caprylate Glyceryl undecylenate	N/A	Emulsification and disruption of the bacterial cell membrane (Hamouda and Baker 2000)

1.3. PRIORITY CONTAMINANT ORGANISMS IN INDUSTRY

Antimicrobial resistance (AMR) is of increasing global concern, especially in the face of a stagnant antibiotic pipeline. The emergence of treatment refractory infections such as the notorious ESKAPE pathogens (*Enterococcus faecium, S. aureus, Klebsiella pneumoniae* (*K. pneumoniae*), *Acinetobacter baumannii, P. aeruginosa*, and *Enterobacter* spp.) has been extensively highlighted, and these organisms contribute greatly to patient morbidity and mortality (Pendleton *et al.* 2013). With the global spotlight on AMR, the epidemiology of such resistant microorganisms is now a priority. Similarly, organisms linked with outbreaks of food-borne disease such as gastroenteritis must be characterised to determine risk levels and sources of contamination. Intrinsically resistant microorganisms are also problematic in manufacturing settings where they can overcome antimicrobial preservatives and contaminate industrial products (Vincze *et al.* 2019; Cunningham-Oakes, Weiser, *et al.* 2020).

The ESKAPE pathogens are so named due to their ability to escape the mechanisms of action of most clinical antibiotics (Boucher *et al.* 2009). As such, these multidrug resistant organisms have been the leading cause of most bacterial, nosocomial infections since the term's inception a decade ago (Tacconelli *et al.* 2018). Currently, the most critical threat to public health is posed by Gram-negative bacteria. In 2017, the WHO reported that carbapenem-resistant *A. baumannii*, *P. aeruginosa*, and *Enterobacteriaceae* (including *Escherichia coli* (*E. coli*), *Enterobacter* spp., *K. pneumoniae*) were the organisms for which novel antibiotics were most critically needed (WHO 2017). Additionally, it was announced in 2018 that carbapenem-resistance has spread globally amongst *K. pneumoniae*, with a prevalence as high as 67% in European countries such as Greece, Italy and Romania (European Centre for Disease Prevention and Control 2018). This could soon make *K. pneumoniae* the most major, yet untreatable cause of bloodstream infections, and noscomially-derived pneumonia.

Gram-positive bacteria also pose a significant threat to public health, though they are not classed as critical with regard to AMR and antibiotic discovery requirements. However, *Mycobacterium tuberculosis* (TB) is no longer included in the WHO's annual review of AMR pathogen prioritisation, as it is globally acknowledged as a priority pathogen (WHO 2017). TB was one of the top ten global causes of mortality, which

was responsible for 1.6 million deaths in 2017 alone (Macneil *et al.* 2019), attributable to its innate antimicrobial resistance, and the combined inadequacy of the only available Bacillus Calmette-Guérin vaccine and anti-TB drugs (Davenne and McShane 2016). Other Gram-positive bacteria of note include Methicillin-resistant, vancomycin-intermediate and vancomycin-resistant variants of *S. aureus* which are no longer considered critical-risk pathogens , but are still recognised as high-risk and also cause problems within the community (WHO 2017).

Table 1.4: Optimum growth conditions for objectionable organisms

Organism	Temperature(˚C)	Minimum water activity (A_W)	pH
Aspergillus brasiliensis	37	0.80	5.0 - 6.5
Candida lipolytica	28-30	0.87	6.0
Enterobacter aerogenes	30	0.94	6.0 – 7.0
Escherichia coli	37-41	0.94	6.0 -7.0
Halophilic bacteria	20-23	0.75	7.0 / 8.5≤
Most spoilage bacteria	20-30	0.90	6.5 – 7.0
Most Yeast	25	0.87	4.0 – 4.5
***Penicillium* spp.**	25	0.75	3.0 – 4.5
Pseudomonas aeruginosa	20-30	0.96	6.6-7.0
Staphylococcus aureus	35-39	0.83	7.0-7.5

Gram-negative bacterial species stand out as the most common contaminants of non-sterile industrial products. Historically, *P. aeruginosa* and *Burkholderia cepacia* complex (Bcc) bacteria, are recognised as problematic, objectionable industrial contaminants, that are identified in a high proportion of recall incidents (Rushton *et al.* 2013; Weiser *et al.* 2019). *P. aeruginosa* accounted for 13% and 6% of recall incidents respectively for 1995 to 2006 and 2004 to 2011 FDA database surveys (Jimenez 2007; Jimenez 2011); over the same time periods Bcc bacteria were identified in 9% and 28% of incidents surveys. *P. aeruginosa* was also the most common contaminant reported in recalled cosmetic products in a short survey (2005 to 2008) of the EU Safety Gate (formerly RAPEX) database (Jimenez 2007; Jimenez 2011). *Enterobacteriaceae* are also prominent contaminants with significant numbers

of *Enterobacter* spp. identified within recalled products in the United States. All of these Gram-negative bacteria are also significant in terms of being recognised ESKAPE pathogens, with high intrinsic and acquired AMR (Pendleton *et al.* 2013). Their role in non-sterile product contaminations is seldom discussed, despite the significant overlap between these priority pathogens, and organisms identified as industrial contaminants (Vincze *et al.* 2019). The microbiological issues, potential to cause infection as a result of non-food product contamination, and methods for accurate identification or tracking of these species in the context of non-sterile food products are discussed further later in the thesis (Bcc bacteria at the beginning of Chapter 3, *P. aeruginosa* in Chapter 4, and *Enterobacteriaceae* in Chapter 5).

1.4. THE IDENTIFICATION GAP: A LACK OF ORGANISM CLASSIFICATION FOR NON-FOOD PRODUCT RECALL INCIDENTS

1.4.1. What are the primary organisms that have been previously found in non-food industrial products?

A meta-analysis of the US FDA non-sterile product recalls up to 2011 had shown that between 16% and 20% of the microorganisms causing contamination are recorded as unidentified (Jimenez 2007; Jimenez 2011). Within the European Union, microbial contamination associated with industrial product recalls is also catalogued to assess potential public health risks. Public databases are available for food products, the European Commission Rapid Alert System for Food and Feed (RASFF) (European Union 2016), as well as non-food products via Safety Gate (European Commission 2015). These databases record a variety of parameters for each recall including the nature and severity of risk, product type, country of origin, notifying country and overall outcomes.

1.4.2. Understanding the challenges of non-food industrial contamination – methodology for an updated analysis

To update the past analyses of recall data and determine the extent of AMR microbial species that are present in European food and non-food product recalls, a meta-analysis from 2005 to 2018 was performed on the RASFF and Safety Gate databases. In relation to contaminated foods, searches of RASFF database were performed for: (i) product type “Food”, and (ii) hazard type “Pathogenic microorganisms” or “Non-pathogenic microorganisms” and each incident

recorded within a spreadsheet. An analogous search of the Safety Gate database for all product categories and the risk type "Microbiological" was also performed. The classification of each reported microorganism was recorded to the family, genus, or species level. Organisms neither classified to at least family level, nor grouped together under collective descriptions such as "coliforms", "moulds" or "fungi" were designated "unidentified". If multiple organisms were found in a product recall, each organism was tallied separately.

1.4.3. Microorganisms in non-food industrial products are antimicrobial resistant Gram-negatives, or remain unidentified

A total of 7577 microorganisms were reported within 1016 recalls in relation to the food products whilst reports of microbial contamination in non-food products were 7-fold lower (378 microorganisms within 243 incidents). To probe the extent of useful epidemiological information within the RASFF and EU Safety Gate databases, information related to the identity of the contaminating organism for each product recall was examined. Proportional analysis of the microbial groups encountered in food and non-food product contamination revealed major differences in the type of microorganisms reported in recalls (Figure 1.1). The top five microorganisms encountered in food product recalls in rank order were *Salmonella*, the Hepatitis A virus, *Listeria*, *E. coli* and moulds (Figure 1.1a). Given the major importance of *Salmonella* as a food-bourne pathogen (Hoffmann *et al.* 2016) it was not surprising that this Gram-negative bacteria was identified in 68.2% of food product recalls. An additional key feature of the recall information in the food product database was the accuracy of identification, with 98% of organisms being identified and only 2% being reported as unidentified (Figure 1.1a).

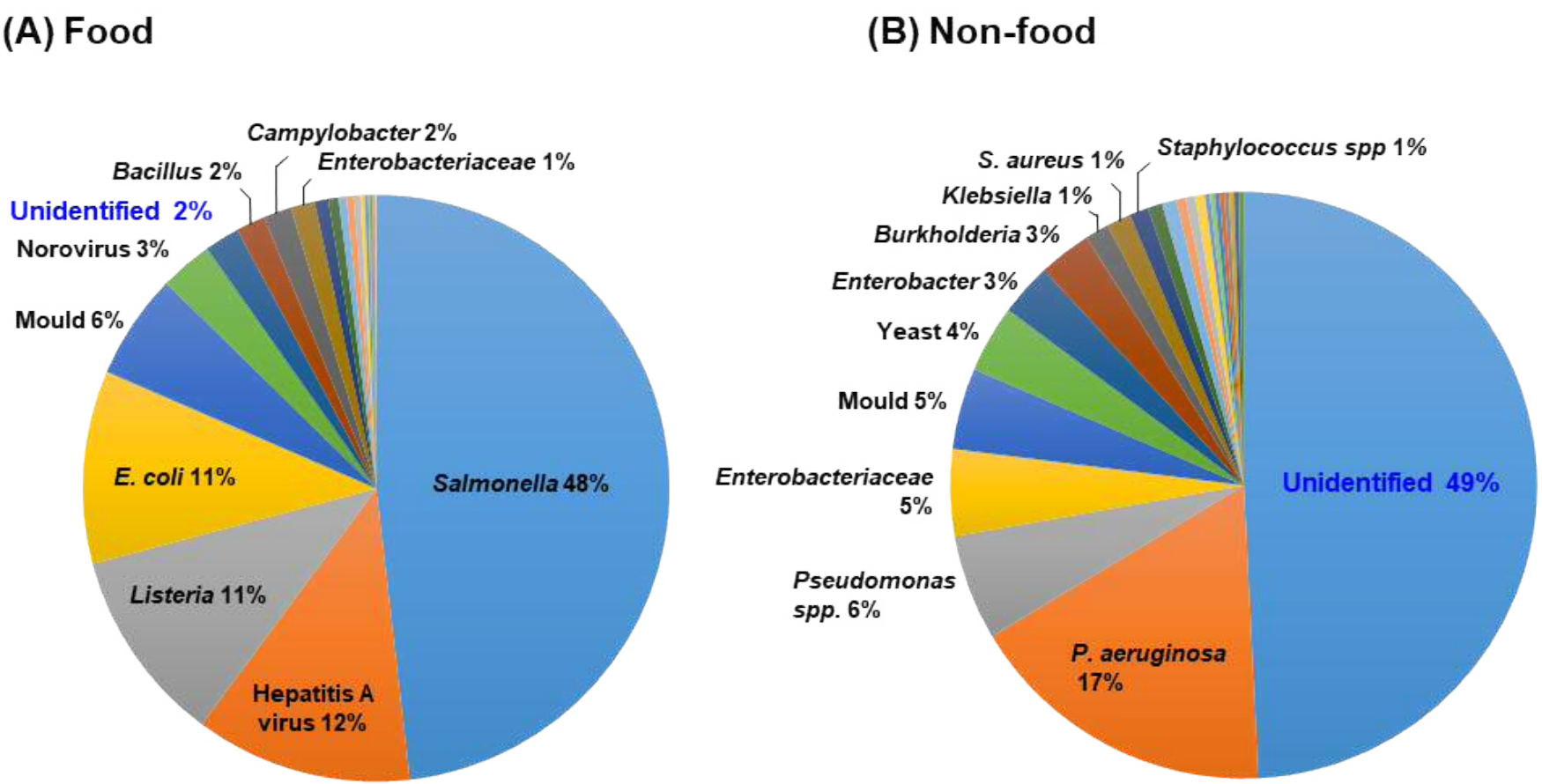


Figure 1.1: **49% of contaminants in non-sterile, non-food products are unidentified** - The lack of microbial species identification in food product recalls (A) compared to non-food product recall (B) from 2005-2018. The proportion of unidentified organisms from each recall database is highlighted in blue font. Each genus was tallied from EU Safety Gate and RASFF public recall databases, and data was used to generate pie charts in Microsoft Excel. Segments representing <1% of recalls are represented by a segment, but not labelled.

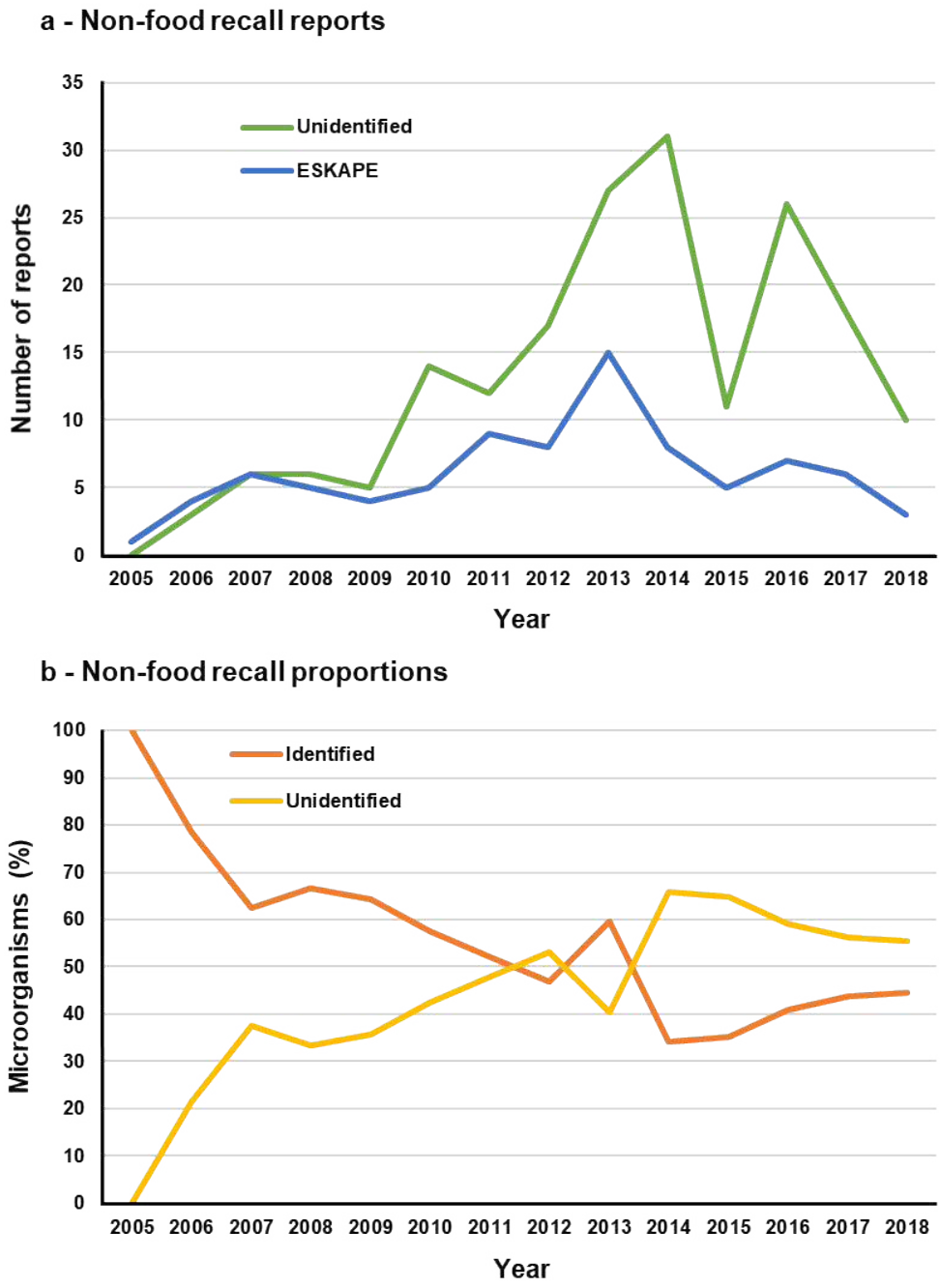


Figure 1.2: Non-identification in industrial products is increasing over time, and trends coincide with reports of ESKAPE pathogens - Trends over time showing (**A**) the similar patterns in the number of reports of ESKAPE pathogens in non-food industrial products, and unidentified organisms, and (**B**) the increase in relative proportion of unidentified organisms over time, from 2005-2018. Unidentified organisms, and organisms known as ESKAPE pathogens were tallied from EU Safety Gate and RASFF public recall databases, and used to generate line graphs in Microsoft Excel.

In striking contrast to the food related recalls, 49% of microorganisms associated with non-food product recalls were unidentified (Figure 1.1b). After this, the most common bacterial taxa in non-food product recalls were *P. aeruginosa*, *Pseudomonas* spp., *Enterobacteriaceae*, *Enterobacter*, *Burkholderia*, *Klebsiella*, and *Staphylococcus* (Figure 1b). Gram-negative bacteria, which are known to be intrinsically antimicrobial resistant, such as *Pseudomonas*, *Enterobacteriaceae*, *Enterobacter*, *Burkholderia*, *Klebsiella* and *Achromobacter*, were associated with 36.7% of non-food product recalls (Figure 1.1b). In addition, the ESKAPE pathogen group of bacteria accounted for 23% of all non-food recalls (data not shown). If the same proportions of antimicrobial resistant ESKAPE species are mirrored in the 49% of recall microorganisms recorded as unidentified, then the non-food product manufacturing industry is considerably under-reporting potential hazards in terms of global health risk and AMR.

Given the recent significant improvements in methods for accurate microbial identification (Maiden *et al.* 2013), it would be expected that more contamination incidents would be accurately reported over time. Between 2005 to 2018 the number of contaminants in non-food product recalls increased, with a concurrent increase in detected ESKAPE pathogens (Figure 1.2a). To examine whether this increasing trend could have been linked to general improvement in incident reporting in the non-food sector, the proportion of unidentified microorganisms by year was also investigated (Figure 1.2b). The lack of identification reporting for non-food product recall microorganisms increased dramatically from 2005 (0%) to 2015 (67%), followed by a slight decrease to its position in 2018 (55%) (Figure 2b). Therefore, although more non-food contaminants were reported, fewer were identified to a surveillance-relevant level (Figure 1.2b).

The large discrepancy in the proportion of identified organisms reported in food and non-food product recall information may reflect several issues with how contaminants are identified and validated for hygienic integrity. Furthermore, the reporting issue is still extremely prevalent to date, as an examination of the EU Safety Gate database conducted after the initial analysis indicates that 23/52 of the microorganisms (44%) reported across 26 recalls were unidentified.

The RASFF and Safety Gate databases do not publish microbial identification methods and there is likely to be considerable variability between reporting sites

and resource limitations when routine testing only detects certain microorganisms. The perceived risk may also influence the extent of reporting, with more stringent quality control being applied to foods that will be ingested, compared to non-food products having more limited direct contact with consumers.

The non-food global ISO 11930 and European Pharmacopeia challenge testing methods require the absences of *E. coli*, *P. aeruginosa*, *S. aureus*, *Candida albicans* and *Aspergillus brasiliensis* isolates; but do not require evaluation against opportunistic pathogens *Enterobacter*, *Burkholderia* or *Klebsiella* species (Siegert 2012). As demonstrated in Figure 1.1a, *Enterobacter*, *Burkholderia* and *Klebsiella* are causing a growing number of non-food recalls and need to be integrated into challenge test methodologies to ensure non-food products are adequately preserved for their intended purpose.

Overall, this analysis highlighted that, beyond the non-identification issue, *Pseudomonas* (both *aeruginosa*, and *non-aeruginosa*), *Enterobacter* spp., *Enterobacteriaeceae* and *Burkholderia* were the primary Gram-negative organisms of interest in terms of their prevalence as industrial contaminants. As such, the majority of this project, which is discussed in the following chapters, was tailored to identifying unique genomic features and elucidating genomic mechanisms of preservative resistance in these organisms.

Overall, the aims and objectives of this thesis were as follows:

1. **To fully characterise organisms present in manufacturing sites of home and personal care product factories** (Chapter 2).

2. **To determine the appropriate techniques for accurate identification of Bcc bacteria as industrial contaminants** (Chapter 3).

3. **To genomically characterise a panel of unique megaplasmid-harbouring *P. aeruginosa* from historical contamination incidents.** (Chapter 4).

4. **To accurately identify, genomically characterise, and determine the presence/absence of megaplasmids in a panel of *Enterobacteriaceae* isolates historically identified as *Enterobacter* spp. by 16S ribosomal RNA sequencing** (Chapter 5).

2. Home and Personal Care Product Factory Ecology

2.1. The importance of characterising microbes in an industrial environment

It is well known that microorganisms display the versatility to survive in both natural and abiotic environments (Castro-González *et al.* 2011; Zhao *et al.* 2017). In the food and beverage industry, it is widely accepted detection of pathogens is paramount to preventing foodborne outbreaks in the general population (Behravesh *et al.* 2011; Centers for Disease Control and Prevention 2011). As such, the microbiota of the food sector industrial environment has been extensively characterised (Pirttijärvi *et al.* 1996; Bagge-Ravn *et al.* 2003; Salo *et al.* 2006). This characterisation is essential to minimise the risk of transferring enteropathic bacteria such as *Enterobacter cloacae*, the *Escherichia coli* (*E. coli*) strain O157:H7 (Wang *et al.* 2013), *Klebsiella pneumoniae* (Davis and Price 2016), and *Salmonella Typhimurium* (Hoffmann *et al.* 2016), all of which have been involved in food contamination incidents.

In contrast, little is known with regard to bacteria present in the manufacturing environment of HPCP industries. This may be attributable in part to the high-throughput nature of the industry, meaning that an organism will only be identified in the event of a contamination incident, when it then may be traced to its source. Identification within the HPCP sector is also rarely carried out beyond basic aerobic culture characteristics, biochemical testing or occasionally 16S ribosomal RNA (rRNA) gene sequencing, though a recent review of recalls has shown that this may not be carried out for ~50% of reports (see Chapter 1 and Cunningham-Oakes, Weiser, *et al.* 2020). Moreover, currently established preservative systems ensure product sterility by exerting bacteriostatic/bactericidal effects via heterogeneous mechanisms of action (Maillard 2002; Lambert 2013a). However, preservation is intended for consumer protection, and cannot be used as a substitute for GMPs (see 1.2.1). By adhering to GMPs, HPCP industries minimise the risk of introducing objectionable organisms into the manufacturing pipeline, thereby pervading contamination and consumer exposure to a plethora of opportunistic pathogens (Moore 2009).

GMP guidelines advise the quantification of aerobic bacteria and fungi in HPCP factories via total viable count, using Typtone Soy Agar (TSA) and Sabaroud Dextrose Agar (SDA) respectively (Maukonen *et al.* 2003). This ensures that key areas of the factory remain within specified limits of sterility, which are often expressed as total viable counts, and are determined by the risk level of a given factory area.

Quantification without additional tests however reveals little about the identity of organisms present in the factory environment and the risk that they pose is unknown. As such, we know very little about the risk that this putative HPCP factory microbiome poses. This is despite reports that food factory surfaces, which are not dissimilar to surfaces in the HPCP factory environment, may harbour opportunistic pathogens such as *Pseudomonas* spp., *Enterobacteriaceae* and *Acinetobacter* spp. (Møretrø and Langsrud 2017). Though normally innocuous, all of the aforementioned organisms are opportunistic pathogens, known HPCP contaminants, and have been evidenced to possess and develop resistance to established preservative systems (Rajamohan *et al.* 2009; Abdel Malek and Badran 2010; Périamé *et al.* 2015). Whilst it is possible to consider new preservation strategies in the face of these and other organisms that display resistance to established preservative systems, the financial impact that the implementation of novel preservative systems bears upon HPCP industries is substantial, and is often impractical within a short timeframe.

2.2. Cost and safety considerations for preservative system implementation

2.2.1. Safety considerations

When implementing a new preservative system, there are a number of safety considerations to take into account, which can make the process innately costly and challenging. One such example of this is the dosing of preservatives into product formulations, which involves the addition of large volumes of powder into vessels (Glor 2006). Prior to this process, the minimum ignition energy (MIE), i.e. the minimum amount of energy required to ignite a combustible vapour, gas, or dust cloud, must be assessed at the most likely points of ignition in the vessel (Glor 2006; Addai *et al.* 2016). This is because the addition of large volumes of powder can lower the MIE of the vessel, which can ultimately result in an explosion if exposed to an ignition source, such as static electricity (Britton 1999a; Bouillard *et al.* 2010). The most likely points of ignition lie close to the surface of the solvent, near the manhole/immediately outside of the vessel, and the upper part of the vessel (Glor 2006). In order to mitigate against this, most manufacturers use solvents with a high flash point temperature (i.e. a high temperature pre-requisite for ignition) to reduce the risk of ignition and collateral damage (Britton 1999b). Flash point temperature is an important consideration as the inside of a vessel may not initially possess a sufficient MIE to trigger an explosion, but solvents with a

low flash point produce large amounts of vapour. This can then oversaturate the atmosphere of the vessel and generate an explosive atmosphere at the manhole (Britton 1999b; Glor 2006).

To prevent ignition risks, HPCP manufacturers must take a number of additional precautions (Eckhoff 2009). One such precaution is strict adherence to Equipment for potentially dangerous atmospheres directives (EU ATEX), which outline equipment and protective measures that must be implemented in explosive environments when using powdered materials and solvents (European Commission 2016). Whilst the measures for the dosing of older preservation systems are well established, newer systems with different physicochemical properties will require consideration on a factory-by-factory basis for their safe implementation (Glor 2006). If a factory lacks the capabilities to safely dose new preservatives with currently available equipment, they will require new dosing systems to be implemented throughout the factory (Eckhoff 2009). The implementation of said systems and even the use of preservatives themselves can prove to be difficult in a factory environment, both in terms of practicality, and cost (Hasnan *et al.* 2014). As such, the implementation of new preservative systems requires tough-decision making to fine-tune the balance between EU legislation, the needs of the consumer, preservative system efficacy, and the safety of those involved in the handling of new materials (Sutton 2006; Thiemann and Jänichen 2014).

2.2.2. Preservative efficacy testing

All new preservative systems must undergo a process known as challenge testing or preservative efficacy testing (PET), the purpose of which is to determine the susceptibility of HPCP preservation systems and raw materials to microbial contamination. HPCP preservation must consider both the efficacy of the preservation system, and in-product levels to prevent microbial viability and resistance development, whilst not exceeding the maximum permitted in-product preservative levels stipulated by the EU for consumer safety (European Commission 2015). The most common protocol to test these criteria involves the introduction of a defined panel of microorganisms to a preservative system followed by removal of samples and enrichment in microbiological media to perform total viable counts (Giorgio *et al.* 2018). With regard to which organisms

should be used for testing, both the United States and British Pharmacopeia suggest a panel consisting of *Pseudomonas aeruginosa* (*P. aeruginosa*), *Staphylococcus aureus* (*S. aureus*), *Aspergillus brasiliensis* and *Candida albicans* (Halla *et al.* 2018).

For cosmetics however, additional organisms are specified. For example, the Cosmetic, Toiletries and Perfumery Association specify that for the testing of eye area cosmetics, panels must contain non-fermenting Gram-negative bacteria (including *P. aeruginosa*), fermenting Gram-negative bacteria, yeast, moulds, in-house isolates (e.g. from historical contamination incidents), and *Bacillus subtilis* (a spore former, optional) (Huang *et al.* 2017). Organisms from the panel are used to challenge a given preservative system via product inoculation, followed by a total viable count, in order to assess organism viability following exposure to a system. This means that the duration of a single PET is highly dependent upon the rate at which an organism grows in the presence of preservative, which can take anywhere up to 7 days (Giorgio *et al.* 2018). Furthermore, PET must account for the physical, chemical, and microbiological properties of each component of a preservative system, and the system as a whole (Sila-On *et al.* 2006). For instance, exposure to single microorganism may be sufficient to demonstrate the efficacy of one preservation system, whilst others may require multiple exposures to single or mixed inoculums to model consumer use, and ensure that they are truly robust (Sutton 2006). As such, PET becomes quite an intricate process, where the appropriate protocol is often determined on a product-by-product basis. A full summary of the acceptable levels of microbial bioburden in non-sterile products is provided in Chapter 1, Table 1.1.

Additionally, all HPCP industries must adhere to certain criteria, regardless of product type. Such criteria include but are not limited to accounting for contamination during processing and consumer use, using product manufactured in the same way as it would be in a factory, and if necessary, performing additional tests to satisfy local markets. Taking all of the aforementioned considerations into account, it is clear that PET is an invaluable tool for HPCP industries, in that it simultaneously enables the comparison of different preservative systems, evaluation of different concentrations for preservative systems and selection of those suitable for final formulation. It is however time-consuming and labour-

intensive (Russell 2003), which can make it challenging for even the most experienced of microbiologists.

2.3. AIMS & OBJECTIVES

Whilst it is clear that implementing the correct preservation system is effective in maintaining product sterility, even the most well-maintained factory environment will never be truly sterile, and preservation is not a substitute for GMP. Furthermore, the constitutive microbial taxa of HPCP factory microbiome (see below for definition) are unknown. Thus, whilst it is clear current challenge testing panels are representative of a number of problematic nosocomial pathogens, it is unclear how representative they are of the factory environment. The overall aim of this chapter is to understand the diversity of microorganisms occurring in HPCP manufacturing environments.

The objectives of this chapter are as follows:

1. **To characterise the diversity and trends of cultivable bacteria and fungi present in HPCP factories in two global manufacturing sites.**

2. **To determine whether objectionable bacteria and fungi could be cultivated from in-line (directly involved in product manufacture) and environmental (not involved in manufacture of product) areas of HPCP factories.**

3. **To achieve genomic characterisation of bacteria of interest isolated from the HPCP factory environment.**

4. **To establish whether panels suggested by the British Pharmacopeia, United States Pharmacopeia and Cosmetic, Toiletries and Perfumery Association are representative of organisms present in manufacturing environments.**

Throughout the chapter, the term microbiota will refer to the assemblage of microorganisms present in an environment, whilst the term microbiome will be used to summarise their genes, genomic material, functions and the surrounding environmental conditions (Marchesi and Ravel 2015). Therefore, by isolating factory

microorganisms and thereby establishing the factory microbiota, downstream genomic analyses will enable us to establish the factory microbiome via taxonomic identification, and the exhibition of putative preservative resistance, akin to antimicrobial resistance seen in nosocomial environments (Oliphant and Eroschenko 2015). Overall, the hypothesis for this chapter was "The HPCP factory environment constitutes a source of microbial species frequently encountered as industrial contaminants".

2.4. Materials & Methods

2.4.1. Sampling the ecology of HPCP factories

Sampling was conducted at two global HPCP manufacturing sites, 1 and 2. Samples were taken from one home care, and one personal care factory at each site. Initial communication with HPCP factory quality hygienists were used to establish sampling points of mutual interest. Following on from these liaisons, 174 samples were taken across factories and included in the study (Site 1, n = 99; Site 2, n = 75). Sampling was conducted using either Fisherbrand™ dry cotton swabs in Peel Packs moistened with 0.9% phosphate buffered saline solution (PBS) (for Site 1) or Transport swabs containing Amie's media without charcoal (for Site 2) depending upon the resources available. Each area was swabbed for 10 seconds, before placing the swab back into a 5 mL tube containing 2 mL 0.9% sterile saline, or Amie's media. Where possible, a sterile 10 cm2 marker was used to control the area swabbed at each sampling point, whilst ensuring side of the swab head gains contact with the surface of the sample. For samples of non-uniform surface, it was not feasible to sample using a marker, for example samples taken from the inside of aqueous product mixers; these are sealed vessels with a relatively narrow aperture for adding raw materials and hence difficult to sample. In such cases, the surface area available was swabbed for 10 seconds instead. For cotton swabbing, swabs were used to mix the saline solution thoroughly for 10 seconds, before disposing of the swab and closing the tube aseptically. Both swabbing and mixing times were established following initial optimisation experiments, which indicated that these times were optimal for the recovery of viable microorganisms.

2.4.2. Isolation of cultivable microorganisms

Transport swabs or 200 µl saline solution were used to inoculate individual plates of Tryptone Soy Agar (TSA), Sabaroud Dextrose Agar (SDA) or Reasoner's 2A agar (R2A). Once dry, were incubated at 28°C for 24 to 96 hours. The exact duration of incubation was dependent upon the rate of biomaterial growth. As such, growth was monitored, and each sample plate was removed from incubation before confluence was reached. Any bacterial colony or fungal isolate deemed determined to have a unique morphological phenotype relative to other colonies was then transferred onto rich media (TSA, SDA) and minimal media (R2A) agar using a sterile plastic loop. Where no growth was observed after one passage on rich media, but growth was observed on minimal media, isolates were enriched on R2A, before transferring to rich media. All isolated morphotypes were then passaged for single colonies on TSA or SDA, before incubating the purified isolates on TSA or SDA for 24 to 48 hours until confluence was reached.

2.4.3. Storage of factory isolates

A swab of each isolate approximately equal to 3-4 McFarland standard was used to create freezer stocks via suspension in cryogenic vials (Pro-Lab Diagnostics), before storing at -80°C. Each vial contains cryoprotective solution, and sterile beads that allow microorganisms to readily adhere to the surface. Beads from each freezer vial were then streaked over the surface of either a TSA or SDA plate to ensure purity, and grown for a maximum of 48 hours at 28°C. Plates were then shipped to Cardiff, and confluence was used to create an equivalent freezer stock in TSB with 8% DMSO as a cryopreservant.

2.4.4. Rapid DNA extraction in Chelex

For rapid DNA extraction, single colony material grown on TSA or SDA agar plates was used to create 3 mL overnight cultures by inoculating TSB or SDB, and incubating for 18 hours at 30°C. 10 µL of overnight culture was then suspended in 100 µL of 5% Chelex (Bio-rad, Hertfordshire; autoclaved prior to use). Each sample was then heated to 95°C for 5 minutes using Bio-rad DNA Engine Dyad Peltier Thermal Cycler, before placing at -80°C for 5 minutes. This process was repeated, before each sample was centrifuged at 1000 g for 10 minutes to remove cellular debris, leaving crude DNA in supernatant. DNA was stored at 4°C, and used on the same day as Chelex resin extraction was prepared.

2.4.5. Maxwell®16 DNA extraction from bacteria with excessive extracellular polymeric substance

For bacteria that produced a large amount of mucosal material and extracellular polymeric substance, crude DNA extraction using Chelex and subsequent PCR reactions were both ineffective. Genomic DNA was therefore extracted from fresh bacterial growth using the Maxwell®16 Tissue DNA Purification Kit and instrument (Promega, Southampton, UK). This involved centrifugation of a fresh 3 mL overnight culture at 4000 x g, followed by resuspension in 1 mL of a solution composed of equal parts TSB and 4M Guanidinium Isothiocyanate (Invitrogen). The resultant suspension was then transferred to FastPrep Lysing Matrix E Tubes (MP Biomedicals). Samples were then homogenised at using a FastPrep®-24 Classic Instrument benchtop homogeniser (MP Biomedicals) before being centrifuged at 10,000 rpm for 10 minutes to remove any excess cellular debris. After centrifugation, genomic DNA was extracted from 400 µL supernatant using the Maxwell®16 Tissue DNA Purification Kit according to the manufacturer's instructions (Promega 2012). All extractions were taken forward for PCR (see section 4.8). DNA was stored at 4°C in experiments, and at -20°C for long term storage.

2.4.6. Maxwell®16 DNA extraction from fungi

Genomic DNA was extracted from fungi grown on SDA agar plates for 24-48 hours using the Maxwell® 16 Tissue DNA Purification Kit and instrument (Promega 2012). This involved suspension of fungal scrapings in 1 mL of Guanidinium Isothiocyanate (4M), before transferring to FastPrep Lysing Matrix E Tubes and homogenising using FastPrep®-24 *Classic* instrument. The resulting lysate was then transferred to the Maxwell® 16 instrument for DNA extraction and purification. All extractions were taken forward for PCR (see section 4.8). Genomic DNA was stored 4°C for experiments, and -20°C for long-term storage

2.4.7. 16S ribosomal RNA gene sequencing and phylogenetic placement for bacteria

DNA extracted as described in 2.4.4 and 2.4.5 was used to perform a partial 16S rRNA gene PCR using universal bacterial primers 27F and 907R (Lane 1991) (see Table 2.3). Successful amplification was confirmed via agarose gel electrophoresis. Purified PCR products were then submitted for Sanger

sequencing via the MWG Eurofins Sanger Sequencing service, using the 'MWG Eurofins Value Read Service in Tubes' submission guide. During initial sample submission, it was determined that there were very few differences between the forward and reverse sequences (obtained using primers 27F and 907R respectively) from Sanger sequencing, and their consensus gene sequences. As such, only the forward primer (27F) was sent with PCR products to accelerate the identification of isolates. The sequence generated from the forward strand was then used to perform initial identification using the Basic Local Alignment Search Tool (BLAST) algorithm. For initial identification, the top hit by nucleotide blast (BLASTN) was taken as the initial genus/species identifier. Objectionable bacteria (*Acinetobacter, Pseudomonas, Staphylococcus*, *Stenotrophomonas* and bacteria within the family *Enterobacteriaceae*) identified via BLASTN were aligned using MAFFT to a manually curated database of 16S ribosomal RNA gene sequences. The database in question was created using The List of Prokaryotic Names with Standing in Nomenclature (LPSN) database (Parte 2018), which includes all validly published names for prokaryotes, along with accession numbers for 16S sequences of type strains. A separate database was generated for each genera of objectionable bacteria, using the 16S ribosomal RNA gene sequences of representative type strains from LPSN, high-quality 16S ribosomal RNA gene sequences from the SILVA ribosomal RNA database (Quast *et al.* 2013), and sequences of several biocide reference strains. Sequences of objectionable isolates were then aligned to the database using MAFFT, before a maximum-likelihood phylogenetic tree was generated using Randomized Axelerated Maximum Likelihood (RAxML) via command line (Stamatakis 2014). The RAxML executable was compiled with multi-thread and SSE3 functionality: raxmlHPC-PTHREADS-SSE3 RAxML was run using a maximum-likelihood with General Time Reversible (GTR) substitution and a GAMMA model of rate heterogeneity supported by 100 bootstraps. Trees generated using RAxML were visualised with phylogeny graphic viewer FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and edited using the graphic editing software Inkscape v0.91 (http://www.inkscape.org/).

2.4.8. ITS sequencing and identification of fungi

DNA extracted as described in 2.2.6 was used to perform PCR amplification of fungal intergenic spacer regions, using primers ITS4 and gITS7 (see Table 2.3),

(White *et al.* 1990; Ihrmark *et al.* 2012). Identification was performed via BLASTN analysis of the resultant PCR products in comparison to the UNITE database for identification of fungal based species (Kõljalg *et al.* 2005). Confirmation of successful amplification and sequencing of purified PCR products using forward primer was then performed as described in 2.2.7.

2.4.9. Whole genome sequencing of objectionable organisms

Genomic DNA was extracted from purified bacterial isolates identified as objectionable to the genus level by 16S (see section 4.8) using the Maxwell®16 Tissue DNA Purification Kit, but was instead suspended in 300µL of 4M Guanidinium Isothiocyanate, as no isolates that were selected required a bead beating step. All *Staphylococci* were sequenced, including those initially identified as "non-*aureus*" species, which are not classed as objectionable organisms. This was used as an opportunity to determine whether non-*aureus* species from the factory environment possess a greater number of antimicrobial and biocide resistance genes than similar organisms isolated from non-factory environments. DNA was quantified using a Qubit Fluorometer, with the double-stranded DNA, broad range assay. In brief, Qubit Fluorometers provide a DNA concentration by measuring fluorescence, emitted when a dye present in the assay binds to DNA. Sequencing libraries were then prepared using a Nextera XT DNA library preparation kit, using Maxwell buffer as a diluent for DNA, and for library normalisation (provided for elution of DNA with the Maxwell kit, contains 0.1mM EDTA). This buffer was used as a diluent to standardise the inhibition caused by EDTA during the library normalisation process across all samples, thereby minimising the impacts on downstream sequencing seen in other chapters (see Chapter 3). Sequencing was performed using an Illumina NextSeq 500 platform.

2.4.10. Short-read (Illumina) assembly for sequenced organisms

Illumina adaptors were trimmed from 150-nucleotide paired-end reads using the TrimGalore, (v0.6.4) script (Krueger 2015), which provides a wrapper around Cutadapt (v2.6) (Martin 2011) and FastQC (v0.11.8) (Andrews *et al.* 2015) to enable automated detection and removal of sequencing adapters, and comparison of read quality comparison before and after trimming. Reads were then assembled using Unicycler (v0.4.8) (Wick *et al.* 2017), which functions as an optimiser for the genome assembler SPAdes (Bankevich *et al.* 2012). The assembled contigs were

screened for contamination using Kraken2 v2.0.8-beta (Wood and Salzberg 2014), a *k*-mer based matching system which provides taxonomic classification to assembled contigs, using the standard Kraken database. Genome quality was assessed using Quast v4.4 (Gurevich *et al.* 2013), which also provided N50 values, alongside other assembly statistics.

2.4.11. Species-level placement using Average Nucleotide Identity

The taxon ID provided by Kraken2 (see section 4.11) for each genome was used to obtain all available reference genomes belonging to the same genus or species, (depending upon the level of resolution provided by Kraken2), from the European Nucleotide Archive (ENA), using the command-line based enaDataGet script from enaBrowserTools v1.5.4 (https://github.com/enasequence/enaBrowserTools). The exceptions to this were *P. aeruginosa*, and *Staphylococcus lugdunensis* (*S. lugdunensis*), for which curated datasets were constructed using past publications, the *Pseudomonas* Genome Database and the National Centre for Biotechnology Information (NCBI) nucleotide database (Winsor et al. 2011; Federhen, 2012), due to the considerable number of *P. aeruginosa* and lack of *S. lugdunensis* genomes, available in ENA. Average Nucleotide Identity (ANI) was then used with reference genomes to determine the species-level identity of factory isolates. ANI was initially conducted using FastANI (Jain et al. 2018), an alignment-free based method which enables fast pairwise comparison between a query genome, and large genomic datasets. To confirm species identity, the top 29 matches for each factory isolate and the representative genome for each dataset were then used to create a genome database. If available, a biocide reference strain from the same genus was included in each database. Databases were then analysed using Python script PyANI v0.2.7 (https://github.com/widdowquinn/pyani), which computes ANI using pairwise alignments to confirm species-level placement of factory isolates. An ANI threshold of 95.0% was used to determine species-level identity. If pairwise ANI was greater than the 95.0% threshold (Jain, 2018) for a factory isolate and reference genome labelled as a particular species, but less than the threshold for the factory isolate and type-strain/representative genome, then the species name for both the factory isolate and reference genome in question was designated as sp. (unknown species).

2.4.12. Antimicrobial Resistance and Biocide Gene Prediction

Antimicrobial, biocide and metal resistance genes were predicted in the factory genome assemblies and reference genomes by employing publically available databases The Comprehensive Antibiotic Resistance Database (CARD) (McArthur *et al.* 2013) and The Antibacterial Biocide and Metal Resistance Database, BacMet2 (Pal *et al.* 2014) respectively through the BLAST-based tool ABRicate (https://github.com/tseemann/abricate). Both databases were chosen for resistance gene predictions from genomes, due to their rigorous curation, and high quality. The CARD database is regularly peer-reviewed and contains resistance gene determinants organised by the Antibiotic Resistance Ontology and AMR gene detection models. BacMet2 possesses databases for both predicted and experimentally validated biocide and metal resistance genes. The BacMet2 experimental database was chosen for resistance gene prediction, as it only consists of experimentally validated genes, all of which are linked to scientific literature. For gene prediction, a coverage cut-off of 75% was used. Boxplots were generated for absolute gene counts using the boxplot package in R. For each boxplot, mean values calculated using the boxplot package were indicated in results text, along with standard deviations, which were calculated using Microsoft Excel. and whiskers were defined using the Tukey HSD test, which extends whiskers to data points that are less than 1.5 x IQR away from 1st/3rd quartile. The purpose of this test is to determine statistically significant differences between mean values within a boxplot. A lack of overlap between the whiskers of two means indicates statistical significance.

2.4.13. SNP analysis and heat mapping of isolates

For isolates from same factory site, with ≥99% ANI (as determined by PyANI, see 2.4.11), and the same taxonomically closest reference genome from ENA, Snippy (Aruhomukama *et al.* 2019) (https://github.com/tseemann/snippy) was used to evaluate the likelihood of the aforementioned isolates shared a common source. Snippy predicts nucleotide polymorphisms by aligning next-generation sequencing reads to a reference genome using BWA MEM, determines the validity of any resultant variants using Freebayes, before using snpeff to determine the effects of any SNPs on the genes of an organism. Variants were called as follows: SNP – Single Nucleotide Polymorphism (e.g. A → T), MNP - Multiple Nucleotide Polymorphism

(e.g. AT → CG), INS – Insertion (e.g. CT → CGT), DEL – Deletion (e.g. CGT → CT), COMPLEX – combination of SNPs and MNPs. This analysis was conducted with the hypothesis that closely related organisms that have originated from the same source would have a similar number of SNPs, but will have accumulated a greater number of polymorphisms if they had travelled further geographically or temporally from their point of origin.

Table 2.1: PCR conditions for amplification of the bacterial 16S region			
PCR step	**Temperature (°C)**	**Time (s)**	**Cycles**
Initial denaturation	98	30	1
Denaturation	98	10	30
Annealing	54	30	
Extension	72	150	
Final Extension	72	300	1
Indefinite hold	12.5	-	-

Table 2.2: PCR conditions for amplification of the fungal ITS region			
PCR step	**Temperature (°C)**	**Time (s)**	**Cycles**
Initial denaturation	94	300	1
Denaturation	94	30	25
Annealing	56	30	
Extension	72	30	
Final Extension	72	420	1
Indefinite hold	12.5	-	-

Table 2.3: Primers used in this study		
Primer	Primer sequence (5' to 3')	Reference
27F	AGAGTTTGATCMTGGCTCAG	(Lane 1991)
907R	CCGTCAATTCMTTTGAGTTT	(Lane 1991)
gITS7F	GTGARTCATCGARTCTTTG	(Ihrmark *et al.* 2012)
ITS4R	TCCTCCGCTTATTGATATGC	(White *et al.* 1990)

2.5. RESULTS

2.5.1. Objectionable organisms are predominantly isolated from environmental, low-risk areas

In total, 86 of 181 (48%) samples taken across HPCP factories displayed viable growth after plating of samples (Site 1 = 42, Site 2 = 44). Of these, 62 of 86 only displayed growth on TSA (Site 1 = 39 of 42, Site 2 = 23 of 44), 5 of 86 displayed growth on SDA (Site 1 = 0 of 42, Site 2 = 5 of 44), and 11 of 86 displayed viable growth on TSA and SDA (Site 1 = 3 of 42, Site 2 = 8 of 42). Additionally, 8 samples from Site 2 required an initial passage on minimal media, followed by a passage onto rich media to display viable growth (TSA from R_2A = 6 of 8, SDA from R_2A agar = 2 of 8). When comparing sites manufacturing different product types, 27 of 64 samples (Site 1 = 6 of 25 Site 2 = 21 of 41) from home care factories displayed viable growth, whilst 58 of 112 samples (Site 1 = 36 of 74, Site 2 = 23 of 41) from personal care factories displayed viable growth. Across both factories, 190 morphologically distinct colonies were identified during CFU counts (Site 1 = 112, Site 2 = 88).

A partial length bacterial 16S rRNA gene or fungal ITS amplicon was successfully amplified using DNA extracted from 170 of 181 isolates (Site 1 = 98 of 107, Site 2 = 72 of 74). BLASTN revealed that cultivated microorganisms from both sites were predominantly Gram-positive (Figure 2.1). Phylogenetic placement of bacterial species that were identified as objectionable via BLASTN revealed that all *Acinetobacter* (*n*=4), *Pseudomonas* (*n*=5) and *Klebsiella* (*n*=1) and 23/24 *Staphylococcus* isolates could be placed with species-level identity (see Figure 2.3 for a representative example). Phylogenetic placement of *Stenotrophomonas* isolates (*n*=6) based upon partial 16S sequences did not elucidate the species-level identity of these isolates. Similarly, 2 of 4 non-Klebsiella *Enterobacteriaceae* could be placed within the genus *Enterobacter*, but no further resolution was provided by phylogenetic placement (Figure 2.2). In contrast, a consistent ID could not be achieved for remaining non-Klebsiella *Enterobacteriaceae*, with numerous inconsistent BLASTN matches to multiple genera within the family *Enterobacteriaceae* (Table 2.4). As such, it was not possible to determine an appropriate reference 16S dataset for these isolates. Overall, the aforementioned

analyses revealed that objectionable organisms (n = 33) are predominantly isolated from environmental locations (n = 19), subject to routine localisation, as opposed to in-line areas directly involved in product manufacture (n = 14).

2.5.2. The genome-level identity of all objectionable bacteria excluding *Pseudomonas* and *Enterobacteriaceae* was consistent with identification by 16s

Prior to analysis, 9 of the 43 genomes sequenced were removed due to contamination, as assessed by comparison of Kraken taxon ID, and the initial identification by 16S rRNA gene sequencing. Three isolates (A55G, P11D and P60C) were removed from further taxonomic placement using FastANI, and any downstream analyses, as FastANI revealed that they were clonal duplicates of A55A, P11C and P60B respectively. For the remaining 31 genomes, species-level identification was achieved for all *Acinetobacter*, *Klebsiella* and *Staphylococcus*.

A summary of the species-level identity of each sequenced factory isolate can be found in Table 2.4, along with their initial identification as proposed by 16S rRNA gene sequencing. Of the isolates initially identified as *Pseudomonas,* 2/4 belong to a yet unnamed species. The closest match to Factory isolate A22 in the constructed database was an unknown species, *Pseudomonas* sp. B4-2017 (Table 2.4). Factory isolate A39 displayed the highest level of similarity to *Pseudomonas parafulva* (*P. parafulva*) NS212, but displayed ANI values below the 95% threshold when compared to the *P. parafulva* type strain NBRC 16636 (84.3%, Table 2.4 and Figure 2.5), and the type strain for the proposed species-level identity by 16S, *Pseudomonas taiwanensis* BCRC 17751 (83.3%, Table 2.4 and Figure 2.5). 2/3 of non-*Klebsiella Enterobacteriaceae* isolates which could not be taxonomically identified via 16S (Figure 2.2), A26A and A55A, could be identified to the species-level by ANI. This analysis revealed that one isolate, A26A, possessed a new genus-level identity in comparison to the initial identification by 16S (*Kosakonia* as opposed to *Siccibacter*), and new species-level identity in comparison to the Kraken taxon ID (*Kosakonia sacchari* by Kraken identification, *Kosakonia oryzendophytica* by ANI). The remaining *Enterobacteriaceae* isolate, A24B, displayed ANI values over the threshold when compared to multiple genomes labelled as *Enterobacter cloacae*, but displayed an ANI value below the threshold

when compared to the type species ATCC 13047 (84.5%). A summary of the taxonomic diversity of all *Enterobacteriaceae* isolates can be found in Figure 2.6. No factory isolates identified as *Stenotrophomonas* could be placed with a named species by ANI, displaying close matches with genomes for isolates labelled as the species *Stenotrophomonas maltophilia*, but displaying ANI values below the threshold in comparison to the type strain ATCC 13637 (see Table 2.4).

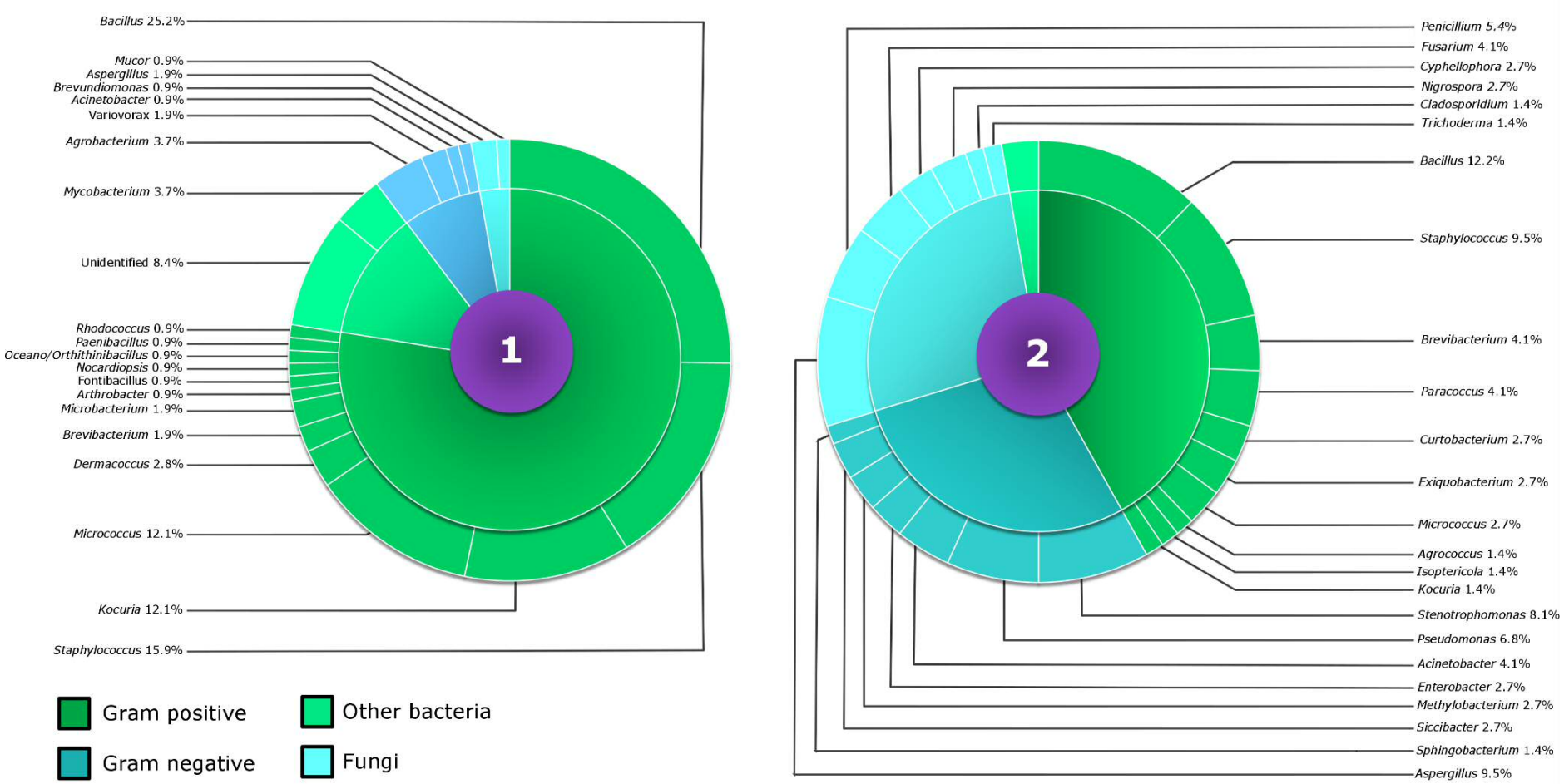


Figure 2.1: Summary of the taxonomic diversity and relative proportion of all factory isolates by genus, across two factory sites (1 and 2). Matches were determined by BLASTN of bacterial 16S or fungal ITS sequences. Absolute numbers of isolates identified as a given genus were calculated and converted to percentages in Microsoft Excel. The same calculations were performed to determine proportions of Gram-positive and Gram-negative bacteria, fungi or other bacteria. The other bacteria category encompassed unidentified isolates, and *Mycobacteria*, which do not traditionally Gram stain. Pie charts generated from these calculations were then overlayed to produce the sunburst charts shown.

Table 2.4: Isolates of interest selected for genome sequencing, and their identification based on 16S rRNA gene sequencing, and average nucleotide identity

				FastANI closest genome			PyANI closest genome		
Strain ID	**Sampling point**	**Site**	**16S ID**	**Species**	**Strain**	**Identity (%)**	**Species**	**Strain**	**Identity (%)**
A14A	Side pot 3*	1	***Stenotrophomonas maltophilia***	*Stenotrophomonas maltophilia*	UBA3513	97.28	*Stenotrophomonas maltophilia*	UBA3513	97.21
A22	Holding tank*	1	***Pseudomonas taiwanensis***	*Pseudomonas* sp.	B4-2017 B4 1	91.00	*Pseudomonas* sp.	FGI182	91.11
A24A	Ultrafiltration sampling tap	1	***Pseudomonas aeruginosa***	*Pseudomonas aeruginosa*	T52373	99.29	*Pseudomonas aeruginosa*	XMG	99.45
A24B	Ultrafiltration sampling tap	1	***Enterobacter* sp.**	*Enterobacter cloacae*	GEN000188	98.92	*Enterobacter cloacae*	GEN000188	99.08
A25B	Post demineralisaiton water	1	***Stenotrophomonas maltophilia***	*Stenotrophomonas maltophilia*	SmF22001	98.78	*Stenotrophomonas maltophilia*	SmF22001	98.83
A26A	Demineralisation sampling tap	1	***Siccibacter* sp.**	*Kosakonia oryzendophytica*	LMG 26432^T	99.38	*Kosakonia oryzendophytica*	LMG 26432^T	99.50
A26B	Demineralisation sampling tap	1	***Stenotrophomonas* sp.**	*Stenotrophomonas maltophilia*	SmF22001	98.73	*Stenotrophomonas maltophilia*	SmF22001	98.80
A33	Sachet laminate roll	1	***Staphylococcus saprophyticus***	*Staphylococcus saprophyticus*	3751	99.57	*Staphylococcus saprophyticus*	Ssapro-055	99.68
A39	Raw material mixer*	1	***Pseudomonas* sp.**	*Pseudomonas parafulva*	NS212	99.14	*Pseudomonas parafulva*	NS212	99.22
A4A	Mixer tank*	1	***Pseudomonas aeruginosa***	*Pseudomonas aeruginosa*	T52373	99.25	*Pseudomonas aeruginosa*	XMG	99.45
A55A	Colour mixer*	1	***Stenotrophomonas maltophilia***	*Stenotrophomonas maltophilia*	SmF22001	98.68	*Stenotrophomonas maltophilia*	SmF22001	98.80
A55B	Colour mixer*	1	***Siccibacter* sp.**	*Siccibacter colletis*	1383^T	98.49	*Siccibacter colletis*	1383^T	98.70
A55E	Colour mixer*	1	***Klebsiella* sp.**	*Klebsiella variicola*	EuSCAPE IT360	99.93	*Klebsiella variicola*	EuSCAPE IT360	99.96
A57A	Colour mixer*	1	***Acinetobacter junii***	*Acinetobacter junii*	AJ 068	98.39	*Acinetobacter junii*	AJ 068	98.46
A72A	Rework container	1	***Acinetobacter indicus***	*Acinetobacter indicus*	DSM 25388^T	98.03	*Acinetobacter indicus*	DSM 25388^T	98.18
A79A	Empty bottle lid*	1	***Staphylococcus saprophyticus***	*Staphylococcus saprophyticus*	3201	99.57	*Staphylococcus saprophyticus*	Ssapro-055	99.69
A7A	Side pot 1*	1	***Pseudomonas stutzeri***	*Pseudomonas stutzeri*	PS 138	98.33	*Pseudomonas stutzeri*	UBA3460	98.30
A7B	Side pot 1*	1	***Acinetobacter junii***	*Acinetobacter junii*	AJ 068	98.29	*Acinetobacter junii*	AJ 068	98.46
N12C	Dye vessel*	2	***Staphylococcus* sp.**	*Staphylococcus epidermidis*	MTCC 3382	99.88	*Staphylococcus epidermidis*	JCM 2414	99.92
N12E	Dye vessel*	2	***Staphylococcus saprophyticus***	*Staphylococcus saprophyticus*	889	99.92	*Staphylococcus saprophyticus*	889	99.97
P11A	Keyboard 1	2	***Staphylococcus hominis***	*Staphylococcus hominis*	DE0425	99.69	*Staphylococcus hominis*	DE0425	99.81
P11C	Keyboard 1*	2	***Staphylococcus pasteuri***	*Staphylococcus pasteuri*	DE0452	99.93	*Staphylococcus pasteuri*	DE0452	99.95
P11H	Keyboard 1	2	***Staphylococcus capitis***	*Staphylococcus capitis*	DE0241	99.62	*Staphylococcus capitis*	DE0241	99.66
P16B	Keyboard 2	2	***Staphylococcus* sp.**	*Staphylococcus aureus*	URU110	98.16	*Staphylococcus aureus*	CHI61	98.12
P59B	Product filler*	2	***Staphylococcus* sp.**	*Staphylococcus saprophyticus*	889 889 1	99.94	*Staphylococcus saprophyticus*	889	99.96
P60A	Keyboard 3	2	***Staphylococcus arlettae***	*Staphylococcus arlettae*	NCTC 12413^T	99.59	*Staphylococcus arlettae*	NCTC 12413^T	99.58
P60B	Keyboard 3	2	***Staphylococcus* sp.**	*Staphylococcus epidermidis*	Scl25 10	99.72	*Staphylococcus epidermidis*	Scl25 10	99.96
P65B	Floor spill	2	***Acinetobacter* sp.**	*Acinetobacter schindleri*	ACE-R	97.67	*Staphylococcus lugdunensis*	ACS-027-V-Sch2	99.43
P68D	Sink	2	***Staphylococcus lugdunensis***	*Staphylococcus lugdunensis*	ACS-027-V-Sch2	99.43	*Staphylococcus lugdunensis*	HKU09 01	99.57
P68C	Sink	2	***Staphylococcus* sp.**	*Staphylococcus saprophyticus*	G764	99.76	*Staphylococcus saprophyticus*	735A001	99.80
P72C	Keyboard 4	2	***Staphylococcus hominis***	*Staphylococcus hominis*	DE0187	99.76	*Staphylococcus hominis*	DE0225	99.86

Footnote. **Summary of all factory isolates of interest taken forward for downstream genomic analysis**. For each isolate, the sampling point (area from which they were isolated in their site of origin) is indicated. Initial identification by BLASTN of partial 16S sequences, alongside closest taxonomic matches by ANI to reference genomes is shown for both alignment-free (FastANI) and alignment-based (PyANI) based methods. Isolates which did not display an ANI value >95.0 % for species-identification when compared to the type strain are highlighted in orange

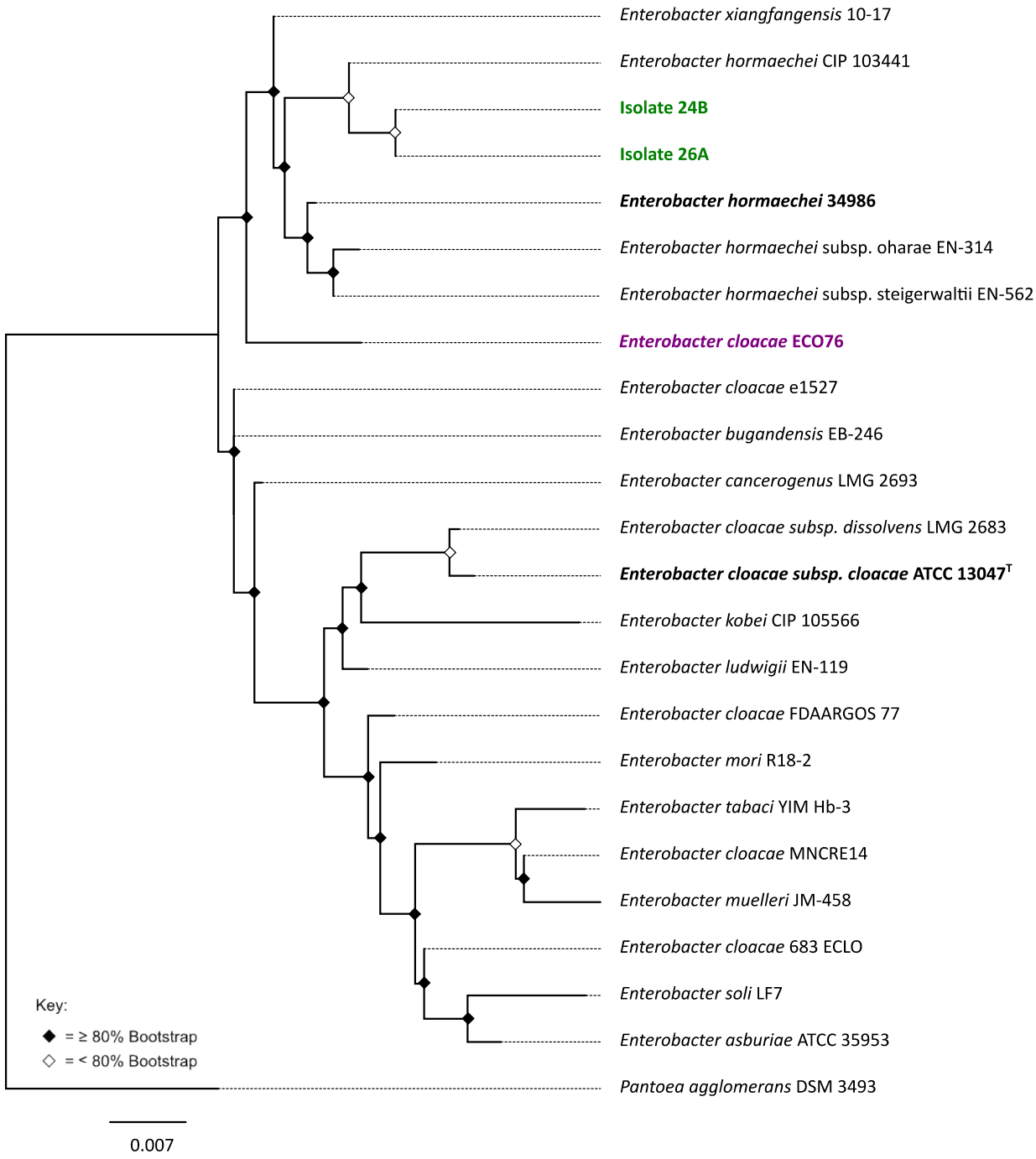


Figure 2.2: Phylogenetic tree of factory isolates identified as *Enterobacter* via BLASTN, *Enterobacter* biocide reference strain, and type strains for *Enterobacter*, based on a partial 16S ribosomal RNA gene sequence (800-1000 bp; all trimmed to 600 bases for alignment). Where species have one representative in the phylogeny, the type strain was used. Where species have more than one representative to confirm phylogenetic placement of factory isolates, type strains are in **bold**, and annotated (T). Biocide strains are shown in purple. Scale bar represents the number of substitutions per base position.

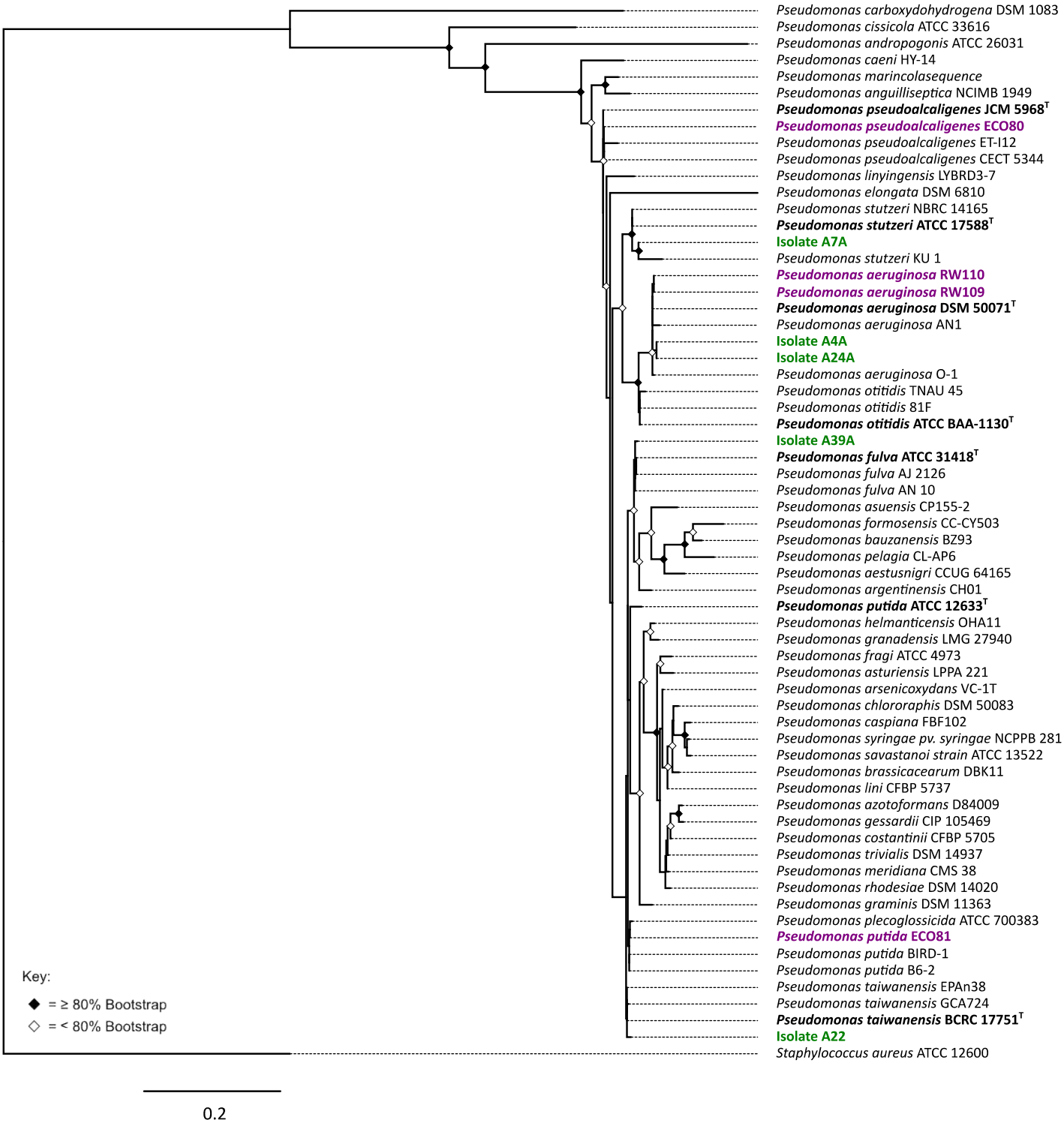


Figure 2.3: Partial length 16S rRNA gene phylogenetic tree of factory isolates identified as *Pseudomonas* via BLASTN, *Pseudomonas* biocide reference strains, and type strains for *Pseudomonas*, based on a partial 16S ribosomal RNA gene sequence (800-1000 bp; all trimmed to 600 bases for alignment). Where species have one representative in the phylogeny, the type strain was used. Where species have more than one representative to confirm phylogenetic placement of factory isolates and biocide strains, type strains are in **bold**, and annotated (T). Biocide strains are shown in purple. Scale bar represents the number of substitutions per base position.

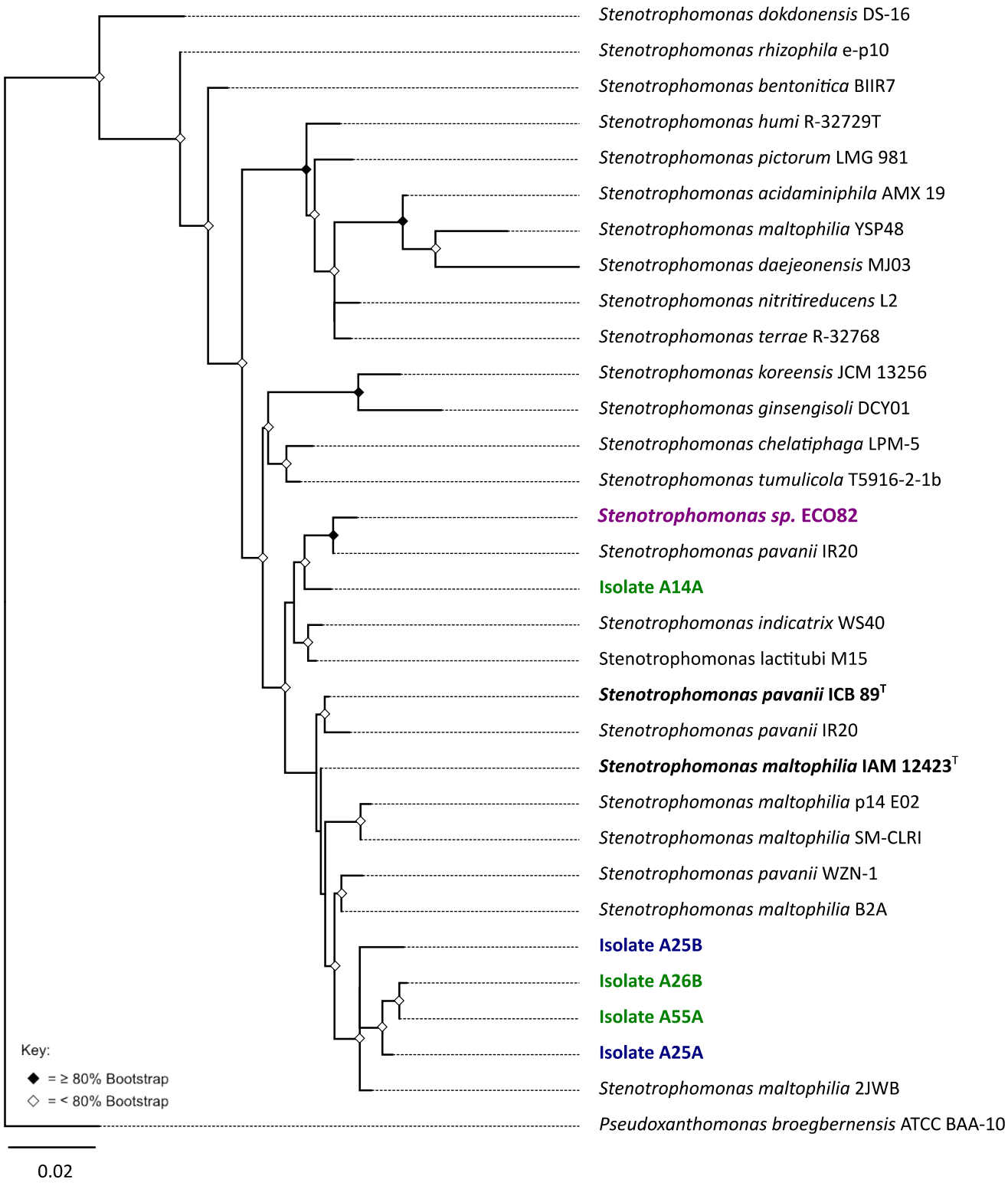


Figure 2.4: Phylogenetic tree of factory isolates identified as *Stenotrophomonas* via BLASTN, *Stenotrophomonas* biocide reference strain, and type strains for *Stenotrophomonas*, based on a partial 16S ribosomal RNA gene sequence (800-1000 bp; all trimmed to 600 bases for alignment). Where species have one representative in the phylogeny, the type strain was used. Where species have more than one representative to confirm phylogenetic placement of factory isolates, type strains are in **bold**, and annotated (T). Biocide strains are shown in purple. Scale bar represents the number of substitutions per base position.

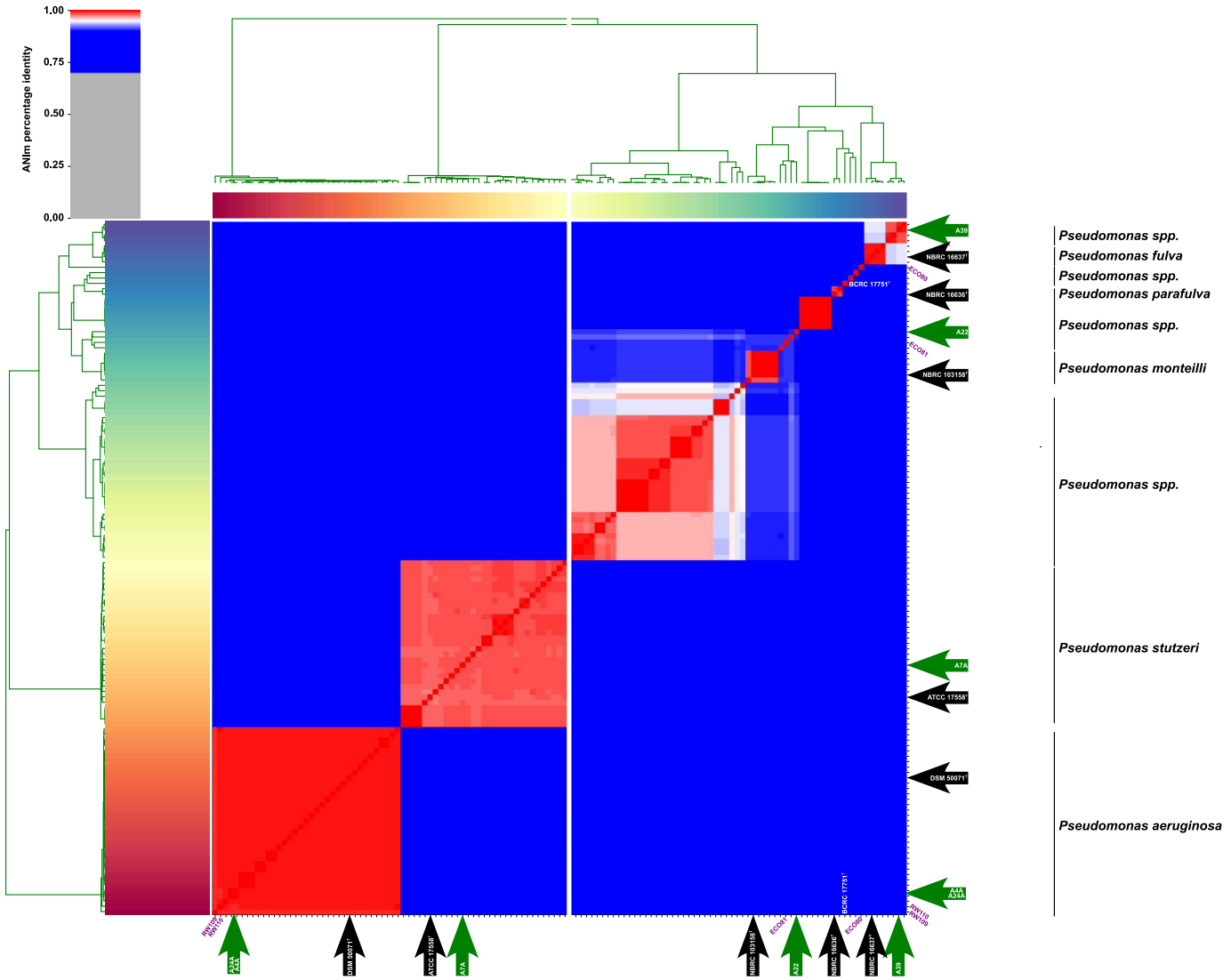


Figure 2.5: Genome sequence taxonomic placement of selected *Pseudomonas* factory isolates. Heatmap generated by the PyANI script, indicating the degree of nucleotide-level similarity between factory-derived *Pseudomonas* isolates, and their closest reference strains. Red areas highlight isolates that possess >95% nucleotide similarity, with darker shades of red indicating greater similarity. Blue indicates <95% nucleotide similarity. All factory isolates are indicated with arrows, with environmental isolates in-line isolates and green arrows, and type or representative genomes denoted by blue, green and black arrows respectively. Biocide reference isolates are denoted by purple text.

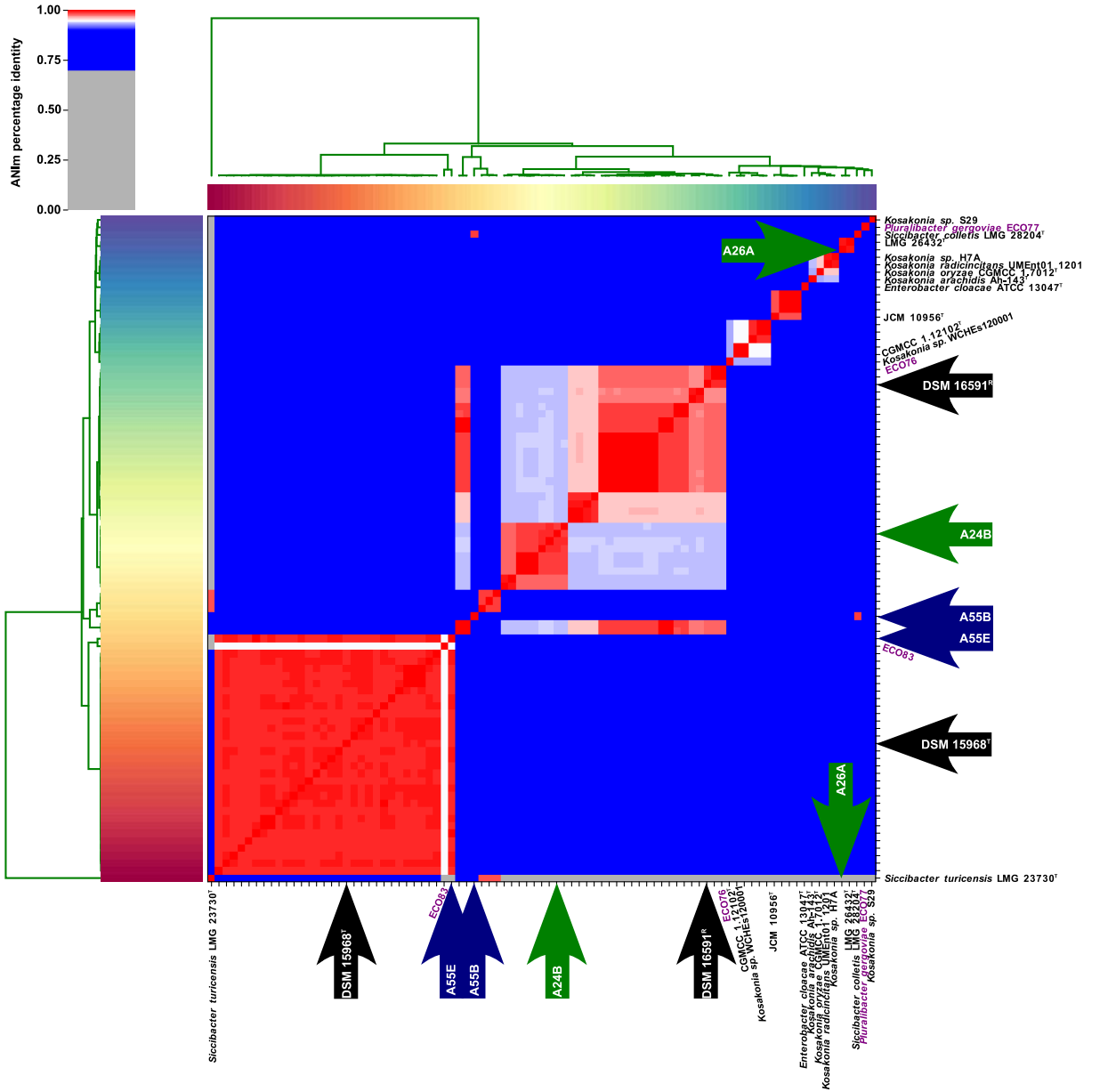


Figure 2.6: Genome sequence taxonomic placement of selected *Enterobacteriaceae* factory isolates. Heatmap generated by the PyANI script, indicating the degree of nucleotide-level similarity between factory-derived *Enterobacteriaceae* isolates, and their closest reference strains. Red areas highlight isolates that possess >95% nucleotide similarity, with darker shades of red indicating greater similarity. Blue indicates <95% nucleotide similarity. All factory isolates are indicated with arrows, with environmental isolates in-line isolates and green arrows, and type or representative genomes denoted by blue, green and black arrows respectively. Biocide reference isolates are denoted by purple text.

2.5.3. Predicted AMR genes in factory isolates are typical of related organisms from a variety of backgrounds

Overall, the mean number of AMR genes for each factory isolates belonging to a given family or genus was highly variable, and comparable to the mean number predicted from the entire dataset, including reference genomes (see Table 2.5). The highest number of predicted genes was in *Pseudomonas* factory isolates (mean = 25 ± 19) and was similar to the number of AMR genes predicted within the closely related reference genomes (mean = 22 ± 16). The high number and variability are largely attributable to the high number of predicted resistance genes in *P. aeruginosa* isolates A4A and A24A (48 predicted genes) in comparison to other factory isolates. In contrast, factory isolates *Pseudomonas stutzeri* A7A (10 predicted genes), *Pseudomonas* sp. A22 (12 predicted genes), and *Pseudomonas* sp. A39 (9 predicted genes) possessed lower numbers of predicted AMR genes. The number of predicted AMR genes in factory isolates was concordant with the number of predicted genes from reference genomes of the same species. Reference *P. aeruginosa* genomes possessed 47 ± 2 predicted AMR genes, including biocide reference strains *P. aeruginosa* RW109 and RW110 (47 and 46 predicted genes respectively). In non-*aeruginosa Pseudomonas* reference genomes, the mean number of predicted AMR genes was 14 ± 5. The number of AMR genes predicted in *Enterobacteriaceae* factory isolates (30 ± 4) was similar to the number of predicted AMR genes in biocide reference strains *Enterobacter* sp. ECO76 (37 predicted genes), *Klebsiella* sp. ECO83 (34 predicted genes), but higher than biocide reference strain *P. gergoviae* ECO77 (24 predicted genes). The lowest number of AMR genes were predicted from the genomes of *Acinetobacter* factory isolates (mean predicted genes = 3 ± 1). A visual representation of the number of predicted AMR genes for each factory isolate in the context of reference genomes is shown in Figure 2.7.

	Table 2.5: Predicted genes in factory isolates of interest						
	Resistance gene type		**Bacterial taxa**				
			Acinetobacter	***Enterobacteriaceae***	***Pseudomonas***	***Staphylococcus***	***Stenotrophomonas***
Dataset	AMR	Mean	5	34	22	7	9
		S.D.	5	6	16	5	3
	Biocide	Mean	96	184	130	72	127
		S.D.	21	42	40	12	20
Factory isolates	AMR	Mean	3	30	25	6	9
		S.D.	1	4	19	2	3
	Biocide	Mean	109	85	155	82	126
		S.D.	22	13	26	20	8

Table 2.5: Predicted AMR and Biocide genes in factory isolates of interest in comparison to their reference datasets. The total number of AMR genes as predicted using the ABRicate tool and CARD database were tabulated and means and standard deviations calculated for the factory isolate genomes and reference genomes together, and for the factory isolate genomes alone.

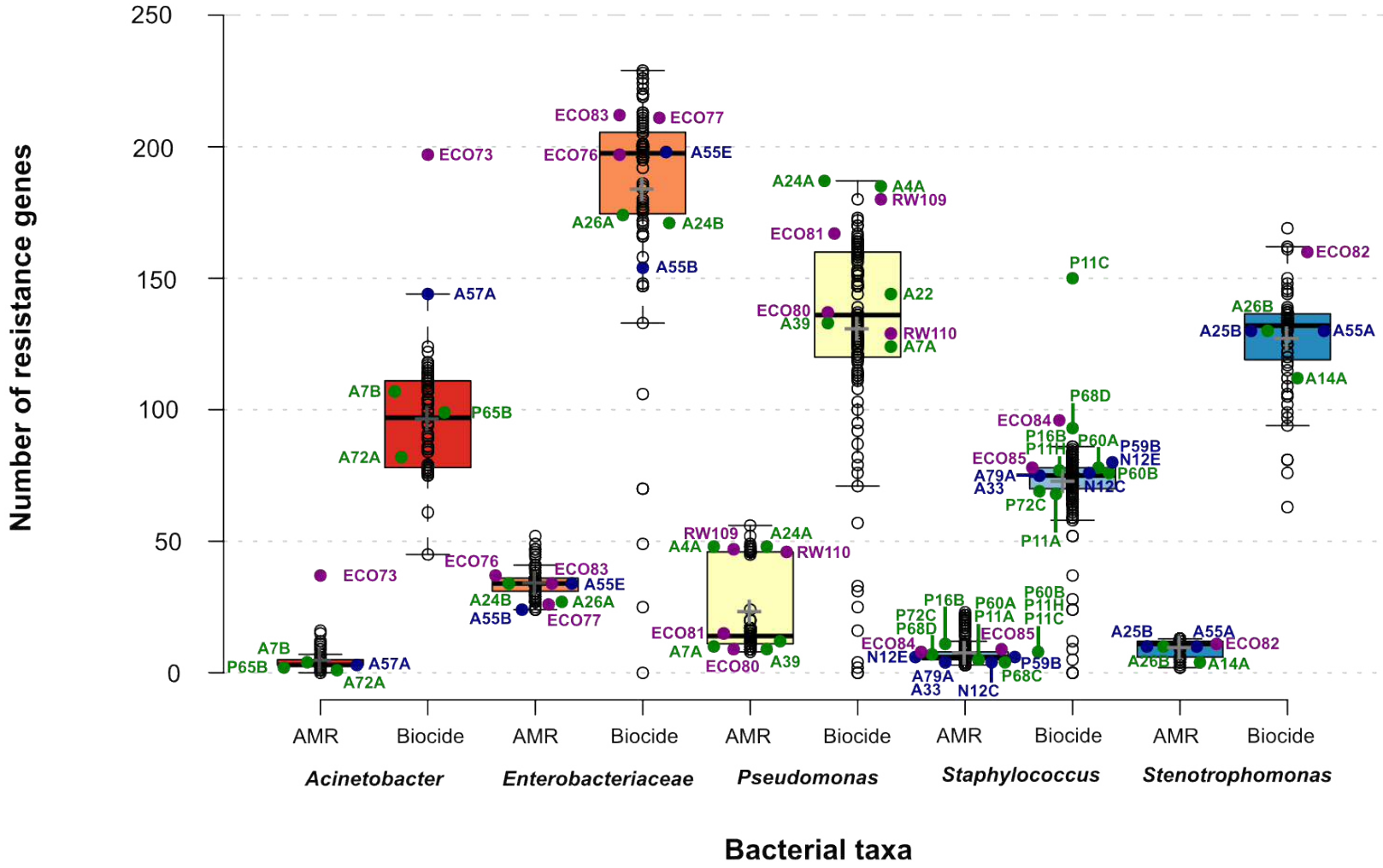


Figure 2.7: Total number of biocide resistance and AMR genes predicted in genomes of factory isolates. Environmental (green) and in-line (blue) factory isolates and biocide reference (purple) isolates of *Acinetobacter*, *Enterobacteriaceae*, *Pseudomonas*, *Staphylococcus* and *Stenotrophomonas* are shown. Genomes of reference isolates are shown in black. The mean is indicated on each boxplot by a grey cross, whilst the median is shown using a black line. Whiskers were defined using the Tukey HSD test.

2.5.4. Factory *Pseudomonas* sp. possess the highest number of predicted Biocide resistance genes, whilst *Staphylococcus* possess the lowest

The mean number of predicted biocide genes was highest for *Pseudomonas* factory isolates and lowest for *Staphylococcus* factory isolates (see Table 2.5). The number of predicted genes was also highly variable and was comparable between reference and factory isolates. Notably, the number of predicted biocide resistance genes was lower from the genomes of factory *Enterobacteriaceae* (85 ± 13) in comparison to reference genomes (184 ± 32). The high number of predicted genes in the *Enterobacteriaceae* reference dataset was largely attributable to the genomes of biocide reference isolates *Enterobacter* sp. ECO76 (197 predicted genes), *Klebsiella* sp. ECO83 (212 predicted genes), and *Pluralibacter gergoviae* (*P. gergoviae*) ECO77 (211 predicted genes). The number of predicted biocide genes was substantially higher in the biocide reference strains than factory isolates, with the exception of *Stenotrophomonas*. Notably, in the *Staphylococcus* dataset, *S. pasteuri* P11C possessed 150 predicted biocide resistance genes, compared to biocide reference isolates *Staphylococcus epidermidis* ECO84 (96 predicted genes) and *S. aureus* ECO85 (74 predicted genes).

2.5.5. Nucleotide variant prediction highlights that organisms from different factory areas may originate from the same source

Two pairs of isolates were identified with ≥99% ANI to each other, and the same taxonomically closest reference genome (see Table 2.4). The pairs of isolates were *Acinetobacter junii* (*A. junii*) A7B, *A. junii* A57A, *P. aeruginosa* A4A and *P. aeruginosa* A24A. A7B and A57A possessed 99.99% identity in comparison to each other, and both possessed 98.46% identity with *A. junii* AJ 068. A4A and A24A possessed 99.99% identity in comparison to each other, and each possessed 99.45 % identity with *P. aeruginosa* XMG. A7B is likely to be ancestral to A57A, as A57A possesses a greater number of polymorphisms than A7B when compared to *A. junii* AJ 068 as follows: 615 more SNPs, 145 more MNPs, 10 more DELs, 7 more INS and 450 more COMPLEX. A24A is likely ancestral to A4A, as A4A possesses a greater number of polymorphisms than A24A as follows: 57 more SNPs, 11 MNPs, 3 more DELs, 1 more INS, and 11 more COMPLEX.

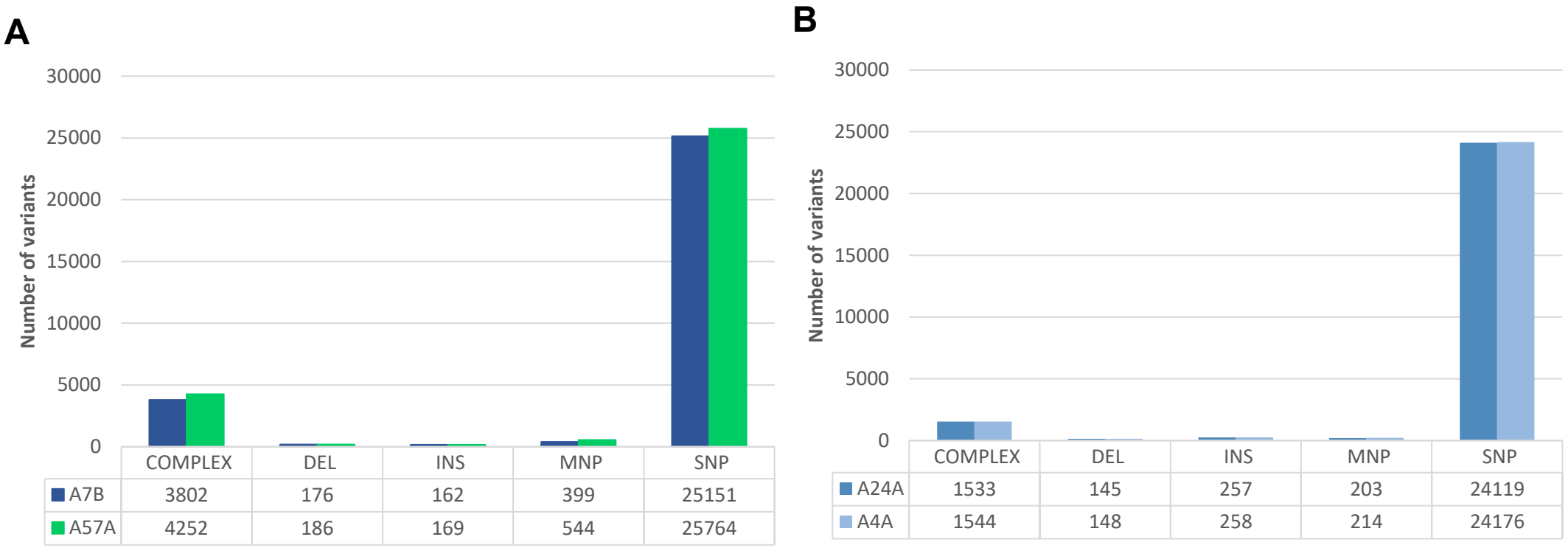


Figure 2.8: Using SNP analysis to compare relatedness of *Acinetobacter* and *Pseudomonas* factory strains. Summary of all predicted nucleotide variants for *A. junii* isolates (A) and *P. aeruginosa* isolates (B) based on the output of Snippy, in comparison to their closest reference genome, as determined by FastANI. Isolates were selected based on having an ANI value of ≥99% in comparison to each other and having a common closest reference genome (individual ANI values are shown in each panel). Variants are as follows: SNP – Single Nucleotide Polymorphism (e.g. A → T), MNP - Multiple Nuclotide Polymorphism (e.g. AT → CG), INS – Insertion (e.g. CT → CGT), DEL – Deletion (e.g. CGT → CT), COMPLEX – combination of SNPs and MNPs.

2.6. Discussion and conclusions

The results from this ecological survey show that organisms isolated from home and personal care factories at sites 1 and 2 are primarily commensal, Gram-positive organisms, which are of little risk to product, and harbour similar numbers of AMR genes to related organisms from non-industrial backgrounds. The predominant organisms isolated from each site were *Bacillus*, (accounting for 25.2% and 12.2% of viable organisms isolated from Site 1 and Site 2 respectively), and *Staphylococcus*, (accounting for 15.9% and 9.5% of isolated viable organisms from site 1 and site 2). To date, this is the first ecological survey of home and personal care product factories, but in the food industry, organisms predominantly isolated are Gram-negatives such as *Pseudomonas* spp., *Enterobacteriaceae* and *Acinetobacter* spp. Overall, a greater diversity of organisms were observed at site 2 in comparison to Site 1. Similar observations have been seen in seminal studies which detailed the improved recovery and consequential diversity of microorganisms when using transport swabs containing Amie's media, similar to those used at Site 2, as opposed to cotton swabs, which were used at site 1 (Barry *et al.* 1972). In particular, the abundance of Gram-negative bacteria was greater at site 2 (21/74 = 28.4%) than site 1 (8/107 = 7.5%). This increase in abundance may be attributable to the use of an initial passage on the minimal media R2A at site 2 prior to passaging onto rich media, which has been evidenced as beneficial for the isolation of bacteria from other routinely treated and sterilised areas of industry, such as treated water sources (Penna *et al.* 2002; Sandle 2015). Isolated organisms of interest were predominantly from environmental areas, (12/17), whilst non-environmental isolates were isolated from non-operational in-line areas, prior to routine sterilisation. The number of predicted AMR genes in factory isolates were extremely similar to related organisms from different environments. *Pseudomonas* factory isolates possessed the highest number of predicted AMR genes, which was attributable to the high number of predicted AMR genes in *P. aeruginosa* (see Table 2.4), which has been seen previously in genomic studies of industrial (Weiser *et al.* 2019) and clinical (Jeukens *et al.* 2017) *P. aeruginosa* strains alike. Similarly, the number of predicted biocide genes were highest in *P. aeruginosa* factory isolates, but the largest number of biocide resistance genes across all bacterial taxa datasets were in *P.*

gergoviae ECO77 and *Klebsiella* sp. ECO83. Though biocide resistance of the *Enterobacteriaceae* has not been studied in a genomic context previously (see Chapter 5 for more on the genomics of *Enterobacteriaceae*). Despite this, the high prevalence of predicted biocide genes from genomes within this group is unsurprising, given that numerous biocide resistance mechanisms have been characterised using an array of phenotypic and molecular methods within *Enterobacteriaceae*, such as *P. gergoviae* (Périamé *et al.* 2014; Périamé *et al.* 2015) previously.

This study however does not come without its caveats. Ultimately, the sampling protocol used in this study may not produce a true measure of organism diversity. One reason for this, is that it is entirely possible for organisms of the same species to possess the same virulence/preservative resistance potential, whilst possessing different phenotypes (Bergmiller *et al.* 2017). In a similar vein, seemingly phenotypically identical colonies can be genotypically variable (Vandamme *et al.* 1997). Thus, several colonies with unique genotypes may not have been isolated in our studies, as they were assumed to be identical to other isolated colonies. The protocol also has the caveat of resulting in the isolation and genome sequencing of clonal isolates, due to a lack of resolution by 16S. Simple PCR profiling strain genotyping approaches such as Random Amplification of Polymorphic DNA (RAPD) which uses short-length primers (8-12 nucleotides) to amplify fragments across the genome to produce a semi-unique profile, could have been a cost-effective screen approach to ensure no clones were subject to sequencing. This technique is however subject to a high level of variability when amplifying DNA from the same organism, and whilst effective, requires optimisation on a genus by genus basis in order to be a reproducible assay, such as those used for *Burkholderia* (Mahenthiralingam, Campbell, Henry, *et al.* 1996) and *Pseudomonas* spp. (Mahenthiralingam, Campbell, Foster, *et al.* 1996).

Furthermore, this protocol does not account for the presence of viable but not culturable (VBNC) cells in factories (Zhao *et al.* 2017). Bacteria undergo a transition to a VBNC state when they are subject to adverse conditions that can be detrimental to their viability. Whilst these bacteria cannot be cultured utilising conventional media, they maintain measurable metabolic activity, any virulence factors that they possess, and may be revived under the correct conditions

(Ramamurthy *et al.* 2014). It is believed that a foodborne outbreak of enterohemorrhagic *E. coli* O157 originated from bacteria in salted salmon roe that were initially undetected in the VBNC state, but were subsequently revived (Zhao *et al.* 2017). As such, it not unreasonable to assume that VBNC bacteria may be present when subjected to the harsh conditions of the HPCP industrial environment.

Additionally, this study relied upon TVCs for the enumeration and isolation of organisms, and used the conventional nutrient rich medium, TSA. The use of this approach may account for the loss of five morphologically distinct colonies identified during initial total viable counts during passage onto TSA, which may have been transitioning into a VBNC state, which is less tolerant to culturing on rich media (Ramamurthy *et al.* 2014).

The conclusions in relation to the objectives of this chapter were:

1. **To characterise the diversity and trends of cultivable bacteria and fungi present in HPCP factories in two global manufacturing sites**. The conclusion was that cultivated organisms from both sites were primarily Gram-positive bacteria. A greater diversity of organisms was cultivated from Site 2 than Site 1, suggesting that the methodology used for isolation at Site 2 was more appropriate for capturing the diversity of microorganisms present at manufacturing sites.

2. **To determine whether objectionable bacteria and fungi could be cultivated from in-line (directly involved in product manufacture) and environmental (not involved in manufacture of product) areas of HPCP factories.** Objectionable organisms were primarily isolated from environmental sources. Objectionable organisms that were cultivated from in-line areas originated from samples taken in areas with higher levels of moisture.

3. **To achieve genomic characterisation of bacteria of interest isolated from the HPCP factory environment**. Of the 31 organisms subject to whole genome analysis, average nucleotide identity placed all organisms

firmly within a specific genus, whilst 25/31 were provided with a confirmed species-level identity. This analysis was largely concordant with initial identification by 16S, with the exception of two *Pseudomonas* species, which were misidentified by 16S at the species-level, and one *Enterobacter*iaceae isolate, which was assigned an incorrect genus-level identity by 16S. Furthermore, nucleotide polymorphism analysis showed that it was possible to determine that two objectionable organisms *A. junii* A7B and *P. aeruginosa* A24A were likely transferred to another area within the same factory, from which their descendents *A. junii* A57A and *P. aeruginosa* A4A were isolated.

4. **To establish whether panels suggested by the British Pharmacopeia, United States Pharmacopeia and Cosmetic, Toiletries and Perfumery Association and supplemented with in-house biocide testing isolates are representative of organisms present in manufacturing environments**. With the exception of *Staphylococcus* and *P. aeruginosa* isolates, all objectionable organisms present in the factory environment are taxonomically distinct from those used in biocide testing panels. All isolates possessed fewer predicted AMR and biocide resistance genes than biocide testing strains within the reference panel, with the exception of one *Staphylococcus*, which was non-*S.aureus*, and from an environmental, low-risk, routinely sterilised location.

Overall, the hypothesis for this chapter "The HPCP factory environment constitutes a source of microbial species frequently encountered as industrial contaminants" was rejected, as objectionable organisms were primarily isolated from environmental locations, subject to routine sterilisation.

3. *Burkholderia cepacia* complex bacteria

3.1. *Burkholderia cepacia* COMPLEX BACTERIA AS INDUSTRIAL CONTAMINANTS

Burkholderia cepacia complex (Bcc) bacteria are historically linked with the ability to cause contamination in both clinical and non-food industrial settings (Torbeck *et al.* 2011). This group of Gram-negative bacteria has undergone multiple taxonomic revisions over the last 20 years with over 20 species defined within the *B. cepacia* complex (Depoorter *et al.* 2016). The historical prevalence of individual *B. cepacia* species in contamination is more difficult to assess since it was reported as "*Pseudomonas cepacia*" up to 1997 prior to taxonomic reclassification (Vandamme *et al.* 1997), and even post 2010, it is still occasionally referred to as "*B. cepacia*" in the published literature (Torbeck *et al.* 2011). Unlike *Pseudomonas aeruginosa*, *B. cepacia* complex bacteria are not straightforward to identify (Table 1), requiring selective media for their enrichment (Henry *et al.* 1999), and molecular tests such as *recA* (Mahenthiralingam *et al.* 2000) gene sequencing and Multilocus Sequence Typing (MLST) (Baldwin *et al.* 2005) for accurate identification. Surveillance and tracking of infections caused by *B. cepacia* complex bacteria has been extensive because of the problematic lung disease they cause in people with cystic fibrosis, and MLST resources for strain molecular epidemiology are comprehensive (Baldwin *et al.* 2005) (https://pubmlst.org/bcc/).

In the last five years, analysis of non-food product contamination caused by *B. cepacia* complex bacteria has been more detailed in relation to reporting taxonomic identity. Species associated with a variety of non-food product sources were accurately identified within a collection of 60 industrial isolates, and *Burkholderia lata* (*B. lata*) *Burkholderia cenocepacia* (*B. cenocepacia*), and *Burkholderia vietnamiensis* (*B. vietnamiensis*) were the top three species accounting for 25%, 18%, and 13% of the collection assessed respectively (Rushton *et al.* 2013). In 2016, a strain of *Burkholderia stabilis* (*B. stabilis*) was shown to cause an outbreak associated with contaminated washing gloves in Switzerland using a comprehensive range of molecular and genomic tests (Sommerstein *et al.* 2017). Whole genome sequencing including whole genome MLST (Maiden *et al.* 2013) was carried out to identify that the *B. stabilis* isolates causing clinical infections in multiple patients at different treatment centres were identical to those recovered from the contaminated gloves (Sommerstein *et al.* 2017). In the United States, a widespread outbreak associated with the laxative docusate

occurred in 2016, and isolates causing the resulting infections in patients shown to be identical using a PCR genotyping method (Marquez *et al.* 2017). However further molecular analysis demonstrated that the docusate outbreak strain belonged to a potentially novel taxonomic group within the *B. cepacia* complex.

Although improvements in the level of resolution being applied to the investigation of *B. cepacia* complex contamination are clearly being made, studies may still fail to follow up on the resources obtained during the investigation of outbreak incidents. For example, while whole genome sequencing was applied to demonstrate that a clonal *B. cepacia* complex strain present in a contaminated octenidine mouthwash had caused an outbreak of clinical infections, further follow-up to identify the strain to a species-level was not performed (Becker *et al.* 2018). The availability of whole genome sequence data offers multiple bioinformatic approaches to accurately assign species status (Bull *et al.* 2012) and excellent comparative resources for sequence data are available at the *B. cepacia* complex MLST site. Using the deposited whole genome sequence of the octenidine contamination isolates (Becker *et al.* 2018), an MLST sequence type was derived, ST-881. This showed that the mouthwash and outbreak strain was a member of the *B. cepacia* complex species *Burkholderia arboris* (*B. arboris*). Although, this retrospective analysis of recently deposited genome sequence data for a *B. cepacia* complex contaminant (Becker *et al.* 2018) was carried out using command-line bioinformatics, the same identification result of *B. arboris* was also obtained when the sequence data was uploaded on the web to the MLST database and a search for matching sequence loci initiated (https://pubmlst.org/bcc/). Beyond phenotypic testing, to date, no extensive genomic characterisation has been conducted for industrial *Burkholderia*, where our understanding of the taxonomic diversity of industrial organisms currently sits at the level of MLST. Similarly our understanding of the genomic basis of their phenotypic preservative resistance in these organisms sits at the level of gene microarrays (Rushton *et al.* 2013), which whilst useful, can only account for a specified repertoire of expressed genes as opposed to evolution of the entire genome.

The aim of this chapter was to gain a genomic understanding of industrial *Burkholderia* strains, with specific objectives investigated as follows:

1. **To characterise a panel of industrial Bcc spp. in terms of their taxonomic diversity.**

2. **To determine the appropriate techniques for accurate identification of closely related Bcc species, with a focus on *B. cenocepacia*, *B. lata* and *B. vietnamiensis* as priority industrial contaminants.**

3. **To determine key genomic features underpinning changes in expression of efflux pump genes both preservative-trained and historically industrially derived *B. lata* isolates.**

The hypotheses for this chapter are as follows:

I. Whole genome analysis can be used to accurately identify Bcc species from historical identification incidents, with a greater level of resolution than MLST.

II. Key mutational signatures can be identified in both preservative-trained and historically industrially derived *B. lata* isolates.

3.2. Materials & Methods

3.2.1. DNA extraction from industrial *Burkholderia*

DNA extraction was performed for a total of 74 isolates. 62 isolates were of industrial environmental origin historically identified as Bcc spp. by sequencing of the 16S ribosomal RNA gene (see 2.4.7), or recombinase A (*recA*) gene (Mahenthiralingam *et al.* 2000), using the Maxwell®16 Tissue DNA Purification Kit and instrument. Isolates were revived via plating onto TSA, followed by inoculation of 3 ml TSB for overnight culture (see 2.4.2), which were pelleted before genomic DNA was extracted. Extractions were also performed for *B. lata* 383-BIT *B. lata* 383-CMIT, *B. lata* 383-MIT, and *B. lata* 383-BC. These isolates are preservative-adapted derivatives of the environmental strain *B. lata* 383, and were specifically adapted to the preservatives Benzisothiazolinone (BIT), Chloromethylisothiazolinone (CMIT), Methylisothiazolinone (MIT) and Benzethonium Chloride (BC), respectively, using an agar passage methodology

(Rushton *et al.* 2013). The isolates were adapted on a minimal media known basal salts medium (BSM), the recipe of which was modified to contain glucose (BSM-Gluc) as a carbon source, as opposed to glycerol. Preservative adapted derivatives were therefore plated from freezer stocks onto BSM-Gluc agar, before using confluent material to create 3 mL overnight cultures in BSM-Gluc broth (18h, 30°C). The same protocol was used to extract DNA from eight intermediate strains generated during stepwise training of 383^T to BIT.

3.2.2. Sequencing of isolates

Sequencing was performed using extracted genomic DNA using an Illumina NextSeq 500 platform using a Nextera XT DNA library preparation kit. Preservative-adapted isolates of *B. lata* 383 (n = 4), including intermediate passaging strains generated during stepwise training to the preservative BIT, were sequenced using the same sequencing platform, but using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® as described previously (see 2.4.9).

3.2.3. Genome assembly and quality-filtering of genomes

Four approaches were evaluated for the assembly of Bcc genomes. The first approach utilised SPAdes v3.14.0, a genome assembler designed for bacteria, fungi, and other small genomes (Bankevich *et al.* 2012). The second approach utilised SPAdes in tandem with Pilon (Walker *et al.* 2014), a genome correction tool which works to correct erroneous bases, gaps and other mis-assemblies in a draft genome. The third used Shovill (Trinetta *et al.* 2019) (https://github.com/tseemann/shovill), an assembly pipeline which utilises SPAdes, but attempts to reduce assembly time by using features such as read sub-sampling, as dictated by an estimated genome size provided by the user. The final approach used Unicycler for assembly (see 2.4.10). During initial genome quality control (QC), assembled contigs were screened for contamination using Kraken2 v2.0.8-beta (see 2.4.10). The N50 of each assembly was generated using Quast v4.4 (see 2.4.10), and used as a quality metric alongside the MLST profile from each genome (see 3.2.4 for more details), when choosing genomes to remove in the QC process and to take forward for further genomic analyses.

3.2.4. Determining the diversity of industrial *Burkholderia* isolates using MLST

MLST profiles were generated using the command-line tool MLSTcheck (Page *et al.* 2017) (https://github.com/sanger-pathogens/mlst_check). This tool uses

BLASTN to compare a query genome to PubMLST databases. PubMLST is a collection of public databases within which exist a variety of individualised schemes, targeting 7 gene loci for molecular typing and determining microbial genome diversity. To date, 134 schemes exist for a number of bacteria, including Bcc spp. MLST has been used for a number of years to determine organism identity down to the species-level (Maiden *et al.* 2013), and was historically determined by means of sequencing of PCR products. A number of tools now exist however which can utilise whole genome data to both automatically detect the appropriate MLST scheme for an organism, and provide a full MLST profile (Page *et al.* 2017). MLSTcheck will query a genome using a user-selected scheme from PubMLST and identify the seven loci within a genome for a given scheme. Each unique sequence at a given locus is then assigned an arbitrary and unique allele number, with the unique assemblage of allele numbers being used to define an overall unique arbitrary number for the organism, called a sequence type (ST).

MLSTcheck also provides the user with a concatenated MLST allele alignment, which can be used for phylogenetic analysis. Where historical MLST profiles had been defined by PCR, command-line generated profiles were correlated to those in Bcc PubMLST database for confirmation of correct genome ID. The taxonomically closest genome for all genomes which passed QC was identified using FastANI (see 2.4.11) by comparing to a local database consisting of all available (n = 3002) *Burkholderia* genomes in ENA (up to May 2019), and additional genomes assembled from all available reads using Shovill (up to October 2019). Correct species-level identity was then confirmed by using FastANI, followed by PyANI, to compare each genome against the type strain or reference genome for the proposed species name from both MLST profiling, and ENA. For example, if the proposed species identification for a genome was *B. lata* based on historical MLST, but the closest matching genome from ENA was labelled as *B. cepacia*, the genome was compared against type strain genomes for both *B. lata* and *B. cepacia*.

3.2.5. Developing curated databases of *B. cenocepacia*, taxon K and *B. vietnamiensis*

The genomes of isolates confirmed as Bcc taxon K, (i.e. *Burkholderia contaminans*, (*B. contaminans*), *B. lata* and undefined taxon K species) (Vanlaere *et al.* 2009), *B.*

cenocepacia and *B. vietnamiensis* by MLST were taken forward for further confirmation of species identity using PyANI (see 2.4.11). These isolates were chosen as organisms of interest based on being the dominant species in previous studies examining the taxonomic diversity and preservative resistance of a panel Bcc spp. isolated from non-food product sources (Rushton *et al.* 2013).

A taxon K database was developed using 60 genomes, including the genomes sequenced in this study, and all available genomes in the ENA database for taxon K isolates and related species. Within this dataset, there were 26 genomes from industrial isolates of interest. Of these, 13 were previously identified by *recA* gene sequencing or MLST as *B. lata*, 5 were classified as *B. contaminans*, and 8 were identified as a novel species subgroup within taxon K. The dataset also contained the complete genomes for type strains of *B. contaminans* (LMG 6992^T), *B. lata* (383^T), and an industrially-relevant *B. lata* isolate (A05) isolated from a mouthwash, preserved with Chlorhexidine (Leong *et al.* 2018).

A *B. cenocepacia* database was developed using 64 genomes from across historical genomovars III-A, III-B, III-C, and III-D. Of these genomes, 11 were derived from industrial isolates of interest. This dataset also included genomes used in recent studies describing the presence of an additional species split within *B. cenocepacia*, supported by phylogenetic and core-gene analysis, and additionally proposing the reclassification of *B. cenocepacia* III-B as *Burkholderia servocepacia* sp. *nov* (Wallner *et al.* 2019).

A *B. vietnamiensis* database was developed using a total of 38 genomes, including one genome sequenced in this study, and a curated panel of 36 isolates. There were five previously industrial isolates within this dataset, making for a total of 6 industrial isolates of interest. All genomes within the dataset were sequenced at Cardiff University (publicly available from ENA under project PRJEB9765), excluding the complete genomes for the type strain LMG 10929^T (accession: GCF_000959445.1) and *B. vietnamiensis* G4 (accession: GCF_000016205.1).

3.2.6. Phylogenetic analysis in *B. cenocepacia*, *B. lata* and *B. vietnamiensis*

Databases constructed as described in 3.2.5 were used to generate MLST and core-gene phylogenies in order to determine the level of genomic resolution required to distinguish closely related Bcc spp. An MLST phylogeny was generated using

MLSTCheck (Page *et al.* 2017), which uses BLASTN to produce an alignment of all detected MLST genes in a set of genomes, using a user specified MLST scheme. Genomes were then annotated using Prokka (v1.14.5) (Seemann 2014) under default parameters. Prokka uses BLASTP, and external feature prediction tools to predict the following genomic features in contigs; coding sequences (Prodigal), transfer RNA gene (Aragorn), signal leader peptides (SignalP), ribosomal RNA genes (RNAmmer), and non-coding RNA (Infernal). Accurate strain-to-strain level relatedness was determined phylogenetically using Roary (v3.12.0) (Page *et al.* 2015) under default parameters. Roary uses sequence annotations to define the core, accessory and pan genome of closely related bacterial isolates, and produce a core-gene alignment. Under default parameters, a core-gene was defined as a gene present within 99% of annotations provided. Alignments were produced using 304 core-genes in *B. cenocepacia*, 48 core-genes in taxon K, and 625 core-genes in *B. vietnamiensis*. Both alignments were then used to generate phylogenies using RAxML and a GAMMA model of rate heterogeneity supported by 100 bootstraps (see 2.4.7). Genomes were then placed amongst custom databases by ANIm using PyANI (see 2.4.11). For ANI, thresholds of 95.0% (ANI95) and 97.0% (ANI97) were investigated to determine which of these boundaries are required to distinguish phylogenetically distinct members of the Bcc using this type of genomic taxonomy analysis. Phylogenetic trees were visualised and edited in FigTree v1.4.2 and Inkscape v0.91 respectively. ANI95 and ANI97 analyses were visualised using the outputs of PyANI in tandem with the heatmap package in R.

3.2.7. Identifying unique genomic features of *B. lata* 383 adapted to historical and in-use preservatives

Sequencing reads generated from *B lata* 383-BIT, intermediate passaging strains, 383-CMIT, 383-MIT and 383-BC were investigated for SNPs using Snippy v4.1.0, by aligning the reads to the complete *B. lata* 383 genome as described previously (see 2.4.13). The tabulated output of Snippy was used to identify variants, which were extracted into a new table, before common variants and hypothetical proteins were manually filtered out. Variants were called and categorised as described in 2.4.13. Snippy outputs were used to describe predicted effects of gene mutations on encoded proteins. Effects described included synonymous mutations with no predicted phenotypic effects, missense mutations to describe mutation

resulting in an amino acid change in the primary structure of an encoded protein, frameshift mutations which shift open-reading frames and alters translation, and stop:gain mutations; amino acid changes which result in the introduction of a stop codon. *B. lata 383* derivatives adapted to BIT (383-BIT) and CMIT (383-CMIT), as well as *B. lata* industrial process contaminant isolates BCC1294, BCC1296 and BCC1406, were chosen for analysis as isolates of interest for further investigation of polymorphisms. Observations in previous studies had shown that B1004, a gene present on chromosome two of *B. lata* 383 as a part of a coregulated RND efflux pump gene cluster, B1004 to B1006, was upregulated in these isolates (Rushton *et al.* 2013). These five isolates were investigated with the hypothesis that the upregulated expression of B1004 could be linked to mutations in upstream regulators. For *B. lata* 383, the nearest upstream regulator was identified as B1003, a *soxR* type regulator encoding gene. Therefore, 383-BIT and 383-CMIT were investigated with the assumption that mutations would be identified in the upstream regulator genes.

For the industrial isolates BCC1294, BCC1296 and BCC1406, the nearest neighbour as determined by core-gene phylogenetic analysis (generated as described in 3.2.6), was used as a reference genome. Once a reference genome was identified for these isolates, B1004 homologues were identified using NCBI BLASTN, before Snippy outputs were queried for mutations in the nearest upstream regulator from this gene. To confirm the synteny of genes between the regulator and the nearest B1004 homologue for BCC1294 BCC1296 and BCC1406, a BLASTN search was performed using nucleotide sequences from the start of the first identified upstream regulator, to 2000 bp downstream of B1004. Gene synteny was then visualised using NCBI coverage maps, and NCBI Sequence Viewer 3.34.0, before editing in Inkscape v0.91.

To determine if the concentration of preservative exposure used in adaptive resistance studies could be associated with specific nucleotide variants, a core-SNP alignment was generated using the snippy-core function of Snippy, using reads from the endpoint BIT-adapted isolate 383-BIT, alongside any isolates generated as intermediates between the parent strain and 383-BIT during stepwise training. A phylogenetic tree was then generated using RAxML and a GAMMA model of rate heterogeneity supported by 100 bootstraps (see 2.4.7). The

phylogenetic tree was visualised using FigTree v1.4.2 and edited using Inkscape v0.91 (see 2.4.7).

3.3. Results

3.3.1. Genome sequences were of variable quality, but successfully filtered down to 51 genomes suitable for analysis

A total of 51 of the 62 of the initially sequenced genomes passed QC and were taken forward for downstream taxonomic and phylogenetic analysis (Table 3.1). The remaining 11 isolates had poor quality sequences due to EDTA in the DNA extraction kit interfering with the library normalisation process. Successful assemblies were obtained for 33 of the 51 by Unicycler, whilst only 18 of 51 were produced using the Spades and Pilon assembly pipeline. 25 of the 51 isolates had ST types previously defined by PCR. The genome derived MLST profiles of 23 of these isolates exactly matched previously defined profiles. Two isolates, BCC1312 and BCC1404, displayed different ST types, (BCC1312 = 425, BCC1404 = 840), to those previously defined (BCC1312 = 328, BCC1404 = 205), but this did not change the species-level identification of these isolates. A total of 26 of the Bcc genomes did not have previously defined MLST profiles.

When analysed using alignment-free ANI, 14 isolates below the default threshold for detection (80%). All isolates that fell below the threshold for detection by alignment-free ANI however displayed ANI values above 80% in comparison to *Burkholderia* type strains when evaluated using alignment-based ANI. Alignment-based ANI was therefore used as a more reliable means of determining species interrelatedness for this dataset. 42 of the 51 isolates could be identified to the species-level by MLST. The MLST species ID was confirmed by ANI for 22 isolates. ANI values fell below the 95% cutoff for species identification for the species ID proposed by both MLST and ENA for the remaining 20 isolates.

3.3.2. Genomic diversity by MLST highlights the presence of 10 named Bcc spp., and one group of unnamed Bcc *spp* in the industrial strain collection.

Predominant isolates identified by MLST were *B. lata* (*n* = 13), *B. cenocepacia* (*n* = 11), followed other Bcc species (*n* = 9). Of the *B. cenocepacia* isolates, the

STs of 4 of the 11 isolates corresponded to members of genomovar III-A, whilst the remaining isolates were of the III-B lineage. All *B. cenocepacia* displayed ANI values above the 95% threshold when compared to type strain J2315^{T} (97.4-98.6%) (see Table 3.1). No *B. lata* isolate displayed genomic ANI values above the 95% threshold when compared to type strain 383^{T}, displaying values of 94.6 – 94.9% (see Table 3.1).

3.3.3. MLST is sufficient for resolution of *B. cenocepacia* III-A and III-B, and is supported by ANI and core-gene analysis

Both MLST and core-gene (*n* = 304 genes) phylogenies supported the ANI classification of 7/11 industrial *B. cenocepacia* isolates as III-B, and 4/11 industrial *B. cenocepacia* isolates as III-A. A clear evolutionary divergence between III-A and III-B was seen in both the MLST and core-gene phylogenies (Figure 3.2). No differences in topology or phylogenetic placement were observed between the 7 gene MLST and 304 core-gene phylogenies. ANI95 was insufficient to support this split, as only 2/35 isolates historically identified as III-B possessed ANI values which fell below the 95.0% for species delineation when compared to the type strain, J2315^{T}, which historically belongs to genomovar III-A (Figure 3.2). However, by increasing the threshold to ANI97, support for the phylogenetic split was observed, as all III-B isolates (*n* = 35) within the dataset fell below the 97.0% threshold for species delineation when compared to J2315^{T}, whilst III-A (*n* = 18) isolates possessed ≥99.0% identity with J2315^{T} (Figure 3.2).

3.3.4. Core-gene phylogenomics reveals incorrect classification of multiple taxon K isolates by MLST

The identification of 14 industrial isolates (including the recently sequenced A05; Leong *et al.* 2018) as *B. lata*, and 5 industrial isolates as *B. contaminans*, (Table 3.1), was supported by MLST phylogeny (Figure 3.5, panel A). This phylogenetically placed the respective type strains for each species alongside the industrial isolates. The MLST profile identification of taxon K isolates was also supported phylogenetically. (Figure 3.5, panel A, clade 2). Whilst core-genome phylogenomics supported the placement of *B. contaminans* and unclassified taxon K spp. as discrete species groups by MLST, this analysis altered the placement of 14 industrial isolates defined as *B. lata* by MLST to a

clade which was phylogenetically distinct clade from the type strain, *B. lata* 383^{T} (Figure 3.5, panel B, clade 1). ANI95 supported the re-classification of the 14 industrial isolates from *B. lata* to a novel species group (Figure 3.3) as observed with the core-gene phylogenetic placement (see Figure 3.5, panel B, clade 1). The ANI of these isolates ranged from 94.6-94.9% in comparison to 383^{T} and ranged from 94.2-94.7% in comparison to *B. contaminans* LMG 23361^{T}. In contrast, the ANI for the 14 isolates ranged from 98.6-99.7% when compared to recently sequenced strain A05 (Leong *et al.* 2018). Similarly, the ANI values fell below 95.0% for all taxon K isolates in core-gene clade 2 (Figure 3.3), when compared to *B. lata* 383^{T} (93.9-94.9%) and *B. contaminans* LMG 23361^{T} (93.4-94.5%), but displayed ANI values ranging from 96.3-99.9% when compared to each other (using BCC1282 as an arbitrary reference).

ANI95 was not sufficient to support the phylogenetic grouping of *B. contaminans*, with 4 of 5 industrial isolates displaying ANI values above the threshold when compared to both *B. lata* 383^{T} (94.8-95.1%) and *B. contaminans* LMG 23361^{T} (97.4-99.9%) (Figure 3.3). ANI95 additionally could not resolve isolates in core-gene clade 1 (see Figure 3.3, panel B, clade 1), and clade 2 (see Figure 3.4, panel B, clade 2), with taxon K isolates in clade 2 possessing ANI values of 94.9-95.5% in comparison to strain A05 in clade 1. ANI97 did however support the distinct phylogenetic grouping suggested by core-gene phylogenomics in all isolates, with the exception of taxon K isolate BCC1410, which possessed 95.0-96.5% identity in comparison to other taxon K isolates.

3.3.5. *B. vietnamiensis* isolates show no major evolutionary divergence, and MLST is sufficient for their identification

Analysis of 38 *B. vietnamiensis* genomes by ANI showed that all genomes were extremely similar to one another, all possessing 97.8-100% identity when compared to the type strain LMG 10929^{T} (Figure 3.4). The grouping of these isolates was thus supported by both ANI95 and ANI97. Both MLST and core-gene phylogenetic analyses showed no evidence of major phylogenetic clades.

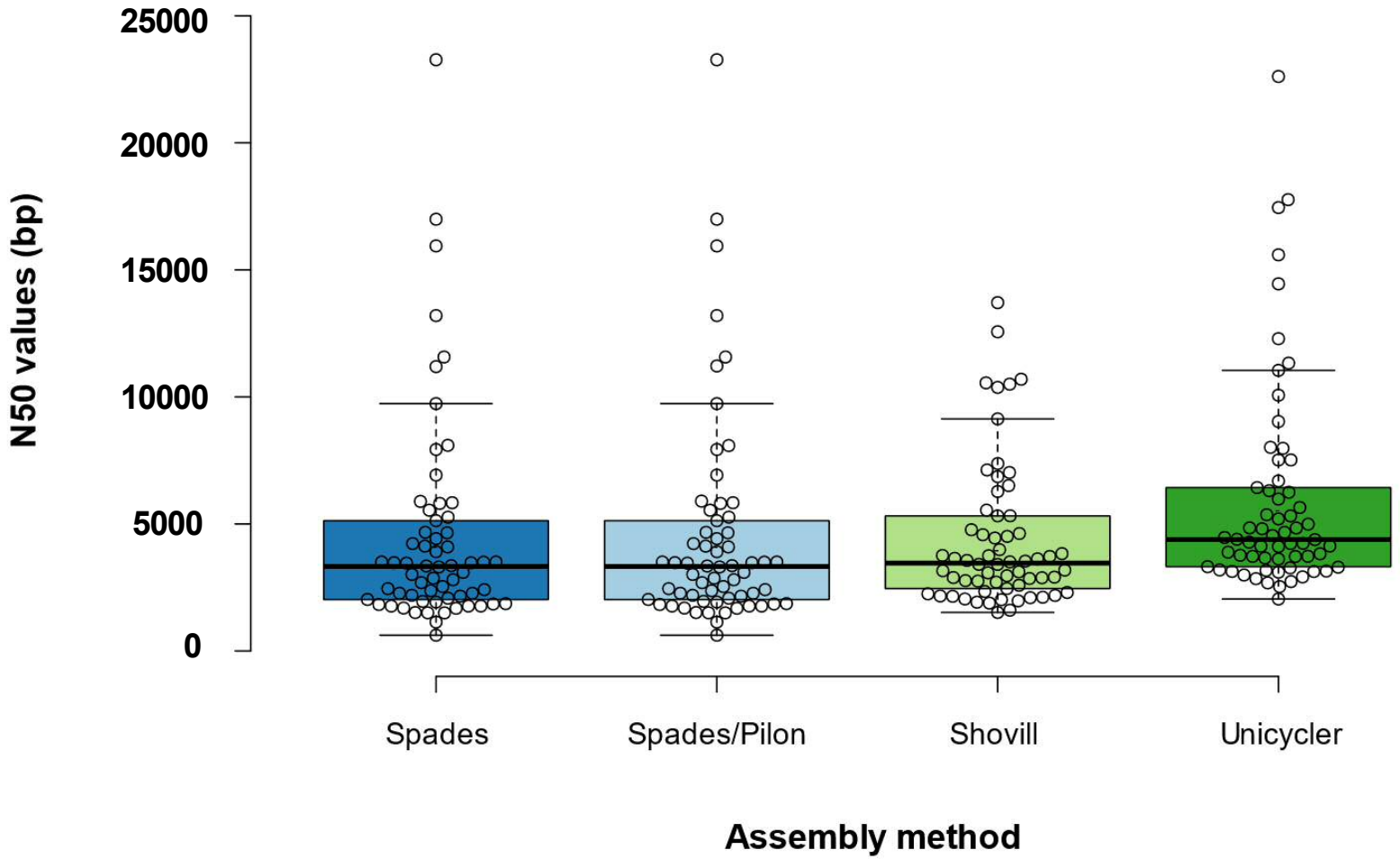


Figure 3.1: Assembly quality N50 values for 62 industrial *Burkholderia* genomes, excluding preservative-adapted *B. lata* isolates. The boxplot shows N50 values for all genomes sequenced in this study, prior to QC and removal of low-quality genomes. The mean N50 value for each assembly method is indicated by a black cross, whilst the median is indicated by a black line. Whiskers were defined using the Tukey HSD test.

Table 3.1: Diversity of 52 industrial Bcc genomes by MLST and ANI

Isolate	Origin	ST	MLST ID	Closest genome ENA		ANI vs type strain (MLST)			ANI vs type strain (ENA)		
				Reference genome	ANI (%)	Strain	F (%)	P (%)	Strain	F (%)	P (%)
BCC1282	IND	333	Other BCC	*B. cenocepacia* SRR3575886	99.1				J2315^{T}	91.8	92.4
BCC1283	IND	250	*B. cenocepacia* III-A	*B. cenocepacia* SRR3578863	99.8	J2315^{T}	97.8	99.1			
BCC1284	IND	98	*B. lata*	*B. lata* BCC0217	99.8	383^{T}	93.2	94.6			
BCC1285	IND	98	*B. lata*	*B. lata* BCC0217	99.9	383^{T}	93.2	94.7			
BCC1288	IND	98~	*B. lata*	*B. lata* BCC0217	99.5	383^{T}	Below	94.8			
BCC1289	IND	98	*B. lata*	*B. lata* BCC0217	99.9	383^{T}	93.4	94.7			
BCC1290	IND	98	*B. lata*	*B. lata* BCC0217	99.7	383^{T}	92.8	94.8			
BCC1292	IND	322	*B. cenocepacia* III-B	*B. cenocepacia* SRR3575867	99.6	J2315^{T}	92.6	95.4			
BCC1293	IND	322	*B. cenocepacia* III-B	*B. cenocepacia* SRR3575867	99.5	J2315^{T}	93.3	95.4			
BCC1294	IND	98	*B. lata*	*B. lata* BCC0217	99.9	383^{T}	93.7	94.7			
BCC1295	IND	241	*B. cenocepacia* III-A	*Burkholderia* sp. 28 3	99.8	J2315^{T}	98.6	99.1			
BCC1296	IND	119	*B. lata*	*B. cenocepacia* SRR3575878	99.0	383^{T}	93.1	94.7	J2315^{T}	90.3	92.5
BCC1297	IND	119	*B. lata*	*B. cenocepacia* SRR3578869	99.1	383^{T}	92.3	94.8	J2315^{T}	89.7	92.6
BCC1298	IND	119	*B. lata*	*B. cenocepacia* SRR3575878	99.1	383^{T}	92.2	94.8	J2315^{T}	89.9	92.7
BCC1299	IND	119	*B. lata*	*B. cenocepacia* SRR3575895	99.1	383^{T}	91.8	94.8	J2315^{T}	89.3	92.8
BCC1300	IND	334~	Other BCC	*B. cenocepacia* SRR3578881	99.1				J2315^{T}	Below	92.9
BCC1302	IND	333	Other BCC	*B. cenocepacia* SRR3575886	99.1				J2315^{T}	89.7	92.7
BCC1303	IND	333~	Other BCC	*B. cenocepacia* SRR3578824	99.0				J2315^{T}	89.7	92.7
BCC1306	IND	325	*B. arboris*	*B. cepacia* ERR2814674	99.2	LMG 24066^{T}	95.0	96.2	ATCC 25416^{T}	91.7	92.6
BCC1307	IND	325	*B. arboris*	*B. cepacia* ERR2820960	99.0	LMG 24066^{T}	94.1	96.2	ATCC 25416^{T}	91.2	92.7
BCC1310	IND	327~	*B. arboris*	*B. cepacia* ERR2820961	98.9	LMG 24066^{T}	95.2	96.3	ATCC 25416^{T}	91.7	92.5
BCC1312	IND	425	*B. arboris*	*B. arboris* BCC0049	99.1	LMG 24066^{T}	94.8	96.3			
BCC1313	IND	335~	Other BCC	*B. cenocepacia* SRR3578869	98.8				J2315^{T}	89.4	92.6
BCC1314	IND	336~	Other BCC	*B. cenocepacia* SRR3578896	99.4				J2315^{T}	87.9	92.8
BCC1316	IND	338~	*B. cenocepacia* III-B	*B. cenocepacia* SRR3575895	99.5	J2315^{T}	Below	95.5			
BCC1318	IND	316~	*B. cenocepacia* III-B	*B. cenocepacia* SRR3578869	99.4	J2315^{T}	Below	95.3			
BCC1319	IND	1089~	*B. dolosa*	*B. cenocepacia* SRR3578898	86.3	AU0158^{R}	Below	81.0	J2315^{T}	Below	83.0
BCC1320	IND	316~	*B. cenocepacia* III-B	*B. cenocepacia* SRR3578852	99.8	J2315^{T}	90.0	95.4			
BCC1321	IND	339~	*B. lata*	*B. cenocepacia* SRR3578896	98.9	383^{T}	91.4	94.9	J2315^{T}	88.5	93.0
BCC1323	IND	323~	*B. contaminans*	*B. cenocepacia* SRR3578819	99.1	LMG 23361^{T}	93.9	97.4	J2315^{T}	89.0	92.9
BCC1395	IND	1677~	*B. arboris*	*B. cenocepacia* SRR3578869	98.4	LMG 24066^{T}	91.6	94.3	J2315^{T}	Below	93.5
BCC1404	IND	840	*B. cepacia*	*B. cenocepacia* SRR3578881	99.4	ATCC 25416^{T}	95.5	97.5	J2315^{T}	89.7	93.0
BCC1406	IND	103~	*B. lata*	*B. cenocepacia* SRR3575886	98.7	383^{T}	Below	94.7	J2315^{T}	Below	93.2
BCC1408	IND	200~	*B. vietnamiensis*	*B. cenocepacia* SRR3575886	100	LMG 10929^{T}	95.7	99.0	J2315^{T}	Below	91.8
BCC1410	IND	455~	Other BCC	*B. cenocepacia* SRR3578824	98.7				J2315^{T}	Below	92.7
BCC1411	IND	597	Other BCC	*B. cenocepacia* SRR3578824	98.8				J2315^{T}	88.9	92.8
BCC1554	IND	119	*B. lata*	*B. cenocepacia* SRR3578896	99.1	383^{T}	93.4	94.7	J2315^{T}	90.4	92.4
BCC1555	IND	653	*B. stabilis*	*B. cenocepacia* SRR3578852	99.1	ATCC BAA-67^{T}	97.5	98.8	J2315^{T}	90.2	92.6
BCC1557	IND	335~	Other BCC	*B. cenocepacia* SRR3578881	99.5				J2315^{T}	93.3	96.4
BCC1558	IND	241	*B. cenocepacia* III-A	*B. cenocepacia* SRR3575878	99.9	J2315^{T}	97.4	99.1			
BCC1559	IND	439	*B. multivorans*	*B. multivorans* BCC0050	99.0	ATCC BAA-247^{T}	Below	97.4			
BCC1560	IND	439	*B. multivorans*	*B. multivorans* BCC0050	99.4	ATCC BAA-247^{T}	95.6	97.4			
BCC1571	IND	202~	*B. cenocepacia* III-B	*B. cenocepacia* SRR3575867	99.5	J2315^{T}	Below	96.6			
BCC1582	IND	102~	*B. contaminans*	*B. cepacia* DHQP2016-12-158	100	LMG 23361^{T}	99.3	99.9	ATCC 25416^{T}	Below	93.9
BCC1590	IND	482	*B. contaminans*	*B. contaminans* ERR1806535	99.6	LMG 23361^{T}	94.7	98.6			
BCC1594	IND	1681~	*B. pyrrocinia*	*B. cenocepacia* SRR3575895	99.1	DSM 10685^{T}	92.8	95.5	J2315^{T}	89.0	93.1
BCC1595	IND	482	*B. contaminans*	*B. contaminans* BCC0123	99.6	LMG 23361^{T}	96.6	98.6			
BCC1596	IND	937	*B. cepacia*	*B. cenocepacia* SRR3575895	99.3	ATCC 25416^{T}	96.1	97.6	J2315^{T}	90.2	92.8
BCC1597	IND	32~	*B. cenocepacia* III-A	*B. cenocepacia* SRR3578881	99.7	J2315^{T}	98.3	99.0			
BCC1598	IND	35	*B. cenocepacia* III-B	*B. cenocepacia* SRR3578837	99.5	J2315^{T}	Below	95.6			
BCC1889	IND	244~	*B. cepacia*	*B. pseudomallei* 7506487	96.3	ATCC 25416^{T}	79.4	83.7	ATCC 23343^{T}	79.0	83.8

Footnote. All isolates are of industrial environmental origin (IND), as indicated in the "origin" column. For each isolate, STs as determined by MLSTCheck and the closest genome in ENA as determined by FastANI are shown. STs highlighted in green match those previously validated by PCR, whilst those in yellow differ but possess the same species-level identification. STs highlighted in grey had not been defined previously. Alignment-free (FastANI, shown as F%) and alignment-based (PyANI, shown as P%) ANI comparisons against various type strains are shown.

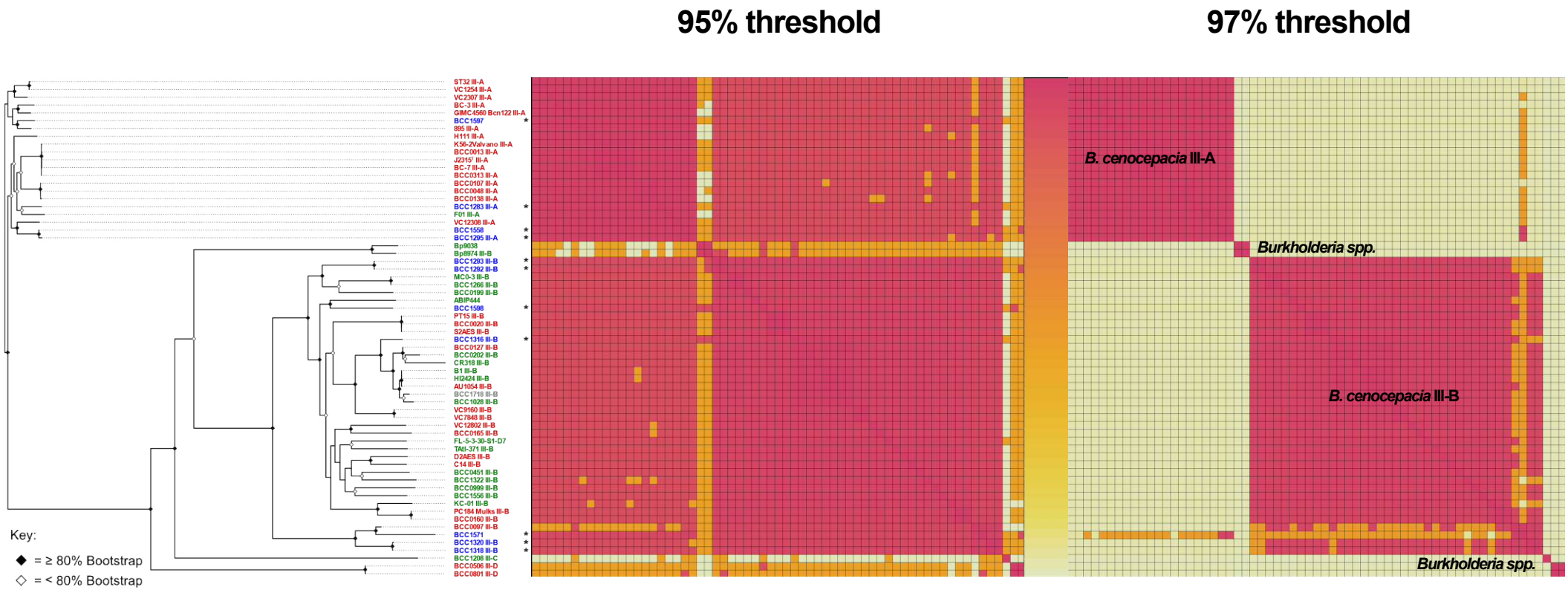


Figure 3.2: Genome sequence and core-gene phylogenetic placement of *B. cenocepacia* industrial isolates. A heatmap was generated by the heatmap package in R, using the outputs of PyANI, indicating the degree of nucleotide-level similarity between industrial isolates initially identified as *B. cenocepacia by recA* or 16S ribosomal RNA genome sequencing, alongside a core-gene phylogeny of representative isolates across the *B. cenocepacia* dataset. The colour scale bar indicated the degree of similarity as follows: AT (red) = above the threshold for species delineation, BO (orange) = borderline species, just above the threshold for species delineation, BT (pale yellow) = below threshold for species delineation. Core-gene phylogeny was generated from an alignment of 304 core-genes, using RAxML with 100 bootstraps. Scale bar represents the number of substitutions per base position. Where historical data was available, the genomovar (III-A, III-B, III-C or III-D) is indicated on the core-gene phylogenetic tree. Clinical isolates are shown in red, environmental isolates are shown green, whilst industrial isolates are shown in blue. Isolates that were sequenced during this study are asterisked. Species groupings as proposed by ANI97 are shown on the ANI97 heatmap.

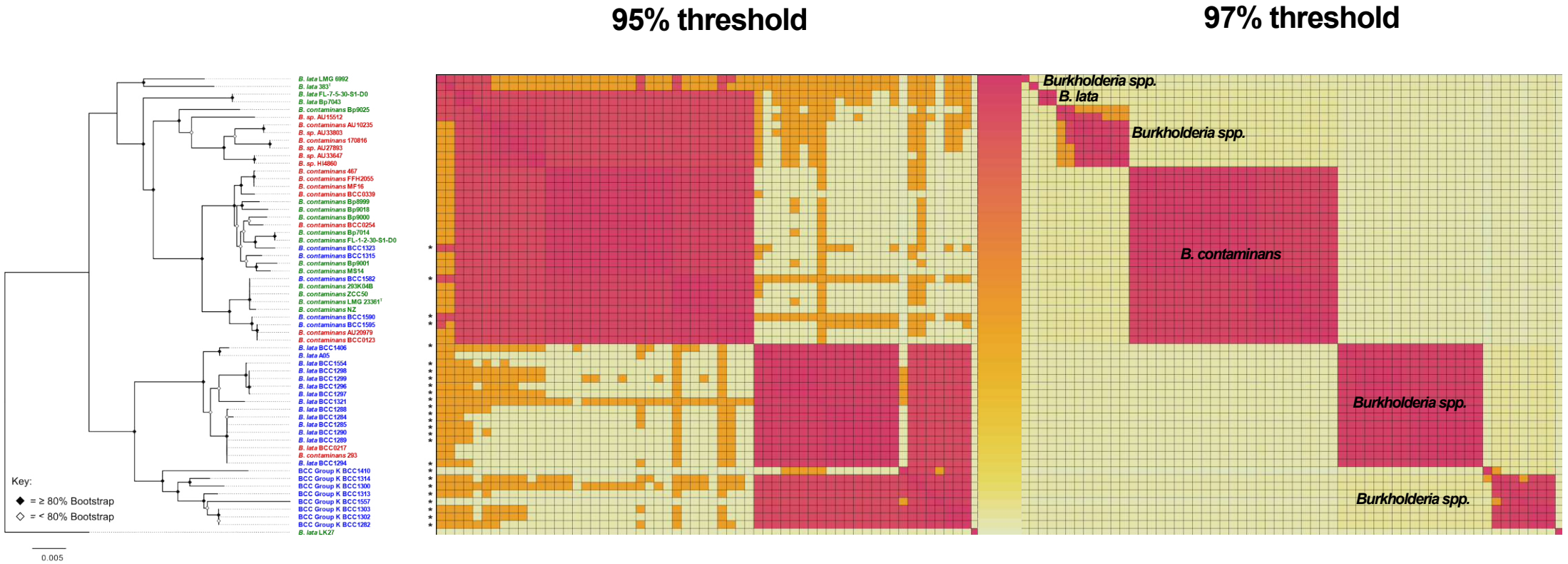


Figure 3.3: Genome sequence and core-gene phylogenetic placement of Bcc taxon K industrial isolates. A heatmap was generated by heatmap in R, using the outputs of PyANI, indicating the degree of nucleotide-level similarity between industrial isolates. All isolates were initially identified as Bcc taxon K *by recA* or 16S ribosomal RNA genome sequencing, alongside a core-gene phylogeny of representative isolates across the Bcc taxon K dataset. The colour scale bar indicated the degree of similarity as follows: AT (red) = above the threshold for species delineation, BO (orange) = borderline species, just above the threshold for species delineation, BT (pale yellow) = below threshold for species delineation. For each organism, the original species-level identification is indicated. The core-gene phylogeny was generated from an alignment of 48 core-genes, using RAxML with 100 bootstraps. Scale bar represents the number of substitutions per base position. Clinical isolates are shown in red, environmental isolates are shown green, whilst industrial isolates are shown in blue. Isolates that were sequenced during this study are asterisked. Species groupings as proposed by ANI97 are shown on the ANI97 heatmap.

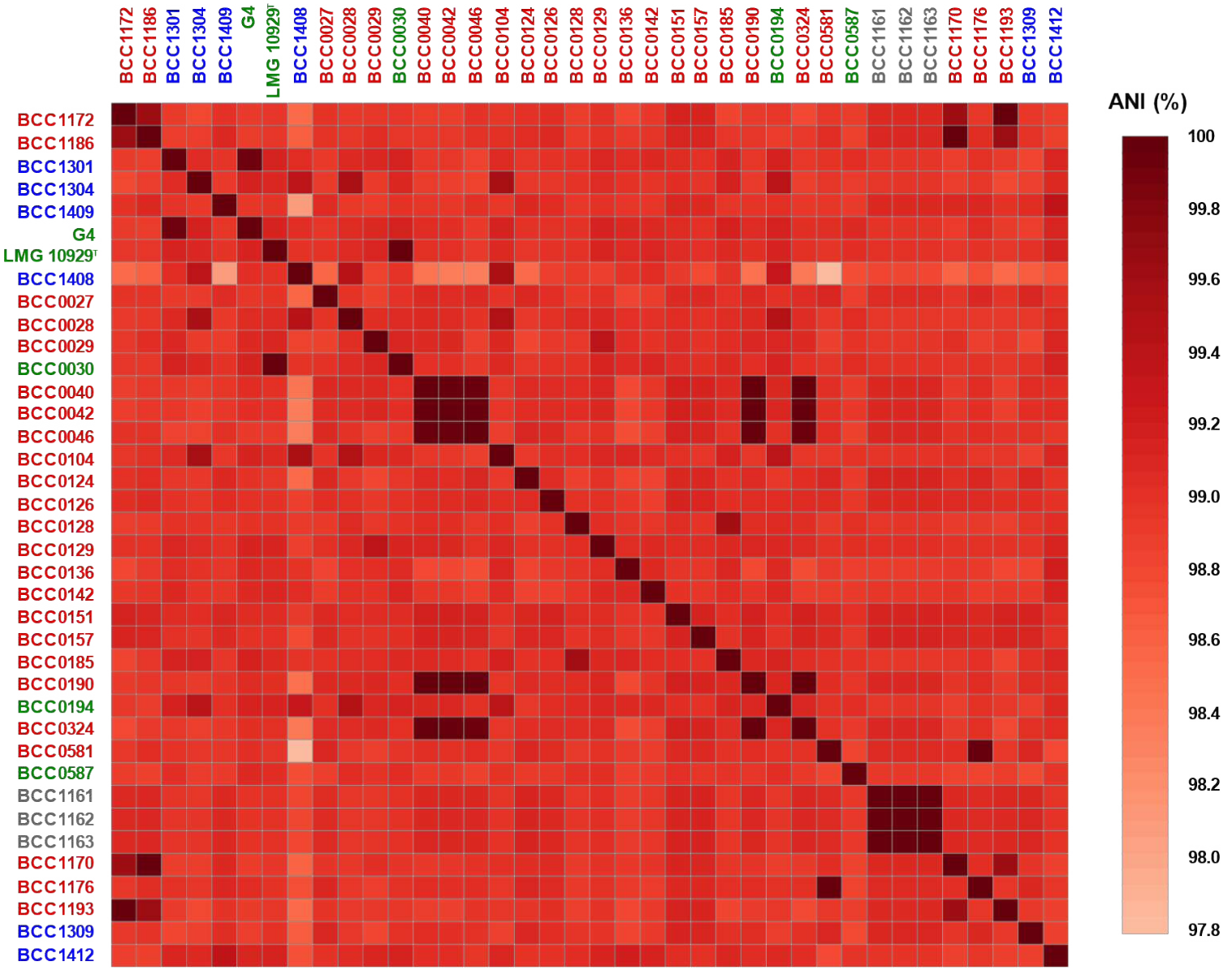


Figure 3.4: Average Nucleotide Identity shows the high level of genetic similarity between strains of *B. vietnamiensis* from a variety of backgrounds. The degree of genome similarity measured as percentage identity by means of pairwise comparison between genomes is indicated by the scale with darker shades of red indicating higher levels of genome similarity. Epidemiological information is also indicated as follows: strains highlighted in red are from a clinical background, strains highlighted in green are environmental, and strains highlighted in blue are industrial.

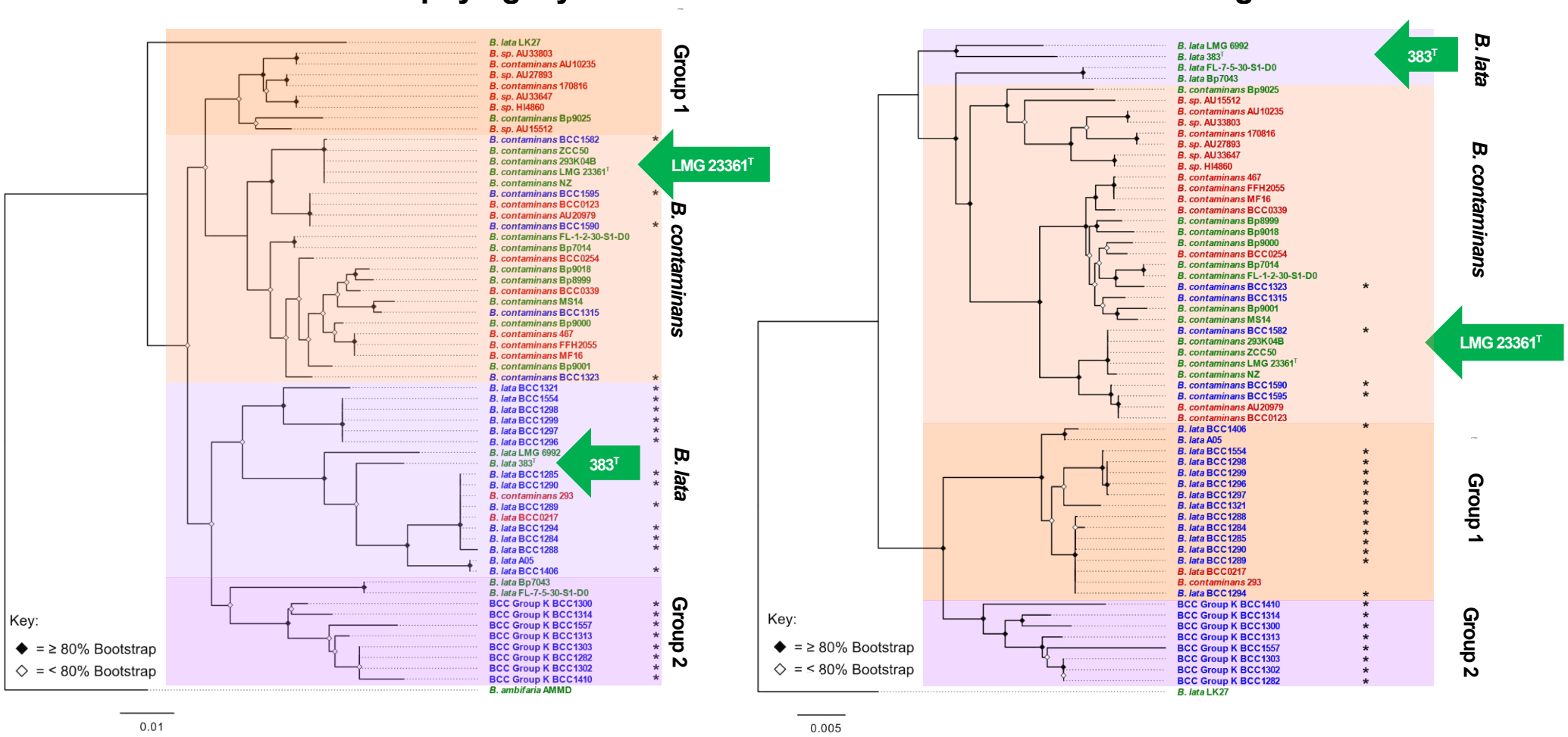


Figure 3.5: MLST and Core-gene phylogenetic placement of *Burkholderia* taxon K industrial isolates – The MLST was generated from the concatenated MLST gene output of MLSTcheck, whilst the core-gene phylogeny was generated from the core-gene alignment output of Roary. Both phylogenies were generated using RAxML with 100 bootstraps and were rooted with *Burkholderia ambifaria* AMMD. Scale bar represents the number of substitutions per base position. Type strains for *B. lata* (383^T) and *B. contaminans* (LMG 23361^T) are denoted by the green arrows. Clinical isolates are shown in red, environmental isolates are shown green, whilst industrial isolates are shown in blue. Isolates that were sequenced during this study are asterisked.

3.3.6. Key mutations identified in preservative adapted isolates

All preservative-adapted *B. lata* 383 derivatives and historical industrial isolates from which DNA was extracted grew well on BSM-Gluc, excluding intermediate isolates generated during stepwise training to BIT, which failed to revive from their freezer stocks. For these isolates, DNA was successfully extracted for sequencing by mixing the original freezer stock in a 1:1 ratio with the lysis agent Guanidinium Isothiocyanate (4M) and transferring this mixture directly to the Maxwell® 16 Tissue DNA Purification Kit and instrument.

The nucleotide sequence variants identified in preservative adapted isolates of *B. lata* 383 are summarised in Table 3.2. Derivative 383-BC possessed a frameshift mutation in *cmpB*, encoding and a stop:gain mutation in *dsbD*, both of which are present in chromosome 1, and encode a bicarbonate transport system permease protein, and a protein disulfide oxidoreductase respectively. Both 383-BIT and 383-CMIT derivatives possessed a missense mutation in the *soxR*, gene of chromosome 2, a known regulator of the B1004-B1006 RND efflux operon. Analysis of 383-BIT and intermediate isolates suggests that this mutation causes a change from guanidine to thymine at position 445, which was introduced following exposure to 0.015% BIT. 383-CMIT also possessed two additional missense mutations in *fcuA* of chromosome 1, which encodes a siderophore receptor, and *bepE* of chromosome 2, which encodes a component of an efflux transporter involved in resistance to antibiotic, dyes and detergents. 383-MIT possessed a missense mutation in *marR* of chromosome 2, a repressor of *marRAB* operon, which is responsible for the expression of numerous antibiotic resistance and oxidative stress genes.

For the industrial isolates of *B. lata,* BCC1294, BCC1296, and BCC1406, the complete genome of *B. lata* A05, was identified as their nearest neighbour by core-gene analysis. A homologue of B1004 corresponding to efflux transporter periplasmic adaptor subunit was identified in the A05 genome. This gene possessed the highest identity (95.87%) of BLASTN hits for all genomes available in NCBI, when the B1004 from *B. lata* 383 was queried against the NCBI nucleotide database (except for 383^T itself). The closest

identified upstream regulator was a *marR* type transcriptional regulator. Missense mutations were present in 2 of the 3 industrial isolates as follows; BCC1294 possessed a missense variant c.91_93delTGCinsGGT resulting in a change at amino acid residue position 31 from cysteine to glycine; and BCC1296 possessed a c.91T>G missense variant, which resulted in an amino acid change at the same position. No mutations were identified in this regulator in BCC1406. Analysis of this regulator by BLASTN showed that the top 13 hits correspond to chromosome 2 of genomes from *B. cepacia*, *B. contaminans* and *B. lata* (Figure 3.7 shows the gene synteny for the top six hits). In these genomes, the MarR family transcriptional regulator and downstream genes were present with 99-100% coverage in 12/13 genomes, and 75% coverage in 1/11 genomes (Figure 3.7). Beyond the top 13 hits, the absence of the MarR family transcriptional regulator also corresponded to the absence of the B1004 homologue (Figure 3.7).

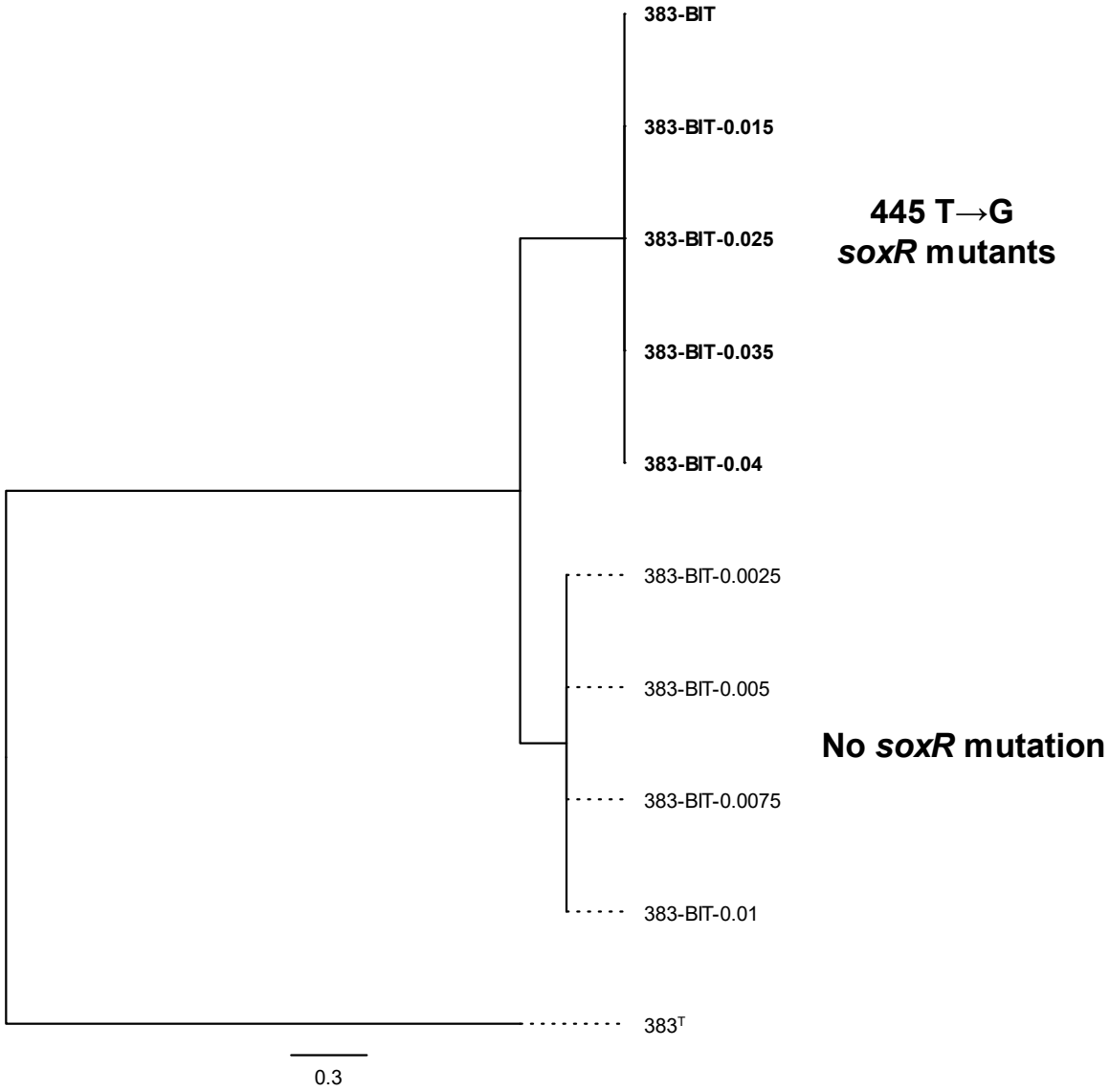


Figure 3.6: Core SNP tree of *B. lata* 383 isolates adapted to increasing concentrations of BIT. For each isolate, the number after the prefix "383-BIT" represents the percentage concentration of BIT to which the isolate was adapted at the point of passage. A core-SNP alignment was generated using Snippy v4.1.0, by aligning reads generated from each adapted isolate to the complete reference genome for the wild-type *B. lata* 383 (**383^{T}**, Accession: GCF_000012945.1). A phylogeny was generated from the core-SNP alignment using RAxML. Mutations underpinning the phylogenetic split were identified from the outputs of Snippy. Scale bar represents the number of substitutions per base position.

Table 3.2: Nucleotide polymorphisms and antibiotic susceptibility profiles for preservative-adapted isolates of *B. lata* 383^{T}											
	Genes and polymorphisms						Mean MIC (µg/ml) value				
Isolate	***cmpB***	***bepE***	***fcuA***	***marR***	***soxR***	***dsbD***	**AMK**	**CIP**	**CHL**	**SXT**	**CAZ**
383-BC	**Del, frameshift Chr 1 939 del C**					**SNP, stop:gain Chr 1 358 C > T**	**64**	**1.25**	**72**	**0.315**	**12**
383-BIT					**SNP, missense Chr 2 445 T > G**		**10**	**>32**	**>256**	**0.25**	**10**
383-CMIT		**SNP, missense Chr 2 1052 T > C**	**SNP, missense Chr 1 793 C >T**		**SNP, missense Chr 2 427 G > T**		**24**	**>32**	**>256**	**0.315**	**12**
383-MIT				**SNP, missense Chr 1 262 G > A**			**128**	**1.5**	**128**	**0.25**	**10**

Footnote. Summary of unique polymorphisms detected within preservative-adapted isolates of *B. lata* 383^{T}, in comparison to the complete reference genome for *B. lata* 383^{T} (accession: GCF_000012945.1). Mutations are described in the following format: mutation type, mutation consequence, chromosome location (1, 2 or 3), and nucleotide position and change (where del = nucleotide deletion, and > represents a change from one nucleotide to another). Antibiotic susceptibility profiles for preservative-adapted isolates, as established in Rushton *et al.* 2013, are also shown. Antibiotics are as follows: AMK, amikacin; CIP, ciprofloxacin; CHL, chloramphenicol; SXT, trimethoprim-sulfamethoxazole; CAZ, ceftazidime.

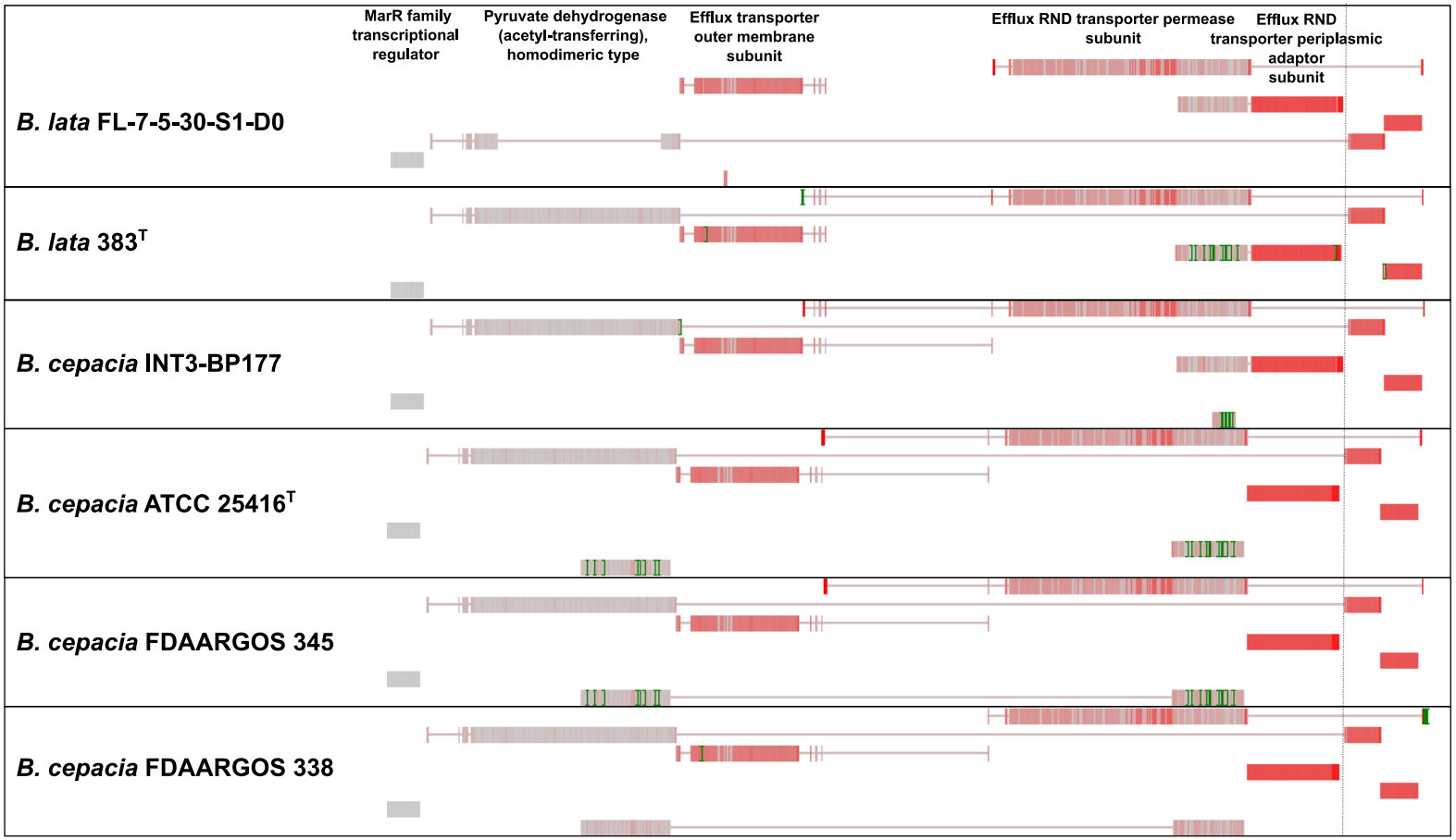


Figure 3.7: Conservation of MarR family transcriptional regulator in *B. lata* A05 and other members of the Bcc. The graphic was generated from a BLASTN search using the multiple sequence alignment viewer from NCBI. The positions of genes downstream from the transcriptional regulator are indicated. The dashed line on the right-hand side of the figure represents end of the B1004 homologue, and the start of the 2000 bp region included in the BLASTN search to confirm similar gene synteny in other genomes.

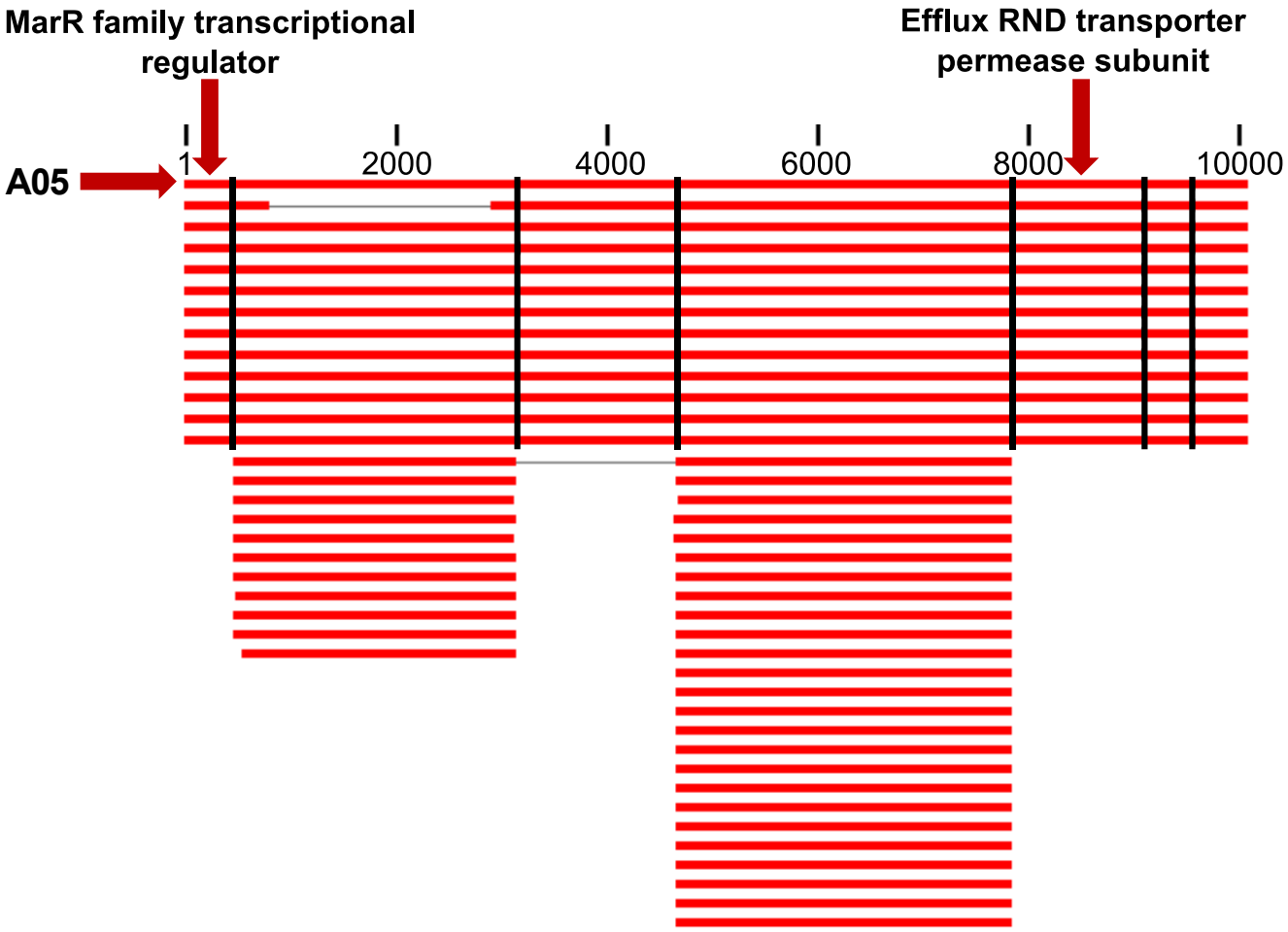


Figure 3.8: Co-absence of MarR family transcriptional regulator and B1004 homologue in Bcc species more distantly related to *B. lata* A05. The graphic was generated from a BLASTN using the graphic viewer of the distribution of the top blast hits in NCBI. Each red line represents BLAST hits to a different genome. The top hit, corresponding to the A05 genome from which the query sequence was derived, is indicated. The positions of the MarR family transcriptional regulator and B1004 homologue encoding an efflux RND transporter permease subunit are indicated. Solid lines represent the boundaries between BLAST hits for individual genes.

3.4. Discussion and conclusions

This chapter presents a clear workflow for the identification of Bcc spp. using genomics. It supports that MLST should be used prior to ANI to generate a preliminary profile of isolates of interest. This MLST sequence type and profile should be used as a basis to pick closely related genomes due to the mislabelling of multiple genomes deposited within the public databases highlighted by this analysis (see 3.3.1). For example, the industrial strain BCC1889 (which was later identified as a novel *Burkholderia* spp.) was found to have a closest match to a strain deposited as *Burkholderia pseudomallei* (*B. pseudomallei*) 7506487. *B. pseudomallei* is a highly virulent *Burkholderia* species known for its potential use in bioterrorism (Galyov *et al.* 2010). However, when BCC1889 was compared to the *B. pseudomallei* type strain, ATCC 23343^{T} the ANI fell below the 95% threshold for species identification, dropping to 83.8%, and was clearly showing the deposited genome was not *B. pseudomallei*. Overall, this illustrates that while genomic resources are becoming more extensive for *Burkholderia* and other industrial contaminant species, the approach to using this form of highly accurate species identification needs to be systematic and fully comparative.

For the key contaminant species *B. cenocepacia*, *B. vietnamiensis*, and taxon K spp., MLST was able to fully distinguish *B. cenocepacia* and *B. vietnamiensis* and could also distinguish members of the *B. cenocepacia* belonging to different lineages within this group. Core-gene phylogenetic analysis is however required for the correct identification of taxon K species, 15 of which were misidentified as *B. lata* by historical MLST analysis. These taxon K isolates clearly belong to a novel species group, which is most closely related to *B. lata* A05, and from an ANI perspective is taxonomically distinct from 383^{T}. This distinction is not shown by BLAST or phylogenetic analysis of MLST genes, both of which place the isolates alongside the *B. lata* type strain, 383^{T}. Core-gene analysis however places these misidentified isolates alongside their closest complete genome, *B. lata* A05, while strain 383^{T} is phylogenetically distant (Figure 3.5, panel B).

Average nucleotide identity analysis also showed that comparison with more stringent thresholds (97%), compared to the currently proposed standard threshold (95%), may serve as an alternative to core-gene analysis for identification. This 97% threshold successfully resolves *B. cenocepacia* genomovars III-A and III-B from one

another and is thus concordant with phylogenetic analysis (Figure 3.2). The use of ANI may be useful in reducing the number of data processing steps required for accurate identification, as this threshold showed a high-level of corroboration with core-gene analysis. Other studies have shown that ANIb provides a level of discrimination between III-A and III-B lineages which supports the designation of III-B as a new species grouping (Wallner *et al.* 2019). It should be noted however that ANIb analyses using whole genomes are highly computationally intensive (Yoon *et al.* 2017), meaning increasing threshold for ANI boundaries to 97% as proposed in our study could provide a much more time effective means for accurate identification of *B. cenocepacia* III-A and III-B. Further studies should investigate the differences in analysis runtime using these two algorithms along with the datasets in this chapter, in order to determine optimal pipeline for the genomic identification of *B. cenocepacia* as industrial contaminants.

The 97% ANI threshold also does not compromise the novel species placement of the misidentified *B. lata* isolates, which display alignment-based ANI values of 94.6-94.9% in comparison to 383^{T}, and 98.6-99.7% when compared to A05. Alignment-free ANI may also be used as a faster, less computationally-intensive alternative to alignment-based ANI (Jain *et al.* 2018), but may fail in the face of poor quality genomic data. This was seen in section 3.3.1 and Table 3.1 where several supposedly fell below the 80% threshold for comparison required by alignment-free ANI but displayed ANI values above 80% when alignment-based ANI was used. This decision should rest with user and should be based upon quality of genomic data available. For all industrial contaminants of interest however, core-gene analysis is essential to determine the true evolutionary interrelatedness of industrial isolates, and any strains of public and industry concern.

Genomic analysis also showed how *B. lata* can genetically adapt to a stable preservative tolerant phenotype. In relation to the mutational signatures behind the preservative tolerant *B. lata* 383 derivatives evolved in a previous study (Rushton *et al.* 2013), it was possible to identify defined polymorphisms. These were present in both adapted-derivatives generated from passaging experiments involving exposure to increasing concentrations of preservatives (Rushton *et al.* 2013), and in BCC1294 and BCC1296, which were originally identified as *B. lata*, but from the current analysis are a part of the novel A05 taxon K clade. In all the industrial isolates and 383

derivatives, polymorphisms were missense mutations in upstream regulators from the efflux pump component B1004, or a homologue thereof.

In adapted isolates 383-BIT and 383-CMIT, this missense mutation was found in *soxR*, a known regulator of the B1004-B1006 RND operon. In 383-BIT, this mutation was introduced following exposure to 0.015% BIT. As a known regulator of the B1004-B1006 operon (Dietrich *et al.* 2008), the *soxR* mutation may be responsible for the upregulation of B1004 of the RND operon observed in 383-BIT and 383-CMIT previously. Further evidence in the literature points to the role of this mutation in the upregulation of the B1004 in comparison to the parent strain, *B. lata* 383^{T}, from which 383-BIT and 383-CMIT are derived. Firstly, it has been previously reported that mutations in *soxR* are responsible for resistance to the antibiotics, chloramphenicol, nalidixic acid, and ciprafloxacin in Gram-negative enterics, such as *Escherichia coli* (Koutsolioutsou *et al.* 2005). Additionally, during the studies in which they were generated, the mean MIC of both 383-BIT and 383-CMIT for the fluoroquinolone ciprofloxacin increased from 1.25 μg/ml to >32 μg/ml. This increase in MIC was reversible with the use of an efflux pump inhibitor, MC-207,110 L-Phe-Arg-β-napthylamide (PaβN), and corresponded with a reduction in expression of the B1004-B1006 operon (Rushton *et al.* 2013). This suggests that identified mutations in *soxR* may be responsible for elevated ciprofloxacin resistance observed in both 383-BIT and 383-CMIT. In industrial strains BCC1294 BCC1296, similar missense mutations were identified in an upstream transcription regulator of the MarR transcriptional-regulator family, whilst no mutations were identified in BCC1406. This corresponds closely with the expression profiles of these isolates (Rushton *et al.* 2013), where BCC1294 displayed the largest change in fold change in expression of B1004 in comparison to 383^{T} (236.8 ± 184.4), followed by BCC1296 (5.7 ± 1), whilst BCC1406 lower, variable changes in expression (3.3 ± 2.4).

Whilst the gene synteny between the MarR family regulator and the identified B1004 homologue has not been characterised previously, the co-absence of the MarR family regulator and B1004 homologue also suggests that these genes tend to be conserved together on an evolutionary basis, indicating that there may be a link between them in terms of their functional role. Additionally, in a somewhat similar vein to *soxR*, there are well-known associations between mutations in MarR family transcriptional regulators, and antibiotic resistance (Bialek-Davenet *et al.* 2011; Warner *et al.* 2013).

For both of the regulators identified in these studies, however, the role of the identified mutations can be inferred, but not validated in terms of their role in antibiotic and preservative resistant phenotypes without further experimental evidence. This could be achieved using a number of gene-knockout systems, (Flannagan *et al.* 2008; Hamad *et al.* 2010), all of which have been shown to be effective in *Burkholderia* spp.

Additionally, future work should look to produce higher-quality draft genomes from the industrial *Burkholderia* collection, in order to enable additional genomic analyses that have not been performed here. Due to the low N50 values of the newly sequenced genomes, in comparison to those in reference databases, genomes were considered of insufficient quality for resistance gene prediction analyses as conducted previously (see 3.3.1). This was because these analyses were reliant upon relative quantification of resistance genes from sequenced genomes, in order to contextualise them within a reference dataset. It is well known that genome quality significantly affects gene prediction (Smits 2019). The number of predicted resistance genes is therefore more likely to be lower in the newly sequenced genomes, (mean N50 = 4978 bp), than in closely related reference genomes, which were of much higher quality (mean N50 values of *B. cenocepacia*, taxon K and *B. vietnamiensis* datasets of 1171560, 1588797, and 241466 bp respectively). As such, if the genomes generated in this study were used for this analysis, there would be no way to tell if differences in resistance gene prediction are artefactual, or of biological significance. These genomes were however, still considered candidates for ANI, which only compares identity between similar regions of genomes (Richter and Rosselló-Móra 2009), and core-gene analysis, which will similarly determine organism relatedness using common genes present in all genomes analysed (Page *et al.* 2015). As such, these analyses are less affected by the relative quality between two genomes.

To summarise the findings behind each of the objectives of the chapter, the following is now clear:

1. **To characterise a panel of industrial Bcc spp. in terms of their taxonomic diversity 51 isolates were successfully identified by MLST, the species identity of which was confirmed by ANI.** The predominant isolates in panel by MLST were *B. lata* (*n* = 13), *B. cenocepacia* (*n* = 11), followed other Bcc species (*n* = 9).

2. **To determine the appropriate techniques for accurate identification of closely related Bcc species, with a focus on *B. cenocepacia*, *B. lata* and *B. vietnamiensis* as priority industrial contaminants.** MLST was sufficient for accurate identification of *B. cenocepacia*, and *B. vietnamiensis*, but resulted in the misidentification of industrial isolates in taxon K. MLST should be used for initial identification of closely related species, before using ANI with a 97% threshold to confirm species identity. Core-gene analysis was required to determine the evolutionary interrelatedness of all isolates.

3. **To determine key genomic features underpinning changes in expression of efflux pump genes both preservative-trained and historically industrially derived *B. lata* isolates.** Missense mutations could be identified in upstream regulators with a known role in antibiotic resistance and efflux pump expression in preservative-trained isolates of 383^T, and isolates from industry historically identified as *B. lata*.

The hypotheses for this chapter were as follows, and were both accepted:

I. Whole genome analysis can be used to accurately identify Bcc species from historical identification incidents, with a greater level of resolution than MLST.

II. Key mutational signatures can be identified in both preservative-trained and historically industrially derived *B. lata* isolates.

4. *Pseudomonas* in industry

This chapter was published in part as a research article (see Weiser *et al.* 2019, Microbial Genomics)

Pseudomonas spp. are motile, non-spore forming, aerobic, Gram-negative rods, which are seemingly ubiquitous in the environment due to their impressive metabolic, stress and xenobiotic-tolerance capabilities, amongst other unique adaptive qualities. Versatile phenotypic properties include, but are not limited to polar flagella, the production of pigments, such as the potent toxin pyocyanine, chelating agents such as pyoverdine, and multiple intrinsic antimicrobial resistance mechanisms. Of all *Pseudomonas* spp., *Pseudomonas aeruginosa* (*P. aeruginosa*) is one of the best studied, due to its problematic nature in the nosocomial setting. *P. aeruginosa* is an extremely versatile bacterium with the ability to survive in diverse habitats including soil, water, plant and animal tissues, community and hospital environments. In particular, its presence within water via natural, tap and potable sources is a recognised problem for manufacturing industries reliant on water as a major raw material (Jimenez 2007). Much to the detriment of the patients and clinicians alike, *P. aeruginosa* often utilises this inherent versatility to become ingrained as an opportunistic pathogen in immunocompromising disorders such as cystic fibrosis (CF). In CF, the lung microbiome is initially predominantly colonised by *Staphylococcus aureus* (*S. aureus*; including methicillin-resistant *S. aureus*) and *Haemophilus influenzae*, whereas *Burkholderia cepacia* complex (Bcc) spp. and *P. aeruginosa* dominate in the later stages of disease (Flight *et al.* 2015). Infections with *P. aeruginosa and* Bcc spp. in late stage CF are often chronic, and have been clearly associated with a more rapid decline in lung function, and higher incidences of mortality (Flight *et al.* 2015).

P. aeruginosa is also problematic in the manufacturing industry, where it is the most frequently encountered microbial contaminant of non-food industrial products, due to its xenobiotic capability and antimicrobial tolerance (Cunningham-Oakes, Weiser, *et al.* 2020). Contamination of taps, surfaces, instruments and patients has been associated with multiple incidents of nosocomial *P. aeruginosa* infections. The presence of the pathogen in home and personal care products has also initiated infections in hospital settings (Cunningham-Oakes, Weiser, *et al.* 2020). A historical case associated with a contaminated hand lotion used by healthcare workers in a neonatal intensive care unit

inadvertently led to vulnerable infants being infected (Becks and Lorenzoni 1995). Interestingly, the *Enterobacteriaceae* species *Enterobacter agglomerans* was also cultured from the contaminated hand lotion bottles, but did not cause infection in the infants (Becks and Lorenzoni 1995). More recently an outbreak with extensively drug-resistant *P. aeruginosa* was associated with sharing of an aromatherapy oil among hospitalised patients in an intensive care unit (Mayr *et al.* 2017). Such localised incidents of contamination leading to infection can normally be controlled by infection control, but outbreaks in wider community settings are problematic to resolve. A national *P. aeruginosa* outbreak in the UK occurred among individuals undergoing ear piercing, which continued for 3 months in 2016 before it was stopped by a source tracing enabled product recall (Evans *et al.* 2018). Using VNTR strain typing and cohort analysis, the outbreak was linked to the use of a cosmetic after-care solution which had become contaminated during production (Evans *et al.* 2018). Within the same industrial sector of beauty and cosmetic products, contaminated tattoo ink has also been implicated in multiple cases of *P. aeruginosa* infection (Høgsberg *et al.* 2013). Overall, *P. aeruginosa* in addition to being frequently described as ubiquitous bacterial species found in multiple niches, is consistently encountered in relation to incidents of non-food product contamination (Jimenez 2007; Sutton and Jimenez 2012).

P. aeruginosa possesses a number of unique characteristics, which facilitate its identification. Firstly, it is a rapidly growing microorganism and is straightforward to culture with a range of positively or negatively selective media (Weiser *et al.* 2014). Standard industrial practices for microbiology and quality control are effective for recovering it as a contaminant from a range of products and industrial settings. Additionally, accurate identification of the species can be achieved using multiple methods including biochemical profiling, MALDI TOF MS, 16S rRNA or *oprL* gene sequence analysis. International databases such as the public multilocus sequencing typing resource (https://pubmlst.org/paeruginosa/) can also enable placement of strains in the context of their wider epidemiological prevalence.

Industrial isolates of *P. aeruginosa* have been shown to be widely-distributed throughout the species population in terms of AT-genotyping and core-gene content, and possess similar numbers of predicted AMR genes to isolates from clinical and environmental backgrounds (Weiser *et al.* 2019). Industrial strains have also been shown to be unique in that they possess considerable lengths of DNA corresponding to a family of

megaplasmids. In this study, one such megaplasmid was resolved in the complete PacBio assembly of *P. aeruginosa* RW109, which consisted of a 7 049 347 bp main chromosome, a megaplasmid (555 265 bp) and a large plasmid (151 612 bp) (Weiser *et al.* 2019). The draft genomes of seven other industrial isolates (RW130, RW131, RW146, RW172, RW176, RW184 and RW204) possessed homology across >250 kbp of the 555 Kbp replicon, and possessed near-identical *parB* genes, which are responsible for correct segregation of DNA into daughter cells during cell division, and are thus well conserved amongst plasmids originating from the family (Weiser *et al.* 2019). The complete genome structure of the industrial isolates (excluding RW109) however has yet to be characterised.

Uniquely, 5 of the industrial isolates, possessed the components of BpeEF-OprC, an efflux pump which confers resistance to trimethoprim, chloramphenicol, tetracyclines and fluoroquinolones was also on the megaplasmid of RW109, for a total of 6 industrial isolates with the efflux pump (Weiser *et al.* 2019). Beyond genomic characterisation however, the phenotypic functionality of this efflux pump, and the other 500+ genes present on the megaplasmid is unknown.

Aims of this chapter were as follows:

1. **To comparatively examine the phenotype of the industrial *P. aeruginosa* RW109 strain and a plasmid-lacking derivative in order to determine the role of large and megaplasmids in the fitness of these isolates.**

2. **To complete as far as possible the genomes of, and characterise large and megaplasmids present in 6 *P. aeruginosa* industrial isolates with evidence of a megaplasmid and efflux pump (RW130, RW146, RW172, RW176 and RW184), and one isolate with evidence of a megaplasmid, but no efflux pump (RW131).**

3. **To determine the changes, if any, in genomic synteny underpinning persistence of three *P. aeruginosa* industrial isolates, RW130, RW131 and RW146, isolated from the same industrial product over a six-year period.**

The overall hypotheses for this chapter were as follows:

I. The megaplasmid of RW109 will play a role in antimicrobial resistance.
II. Megaplasmids will be present throughout all of the sequenced *P. aeruginosa* industrial isolates.
III. Changes in the genome structure will be seen when comparing the industrial *P. aeruginosa* strain RW130 to isolates RW131 and RW146.

4.1. Materials & methods

4.1.1. Calgary (Peg-lid) biofilm model of *P. aeruginosa* RW109

Cells were incubated at 28^{0}C for 18 hours in a Calgary biofilm device (Macià *et al.* 2014) consisting of 96-well microtitre plate with a lid containing the same number of removable pegs, before being transferred to fresh media. *P. aeruginosa* strain RW109 was initially passaged in TSB containing 0.05% Sodium Benzoate (NaB), and TSB as a control. RW109 was incubated in the same concentration of NaB until confluent growth was observed at a given concentration (typically 5 days). If growth was poor, RW109 was transferred to fresh media containing the same concentration of NaB. Once confluence was reached, cells were transferred to a 96-well plate containing a higher concentration of NaB in TSB. This methodology was conducted in two technical replicates, over three biological replicates. It was carried out by Stuart Campbell-Lee and Dr Tom Pointon at Unilever Research and Development, Port Sunlight, UK.

4.1.2. Robbins device biofilm model of *P. aeruginosa* RW109

Cells were incubated at 28^{0}C and passed through an artificial multiport sampling catheter known as a Robbins device (see schematic in Figure 4.1) (Kharazmi *et al.* 1999). The device was inoculated using exponential phase *P. aeruginosa* RW109, and grown for 5 days, in order to initially establish a biofilm. This was then followed by a medium containing a given concentration of NaB, but no RW109. For both of the aforementioned phases, a flow rate of 10 mL/min was used and maintained using a peristaltic pump. Evenly spaced sampling ports positioned at the end of the device contained discs, which could be removed aseptically to sample the biofilm, once it had developed. The sampling discs (13.2 mm diameter and 0.8 mm thick) were made of stainless steel 314, in order to mimic biofilm development and preservative exposure in a factory environment where this

grade of stainless steel is commonly used. The Robbins device was run until biofilm coated the sampling disc. The time taken for biofilm development sufficient for sampling was recorded for each concentration of NaB. Once sufficient biofilm was present on the sampling disc, the disc was then sonicated into TSB with 8% (v/v) DMSO, in order to create freezer stocks of NaB adapted strains. Each freezer stock was then used to generate exponential phase medium of an adapted strain and begin the process again. This methodology was conducted in two technical replicates within one biological replicate. The experiment was carried out in parallel to the peg-lid assay above by Stuart Campbell-Lee and Dr. Tom Pointon at Unilever Research and Development, Port Sunlight, UK.

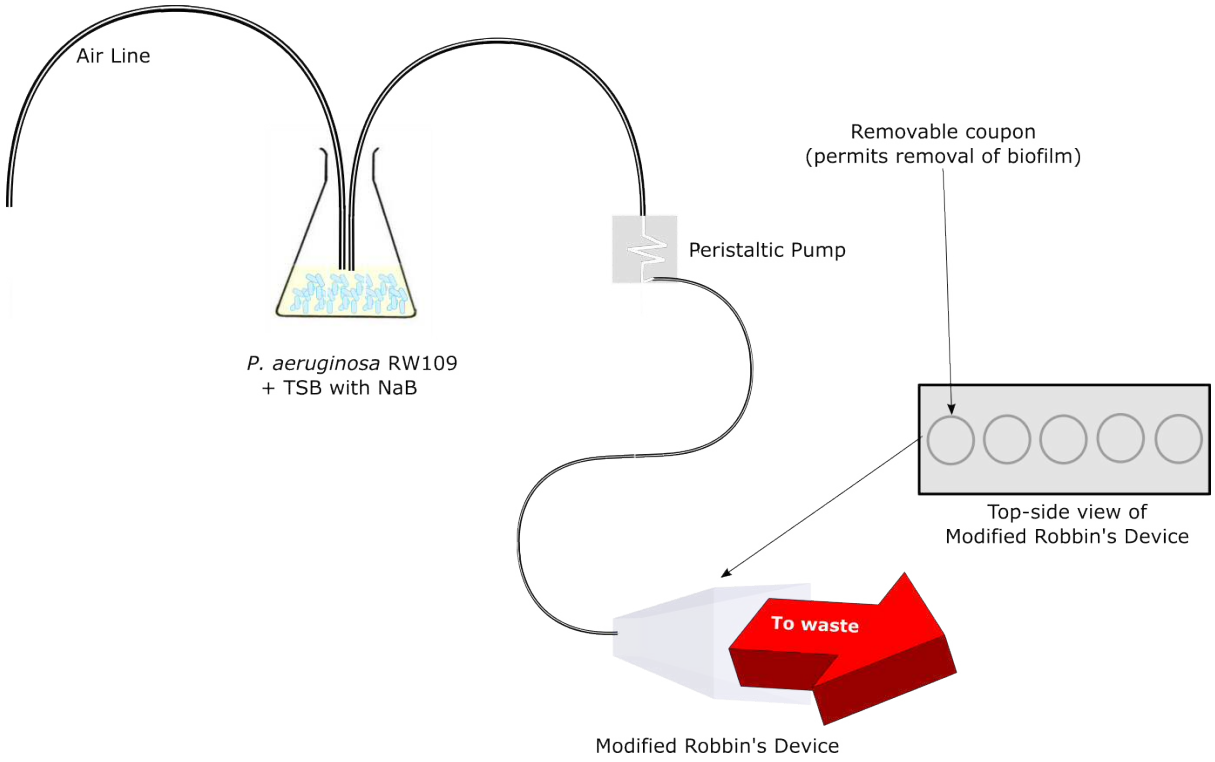


Figure 4.1: Schematic of the modified catheter (Robbins) device used for the generation of adapted derivatives of *P. aeruginosa* RW109: The Robbins device consists of an line feeding air directly into a bacterial culture for aeration, which is then fed through the device by means of the action of a peristaltic pump. A view from above the stainless steel 314 coupons, used for the removal of biofilm of biofilms, is shown.

4.1.3. Genome sequencing of adapted *P. aeruginosa* RW109

Genomic DNA was extracted from overnight cultures of Robbin's device adapted *P. aeruginosa* RW109 and quantified as described in previous sections. Libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, Massachusetts) (see 2.4.9). Libraries were then quantified, normalised and sequenced as described previously (see 2.4.9). Genomes were assembled using Unicycler v0.4.8 as described previously (see 2.4.10).

4.1.4. Read mapping to verify absence of large and megaplasmids in industrial *P. aeruginosa*

The complete genome assembly for RW109 was obtained from Genbank using ncbi-genome-download (https://github.com/kblin/ncbi-genome-download). The three-replicon assembly for this isolate was then separated into individual replicons. Sequencing reads from each biofilm adapted isolate of RW109 were mapped to each replicon of the wild-type RW109 assembly using Minimap2 v2.17 (Li 2018). Mapping was conducted to determine whether adaption to NaB involved loss of the large plasmid and megaplasmid.

4.1.5. Phenotypic assessment of plasmid loss in *P. aeruginosa* RW109

A MASTRING urine ring device (Mast Group, United Kingdom) carrying 8 antibiotic impregnated tips; (ampicillin 25µg, chloramphenicol 50µg, colistin sulphate 100 µg, kanamycin 30 µg, nalidixic acid 30 µg, nitrofurantoin 50 µg, streptomycin 25 µg, tetracycline 100 µg) was used to initially compare the antibiotic susceptibility profiles of three derivatives of RW109. The derivatives tested were the RW109 parental strain (RW109), RW109 Robbins device control (RW109-RB), and RW109 Robbins device 0.175 % NaB (RW109-RB-NaB) adapted isolate. RW109-RB and RW109-RB-NaB were both generated during the biofilm assays using the Robbins device, described in 4.1.2. The susceptibility test was conducted by inoculating a plate by spreading ~10^8 cells on a 30 ml plate of Mueller Hinton agar, before placing a MASTRING on the surface of the plate, and incubating for 24 hours at 30°C. Zones of clearing were then measured to give an initial insight into the differing susceptibilities of the strains. Trimethoprim (TRI) and chloramphenicol (CHL) were identified as antibiotics of interest from MASTRING assays, read mapping (see 4.1.4 and 4.2.1) and previous studies

highlighting the presence of a BpeEF-OprC type efflux pump in the genome of RW109. This efflux pump is known to confer resistance to TRI and CHL (Weiser *et al.* 2019) but a phenotype is not known in RW109. In order to determine more accurate susceptibility profiles for TRI and CHL, a two-fold microbroth dilution method was used, where a range of antibiotics concentrations were challenged with approximately 10^6 viable *P. aeruginosa* cells. For these assays, the Minimum Inhibitory Concentration (MIC) for TRI and CHL was defined for RW109, RW109-RB, RW109-RB-NaB, and *P. aeruginosa* ATCC 19429, an antibiotic-testing reference strain specified by Clinical and Laboratory Standards Institute methodology (CLSI, 2018). The MIC was determined by identifying the concentration at which growth is reduced (MIC breakpoint) after 24 incubation at 30°C, as determined by optical density readings at 600 nm. Specifically, the first concentration at which optical density was reduced by at least 80% (MIC_{80}) was chosen as the MIC breakpoint. In order to corroborate a phenotypic role for the megaplasmid in the TRI and CHL resistance of RW109, the MIC_{80} for TRI and CHL was also determined in the presence of 50μg/ml of the broad-spectrum efflux inhibitor MC-207,110 l-Phe-Arg-β-napthylamide (PAβN) (Sigma-Aldrich, United Kingdom).

4.1.6. Initial characterisation of megaplasmid families in industrial *P. aeruginosa*

As a preliminary analysis investigating the evolutionary diversity of megaplasmids related to the RW109 megaplasmid, the RW109 plasmid partition gene, *parB*, was used to identify and extract homologous sequences from Genbank in NCBI, and from the short-read genome sequence assemblies for eight *P. aeruginosa* isolates (Weiser *et al.* 2019). Plasmid partition genes can be used to classify and to determine evolutionary interrelatedness of plasmids, such as the RW109 megaplasmid (Weiser *et al.* 2019), due to their unique conservation within mobile elements (Petersen 2011), which are otherwise promiscuous in terms of their gene synteny (Sevastsyanovich *et al.* 2008; Liapis *et al.* 2019). A maximum-likelihood phylogenetic tree was then generated based on MAFFT alignment on the extracted *parB* sequences using RAxML, as described in 2.4.7.

4.1.7. Complete genome sequencing of a panel of industrial *P. aeruginosa*

Short-read sequencing was performed previously at the Oxford Genomics Centre of the Wellcome Trust Centre for Human Genetics, using a 100 bp paired-end Illumina HiSeq 2000 sequencing approach (Weiser *et al.* 2019), with DNA

extraction conducted using the Maxwell DNA extraction system as described previously (2.4.5). Long-read genome sequencing was performed using the PacBio Sequel System (Pacific Biosciences) via the Novogene sequencing company. DNA was extracted from 3 mL of overnight cultures of RW130, RW131, RW146, RW172, RW176, and RW184 grown in TSB at 30°C (Promega) using the Wizard® Genomic DNA Purification Kit (Promega), which is designed for isolation of high molecular weight DNA with large fragment sizes suitable for long-read sequencing. This was then supplemented with DNA extracted from another 3 mL overnight culture, using the Maxwell®16 Tissue DNA Purification Kit, which produces high-concentration DNA with slightly greater degree of shearing than the Wizard® Genomic DNA Purification Kit (Promega). DNA samples were combined in order to increase the input concentration to >80 ng/μL for sequencing using the PacBio Sequel System, as required by Novogene, Cambridge, UK.

4.1.8. Completion of industrial *P. aeruginosa* genomes

Long-reads from PacBio sequencing were used to produce initial genome assemblies using Unicycler v0.4.8 (Wick *et al.* 2017) under default settings. Genomes were then completed using MeDuSa (Bosi *et al.* 2015), a draft genome scaffolder which uses the whole genome aligner MUMmer alongside one, or multiple reference genomes in a graph-based approach, to complete draft genomes. For all draft genomes produced using Unicycler excluding RW184, the complete sequence for the industrial *P. aeruginosa* isolate RW109 was used as a reference in a single round of genome scaffolding. RW184 was initially scaffolded to RW109, resulting in three contigs. The third contig of ~7 kBp was identified as a misassembled contig by BLASTN, with greatest homology to the largest contig of RW172. As such, the largest contig of RW184 and the misassembled contig were passed through a second round of scaffolding, using the largest contigs of both RW172 and RW184 as a reference, resulting in a single contig. The newly formed single contig of RW184 and second contig produced during the first round of scaffolding to RW109 were then concatenated to produce the final RW184 scaffold. The contigs of genome assemblies were polished with Pilon v1.23 (Walker *et al.* 2014) using Illumina reads to identify and correct any misassembled contigs. Genome quality was assessed using Quast v4.4. FastANI was then used to

compare each isolate against the *P. aeruginosa* type strain ATCC 10145^{T} to confirm species identity.

4.1.9. Using the homology of the origin of replication gene, *parB* to identify related chromosomes and plasmids and develop custom databases

Each contig produced during the genome assembly process was isolated using Samtools via command-line and was interrogated using BLASTN against custom plasmid partition gene BLAST databases, generated via command-line. The constructed databases contained the *parB* plasmid partition genes of both the RW109 main chromosome and RW109 megaplasmid (see 4.1.6), and *parA* plasmid partition gene identified in the large plasmid of RW109. Individual contigs from each industrial *P. aeruginosa* assembly were then isolated and sorted into groups based upon the partition gene with which they shared the highest percentage identity. The same partition genes used to construct the custom database were then queried using BLASTN against the NCBI database, in order to identify genomic sequences with closely related origins to each replicon of RW109. Genbank annotations and fasta files for the whole genome assemblies corresponding to the first 100 hits for each gene were then obtained using ncbi-genome-download (see 4.1.4). Genbank annotations were used to create custom databases for the reference main chromosome, megaplasmid and large plasmid, which were then used to annotate their corresponding fasta files and any contigs with similar plasmid origins, from the *P. aeruginosa* isolates sequenced in this study, using Prokka (see 3.2.6) (Seemann 2014) with the proteins flag.

4.1.10. Evolutionary, AMR and Biocide gene analysis of *P. aeruginosa* genomes and plasmids

Due to their considerable size, the plasmids of RW109 can be effectively considered as small bacterial genomes. As such, the core-gene content of the chromosomes and plasmids was determined using Roary (as described in 3.2.6) (Page *et al.* 2015) in order to determine the evolutionary interrelatedness of the chromosomes and plasmids of the sequenced industrial *P. aeruginosa* isolates, and their isolates possessing *parB* genes with which they share the greatest homology. In order to visualise the evolutionary interrelatedness of industrial *P. aeruginosa*, core-gene alignments produced by Roary were then used by RAxML and a GAMMA model of rate heterogeneity supported by 100 bootstraps to generate a

phylogenetic tree (see 2.4.7). Phylogenetic trees were visualised in FigTree v1.4.2 and Inkscape v0.91 as described previously (see 2.4.7). In order to determine the number of predicted AMR, biocide and metal resistance genes conferred by megaplasmids, AMR, biocide and metal resistance genes were predicted for the megaplasmids and whole genome of industrial isolates and references where both the megaplasmid and whole genome assemblies were available, using ABRicate. For AMR prediction, a coverage cut-off of 75% was used, whilst it was not possible to set a coverage cutoff when using Bacmet2 for biocide gene prediction, which utilises a protein database in tandem with TBLASTX (Ladunga 2017) for predictions. Absolute gene counts were then plotted using the boxplot package in R as described previously (see 2.4.12). Means and standard deviations were also calculated as described previously (see 2.4.12). Where appropriate, gene prediction outputs were interrogated to determine key resistance determinants predicted on megaplasmids.

4.1.11. Assigning KEGGs to replicons from the same family as the RW109 megaplasmid and RW109 large plasmid

The core-gene analysis output of Roary was interrogated to identify core-genes with the megaplasmid, and large plasmid families of the industrial *P. aeruginosa* isolates. Additional genes that were exclusive to the industrial isolates sequenced in this study, and RW109 were also identified. The BlastKOALA automatic annotation server (Kanehisa, Sato and Morishima 2016) for genome sequences was used to perform KO (KEGG Orthology) assignments to characterise individual gene functions and reconstruct KEGG pathways for the newly sequenced megaplasmids of *P. aeruginosa* RW131, RW146, RW172, RW176 and RW184, as well as the previously sequenced megaplasmid of RW109. The amino acid fasta file produced by Prokka was used for the analysis. The purpose of this analysis was to provide an understanding and context of high-level functions and utilities of the biological systems underpinned by the genes on each megaplasmid (Kanehisa, Sato, Kawashima, *et al.* 2016). The results of the outputs were compared to those of the RW109 megaplasmid, for which similar KO assignments highlighted the presence of BpeEF-OprC efflux pump on the megaplasmid (Weiser *et al.* 2019). Whilst the presence of components of this efflux pump have been previously inferred BLASTN in all of the aforementioned isolates previously, the

genomic location of this efflux pump could not be confirmed in isolates other than RW109, due to the incomplete status of these genomes. As such, the outputs of BlastKOALA were also interrogated for the presence of BpeEF-OprC in all megaplasmids. KO assignments were then visualised, using the radar chart package in R, and edited using Inkscape v0.91.

4.1.12. Investigating genomic changes in three industrial isolates of *P. aeruginosa*, isolated from the same industrial product type over six years

Previous studies have highlighted that three industrial isolates in the current dataset, RW130, RW131 and RW146, for which less-complete genomes were produced (Weiser *et al.* 2019), which were isolated from a household cleaning product at the same manufacturing location over several years. These isolates were isolated sequentially as follows: RW146 in 2004, and RW131 and RW130 in 2010. All of the aforementioned isolates shared the genotype, as determined by AT-genotyping, a microarray which provides a unique genotype for *P. aeruginosa* isolates subject to the assay. This assay represents the core-genome of *P. aeruginosa* by means of SNPs, and the accessory genome by utilising genomic island and islet markers. In order to confirm the interrelatedness RW130, RW131 and RW146, and investigate the co-occurrence of underlying genomic changes with strain persistence in industrial products, replicons with similar partition genes were then compared using the Circular Genome Viewer (CGView) server (Grant and Stothard 2008). The CGView server BLASTN with an expect (E) value of 0.1 to compare input against Genbank annotation, generated using Prokka as described in 3.2.6. The CGView server then produced a graphical output of these results, which were merged and edited in Inkscape v0.91 to produce a final figure. In the case of RW130, where main chromosome and megaplasmid *parB* genes localised to the same replicon, the same replicon was used as a reference for both megaplasmid and main chromosome analyses.

Table 4.1: Short-read mapping of RW109 isolates adapted to Sodium Benzoate (NaB) in the Robbin's device biofilm model					
			Mapping to *P. aeruginosa* RW109 replicon (%)		
NaB (%)	**Biological Replicate**	**Technical Replicate**	**Main chromosome**	**Megaplasmid**	**Large Plasmid**
0.05	1	1	89.60	7.54	2.46
0.05	1	2	89.60	8.31	2.70
0.1	**1**	**2**	**91.68**	**0.30**	**0.02**
0.1	**1**	**1**	**90.74**	**0.31**	**0.00**
0.175	**1**	**1**	**90.74**	**0.31**	**0.00**
0.175	**1**	**2**	**91.68**	**0.30**	**0.02**
CTRL	1	2	89.49	8.16	2.64
CTRL	1	1	88.91	8.75	2.94

Table 4.2: Short-read mapping of RW109 isolates adapted to Sodium Benzoate (NaB) in the Calgary device biofilm model					
			Mapping to *P. aeruginosa* RW109 replicon (%)		
NaB (%)	**Biological Replicate**	**Technical Replicate**	**Main chromosome**	**Megaplasmid**	**Large Plasmid**
0.2	1	1	85.47	11.77	4.20
0.2	1	2	86.06	11.29	4.02
CTRL	**1**	1	85.44	11.77	4.12
CTRL	**1**	2	85.66	11.58	3.94

Footnote. The percentage of NaB to which the isolate was exposed prior to isolation and downstream sequencing is indicated, alongside the biological replicate and technical replicate for each sequenced isolate form the experiment. The percentage mapping of reads to each replicon to *P. aeruginosa* RW109 is indicated.

Table 4.3: Genome quality metrics for industrial *P. aeruginosa* isolates assembled using two assembly methods														
	Hybrid assembly							Scaffolded PacBio assembly with Illumina read polish						
		Number of contigs							Number of contigs					
Strain	**Genome size (Mbp)**	**>1000bp**	**>50000bp**	**GC (%)**	**N's per 100Kbp**	**N50 (Mbp)**	**Coverage**	**Genome size (Mbp)**	**>1000bp**	**>50000bp**	**GC (%)**	**N's per 100kBp**	**N50 (Mbp)**	**Coverage**
RW130	7.78	22	6	65.23	0	1755337	68.5x	7.84	2	2	65.29	2.68	7699334	67.7x
RW131	7.79	10	3	65.25	0	7356863	46.7x	7.90	3	3	65.23	0.13	7443836	46.3x
RW146	7.85	12	3	65.21	0	7348781	64.4x	7.87	3	3	65.26	3.81	7486930	63.7x
RW172	7.59	12	3	65.27	0	3211311	64.3x	7.65	2	2	65.24	3.92	7175507	63.1x
RW176	7.51	14	3	65.29	0	7049023	70.5x	7.74	2	2	65.15	6.46	7234967	69.2x
RW184	7.29	21	4	65.39	0	3241870	92.0x	7.50	2	2	65.30	10.66	7043258	88.6x

Footnote. All genome statistics are based on contigs > 1000bp, unless stated otherwise. Coverage is shown as an average across the genome.

Table 4.4: ANI, and partition gene matches for industrial *P. aeruginosa* isolates				
	Replicons with BLASTN matches (contig, contig size in bp, % identity to partition genes)			**ANI of whole genome vs ATCC 10145^T (%)**
	RW109 main chromosome *parB*	**RW109 megaplasmid *parB***	**RW109 large plasmid *parA***	
RW130	Contig 1, 7699334bp, 100%	Contig 1, 7699334bp, 100%	Contig 2, 136135bp, 100%	99.3
RW131	Contig 1, 7443836bp, 100%	Contig 2, 316352bp, 100%	Contig 3, 137346bp, 100%	99.3
RW146	Contig 1, 7486930bp, 100%	Contig 3, 284977bp, 100%	Contig 2, 138780bp, 100%	99.3
RW172	Contig 1, 7175507bp, 99.1%	Contig 2, 472591 bp, 100%		98.8
RW176	Contig 1, 7234967bp, 99.7%	Contig 2, 503458 bp, 99.1%		99.2
RW184	Contig 1, 7043258bp, 99.7%	Contig 2, 461163bp, 100%		99.3

4.2. RESULTS

4.2.1. RW109 exposed to concentrations of 0.1% NaB and above in a Robbins device biofilm model shows evidence of megaplasmid and large plasmid loss

RW109 displayed the ability to grow in the presence of NaB up to a concentration of 0.2% in the Calgary device, and 0.175% in the Robbins device. Mapping of sequenced short-reads from the endpoint 0.2% NaB adapted isolate from the Calgary device showed little difference in the number of mapped reads to the main chromosome, large plasmid and megaplasmid of RW109 in comparison to the control (Table 4.2). Similarly, little difference was observed between the number of mapped reads from isolates adapted to concentrations of NaB below 0.1%, with isolates exposed to 0.05% NaB (the highest concentration of NaB used below 0.1%) displaying a similar number of mapped reads to all three RW109 replicons in comparison to the control (Table 4.1). In contrast, isolates exposed to concentrations of 0.1% and 0.175% NaB displayed a drastic reduction in reads mapped to the RW109 megaplasmid, with the percentage of mapped reads ranging from 0.30-0.31%, in contrast to 8.16-8.75% in the Robbins device control and 7.54-8.31% in the isolate exposed to 0.05% NaB (Table 4.1). A similar phenomenon was observed when mapping reads to the RW109 large plasmid, with the percentage of mapped reads decreasing from 2.64-2.94% in the Robbins device control and 2.46-2.70% in the isolate exposed to 0.05% NaB, to 0-0.02% mapped reads in both isolates exposed to NaB concentrations of 0.1% and above (Table 4.1).

4.2.2. Reduction of MICs in presence of PAβN suggests that the efflux pumps on plasmids contribute resistance to TRI and CHL resistance in RW109

Microbroth dilution assays were performed over three biological replicates, with the hypothesis that a greater fold-reduction in TRI and CHL MIC_{80} would be observed in *P. aeruginosa* RW109-RB-NaB, for which no sequencing reads mapped to the RW109 megaplasmid, than RW109 or RW109-RB. The MIC_{80} of the *P. aeruginosa* reference strain ATCC 19429 was consistent for both TRI (1024μg/ml) and CHL (512 μg/ml). MICs for both CHL (32-64 μg/ml) and TRI (8 μg/ml) were also consistent in the presence of PaβN. The MIC_{80} values for RW109 and RW109-RB in the presence of TRI were 1024 μg/ml and 512 μg/ml respectively. These values decreased to 8-16 μg/ml for RW109, and 8 μg/ml for RW109-RB in the

presence of PaβN. In contrast, the initial MIC_{80} for RW109-RB-NaB in the presence of TRI was lower, at 256 μg/ml, but similarly to RW109 and RW109-RB, was reduced to 8 μg/ml in the presence of PaβN. In the presence of CHL, the MIC_{80} value for both RW109 and RW109-RB was 2048 μg/ml, whilst the MIC_{80} ranged from 512-1024 μg/ml for RW109-RB-NaB. In the presence of PaβN, the MIC_{80} was reduced to 64 μg/ml for both RW109 and RW109-RB, and to 32-64 μg/ml for RW109-RB-NaB.

4.2.3. The RW109 megaplasmid belongs to a single conserved family of megaplasmids

Analysis of the *parB* region revealed the existence of a conserved plasmid of megaplasmids. In the phylogenetic analysis, RW109 megaplasmid *parB* clustered with the *parB* genes from seven of the eight additional industrial strains included in the analysis, and two clinical strains. The same analysis highlighted the close phylogenetic relationship of the *P. aeruginosa* RW109 megaplasmid to characterised plasmids such as pJB37 and pOZ176, indicating that this plasmid belongs to the Inc-P2 incompatibility group.

4.2.4. All industrial isolates in the panel produced multi-replicon assemblies, for which scaffolding produces the fewest contigs

A summary of industrial *P. aeruginosa* genome quality metrics for both hybrid assembly and long-read contig scaffolding with short-read polishing is shown in Table 4.3. Of the two assembly methods evaluated, hybrid assembly produced genomes with a marginally higher coverage, and fewer nucleotides designated as N (unknown nucleotides) than scaffolded assemblies (Table 4.3). The hybrid assemblies were however much more fragmented than those produced by scaffolding, with hybrid assemblies possessing a greater number of contigs, and lower N50 values than scaffolded assemblies (Table 4.3). Furthermore, the use of Illumina reads for polishing in scaffolded assemblies meant that both hybrid and scaffolded assemblies took the more accurate chemistries of Illumina short-read sequencing into account. As such, the scaffolded assemblies were chosen for any downstream analysis.

All scaffolded assemblies were confirmed as *P. aeruginosa*, via ANI comparison to *P. aeruginosa* ATCC 10145^T (98.8-99.3% identity, see Table 4.4). The scaffolded assemblies of RW130, RW172, RW176 and RW184 assemblies contained two

contigs. RW130 possessed the RW109 main chromosome, and megaplasmid *parB* genes respectively whilst RW172, RW176 and RW184 had one match for either *parB* gene per contig (see Table 4.4 for more details). RW131 and RW146 possessed three replicons, with two contigs possessing a BLASTN match to either the RW109 main chromosome *parB*, or RW109 megaplasmid *parB*. The second contig of RW130 and third contig of both RW131 and RW146 possessed a BLASTN match with 100% identity to the *parA* gene of the RW109 large plasmid. Based on BLASTN matches to RW109 partition genes, each contig was labelled as individual replicons as follows:

- **Main chromosomes (MCs):** Contig 1 of all isolates
- **Megaplasmids (MGs/Plasmid 1):** RW131 contig 2, RW146 contig 3, RW172 contig 2, RW176 contig 2 and RW184 contig 2
- **Large plasmids (LGs, Plasmid 2):** RW130 contig 2, RW131 contig 3 and RW146 contig 3

A summary of the BLASTN matches and contig sizes for each the genome assembly of each isolate are shown in Table 4.3 and Table 4.4.

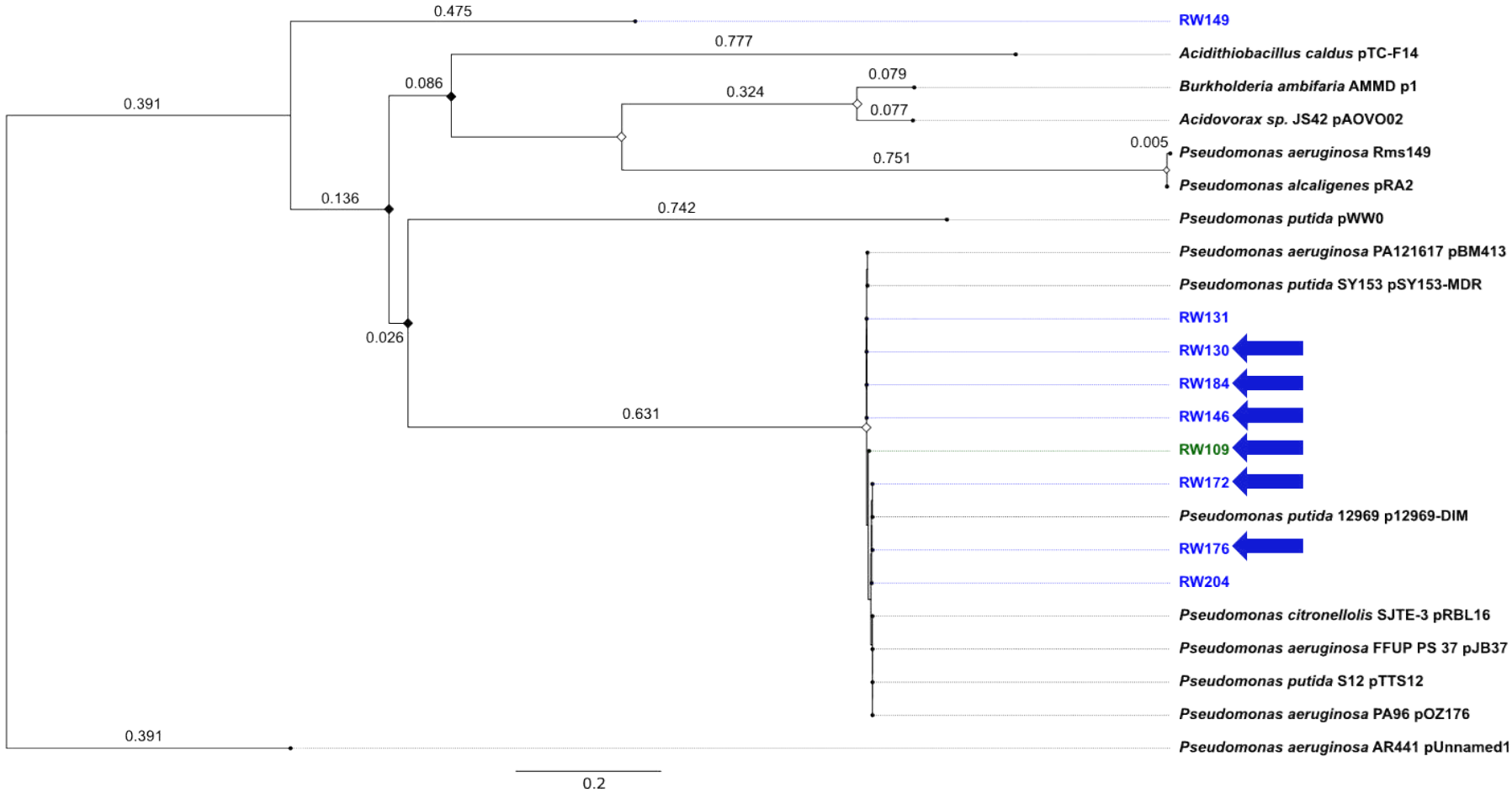


Figure 4.2:. A maximum-likelihood phylogenetic tree of the RW109 megaplasmid *parB* gene (green text) aligned against homologous sequences from the *P. aeruginosa* industrial strains (blue text) and other related *Pseudomonas* plasmids. Industrial strains uniquely encoding components of the BpeEF-OprC efflux pump are indicated by the blue arrows. Bootstrap values are indicated at the node (diamonds; white nodes ≥80 % and black nodes <80 %) and the genetic distance (the number of base substitutions per site) is indicated on the branches and by the scale bar. The tree was rooted with the *parB* gene from an unnamed plasmid in *P. aeruginosa* strain AR441.

4.2.5. Industrial *P. aeruginosa* megaplasmids possess 81 conserved genes, no AMR genes and differing functional gene annotation profiles

BLASTN of the partition genes for the main chromosomes, megaplasmid and large plasmid of the isolates sequenced in this study revealed 100 main chromosome assemblies, 21 reference megaplasmid assemblies, and 2 large plasmid assemblies for use in custom databases (as described in 4.1.9). The returned search results for each of the replicons included either the main chromosome, megaplasmid or large plasmid of the previously sequenced RW109. This analysis also revealed the deposited draft genome sequences for five additional industrial *P. aeruginosa* isolates four of which, namely AK6U (Accession: GCA_002843285.1), M8A1, M8A4 and Pb18 (Vives-Flórez and Garnica 2006) were associated with crude oil, whilst KRP1 (Rabaey *et al.* 2004) was isolated from the methanogenic sludge of a potato processing plant.

Core-gene analysis revealed 2420 core-genes amongst the industrial *P. aeruginosa* main chromosomes belonging to the same family as those of industrial isolates. Phylogenetic analysis based on the alignment of these genes (Figure 4.3) revealed that the main chromosome of RW109 was extremely similar to that of *P. aeruginosa* F5677, both of which share a common ancestor. The main chromosomes of *P. aeruginosa* RW130, RW131 and RW146 isolated over multiple years from the same manufacturing facility were extremely similar, but show some evidence of core-gene divergence, as evidenced by phylogenetic tree branching (Figure 4.3). The nearest neighbour of these main chromosomes was that of *P. aeruginosa* H47921, though the deep branching of this isolate within the phylogenetic tree suggests significant evolutionary distance between the H47921 main chromosome and those of RW130, RW131 and RW146 (highlighted by the dashed box in Figure 4.3). The RW172 main chromosome was placed uniquely within the phylogenetic tree with deep branching, amongst nearest neighbours *P. aeruginosa* isolates B10w, Pa58 and WPB100. Similarly, RW176 placed uniquely within the tree with very few similar isolates. The main chromosome of RW184 was most similar to those of *P. aeruginosa* isolates Pa1027 and ATCC 27853. The relative phylogenetic distances and evolutionary divergence of all PacBio sequenced (see 4.1.7) industrial *P. aeruginosa* included in this analysis were

comparable to those observed in previous analyses using short-read sequencing alone (Weiser *et al.* 2019).

Core-gene analysis also revealed the conservation of 61 genes amongst 27 sequences, which were determined as a part of the same megaplasmid family, based on the *parB* plasmid partition gene. The main chromosome of RW130 was included in this analysis, due to the co-localisation of main chromosome and megaplasmid *parB* genes on the main chromosome (see 4.2). Fourteen of the identified core-genes possessed annotations (Table 4.5) An additional 20 core-genes were identified as core to industrial *P. aeruginosa* megaplasmids, (inclusive of RW109 and the RW130 main chromosome), six of which possessed annotations (Table 4.5). Core-gene phylogenetic analysis revealed that the megaplasmid of RW109 is most closely related to plasmid pPABL048 of *P. aeruginosa* PABL048 (Figure 4.4), whilst in similar vein to its main chromosome, the megaplasmid of RW172 placed with a significant phylogenetic distance between itself and other plasmids in the phylogeny (Figure 4.4). The megaplasmids of RW131, RW146, RW176 and RW184 were all closely related, with the RW176 megaplasmid appearing ancestrally to the other megaplasmids. Interestingly, the main chromosome of RW130, and the megaplasmids of RW131 and RW146 possessed identical core-gene content.

All industrial megaplasmids from the isolates sequenced in this study and RW109 possessed no predicted AMR genes, with all predicted AMR genes instead being encoded by other regions of the genome (Figure 4.5). In contrast, biocide and metal resistance genes were predicted on all industrial megaplasmids, with the greatest number of biocide and metal resistance genes predicted on the megaplasmid of RW176 (37), and the fewest predicted on the megaplasmid of RW146 (4) (Figure 4.5). Of note, the megaplasmid of RW176 possessed the highest number of predicted biocide and metal resistance genes amongst the entire dataset, including reference megaplasmids (Figure 4.5). The most common biocide and metal resistance genes predicted amongst megaplasmids were *terA*, *terC*, *terD* and *terZ*, which are associated with tellurium resistance, and were predicted in 15 out of 17 megaplasmids within the dataset, excluding those of RW131 and RW146. Another gene associated with tellurium resistance, *terE* was predicted in 14 out of 17 megaplasmids, excluding those of RW131, RW146 and

RW172. The mercury resistance gene, *merT*, was predicted 15 out of 17 megaplasmids, including all industrial megaplasmids. The *merT* gene was not predicted in the reference plasmids pBM413 from *P. aeruginosa* PA121617, and pRBL16 from *Pseudomonas citronellolis* SJTE-3.

The functional annotations of all megaplasmids excluding those of RW146 and RW184 confirmed the presence of the complete BpeEF-OprC efflux pump, detected in RW109 previously. This included the megaplasmid RW131 for which the BpeEF-OprC was not detected via BLASTN of the draft, short-read assembly previously. Similarly, whilst the efflux pump was not detected by KEGG orthology on the RW146 megaplasmid, two components of the efflux pump were *bepE* (74.2 – 80.7% identity) and *dmlR_1* (94.4 – 100% identity) were detected via BLASTN on the RW146 main chromosome. The functional gene categorisation for all megaplasmids suggested predominant roles in genetic information processing, signalling and cellular processes and environmental information processing, though no megaplasmid possessed identical functional gene annotation profiles (see Figure 4.6). Notably, RW109 possessed a large number of gene annotations associated with carbohydrate metabolism (n = 13). The number of carbohydrate metabolism genes on the RW109 megaplasmid far exceeded other megaplasmids in the dataset with the second greatest number of genes in this category (RW176 and RW184, 4 genes). The RW172 megaplasmid was unique in possessing genes for Glycan biosynthesis and metabolism, and genes with a role in human disease (n = 1 and n = 2 respectively), whilst RW176 was the only megaplasmid possessing genes involved in nucleotide metabolism (n = 2).

Within the large plasmid family, 116 core-genes were identified. Of these, twenty-one possessed defined gene annotations, and whilst nine additional core-genes were identified when looking at the industrial *P. aeruginosa* plasmids alone, no additional genes with annotations were identified. The industrial large plasmids displayed little to no variation in functional gene profiles, with the majority (ranging from 17 to 21) of gene annotations associating with a role in genetic information processes, and the rest (between 16 and 19) playing a role in signalling and cellular processes.

Table 4.5 – Annotated core-genes in the RW109 megaplasmid family		
	Gene	**Annotation**
Megaplasmid family	*noc_2*	Nucleoid occlusion protein
	outO	Type 4 prepilin-like proteins leader peptide-processing enzyme
	xerC_1	Tyrosine recombinase XerC
	rpoS	RNA polymerase sigma factor RpoS
	sbcB	Exodeoxyribonuclease I
	dnaN	Beta sliding clamp
	wspC	putative biofilm formation methyltransferase WspC
	cobS	Aerobic cobaltochelatase subunit CobS
	pilT	Twitching mobility protein
	walR_1	Transcriptional regulatory protein WalR
	parA	Partition protein A
	dksA	RNA polymerase-binding transcription factor DksA
	atsB	Anaerobic sulfatase-maturating enzyme
	hupB	DNA-binding protein HU-beta
Industrial megaplasmids only	*smc*	Chromosome partition protein Smc
	xerC_4	Tyrosine recombinase XerC
	kdpE	DNA-binding transcriptional activator
	noc_1	Nucleoid occlusion protein
	recA	Protein RecA
	cheB	Protein-glutamate methylesterase/protein-glutamine glutaminase

Table 4.6 – Core-genes in the RW109 large plasmid family	
Gene	**Annotation**
topB	DNA topoisomerase 3
tcpE	Toxin coregulated pilus biosynthesis protein E
topA	DNA topoisomerase 1
pdeF	Cyclic di-GMP phosphodiesterase PdeF
dnaB	Replicative DNA helicase
rdgC	Recombination-associated protein RdgC
dsbG	Thiol:disulfide interchange protein DsbG
xerC_2	Tyrosine recombinase XerC
chpB	Endoribonuclease toxin ChpB
chpS	Antitoxin ChpS
oatA	O-acetyltransferase
pilT	Twitching mobility protein
xerC_1	Tyrosine recombinase XerC
yafQ	mRNA interferase toxin YafQ
noc	Nucleoid occlusion protein
dinG	3'-5' exonuclease DinG
rep	ATP-dependent DNA helicase Rep
ssb	Single-stranded DNA-binding protein
tnsB	Transposon Tn7 transposition protein TnsB
tnsC	Transposon Tn7 transposition protein TnsC
ompP1	Outer membrane protein P1

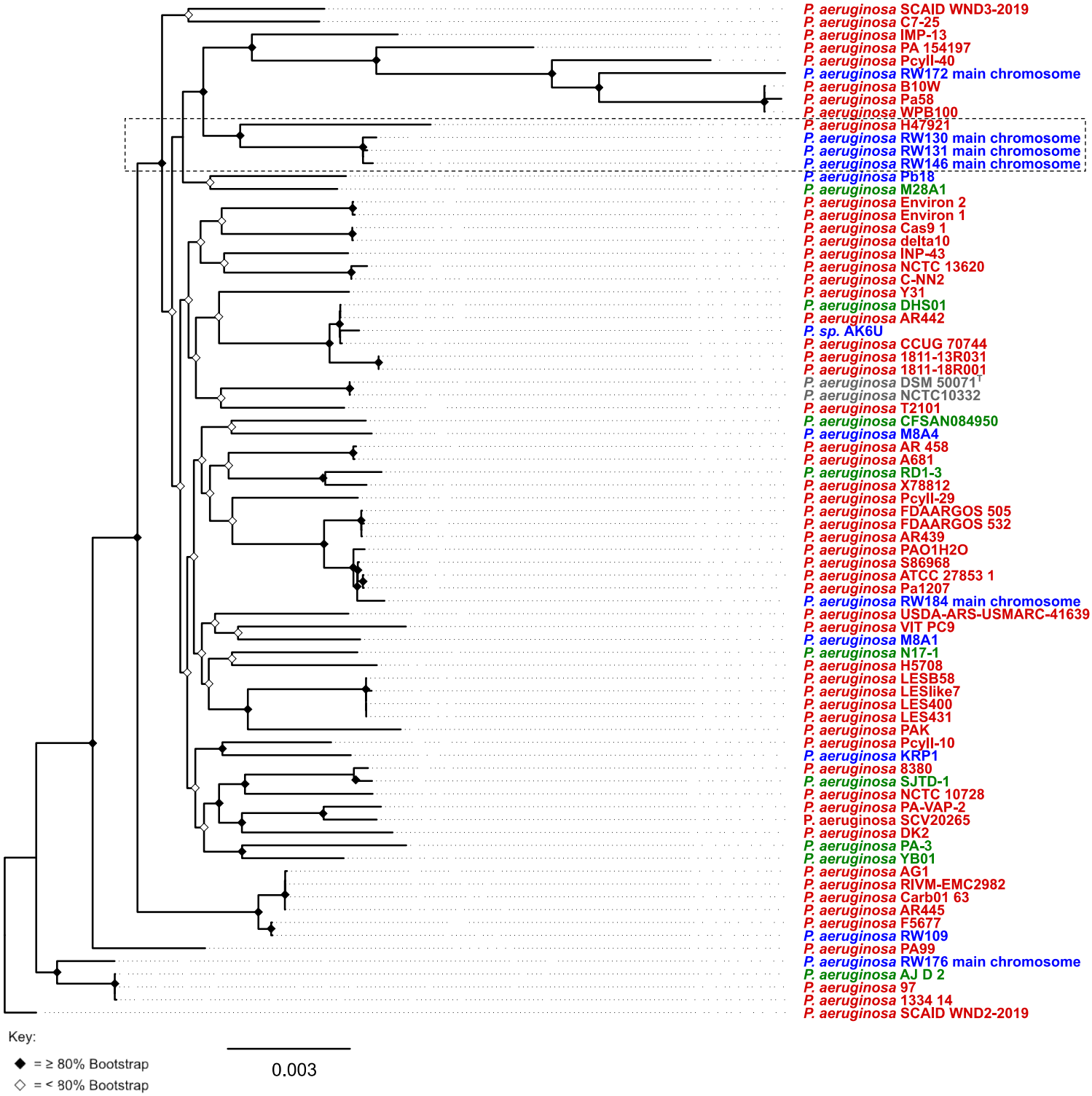


Figure 4.3: Core-gene phylogenetic placement of industrial *P. aeruginosa* main chromosome assemblies amongst related genomes. A core-gene phylogeny of the main chromosomes of 79 *P. aeruginosa* strains was generated from an alignment of 2420 core-genes, using RAxML with 100 bootstraps. Scale bar represents the number of substitutions per base position. Clinical isolates are shown in red, environmental isolates are shown green, whilst industrial isolates are shown in blue. Isolates that were sequenced are those highlighted as industrial using the aforementioned colour scheme and are additionally labelled with "main chromosome". The dashed box highlights the phylogenetic position of the main chromosomes of three industrial *P. aeruginosa* strains, RW130, RW131 and RW146, and their nearest neighbour, the clinical *P. aeruginosa* strain H47921.

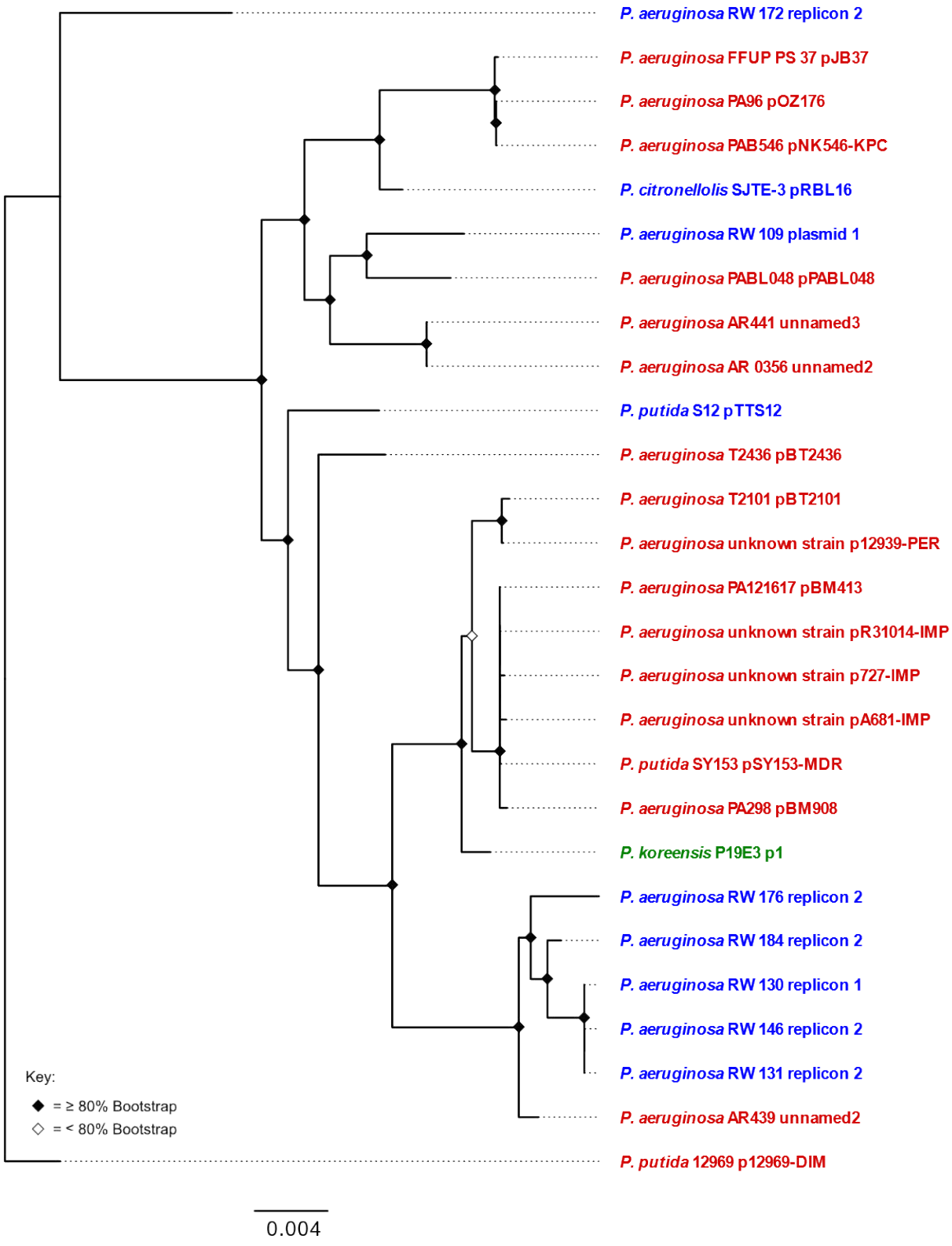


Figure 4.4: Core-gene phylogenetic placement of industrial *P. aeruginosa* megaplasmids within a *Pseudomonas* megaplasmid family A core-gene phylogeny of 26 megaplasmids, and one main chromosome was generated from an alignment of 65 core-genes, using RAxML with 100 bootstraps. Scale bar represents the number of substitutions per base position. Clinical isolates are shown in red, environmental isolates are shown green, whilst industrial isolates are shown in blue. Isolates that were sequenced during this study are designated as main chromosomes, or megaplasmids (MG) as appropriate.

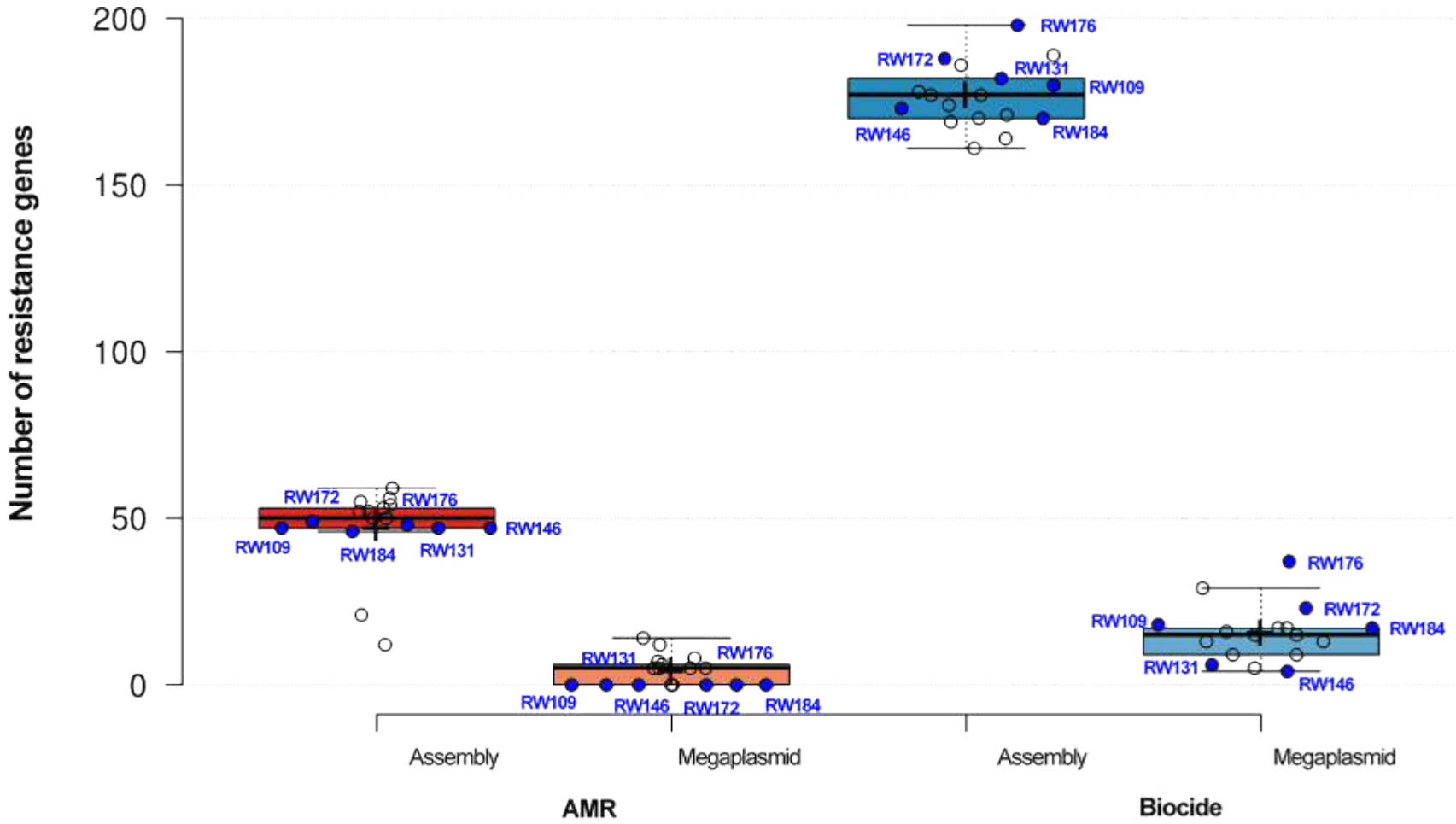


Figure 4.5: Distribution of AMR, biocide resistance, and metal resistance genes amongst industrial *P. aeruginosa*. The number of predicted genes in genomes and plasmids of industrial *P. aeruginosa* isolates are shown alongside those of reference isolates, Industrial isolates are shown in blue. For all boxplots, *n* = 17. The mean is indicated on each boxplot by a black cross, whilst the median is shown using a black line. Whiskers were defined using the Tukey HSD test.

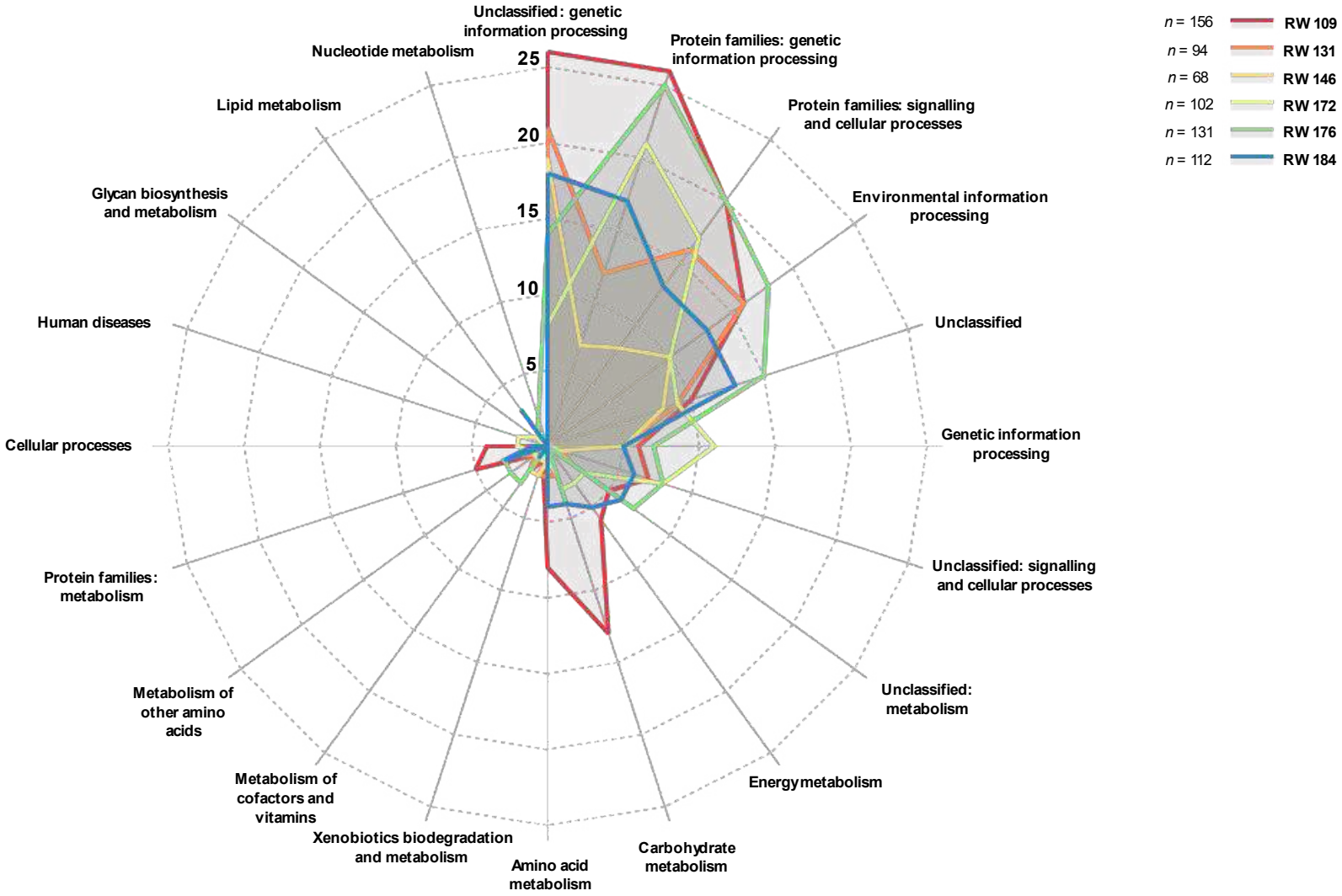


Figure 4.6: KEGG functional annotation of six industrial *P. aeruginosa* megaplasmids: Each shaded area represents the full functional gene profile of the megaplasmid of a given isolate, as indicated by the key (top right), with each radially extending line representing a different functional gene category (annotated). The total number of returned gene annotations from the queried plasmid annotation are shown by the *n* numbers, next to each isolate labelled in the key. Concentric dashed lines are used to measure the number of genes in each category, with each concentric line representing five genes.

4.2.6. Partition gene and BLASTN analysis of genome replicons suggest the integration of the RW130 megaplasmid into the main chromosome

The combined results of genome assembly and localisation of partition gene copies suggest that in the assembly of RW130, the megaplasmid of RW130 has become integrated into the main chromosome. The first evidence of integration is shown by the number of replicons in the genome assemblies of RW131 and RW146, both of which possess three replicons as follows; RW131 MC = 7.44Mbp, RW131 MG = 0.32 Mbp, RW131 LG = 0.14 Mbp; RW146 MC = 7.49 Mbp, RW146 = 0.29 Mbp, RW146 LG = 0.14 Mbp (Table 4.3). In contrast, the assembly of RW130 possessed two replicons (Table 4.3) consisting of the 7.70 Mbp RW130 MC, and the RW130 LG of 0.14Mbp. Notably, the combined assembly size of the main chromosome and megaplasmid of both RW131 and RW146 is similar to the size of the RW130 main chromosome. Additionally, two copies of the *parB* gene corresponding to the main chromosome and megaplasmid of RW109 respectively could be identified on the RW130 MC, whilst the same main chromosome and megaplasmid *parB* copies are localised on the main chromosome and megaplasmid respectively in both RW131 and RW146.

Further evidence of integration of the RW130 megaplasmid is shown by the circular genome plots of BLASTN results, where regions of comparatively lower identity (Figure 4.7, Panel A) between the RW130 main chromosome and those of the other isolates, display high levels of identity (Figure 4.7, Panel B) when compared to the megaplasmids of RW131 and RW146. The content of megaplasmids remained relatively consistent between RW146 and RW131 (Figure 4.7, Panel C), despite the acquisition of 31Kbp between these megaplasmid. The large plasmids of RW130, RW131 and RW146 displayed extremely high levels of identity between all three isolates (Figure 4.7, Panel D). The branch lengths of core-gene phylogenetic analysis (described in 4.2.4, Figure 4.3) also suggests integration of the RW130 megaplasmid into the main chromosome, due divergence of core-gene content in the RW130 main chromosomes when compared to the RW131 and RW146 main chromosomes, whilst possessing the same 198 core-genes as the megaplasmids of RW131 and RW146.

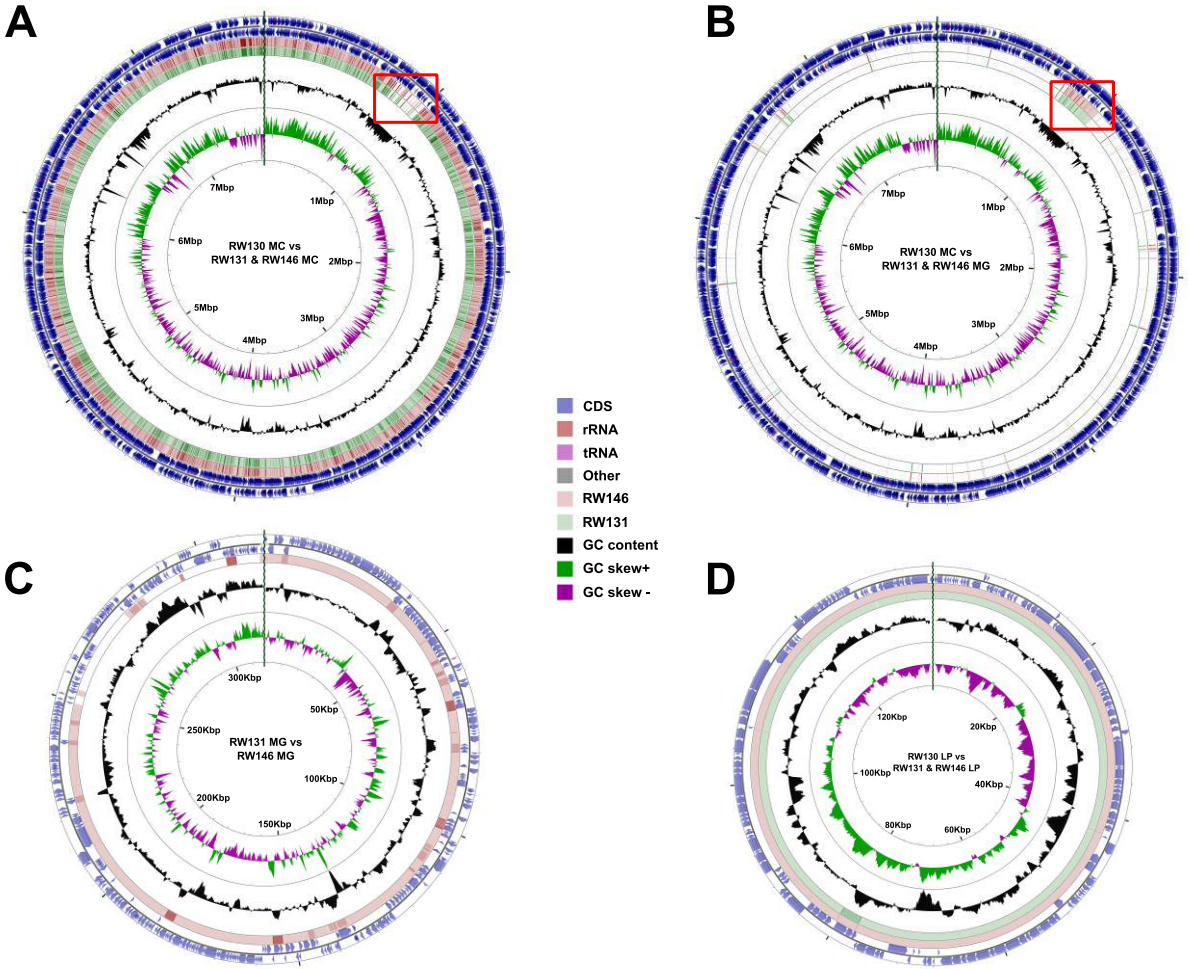


Figure 4.7: Circular genome plot of homology of main chromosomes (MCs), megaplasmids (MPs) and large plasmids (LPs) of *P. aeruginosa* RW130, RW131 and RW146. Graphical plots of the homology were generated using CGView. An individual replicon (MC, MG or LP) is shown in each panel. Concentric pink and green bands forming peripheral circles around the replicon in each panel indicate regions of homology, as determined by BLASTN, for RW146 and RW131 respectively. White areas within the same circle indicate little to no homology. Replicons and homology comparisons are shown as follows: A) MCs in comparison to the MCs of RW131 and RW146 B) RW130 MC in comparison to the MGs of RW131 and RW146. C) RW131 MG in comparison to RW146 MG. D) RW130 LP in comparison to the LPs of RW131 and RW146. Features such as coding sequences (CDS), ribosomal RNA (rRNA), transfer RNA (tRNA) and GC content are indicated by the key in the centre of the figure.

4.3. DISCUSSION AND CONCLUSIONS

The genomic and phenotypic data in this chapter highlight a clear role for the RW109 megaplasmid in resistance to chloramphenicol and trimethoprim, and that a proportion of resistance is mediated by efflux pumps. This is unsurprising, given that megaplasmids have an extensive history in conferring antibiotic resistance amongst Gram-negative bacteria, as an inevitable consequence of their phenomenal genomic capacity. A recent study details a compendium of antibiotic resistance genes located on the megaplasmids of multidrug resistant isolates in Thailand, many of which have been disseminated amongst members of the genus since the 1970s (Cazares *et al.* 2020). These findings support the role of this megaplasmid family in antimicrobial resistance, but do not provide evidence for the phenotypic role of this megaplasmid family in AMR, which is provided by the MIC results of this chapter. This study does note however, that transconjugation of megaplasmids from *Pseudomonas koreensis* and *P. aeruginosa* to *Pseudomonas fluorescens* SBW25 did not confer any fitness advantage in direct competitive fitness assays against plasmid-free competitors (Cazares *et al.* 2020). Elsewhere in the literature, antibiotic resistance dissemination by megaplasmids is well evidenced, where these large genomic elements are a well-known driver of AMR in multiple Gram-negative bacteria, including carbapenem resistance (Desmet *et al.* 2018; Lü *et al.* 2019) amongst numerous members of the family *Enterobacteriaceae*. More widely, in *Salmonella enterica*, megaplasmids have also been shown to confer resistance to antibiotics such as trimethoprim, tetracycline and sulfamethoxazole (Nicoloff *et al.* 2019).

This study only addressed MIC differences, and we were not able to address the phenotypic consequences the absence of the large and megaplasmids have in terms of the growth dynamics, both in the presence and absence of CHL and TRI antibiotics. Such an assay could be performed by performing growth curve assays by means of optical density, over a 24-hour time period (Lou *et al.* 2015). Future work should also confirm that the optical density reductions observed in this study are indeed MIC breakpoints, and not bactericidal. This can be achieved by conducting the same MIC assay for both trimethoprim and chloramphenicol and plating small volumes from the chosen MIC breakpoint and higher concentrations onto media such as Tryptone Soy Agar, before incubating (Rushton *et al.* 2013). This would both confirm cell viability

at the MIC and determine the bactericidal concentrations of TRI and CHL in relation to RW109, RW109-RB and RW109-RB-NaB.

Additionally, this study does not account for the role of each plasmid in isolation, as the results suggest the absence of both the large and megaplasmids in RW109-RB-NaB. Progressive knockout of each plasmid in isolation could be achieved by the bioinformatic identification of toxin and antitoxin systems, e.g. using BLAST based tools such as ABRicate (see 2.4.12), and knock-out of any toxin-antitoxin system and partition genes via the introduction of a stop-codon by means of CRISPR/Cas9-based Genome Editing (Chen *et al.* 2018). A similar protocol could also be used to determine the role of large and megaplasmids in the other industrial *P. aeruginosa* isolates sequenced in this study. In somewhat similar vein however, short-read mapping to the RW109 alone does not confirm the complete absence of the large and megaplasmids in the adapted derivative. This could be achieved by completing the genome of RW109-RB-NaB using long-read sequencing, such as those utilised in this study to provide near-complete genomes of the industrial panel (Rhoads and Au 2015). Once plasmid loss is confirmed, future studies should look at the consequences of plasmid loss on resistance to Sodium Benzoate, to which the plasmid lacking isolate was progressively exposed prior to plasmid loss. It would be of interest to determine whether this plasmid loss confers additional resistance to NaB.

With regard to genomic analyses, partition gene and core-gene analyses indicate that other plasmids within this emerging megaplasmid family play a role in tolerance to maintaining fitness traits in the *Pseudomonas* genus in a variety of stressful environments. This role is supported by the AMR and biocide resistance gene analyses in this study, which demonstrate the ubiquity of predicted metal resistance genes amongst the megaplasmids. Traits conferred which contribute to tolerance of stressful environment, as reported in the literature, include antibiotic resistance for pSY153-MDR, pBM413, p12969-DIM, pOZ176 (Yuan *et al.* 2017) and pJB37 (Botelho *et al.* 2017); mediating biodegradation processes in wastewater treatment (pRBL16) (Zheng *et al.* 2016) and solvent tolerance (pTTS12, Kuepper *et al.* 2015). However, whilst the genomes presented and chosen for analyses are near complete, assembly errors do exist, as evidenced by the complexity of genome assembly, which required scaffolding, and resulted in final genomes with undetermined nucleotides. Future work should look to examine read mapping coverage on a region by region basis for

the genomes presented, and look to resolve any regions of low confidence/coverage by means of further next generation sequencing (Deamer *et al.* 2016), or sanger sequencing (Pareek *et al.* 2011).

Overall, the aims and objectives of this chapter were addressed as follows:

1. **To comparatively examine the phenotype of RW109 and a plasmid-lacking derivative in order to determine the role of large and megaplasmids in RW109.** Efflux plays a significant role in TRI and CHL resistance in RW109, a proportion of which is mediated by the presence of the large and megaplasmid, as evidenced by the higher baseline MICs for TRI and CHL in the parental, and Robbins device control strains of RW109, compared to RW109-RB-NaB. The MIC of all isolates were reduced to similar concentrations in the presence of the efflux pump inhibitor PaβN.

2. **To complete the genomes of, and characterise large and megaplasmids present in 6 isolates with evidence of a megaplasmid and efflux pump (RW130, RW146, RW172, RW176 and RW184), and one isolate with evidence of a megaplasmid, but no efflux pump (RW131)**. The assemblies of RW130, RW172, RW176 and RW184 assemblies possessed two replicons, consisting of a main chromosome and megaplasmid for all isolates, excluding RW130, where the assembly consisted of a main chromosome and large plasmid. RW131 and RW146 possessed a main chromosome, megaplasmid and large plasmid. The main chromosomes, megaplasmids and large plasmids of all industrial isolates belonged to the same families as those of RW109.

3. **To determine the changes in genomic synteny underpinning persistence of three isolates, RW130, RW131 and RW146, isolated from the same industrial product over a six-year period.** The genome structure of isolates and RW146 and RW131 are largely similar, but the RW130 genome assembly has evidence of megaplasmid integration into its main chromosome.

Overall, this chapter addresses, and confirms the following hypotheses:

I. The megaplasmid of RW109 plays a role in resistance to trimethoprim and chloramphenicol.
II. Megaplasmids were present throughout all of the sequenced industrial isolates.
III. Significant differences in the genome structure were seen when comparing RW130, RW131 and RW146, namely in the integration of a megaplasmid into the main chromosome of RW130.

5. The genomics of industrial *Enterobacteriaceae*

5.1. *Enterobacteriaceae* - FREQUENTLY RECLASSIFIED, OFTEN MISIDENTIFIED

Enterobacter spp. are a genetically diverse group of Gram-negative, aerobic, motile bacilli, bacteria which are isolated from a variety of environments, including the soil (Khalifa *et al.* 2016), plant phyllosphere (Williams and Marco 2014), water (Davin-Regli and Pagès 2015), processing facilities (Mullane *et al.* 2007; Møretrø and Langsrud 2017), outer space, and the nosocomial environment (Noël *et al.* 2019), as well as being components of the human gut microbiota. In the nosocomial setting, the recently reclassified *Klebsiella aerogenes* (Malek *et al.* 2019) and *Enterobacter cloacae*, (*E. cloacae*) are amongst the most frequently isolated *Enterobacteriaeceae* from immunocompromised patients (Davin-Regli and Pagès 2015). Their frequency of isolation is a consequence of their intrinsic antimicrobial resistance (Eichenberger and Thaden 2019), highly virulent/transmissible nature (Azevedo *et al.* 2018), and subsequent propensity to cause secondary infections as opportunistic pathogens (Fei and Zhao 2013). The aforementioned characteristics have resulted in the placement of *Enterobacter* spp. and *Klebsiella pneumoniae* (*K. pneumoniae*) amongst the ESKAPE pathogens (Mulani *et al.* 2019), six bacterial groups which are recognised as the most troublesome pathogens in the context of clinical antimicrobial resistance.

There are currently 22 species within the *Enterobacter* genus (Davin-Regli *et al.* 2019), many of which have emerged from previously established species. This emergence owes to a transition from traditional phenotypic and molecular techniques for taxonomic delineation, such as biochemical profiling and wet-lab DNA-DNA hybridisation (Izard *et al.* 1983; Farmer *et al.* 1985), to whole-genome comparison tools such as average nucleotide identity (Potter *et al.* 2018), in-silico DNA-DNA hybridisation (Hardoim *et al.* 2013) and core genome analysis (Caputo *et al.* 2019). Numerous examples of this exist, such as the use of multilocus sequence based methods to re-classify *Enterobacter* species (Brady *et al.* 2013). An investigation of a contaminated batch of liquid hand soap identified a single strain of *Klebsiella oxytoca* was present in the product bottles and clearly illustrates that accurate species identification can be achieved in the non-food product sector (Dieckmann *et al.* 2016). A combination of rapid phenotypic methods (MALDI TOF MS and infrared spectroscopy) and genotypic methods (pulsed-field gel electrophoresis, MLST and

whole genome sequencing) was used to accurately identify the soap contaminant to both the species and strain level. Multiple media for culture and putative identification of *Enterobacteriaceae* are available (Wilkinson *et al.* 2012); extensive genomic databases such as Enterobase are also available for detailed strain characterisation (Alikhan *et al.* 2018). Overall, despite the widespread availability of the aforementioned methods, many members of the *Enterobacteriaceae* remain subject to a taxonomic ambiguity (Brady *et al.* 2013), and the wider application of multiple identification methods should be encouraged.

The risks posed by the *Enterobacteriaceae* are not limited to the clinical environment, as this family are also frequent contaminators of industrial products (Périamé *et al.* 2014; Neza and Centini 2016). Reports from the EU Safety Gate (formerly RAPEX) database between 2005-2019 (see 1.4.1) highlight that the most frequently contaminated non-sterile consumer products are cosmetics and toys. Moreover, the majority of microorganism related contamination is constituted by *Pseudomonas* spp. (particularly *Pseudomonas aeruginosa*; *P. aeruginosa*), closely followed by the *Enterobacteriaceae* (Vincze *et al.* 2019) (see 1.4.1). Multiple *Enterobacteriaceae* species have been encountered as non-food product contaminants, and recent reports show certain species within this highly related group of Gram-negative bacteria have become quite problematic, and are even found alongside other Gram-negative bacteria in over 90% of in-use cosmetics (Bashir and Lambert 2020).

Enterobacter gergoviae (*E. gergoviae*) is a well-known, intrinsically preservative-resistant contaminant of food and cosmetic products. Recent taxonomic reclassification of several *Enterobacter* species placed *E. gergoviae* isolated within a new genus *Pluralibacter* (as *P. gergoviae*) using multilocus sequence based methods (Brady *et al.* 2013). A recall of 15,000 tubes of a best-selling skincare product because of high levels of *P. gergoviae* was reported in the popular press (Cherrington 2016). This brought public attention to the issue of non-food product contamination. The home and personal care industries however, are no strangers to *P. gergoviae*, as this species has been responsible for 8 major European recall incidents in the past decade alone (Cunningham-Oakes, Weiser, *et al.* 2020). This is unsurprising, given the numerous biocide and preservative resistance mechanisms that have been identified in *P. gergoviae*, including outer membrane modifications such as efflux, which play a role in resistance to methylparabens and methylisothiazolinones (Périamé *et al.*

2015). Overall, the presence of a myriad of biocide and antibiotic resistance mechanisms appears to be a ubiquitous trait in the *Enterobacteriaceae* (Périamé *et al.* 2014; Davin-Regli *et al.* 2019).

My previous analysis of factory isolates and biocide reference strains of *Enterobacteriaceae* highlighted that *Enterobacteriaceae* from multiple backgrounds possessed greater numbers of predicted biocide resistance genes than *Pseudomonas* spp.*,* including *aeruginosa* (Figure 2.7). Furthermore, this analysis highlighted that the industrial biocide reference strain *P. gergoviae* ECO77 has multiple (211) predicted biocide and metal resistance genes. This was comparable to that of the industrial biocide reference strain *P. aeruginosa* RW109, which possesses 180 predicted genes. It has also been shown phenotypically that RW109 possesses elevated preservative tolerance in comparison to other *P. aeruginosa* from non-industrial environments (Weiser *et al.* 2019). RW109 and other *P. aeruginosa* from the industrial environment uniquely possess large segments of DNA known as megaplasmids, conferring extra genomic capacity with 500+ genes, and may thus play a role in preservative resistance (Weiser *et al.* 2019). Furthermore, *Enterobacteriaceae* possess multiple mobile genetic elements and mechanisms for transposition, and multiple studies suggest clinical *Enterobacteriaceae* harbour megaplasmids, which are contributing to the global spread of antibiotic resistance amongst this family of bacteria (Noël *et al.* 2019).

The aforementioned evidence from industrial *P. aeruginosa* strains begs the question; how prevalent are large and megaplasmids amongst contaminant strains of *Enterobacteriaceae*, and do they contribute to the presence of biocide and metal resistance genes in these strains?

To investigate this question, the objectives of this chapter were as follows:

1. **To utilise whole genome sequencing to determine the identity and establish the taxonomic diversity of a panel of organisms, all of which were isolated from industrial contamination incidents, and were initially identified as *Enterobacter* sp. by 16S rRNA sequencing**.

2. **To determine the presence of large plasmids and megaplasmids amongst these organisms using short read sequencing data in this panel, in the context**

of a curated genome panel of *Enterobacteriaceae* from a variety of backgrounds.

3. **To determine the validity of plasmid prediction tools for detecting large and megaplasmids in short read assemblies, by completing the genome of two *P. gergoviae* isolates, ECO77 and ECO124, using long-read sequencing technologies.**

4. **To determine the relationship between the presence of large and megaplasmids, and the presence of biocide and metal resistance genes.**

Based on previous literature and the findings of this chapter, a large plasmid was defined as a plasmid of 60 to 200Kbp, whilst a megaplasmid was defined as a plasmid of >200Kbp. The overall hypothesis for this chapter was "*Enterobacteriaceae* isolated from industrial products will display wide taxonomic diversity, and harbour large and/or megaplasmids".

5.2. Materials & methods

5.1.1. Isolation and long-term storage of organisms

16 isolates were historically isolated from contaminated industrial products by first neutralising the preservatives in the product using Tween 80 as a neutralising agent, before enriching on TSA. Growth from TSA plates was then used to create freezer stocks in cryogenic vials, as described in 2.4.3. Isolates were revived by inoculating TSA with a single bead, before incubating at 30°C for 16-18 hours. A swab of confluent material was then used to inoculate 3 mL of TSB, which was grown overnight for 16-18 hours on an orbital shaker at 150rpm. Confluent overnight culture was used to create freezer stocks of individual isolates with 8% (v/v) DMSO before storing at -80°C.

5.1.2. Short-read sequencing, assembly, taxonomic placement and gene prediction for draft *Enterobacteriaceae* genomes

DNA extraction for genome sequencing of the aforementioned *Enterobacteriaceae* isolates was performed as described previously (section 2.4.5) (Weiser *et al.* 2019). Libraries for short-read sequencing were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs), and were

quantified and sequenced as described previously (see 2.4.9). Reads were quality trimmed, then assembled using Unicycler (v0.4.8) (see 2.4.10), and assessed for quality using Quast v5.0.2, which provides quality metrics such as the total number of contigs, N50 values and GC content for each genome. For genome quality assessment, only contigs ≥1000bp were considered. Genomes were assessed for contamination using Kraken2 v2.0.8-beta (section 2.4.10) and placed against reference genome datasets by ANI, which were obtained using enaDataGet as described previously (section 2.4.11). Antimicrobial, biocide, and metal resistance genes were predicted using ABRicate as described previously (see 2.4.12). Plasmids were predicted from short reads using ABRicate with the PlasmidFinder database (Carattoli *et al.* 2014).

5.1.3. Hybrid genome assembly using Illumina and MinION reads to produce a near-complete genome for *P. gergoviae* ECO77 & ECO124

Illumina reads generated as described above (see 5.2.2) and long-reads generated by Oxford Nanopore Technologies (ONT) MinION (MIN-101B), using a MinION fluidics module R9 version flow cells (FLO-MINSP6, R9.4.1, Oxford Nanopore Technologies) were used to generate hybrid assemblies using Unicycler (v0.4.8) (Wick *et al.* 2017). When utilising long reads to produce hybrid assemblies, Unicycler incorporates Racon, an ultrafast consensus module for raw de novo genome assembly from long-reads (https://github.com/isovic/racon), with the aim of producing a complete genome, with each replicon in a single contiguous sequence (contig). Long-read sequencing libraries were generated using a bespoke version of the Nanopore rapid barcoding kit (SQK-RBK004) protocol. DNA for sequencing was extracted twice using the Maxwell®16 Tissue DNA Purification Kit as described previously (see 2.4.5) and pooled for each sample. DNA concentration and fragment size were assessed using a Qubit fluorometer and Agilent Tapestation 2100 (Tapestation) respectively, before 15 µg of genomic DNA per sample was sheared to ~20 kbp fragments using the g-TUBE Covaris instrument (5,800 rpm, 60 seconds). Fragment quality and size was then confirmed using the Tapestation. Fragment-size exclusion was then performed in two stages to remove small DNA fragments prior to sequencing. A 0.4x volume of SPRI beads was first used to reduce the number of fragments <1500bp. SPRI beads and AMPure beads were then used as per manufacturer's instructions (Beckman

Coulter, Indianapolis, USA) to remove fragments <1000bp. Finally, the size-excluded DNA was re-quantified by Qubit fluorometry and diluted to 400ng in 7.5ul of molecular grade water as per the manufacturer's instructions (Oxford Nanopore Technologies, Oxford, UK).

DNA samples were pooled in ratios based upon draft genome sizes (e.g. 5.5Mbp = 5.5µl), before the pool was cleaned using AMPure beads. The concentration of the final pool was assessed using Qubit fluorometry, which was used to load 650ng of library onto a MinION flow cell, which possessed 375 active pores prior to sequencing, and ran until 100% base-calling was achieved using the MinKNOW base-calling software. Reads exported from the software were then quality-filtered and underwent adapter sequence removal using Porechop version 0.2.4, (https://github.com/rrwick/Porechop). Quality-trimmed long and short-reads were then used by Unicycler to produce hybrid assemblies, which were assessed for replicon contiguity using Bandage v0.8.0 (Wick *et al.* 2015).

5.1.4. Inferring the ancestry of large and megaplasmids in *P. gergoviae* ECO77 & ECO124

Once assembled, the largest complete contigs of each replicon (equal to the N50 of that replicon) of the near-complete ECO77 and ECO124 genomes were initially annotated using Prokka (v1.14.5) (Seemann 2014) under default parameters. Prokka uses BLAST+ BLASTP, which uses external feature prediction tools to predict the following genomic features in contigs; coding sequences (Prodigal), transfer RNA gene (Aragorn), signal leader peptides (SignalP), ribosomal RNA genes (RNAmmer), and non-coding RNA (Infernal). However, as Prokka is designed for whole bacterial genomes, it does not possess a well-curated database for plasmids beyond the annotation of genes that are also present in bacterial chromosomes, such as replication initiator proteins, which are responsible for plasmid copy number. To remedy this, a curated database was developed by identifying the replication initiator protein gene *repA* in the annotations of each replicon and extracting the corresponding sequence via command line. The extracted sequence was then used to infer the origin of each large and megaplasmid by identifying homologous *repA* genes by BLASTN from GenBank. The accession numbers of assemblies corresponding to the top 20 hits were used to obtain curated genome assemblies and protein annotations for closely related large and megaplasmids by

via ncbi-genome-download (see 4.1.4). The protein annotations were then used to generate a custom database, which was then used re-annotate ECO77 and ECO124 and all assemblies for closely related large and megaplasmids, by using Prokka with the proteins flag. Roary, (v3.12.0) (Page *et al.* 2015) was then used as described previously (see 4.1.9) to determine the relatedness of *Enterobacteriaceae* megaplasmids. For this analysis, default parameters were used for plasmid family 3, which was only present in ECO124 (see 5.3.5.3), which define a core-gene as being present in 99% of input sequences. Due to the high variability of plasmid families 1 & 2 however, parameters were altered to define a core-gene as a gene present in 90% of plasmid and subsequently produce an alignment. MAFFT was then used for core-gene sequence alignment, and phylogenetic trees were generated using RAxML. Phylogenetic trees were visualised as described previously (see 2.4.7).

5.1.5. Determining the core, antibiotic, biocide and metal resistance genes conferred by megaplasmids

To determine the core-genetic constituents within the large and megaplasmids of ECO77 and ECO124, core-genes within each large and megaplasmid family of ECO77 and ECO124 were extracted from the outputs of Roary. In order to determine the level of predicted resistance conferred by large and megaplasmids, antimicrobial, biocide, and metal resistance genes were predicted in each megaplasmid family, and the whole bacterial genome assembly from which a given plasmid was extracted as described previously (see 4.1.10). Boxplots were generated for absolute gene counts using the boxplot package in R as described previously (see 2.4.12). Means and standard deviations were also calculated as described previously (see 2.4.12).

5.3. Results

5.1.6. Genome quality of Illumina drafts and near-complete hybrid assemblies

Genome sequence analysis of the 16 industrial *Enterobacteriaceae* isolates were of good quality to support further analysis and the metrics are summarised in Table 5.1. Genome assembly size of the Illumina drafts from short reads ranged from 4.80 Mbp (ECO76) to 5.99 Mbp (ECO128). The number of contigs per assembly ranged from 32-157 per genome. The %GC content for each organism ranged from 55.4 – 58.6%. Following hybrid assembly, the genome assembly size of ECO77

increased from 5.96 to 6.15Mbp, and ECO124 increased from 5.91 to 6.20 Mbp. The number of contigs for each genome was reduced from 137 to 15, and 157 to 14, whilst %GC content was decreased from 58.15% to 57.96% and from 58.04% to 57.84% for ECO77 and ECO124 respectively.

5.1.7. Average Nucleotide Identity reclassifies 5 incorrectly identified *Enterobacteriaceae* isolates, and confirmed the identity of *P. gergoviae* isolates

Of the 16 isolates subject to genome sequencing, 3 genomes were removed from the dataset and further analysis due to contamination. Average Nucleotide Identity provided species-level placement for all *Enterobacteriaceae* genomes. This revealed that 8 isolates were all correctly identified by 16S rRNA gene sequencing as *P. gergoviae*, and that five isolates had been incorrectly identified by 16S rRNA gene sequencing. ECO126 was re-identified as *Klebsiella variicola* (*K. variicola*) after being historically identified as *E. cloacae* (99.2% ANI with the type strain DSM 15968^{T}). ECO118 was historically matched with multiple *Enterobacteriaceae* by 16S rRNA gene sequencing but was re-identified as *Leclercia adecarboxylata* (*L. adecarboxylata*; 98.8% ANI identity when compared to the representative genome, USDA-ARS-USMARC-60222^{R}). ECO76 and ECO127 were previously identified as *E. cloacae* but were found to belong to the species *Enterobacter hormaechei* (*E. hormaechei*; 95.2% ANI identity when compared to the type strain ATCC 49162^{T}). ECO120 was previously identified as *E. cloacae* but was reclassified as *Enterobacter soli* (*E. soli*, 97.5% ANI identity with the type strain, LMG 25861^{T}). A summary of these taxonomic changes after ANI analysis is shown in Table 5.2.

5.1.8. The number of predicted AMR genes were consistent throughout the *Enterobacteriaecae,* whilst predicted biocide genes were lower in *Leclercia* sp.

For each industrial isolate, and each of their nearest neighbours within the same genera, the number of predicted AMR gene content remained fairly consistent, with a mean number of predicted genes as follows: *Enterobacter* = 37 ± 4, *Klebsiella* = 35 ± 2, *Leclercia* = 30 ± 3 *Pluralibacter* = 28 ± 4. The number of predicted biocide genes also remained relatively consistent between *Enterobacter*, *Klebsiella*, and *Pluralibacter*, with mean numbers of predicted genes at 79 ± 12, 79 ± 6, and 93 ± 9, respectively. The number of predicted biocide genes in *Leclercia*

spp. was lower at 62 ± 7. The industrial isolate *L. adecarboxylata* ECO118 was an outlier in the *Leclercia* dataset, with 80 predicted biocide resistance genes. All data is shown in Figure 5.5.

Table 5.1: Genome quality of industrial isolates initially identified as *Enterobacter* spp. by 16S rRNA gene sequencing						
Strain	**Product type**	**Total Genome Size (Mbp)**	**Contigs > 1000bp**	**Contigs > 5000bp**	**GC (%)**	**N50 (bp)**
ECO76	Fabric Conditioner	4.80	45	23	55.35	218938
ECO77[H]	Skin care	6.15	15	5	57.96	5368409
ECO77[I]	Skin care	5.96	137	30	58.15	168916
ECO118	Raw Material	5.10	80	27	55.12	218528
ECO120	Unknown	5.21	80	33	53.45	137654
ECO121	Skin care	5.77	110	35	58.55	125990
ECO122	Skin care	5.92	152	33	58.02	164101
ECO123	Skin care	5.78	86	36	58.49	140803
ECO124[H]	Skin care	6.20	14	5	57.84	5368697
ECO124[I]	Skin care	5.91	157	36	58.04	136601
ECO125	Skin care	5.81	75	25	58.41	259693
ECO126	Dish Wash Liquid	5.47	32	18	57.46	344325
ECO127	Fabric Wash Liquid	4.81	43	22	55.35	278377
ECO128	Skin care	5.99	135	30	58.02	217534
ECO129	Skin care	5.79	100	25	58.33	288226

Footnote. For each genome sequenced strain, the product type from which the strain was isolated is indicated. For each genome, the total assembly size, number of contigs > 1000bp, number of contigs > 5000bp, percentage GC content, and N50 values are shown. All metrics excluding the number of contigs > 5000bp, are based on contigs > 1000bp. For genomes that were sequenced by both Oxford Nanopore and Illumina technologies, [I] = assembly generated using Illumina reads, [H] = hybrid assembly generated using Illumina and Oxford Nanopore reads.

Table 5.2: Genome-derived IDs of industrial isolates initially identified as *Enterobacter* spp. by 16S rRNA gene sequencing.			
Strain	**16S rRNA gene sequencing identification**	**ANI identification**	**Percentage ANI (%) to Type strain**
ECO76	*Enterobacter cloacae*	*Enterobacter hormaechei*	95.2
ECO77	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	98.0
ECO118	*Enterobacter* (*cloacae*/*kobei*/*ludwigii*) / *Leclercia adecarboxylata*	*Leclercia adecarboxylata*	98.8
ECO120	*Enterobacter cloacae*	*Enterobacter soli*	97.5
ECO121	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	99.0
ECO122	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	98.9
ECO123	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	99.9
ECO124	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	98.9
ECO125	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	99.0
ECO126	*Enterobacter* sp.	*Klebsiella variicola*	99.2
ECO127	*Enterobacter cloacae*	*Enterobacter hormaechei*	95.2
ECO128	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	98.9
ECO129	*Enterobacter gergoviae*	*Pluralibacter gergoviae*	99.0

Footnote. For each isolate, the original identification by 16S rRNA gene sequencing, the species-level identification by average nucleotide identity, and the percentage identity in comparison to the type or reference genome as determined by ANI analysis is shown.

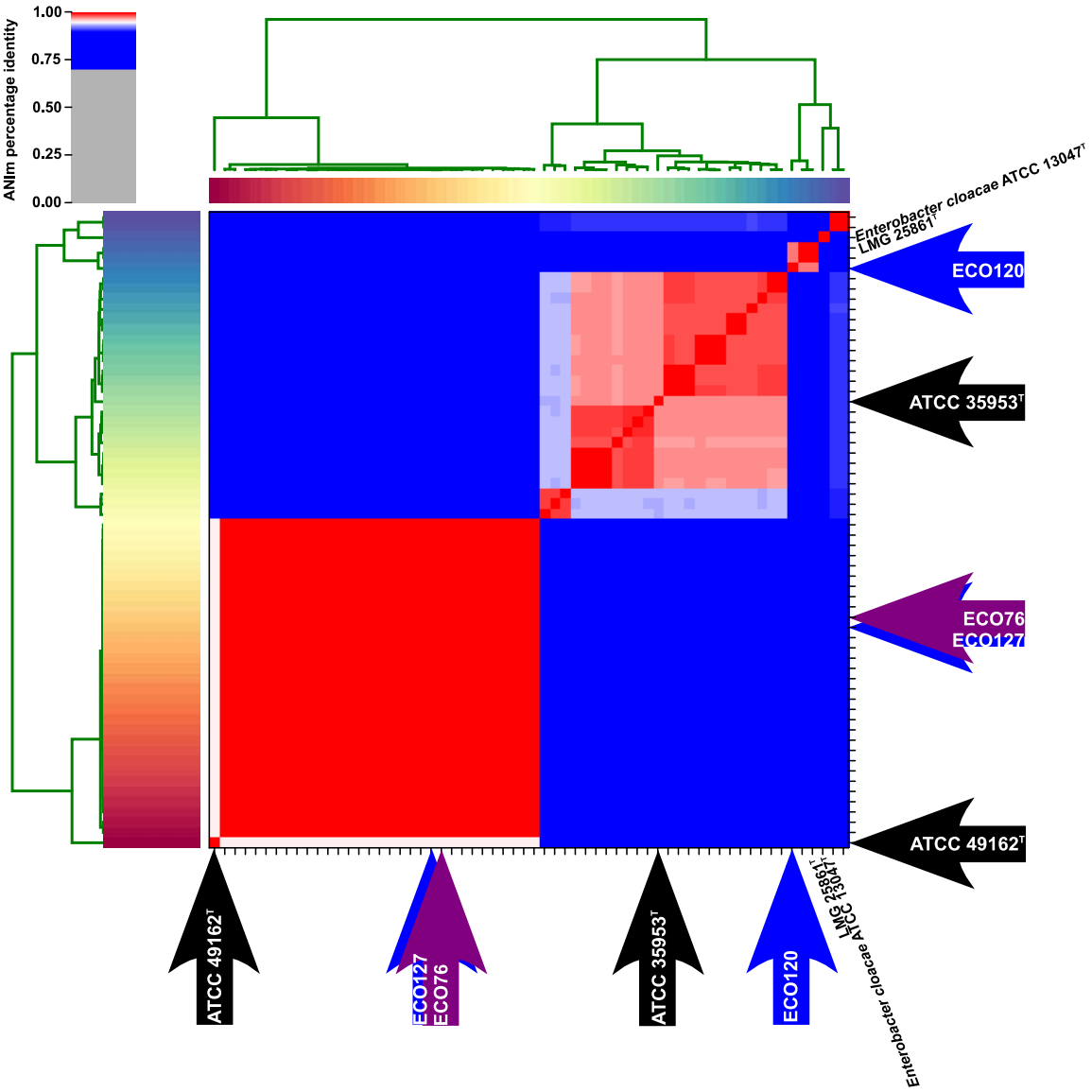


Figure 5.1: Average nucleotide identity analysis accurately speciates the industrial Enterobacteria isolates. Heatmap generated by the PyANI script, placing the genomes of several industrial isolates (denoted by blue arrows) and biocide reference strain ECO76 (purple arrow) amongst closely related reference genomes of *Enterobacter* of various species. Red areas highlight isolates that possess >95% nucleotide similarity, with darker shades of red indicating greated similarity. Blue indicates <95% nucleotide similarity. Type strains are denoted by black arrows, and are annotated (T)

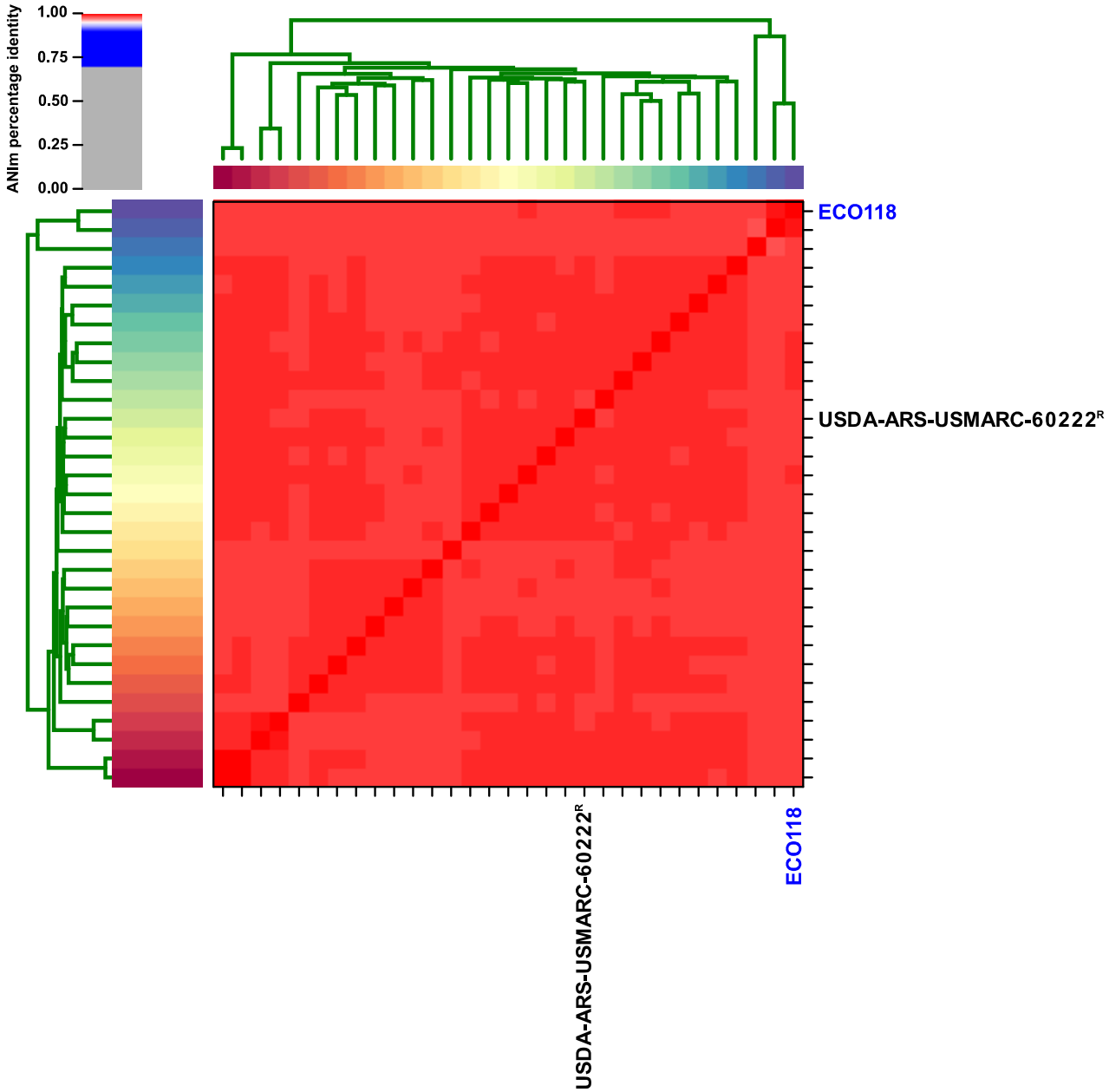


Figure 5.2: Average Nucleotide Identity places ECO118 as *L. adecarboxylata*. Heatmap generated by the PyANI script, placing the genomes of industrial isolate ECO118 (blue) amongst closely related reference genomes of *L. adecarboxylata* of various species. Red areas highlight isolates that possess >95% nucleotide similarity, with darker shades of red indicating greated similarity. Blue indicates <95% nucleotide similarity. The representative genome for this species USDA-ARS-USMARC-602222 is annotated (R).

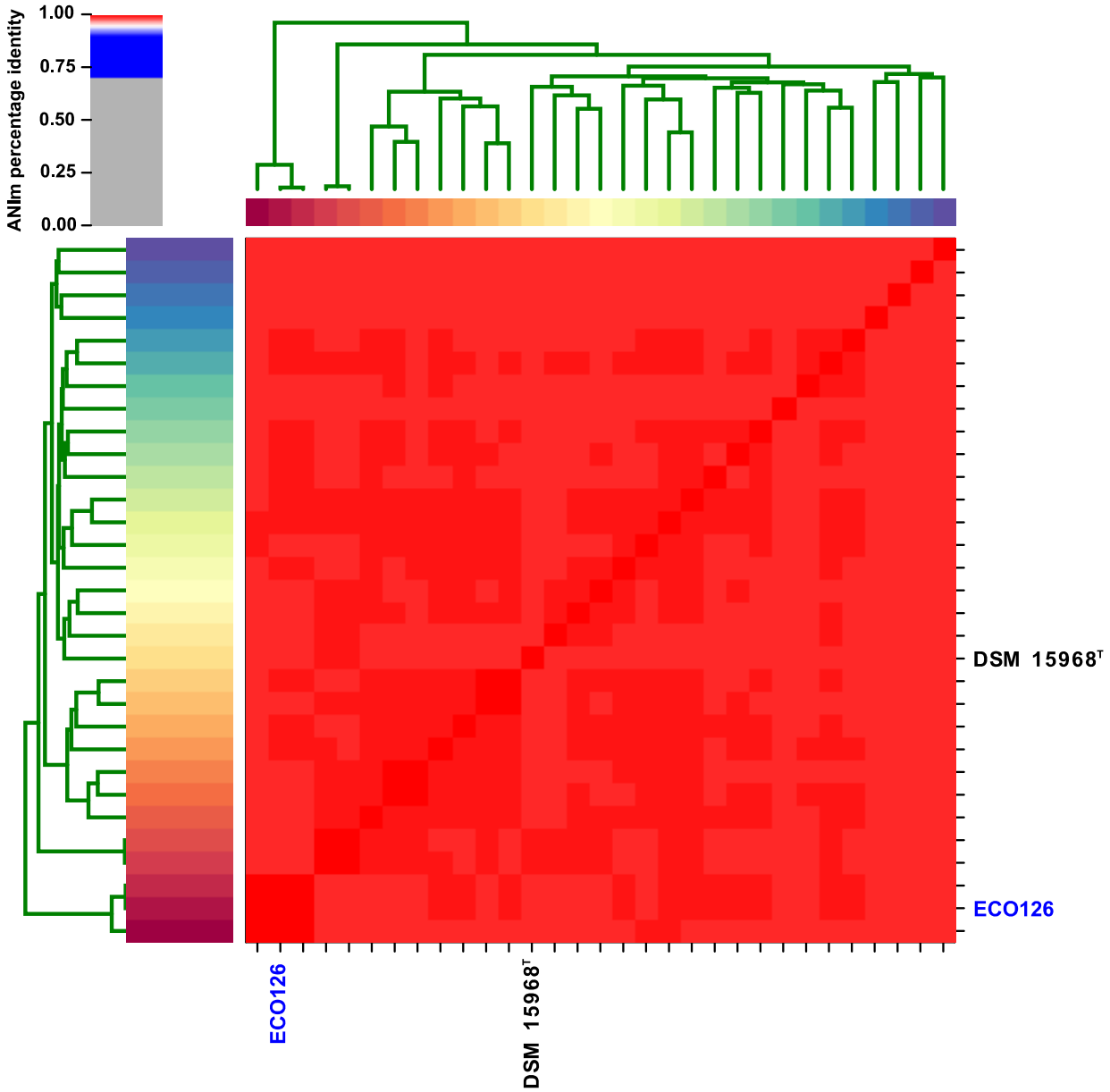


Figure 5.3: Average Nucleotide Identity places ECO126 taxonomically as *K. variicola*. Heatmap generated by the PyANI script, placing the genome of industrial isolate ECO126 amongst closely related reference genomes of *K. variicola*. Red areas highlight isolates that possess >95% nucleotide similarity, with darker shades of red indicating greated similarity. Blue indicates <95% nucleotide similarity. The industrial isolate ECO126 is shown in blue, whilst the type strain DSM 15968 is annotated (T).

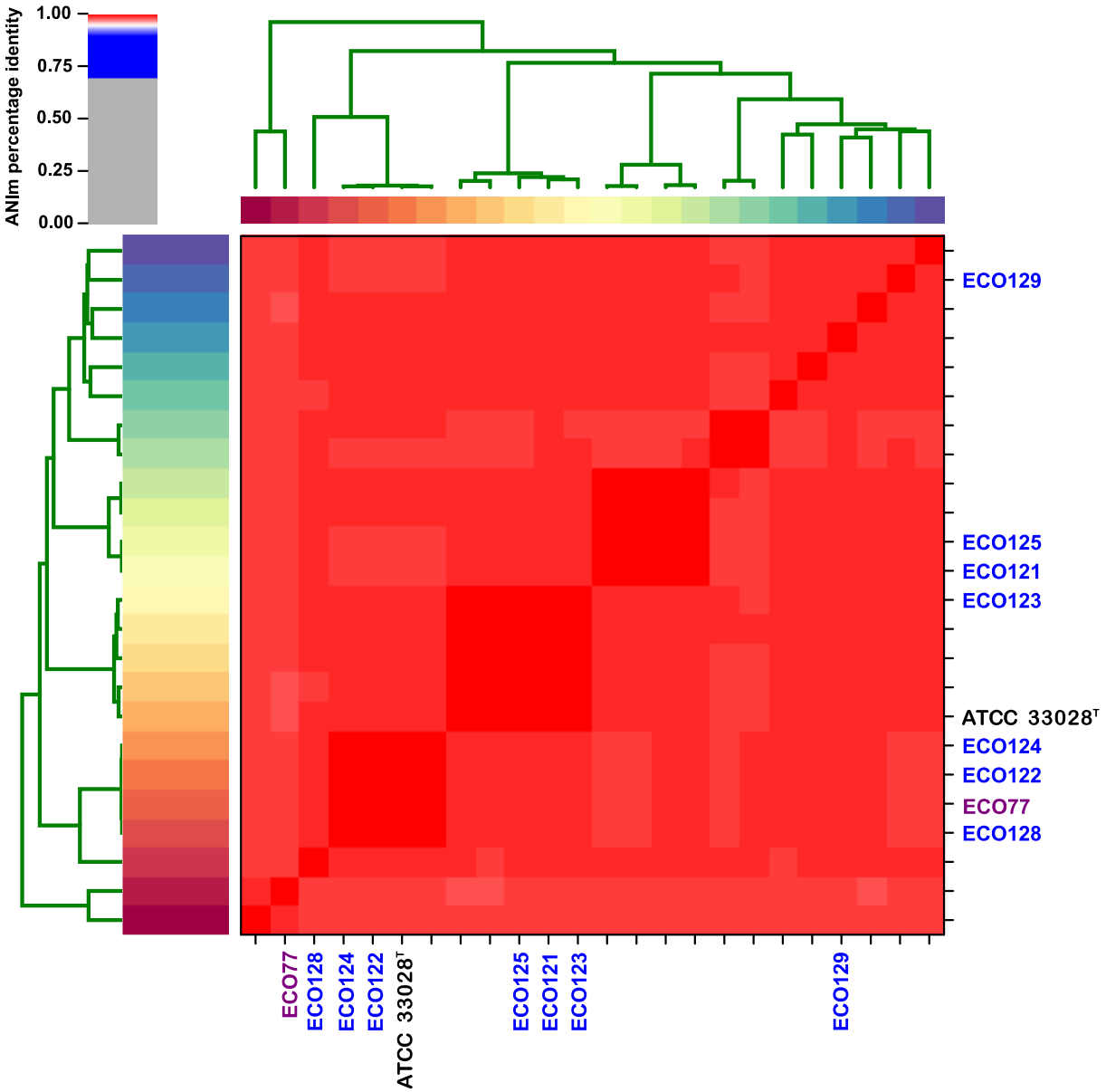


Figure 5.4: Average Nucleotide Identity taxonomically places 8 industrial *Enterobacteriaceae* as *P. gergoviae:* Heatmap generated by the PyANI script, placing the genomes of several industrial isolate (shown in blue) amongst closely related reference genomes of *P. gergoviae.* Red areas highlight isolates that possess >95% nucleotide similarity, with darker shades of red indicating greated similarity. Blue indicates <95% nucleotide similarity. The type strain ATCC 33028 is annotated (T).

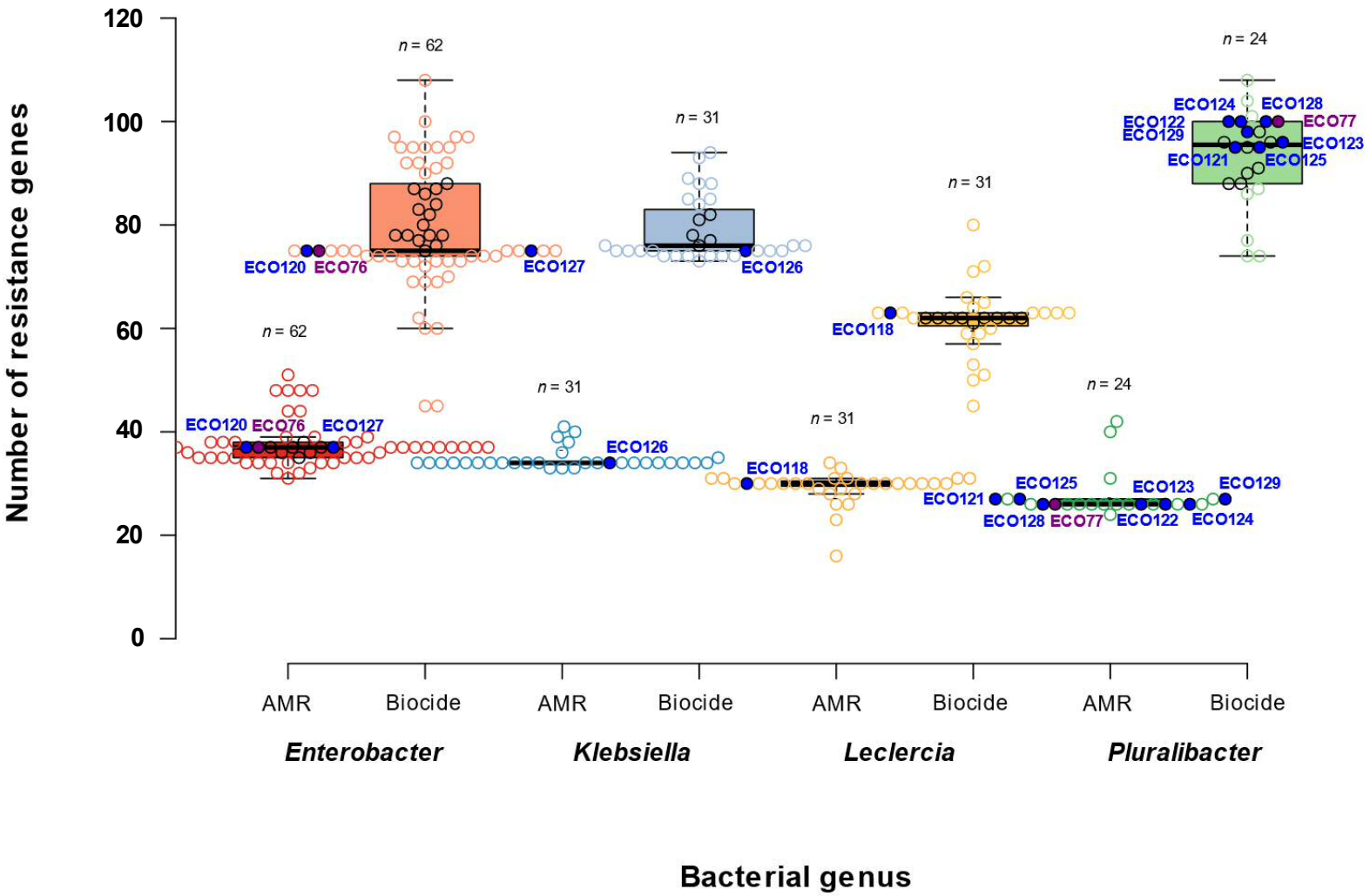


Figure 5.5: Prediction of antimicrobial resistance (AMR) and Biocide resistance genes present in industrial *Enterobacteriaceae* The number of predicted genes is shown for both sequenced industrial isolates, and their most taxonomically related reference genomes. Isolates and reference genomes were analysed for AMR and biocide resistance genes by genus. Isolates derived from industrial product are highlighted in blue, whilst isolates of industrial origin which are now used in biocide testing panels are shown in purple. For visual simplicity, means are indicated in the main body of the results text, whilst the median is shown using a black line on the boxplot. Whiskers were defined using the Tukey HSD test.

5.1.9. Megaplasmids and large plasmids are predicted widely across *Enterobacteriaceae*

Large and megaplasmids were predicted widely across the *Enterobacteriaceae*, apart from *L. adecarboxylata*, as follows:

- ***Enterobacter* spp.:** 36 plasmid hits were produced across the 62 assemblies. Four pairs of hits linked back to the same accession, making for a total of 32 unique plasmid accessions. Of these unique accessions, 4 were associated with megaplasmids, and 14 were associated with large plasmids. No megaplasmids were predicted in industrial *Enterobacter* spp., but two large plasmids were predicted in *E. soli* ECO120; *E. cloacae* subsp. *cloacae* ATCC 13047 plasmid pECL_A (199562 bp, Accession: CP001919.1), and *E. soli* strain LF7a plasmid pENTAS01 (166725 bp, Accession: CP003027.1). The BLAST hits for these plasmids were met with 100% coverage, and hits associated with their accessions were produced from 12 and 21 genomes within the dataset respectively.
- ***K. variicola*:** 12 plasmid hits were produced across 31 assemblies, all of which were associated with unique accessions. Two of these accessions were associated with megaplasmids, and four with large plasmids. No megaplasmids or large plasmids were predicted within the sole industrial isolate, ECO126 (see Table 5.3).
- ***L. adecarboxylata*:** Plasmids hits were only produced for the genome assemblies of industrial isolate ECO118, and reference UBA902. The hits for each isolate were unique to each isolate. Plasmid hits in industrial isolate ECO118 did not correspond to megaplasmids, instead corresponding to two small plasmids and two large plasmids, of 199562 bp and 99159 bp originating from the *E. cloacae* subsp. *cloacae* type strain ATCC 13047 (Accession: CP001919.1) and *Escherichia coli* (*E. coli*) K12 (Accession: AP001918.1) respectively.
- ***P. gergoviae*:** 19 plasmid hits were produced across the 24 assemblies. Within these hits, 3 pairs of hits linked back to the same accession, for a total of 16 unique plasmid accessions (see Table 5.4). Of these accessions, 5 were for megaplasmids, and 4 for large plasmids. Three out of four of the predicted megaplasmids were present within industrial isolates as follows: *K.*

pneumoniae strain ST258 plasmid pKPN-IT (Accession: JN233704.1) in 4/8 isolates, *K. pneumoniae* plasmid pNDM-MAR (Accession: JN420336.1) in 6/8 isolates, and *Citrobacter freundii* strain CAV1321 plasmid pKPC CAV1321-244 (Accession: CP011611.1) in 4/8 isolates, with 100% coverage for all hits. Three out of four of the predicted large plasmids were present within industrial isolates: *K. variicola* strain 342 plasmid pKP91 (Accession: CP000966.1) and *K. pneumoniae* strain NK245 plasmid pK245 (Accession: DQ449578.1) in 2/8 isolates, and *E. cloacae* strain T5282 plasmid pT5282-CTXM (Accession: MF062700.1) in 4/8 isolates. Notably, the prediction of pT5282-CTXM was unique to the industrial isolates within the *P. gergoviae* dataset but was predicted in one reference genome (*E. hormaechei* TUM10887) in the *Enterobacter* sp. dataset. All isolates and their associated accessions are shown in Table 5.4.

Table 5.3: Predicted plasmids within the draft genome assemblies of industrial *K. variicola*								
			Plasmid gene hits (+ % query coverage)					
Isolate	Plasmid hits	Accession	AF250878.1	CP000966.1	CP003027.1	CP003223.1	JN233704.1	JN420336.1
		Assembly size (bp)	180461	91096	166725	122799	208191	267242
		# isolates with hit	5	11	1	4	16	1
A55E	0		.	.	.	.	.	.
BD DM 395	3		.	100	.	.	100	.
DSM 15968^{T}	0		.	.	.	.	.	.
ECO126	0		.	.	.	.	.	.
EuSCAPE AT005	4		100	.	.	100	100	.
EuSCAPE BE101	1		.	.	.	.	.	.
EuSCAPE DE018	4		.	100	.	.	100	.
EuSCAPE EE016	2		.	.	.	.	100	.
EuSCAPE HR033	5		99.74	.	.	.	100	.
EuSCAPE IT159	2		.	.	.	.	100	.
EuSCAPE IT251	2		.	.	.	100	100	.
EuSCAPE IT344	2		.	.	.	.	.	.
EuSCAPE IT360	3		.	100	.	100	100	.
EuSCAPE IT389	0		.	.	.	.	.	.
EuSCAPE RO063	4		.	100	.	.	100	.
EuSCAPE RO090	0		.	.	.	.	.	.
EuSCAPE RS065	1		.	.	98.75	.	.	.
EuSCAPE RS096	4		.	100	.	.	100	.
EuSCAPE TR218	5		.	100	.	.	100	100
EuSCAPE UK124	0		.	.	.	.	.	.
K2	0		.	.	.	.	.	.
KCRI-152E	0		.	.	.	.	.	.
KP007 1	5		100	97.83	.	.	100	.
TUM14025	0		.	.	.	.	.	.
TUM14035	0		.	.	.	.	.	.
TUM14078	4		.	100	.	100	100	.
TUM14134	0		.	.	.	.	.	.
ID 2	6		100	95.65	.	.	100	.
ID 6	6		100	95.65	.	.	100	.
SB3295	2		.	96.09	.	.	100	.
UBA694 22	0		.	.	.	.	.	.

Footnote. Industrial isolate ECO126 is shown in dark blue, in the context of reference genomes. The total number of BLASTN plasmid hits are indicated, but only large (light blue) and megaplasmid (orange) are shown for simplicity. The size of the assembly with which the accession number is associated and number of isolates with a given plasmid hit (# isolates with hit) is also indicated. Numbers within the table indicate the percentage coverage for each hit.

Table 5.4: Predicted plasmids within the draft genome assemblies of industrial *P. gergoviae*

Isolate	Plasmid hits		Plasmid hits (+ % query coverage)											
		Accession	**JN233704.1***	**JN420336.1***	**JN233704.1***	CP000670.1	AP012055.1	CP000966.1	**JN420336.1***	**BX664015.1***	**BX664015.1***	DQ449578.1	MF062700.1	CP011611.1
		Assembly size (bp)	208191	267242	208191	137010	250444	91096	267242	274762	274762	98264	60206	243709
		# isolates with hit	11	3	1	1	3	4	5	2	2	4	4	7
ATCC 33028^T	2		100	.	.	.	.	.	.	.	.	.	.	.
ECO121	3		.	.	.	.	.	100	.	.	.	100	.	.
ECO122	4		100	.	.	.	.	.	100	.	.	.	100	100
ECO123	1		.	100	.	.	.	.	.	.	.	.	.	.
ECO124	4		100	.	.	.	.	.	100	.	.	.	100	100
ECO125	3		.	.	.	.	.	100	.	.	.	100	.	.
ECO128	4		100	.	.	.	.	.	100	.	.	.	100	100
ECO129	1		.	100	.	.	.	.	.	.	.	.	.	.
ECO77	4		100	.	.	.	.	.	100	.	.	.	100	100
40874	1		.	.	.	.	.	.	100	.	.	.	.	.
C3	3		100	100	.	100	.	.	.	.	.	.	.	.
C7B	2		.	.	.	.	.	.	.	.	.	.	.	100
DL84A27	1		.	.	.	.	.	96.09	.	.	.	.	.	.
DS82E24	0		.	.	.	.	.	.	.	.	.	.	.	.
FDAARGOS 386	1		100	.	.	.	.	.	.	.	.	.	.	.
JB83E35	0		.	.	.	.	.	.	.	.	.	.	.	.
JP84E9	1		.	.	.	.	100	.	.	.	.	.	.	.
JS81F13	1		100	.	.	.	.	.	.	.	.	.	.	.
MGH173	7		.	.	.	.	100	.	.	100	100	100	.	99.66
MGH183	6		.	.	.	.	100	.	.	100	100	100	.	99.66
NCTC11434	1		100	.	.	.	.	.	.	.	.	.	.	.
SF84F43	2		100	.	.	.	.	.	.	.	.	.	.	.
SN86A38	5		.	.	100	.	.	96.09	.	.	.	.	.	.
SS84A28	1		100	.	.	.	.	.	.	.	.	.	.	.

Footnote. Isolates of industrial origin are shown in dark blue, whilst isolates of industrial origin which are now used in biocide testing panels are shown in purple, within the context of genomes of reference isolates. The total number of BLASTN plasmid hits is indicated, but only large (light blue) and megaplasmid (orange) are shown for simplicity. The size of the assembly with which the accession number is associated and number of isolates with a given plasmid hit (# isolates with hit) is also indicated. Duplicate accessions are in bold, and asterisked. Numbers within the table indicate the percentage coverage for each hit.

5.1.10. Hybrid complete and new complete genome assembly confirms prediction of large and megaplasmids within ECO77 and ECO124

Assembly of ECO77 using Illumina and Nanopore reads resulted in an assembly consisting of two complete replicons of a 5,368,409 bp main chromosome and one large plasmid of 182,007 bp (herein referred to as ECO77 plasmid 1). The third incomplete replicon was a megaplasmid consisting of 605,666 bp, in 36 contigs (ECO77 plasmid 2). The N50 of this incomplete replicon was 172,670 bp. The presence of a megaplasmid and large plasmid was concordant with BLAST hits from the ECO77 draft assembly when compared to the PlasmidFinder database, three out of four of which were for megaplasmids. The megaplasmids for which the draft assembly produced hits were as follows: *K. pneumoniae* plasmid pNDM-MDR (267,242 bp, Accession: JN420336.1), *K. pneumoniae* strain ST258 plasmid pKPN-IT (208, 191 bp, Accession: JN233704.1), and *Citrobacter freundii* (*C. freundii*) strain CAV1321 plasmid pKPC_CAV1321-244 (243,709 bp, Accession: CP011611.1).

ECO124 also consisted of two complete replicons, consisting of a 5,368,697 bp main chromosome, and one large plasmid of 183,872 bp (ECO124 plasmid 3). The incomplete third and fourth replicons were two megaplasmids of 329,569 bp in 15 contigs (ECO124 plasmid 1) and 264,578bp in 18 contigs (ECO124 plasmid 2), with N50 values of 305,268bp and 248,071bp respectively. This was again concordant with BLAST hits from the ECO124 assembly, which were identical to those produced from the ECO77 draft assembly.

Extraction of the *repA* gene from the annotations of each of these plasmids revealed the presence of distinct families of plasmids, the properties of which are discussed below.

5.1.10.1. PLASMID FAMILY 1

On initial examination of database entries, this plasmid family is widely distributed across the *Enterobacteriaceae*, with representative plasmids from *E. coli*, *Enterobacter*, *Klebsiella* spp., *Metakosakonia* spp., *Pantoea*, *Raoultella* and *Serratia*. Phylogenetic analysis of the core-gene alignment produced from plasmid family 1 (Figure 5.9) confirmed that ECO77 plasmid 1 (megaplasmid) and ECO124 plasmid 3 (large plasmid) are identical in terms of core-gene content. Of the predicted core-genes, 4/6 possessed functional annotations;

repA, encoding the RepFIB Replication initiation protein A, *sopB*, which encodes a protein involved in plasma membrane alterations, *umuC*, a DNA polymerase involved in UV protection and mutation, and *xerC,* which plays a role in the stability of plasmids during cell-division. When examined in the context of other closely related plasmids (Figure 5.7), the majority of plasmids from this family, including those of ECO77 and ECO124 possessed no AMR genes, despite the whole genome assemblies from all organisms with this plasmid possessing predicted AMR genes. Exceptions to this were the *Serratia marcescens* derived pCAV1761-205, with 1 predicted AMR gene, *C. freundii* pOZ172 and *Klebsiella michiganensis* pKOCBH-B which both possessed 3 predicted AMR genes, and *K. pneumoniae* pKPN-431cz which possesses 11 AMR genes. In terms of absolute numbers, a significant proportion of biocide genes predicted from the whole genome assembly present on this plasmid – 29/212 (13.7%) for ECO77, 29/210 (13.8%) for ECO124. Furthermore, the same biocide resistance genes were predicted genes in both ECO77 plasmid 1 and ECO124 plasmid 3, predominantly conferring metal resistance, including the complete presence of the sil cation-efflux system *silABC* (silver resistance) and copper resistance determinant system (*pcoABCDRSE*). 22/29 of the predicted biocide genes were reported in the Bacmet2 database as being observed on plasmids previously. All genes had a percentage coverage ≥81.99%. A visual summary of the localisation of AMR and biocide genes in the genomes and isolates harbouring plasmids from plasmid family 1 are shown in Figure 5.6 and Figure 5.7.

5.1.10.2. PLASMID FAMILY 2

This family of plasmids was unique *K. pneumoniae* and related species of *Klebsiella*. In a similar vein as their respective plasmids from plasmid family 1, ECO77 plasmid 2 and ECO124 plasmid 2 of plasmid family 2 (both megaplasmids) are identical in terms of core-gene content (Figure 5.10). Of these core-genes, 6/12 possessed a functional annotation: *repA* encoding RepFIB replication protein A, HhA (involved in persister cell formation), Plasmid transfer protein, Tyrosine-type recombinase/integrase, *yadA*, encoding an outer membrane protein promoting cell adhesion), and a Hemolysin expression-modulating protein. AMR genes were predicted in the

majority of plasmids (16/23) from this family (the fewest predicted AMR genes above zero was 3, and the highest was 22). ECO77 plasmid 2, ECO124 plasmid 2, and 5 reference plasmids from this family possessed no predicted AMR genes. Biocide and metal resistance genes were predicted in both ECO77 plasmid 2 and ECO124 plasmid 2; *adeL* (Organo-sulfate, Phenanthridine and Acridine resistance) *dpsA* (oxidative stress tolerance) were predicted in both plasmids, with 69.14% and 93.83% coverage respectively, and were reported as chromosomally derived in previous database entries. An additional 7 genes, all of which confer metal resistance, and 6/7 of which have been reported on plasmids previously were predicted in ECO124 plasmid 2; *arsH* (arsenic resistance) *pbrA*, *pbrR* (lead resistance), *arsB*, *arsC* and *arsR* (arsenic and antimony resistance). A visual summary of the localisation of AMR and biocide genes in the genomes and isolates harbouring plasmids from plasmid family 2 are shown in Figure 5.6 and Figure 5.7.

Plasmid family 3

A large plasmid belonging to this family was detected in ECO124 (plasmid 1). This family of plasmids was found to be predominantly derived from *C. freundii* and related genera, namely *Enterobacter*, *Klebsiella* and *Raoutella*. Core-gene analysis revealed that the most closely related plasmids to ECO124 plasmid 1 were derived from *Raoutella electrica* (*R. electrica*) and *C. freundii complex* spp. Within plasmid family 3, 12/43 core-genes possessed a functional annotation. Furthermore, within this dataset, only 8/21 in plasmids possessed predicted AMR genes; no AMR genes were predicted within ECO124 plasmid 1. Where AMR genes were predicted, the lowest number of predicted genes was 4, and the most was 13. The predicted biocide and metal resistance genes on the plasmid in ECO124 are predominantly involved in metal resistance, including the complete *ncrABCY* complex (nickel resistance), *arsA*, *arsB*, *arsC* and *arsR* (arsenic resistance) and *merA*, *merD*, *merE*, *merP*, *merT* and *merR* conferring mercuric resistance. Of these genes, database entries confirmed that 13/21 have been seen on plasmids previously. A visual summary of the localisation of AMR and biocide genes in the genomes and isolates harbouring plasmids from plasmid family 3 are shown in Figure 5.6 and Figure 5.7.

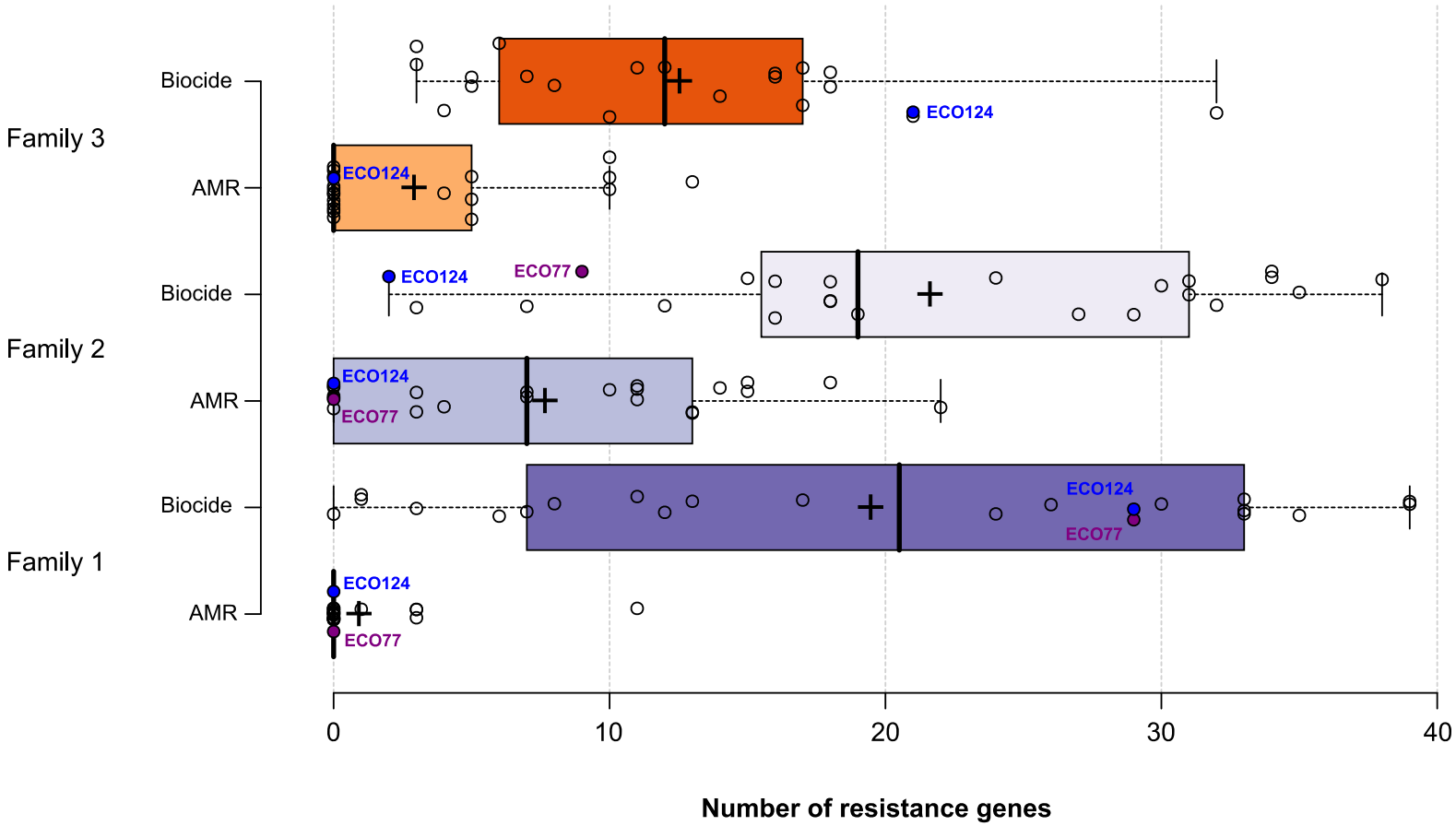


Figure 5.6: Predicted AMR and biocide resistance genes in three *Enterobacteriaceae* large and megaplasmid families. Number of antimicrobial resistance (AMR) and biocide resistance genes present in two families of megaplasmids and one family of large plasmids found within industrial strains of *P. gergoviae*, and reference plasmids chosen on the basis of *repA* sequence similarity. Isolate derived from industrial product (ECO124) is highlighted in blue, whilst the isolate of industrial origin which is now used in biocide testing panels (ECO77) is shown purple. The mean is indicated on each boxplot by a black cross, whilst the median is shown using a black line. Whiskers were defined using the Tukey HSD test. For all datasets, $n = 22$.

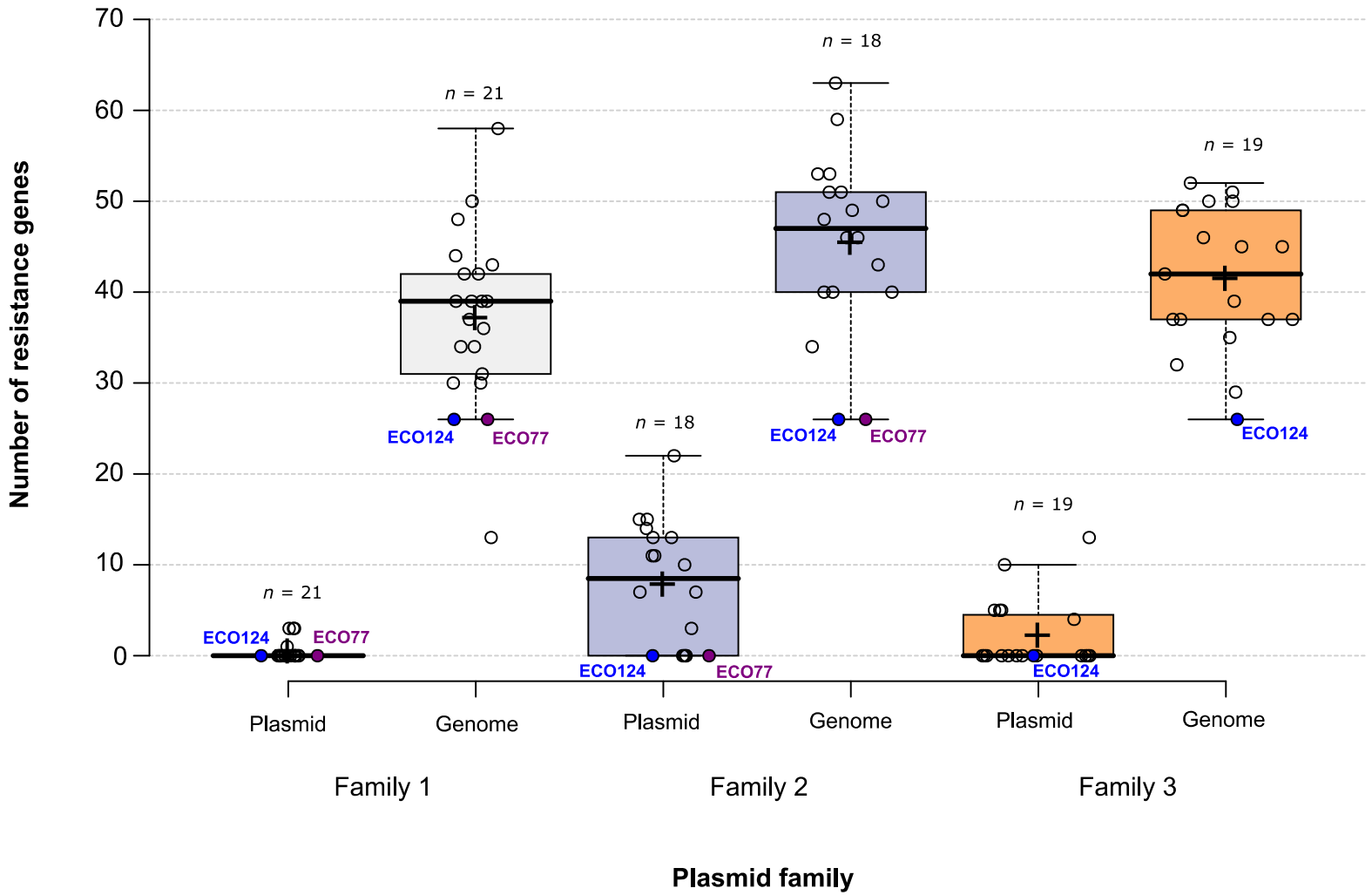


Figure 5.7: Localisation of predicted AMR genes to the genomes of ECO77 and ECO124. The number of predicted genes in genomes and plasmids of ECO77 and ECO124 are shown alongside reference plasmids, and the whole genome assembly from which they are derived. Isolate derived from industrial product (ECO124) is highlighted in blue, whilst the isolate of industrial origin which is now used in biocide testing panels (ECO77) is shown purple. The mean is indicated on each boxplot by a black cross, whilst the median is shown using a black line. Whiskers were

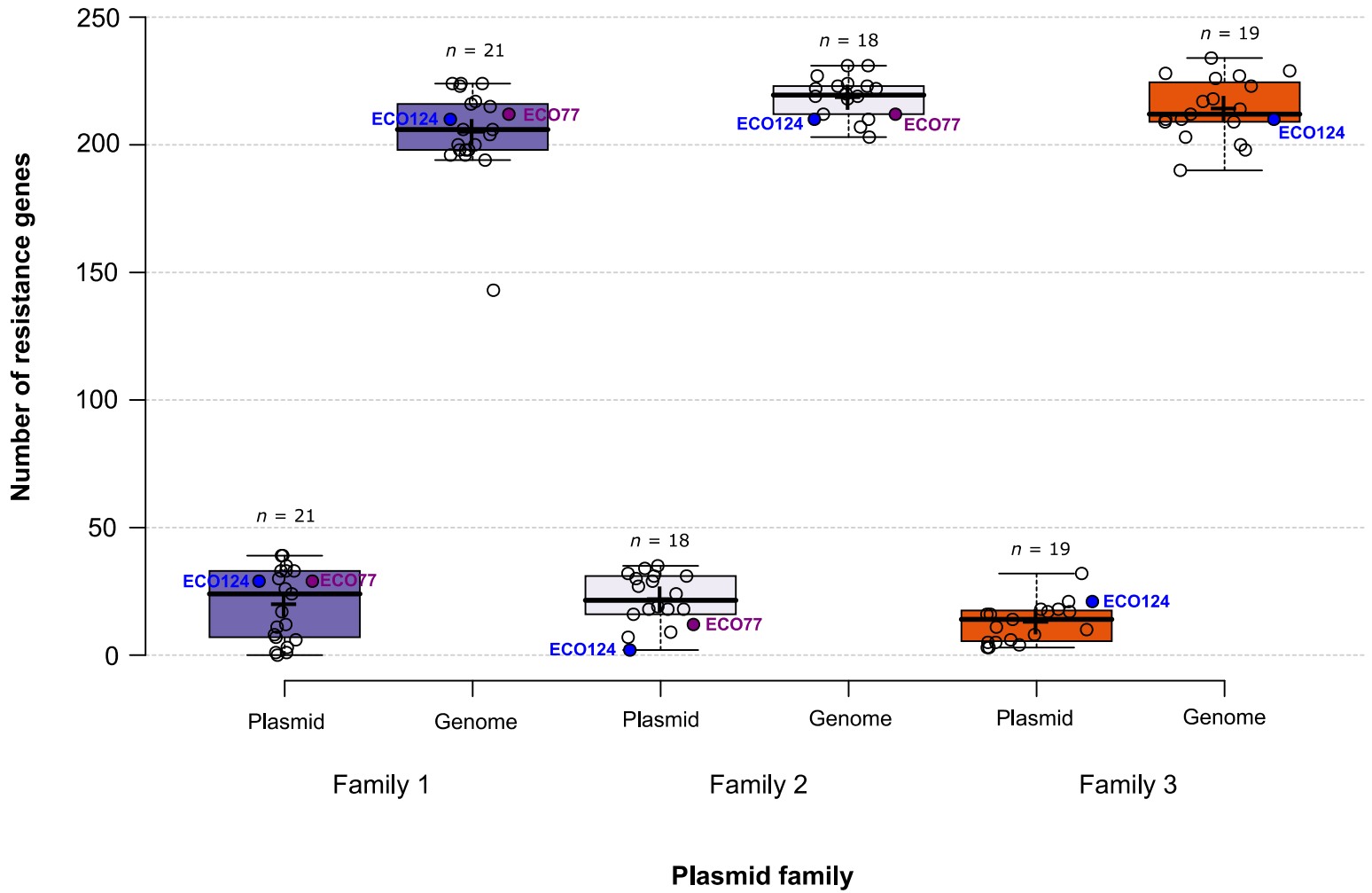


Figure 5.8: Wide distribution of predicted biocide genes across the genomes and plasmids of ECO77 and ECO124. The number of predicted genes in genomes and plasmids of ECO77 and ECO124 are shown alongside reference plasmids, and the whole genome assembly from which they are derived. Isolate derived from industrial product (ECO124) is highlighted in blue, whilst the isolate of industrial origin which is now used in biocide testing panels (ECO77) is shown in purple (ECO77). The mean is indicated on each boxplot by a black cross, whilst the median is shown using a black line. Whiskers were defined using the Tukey HSD test.

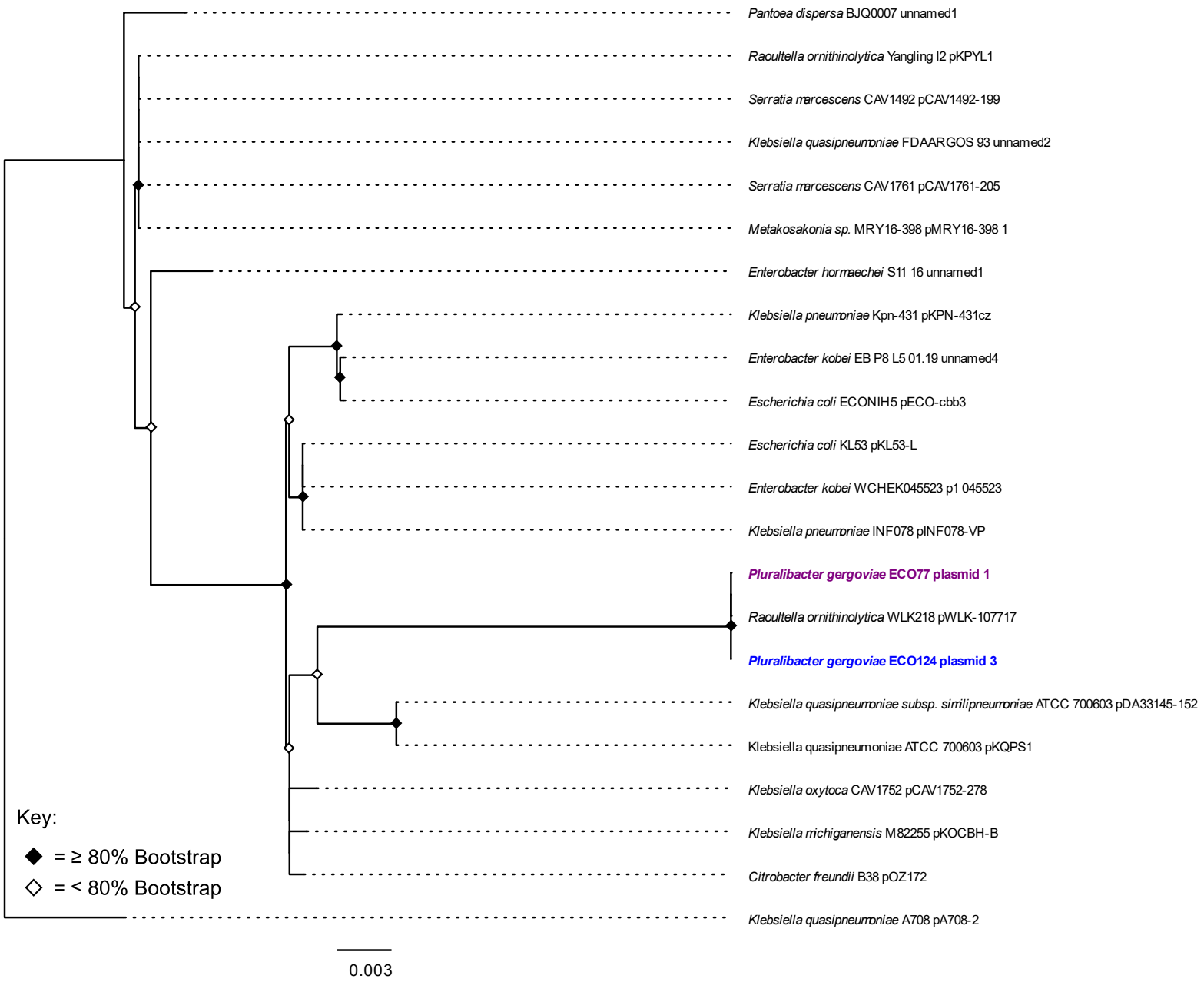


Figure 5.9: Evolutionary relatedness of *P. gergoviae* plasmid family 1. Maximum-likelihood tree of 6 core-genes identified in *P. gergoviae* plasmid family 1, consisting of ECO77 plasmid 1, ECO124 plasmid 3, and 20 reference plasmids chosen on the basis of *repA* sequence similarity. Isolate derived from industrial product (ECO124) is highlighted in blue, whilst the isolate of industrial origin which is now used in biocide testing panels (ECO77) is shown purple. The scale bar represents the number of substitutions per base position.

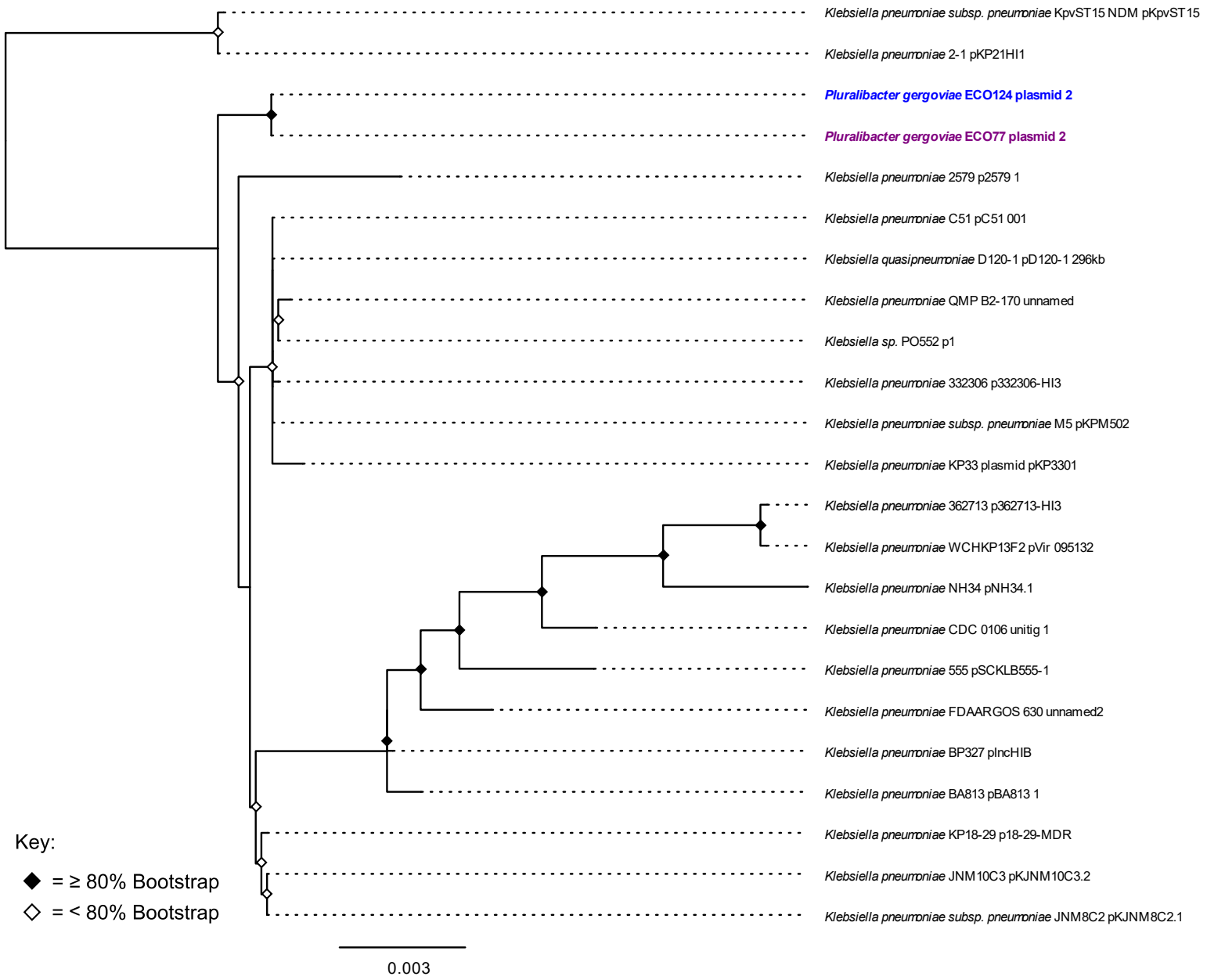


Figure 5.10: *P. gergoviae* plasmid family 2 is unique to *Klebsiella* species. Maximum-likelihood tree of 12 core-genes identified in *P. gergoviae* plasmid family 2, consisting of ECO77 plasmid 2, ECO124 plasmid 2, and 20 reference plasmids chosen on the basis of *repA* sequence similarity. Isolate derived from industrial product (ECO124) is highlighted in blue, whilst the isolate of industrial origin which is now used in biocide testing panels (ECO77) is shown purple.

5.4. Discussion and conclusions

Overall, the results from this chapter show that sequencing of 16S rRNA alone is insufficient for accurate identification of multiple industrial *Enterobacteriaceae*. This is consistent with previous studies which show that *Enterobacteriaceae* share up to 99% sequence similarity across the full length of the 16S ribosomal RNA gene (Jovel *et al.* 2016), thus rendering the family susceptible to misidentification when sequencing this gene.

The results also show that the *Enterobacteriaceae* possess similar absolute numbers of predicted AMR genes within their genomes, with the genus *Pluralibacter* possessing the fewest mean predicted AMR genes. The *Enterobacteriaceae* family also appears to harbour large numbers of predicted biocide genes, with the absolute numbers of predicted genes that are relatively consistent between the *Enterobacter, Klebsiella*, and *Pluralibacter* genera, but are slightly lower for the *Leclercia* genus. Together, these results suggest that all of the aforementioned genera are likely to possess intrinsic AMR and Biocide resistance.

The analyses of the near-complete assemblies in this chapter show that ECO77 and ECO124 harbour three families of plasmids. Plasmid family 1 appears to be variable in size, appearing as a large plasmid in ECO77 (182,007 bp), but as a megaplasmid in ECO124 (329,569 bp). Whilst the most closely related plasmid to ECO77 plasmid 1 and ECO123 plasmid 3, pWLK107757, is derived from a BlaNDM-1 and BlaKPC-2 positive (AMR) strain of *Raoutella orthinolytica* (based upon the NCBI entry for this plasmid, accession: , CP038276.1). In this analysis, AMR genes were not detected on plasmid, but were detected whilst screening entire genome for AMR. The next nearest-neighbouring plasmids, pDA33145-152 and pKQPS1 which possessed identical core-gene content, were both derived from the same strain of *Klebsiella quasipneumoniae* subsp. *similipneumoniae*, ATCC 700603, an antimicrobial-resistant clinical isolate from the urine of a hospitalised patient. Studies in which pDA33145-152 was characterised found that antibiotic heteroresistance in subpopulations of clinical isolates was caused by gene amplification around the genome in discrete clusters of antimicrobial resistance genes (Nicoloff *et al.* 2019). This plasmid, however, was not identified as a determinant of antibiotic resistance. The study in which pKQPS1 was identified similarly concluded that AMR genes were present on the main chromosome, and another plasmid, but not pKQPS1 (Elliott *et al.* 2016). Similarly, in analysis of

industrial plasmids, AMR genes were not detected on either of these plasmids. Overall, the absence of predicted AMR genes, and presence of predicted metal resistance genes in the vast majority of plasmids from this family (29 in both ECO77 and ECO124), suggests that plasmid family 1 primarily play a role in metal and/or stress tolerance in ECO77 and ECO124, and other strains with plasmids from this family.

Plasmid family 2 is present within ECO77 (605,666 bp), and ECO124 (264,578bp) and exist as megaplasmids in both isolates. This plasmid family has a number of predicted AMR, biocide resistance and metal resistance genes across the family. Furthermore, the presence of AMR in these plasmids is reported widely across the literature. One of the nearest neighbouring plasmids, pKP21HI1 (see Figure 5.10), is derived from *K. pneumoniae*, from the stool of a cancer patient from Shandong province, China (Li *et al.* 2019). This plasmid was found to be conjugation defective, of the HI1B family, possess 18 AMR genes, and to confer AMR to a previously-susceptible *E. coli* DH5α (Li *et al.* 2019). ECO77 plasmid 2 and ECO124 plasmid 2 however, were relatively unique members of this family in that they possessed no predicted AMR genes.

Family 2 plasmids within ECO77 and ECO124 possessed predicted genes conferring resistance to oxidative stress and metals. The presence of such genes has been observed in related plasmids previously; pKpvST15, a near neighbour (see Figure 5.10) from a *K. pneumoniae* isolated from the throat of a patient in London was found to carry various metal resistance genes including copper (*pcoABCDERS*), lead (*pbrABCR*), and silver (*silCERS*) (Turton *et al.* 2019). The presence of these genes on the plasmid is consistent with this analysis, with both *pbrR* and *pbrA* predicted in ECO124 plasmid 2. Overall, this suggests that plasmid family 2 plays a role in conferring resistance to oxidative stress and metals in ECO77 and ECO124 but may not contribute to the antibiotic resistance of either isolate, despite other related plasmids possessing AMR genes.

Plasmid family 3 was unique to ECO124, existing in the form of a complete 183,872 bp plasmid. In a somewhat similar vain to plasmid family 2, certain members of family 3 within the reference dataset possessed predicted AMR genes. This is consistent with previous studies, in which these plasmids have been isolated from strains reported as multi-drug resistant, with multiple AMR genes in report in the genome, but not the plasmid themselves. For example, the nearest neighbour,

R. electrica strain DSM 102253 plasmid unnamed 1, isolated from biofilms of a glucose fed microbial fuel cell inoculated with sewage waste, contains genes for penicillin binding proteins, chloramphenicol acyltransferase and spectinomycin tetracyclin efflux pump (Thiel *et al.* 2019). This strain also possessed multiple mercuric and tellurite resistance proteins encoded by the genome and plasmids (Thiel *et al.* 2019). In this analysis, ECO124 similarly did not possess any predicted AMR genes, but did possess 21 biocide resistance genes. Additionally, the analysis in this chapter revealed that all of the AMR genes described in nearest neighbour *R. electrica* strain DSM 102253 were detectable in the whole genome assembly, but not on the plasmid of this isolate. This suggests that in any AMR strains with plasmids from this family, AMR is primarily intrinsic, and likely to exist on a main chromosome, as opposed to mobilisable elements such as plasmids.

In summary, large and megaplasmids appear to be present throughout the family *Enterobacteriaceae* and are not unique to industrial isolates. Though megaplasmids, including those analysed in this chapter, have been reported as conferring AMR previously, the role of plasmid families in the industrial isolates described in this chapter appears to be geared towards metal and oxidative stress tolerance on the genetic level. Furthermore, a significant majority of these metal and oxidative stress genes in industrial isolates are localised to the plasmids, with 17.9% (38/212) resistance genes on the plasmids of ECO77, and 24.8% (52/210) present on the plasmids of ECO124. The following genes were present in all three plasmid families: *adeL* (organo-sulfate, phenanthridine, azin and acridine resistance) *arsA*, *arsB*, *arsC*, *arsD*, *arsH*, *arsP* (arsenic resistance) *copG* (copper and zinc resistance), *merT* (mercury resistance), *ncrY* (nickel resistance), *pbrR* (lead resistance), and *silS* (silver resistance).

Future studies should examine the phenotypic role of these plasmids in the overall fitness of these organisms and determine the prevalence of these plasmids in other industrial isolates.

The objectives of this chapter were addressed as follows:

1. **To utilise whole genome sequencing to determine the identity and establish the taxonomic diversity of a panel of industrial contaminant organisms, initially identified as *Enterobacter* sp. by 16S rRNA sequencing.** Industrial

contaminant organisms spanned four genera (*Enterobacter*, *Klebsiella*, *Leclercia* and *Pluralibacter*. Average Nucleotide Identity analysis confirmed the taxonomic placement of 8 isolates initially identified by 16S rRNA gene sequencing as *P. gergoviae* and resulted in taxonomic reclassification of 5 other isolates.

2. **To determine the presence of large plasmids and megaplasmids amongst these organisms using short read sequencing data in this panel, in the context of a curated genome panel of *Enterobacteriaeceae* from a variety of backgrounds.** The prediction of megaplasmids and large plasmids was ubiquitous throughout the analysed datasets of *Enterobacter* sp., *K. variicola* and *P. gergoviae*. In contrast, very few plasmids of any variety were predicted within *L. adecarboxylata*. Within the *P. gergoviae* dataset, the prediction of *E. cloacae* strain T5282 plasmid pT5282-CTXM was unique to industrial isolates but was predicted in one other genome in the *Enterobacter* sp. dataset.

3. **To determine the validity of plasmid prediction tools for detecting large and megaplasmids in short read assemblies, by completing the genome of two *P. gergoviae* isolates, ECO77 and ECO124, using long-read sequencing technologies.** The prediction of large and megaplasmids within the short-read draft assemblies correlated with the presence of multiple large and megaplasmid families within the near-complete assemblies of ECO77 and ECO124. BLAST hits from short read assemblies correctly predicted the presence of megaplasmids within the draft assemblies of ECO77 and ECO124. The predictions from the draft assembly, however, did not predict members from all plasmid families, and those that were detected were not necessarily the nearest neighbours to plasmids revealed by hybrid assembly.

4. **To determine the relationship between the presence of large and megaplasmids in industrial isolates, and the presence of predicted AMR, biocide and metal resistance genes.** Industrial isolates ECO77 and ECO124 possess multiple large and megaplasmids, belonging to three different families, two of which they share, and one of which is unique to ECO124. The other members of the families against which the plasmids of ECO77 and ECO124 were compared possess a myriad of predicted AMR, biocide and metal resistance genes.

These isolates however possess no predicted AMR genes on their large and megaplasmids but do possess large numbers of genes conferring metal resistance.

The overall hypothesis for this chapter "*Enterobacteriaceae* isolated from industrial products will display wide taxonomic diversity, and harbour large and/or megaplasmids" was accepted.

6. Discussion and Future Perspectives

6.1. Genomics of industrial bacteria

Overall, this work has identified the predominant Gram-negative bacteria responsible for industrial recalls, (see 1.4.1), with *Pseudomonas aeruginosa* (*P. aeruginosa*) being the most prevalent (17% of recalls), followed by *Pseudomonas* spp. (6%), the *Enterobacteriaceae*, (other *Enterobacteriaceae* (5%), *Enterobacter* (3%)), and *Burkholderia* (3%). Bacteria from all the aforementioned taxonomic groups, excluding *Burkholderia*, were isolated from low-risk areas in the factory environment (see 2.5.2, Table 2.4). Furthermore, industrial strains of all these bacteria possessed multireplicon genomes with large replicons (>250 kbp), which are ubiquitous in *Burkholderia* (Agnoli *et al.* 2014), but arguably less common in the *Enterobacteriaceae* (Cunningham-Oakes *et al.* 2020), and *Pseudomonas* (Weiser *et al.* 2019).

The multireplicon nature of isolates from industrial products appears to be attributable to the acquisition of large plasmids and megaplasmids (see Chapter 4 and Chapter 5). Whilst the precise role of these large genomic elements in terms of their contribution to a preservative resistant phenotype remains to be seen, a myriad of evidence exists within the work presented and the literature to shed some light upon their potential phenotypic roles. Firstly, megaplasmids and similar genomic elements such as the third replicon in *Burkholderia* are widely associated with stress tolerance (Agnoli *et al.* 2014). In *Pseudomonas*, stress tolerance frequently takes the form of antibiotic or antimicrobial resistance (Yuan et al. 2017) (Botelho *et al.* 2017), mediating biodegradation processes in wastewater treatment (Zheng *et al.* 2016) and solvent tolerance (Kuepper *et al.* 2015). Additionally, the work presented here suggests that megaplasmids contribute to antibiotic resistance in Pseudomonas strains from historical contamination incidents via RND efflux mechanisms, which is partially inhibited using RND efflux pump inhibitors (see 4.2.2).

RND efflux pumps are also upregulated in industrial strains of *Burkholderia lata* (*B. lata*) with a co-occurrence of elevated fluoroquinolone resistance, both of which are reversible with the use of efflux pump inhibitors (Rushton et al. 2013). The genetic basis for this upregulated of efflux in *B. lata* was the identification of missense mutations in key regulators upstream of the efflux pump (see 3.3.6). In *Enterobacteriaceae*, megaplasmids have been shown to contribute to enhanced biofilm

formation, pathogenicity (Aviv *et al.* 2014), and the global dissemination of carbapenem resistance (Desmet *et al.* 2018; Lü *et al.* 2019). The genomic analyses in this study also suggest that megaplasmids with similar evolutionary origins to those from industrial isolates possess are able to acquire AMR genes, but megaplasmids from industrial isolates do not possess AMR genes. Industrial megaplasmids instead possess a multitude of predicted biocide and metal resistance genes (see 4.2.4 and 5.3.3). This, alongside evidence that industrial isolates possess higher levels of preservative resistance (Weiser *et al.* 2019) suggest that the predominant role of megaplasmids in industrial isolates is survival in harsh abiotic industrial environments. Further phenotypic characterisation will be required to determine the role of regulator mutations and megaplasmids in the preservative resistant phenotypes of industrial isolates. A similar methodology can be used to investigate both roles, as single-base editing can be used to either correct missense mutations in regulators, or introduce stop codons into megaplasmid replication origin genes, resulting in megaplasmid removal (Chen *et al.* 2018).

Whilst the role of megaplasmids has not been fully elucidated, the requirement of genomics for the accurate identification of industrial contaminants is now abundantly clear. Whilst PCR based methodologies such as 16S rRNA gene sequence analysis are sufficient for *P. aeruginosa* and most other *Pseudomonas* spp. (see 2.5.2, Table 2.4), this is insufficient for both *Enterobacteriaceae* and closely related members of the *Burkholderia cepacia* complex (Bcc) spp. For *Burkholderia*, this is evidenced by misidentification of a strain deposited as a strain of the bioterrorism species *Burkholderia pseudomallei*, (see 3.4) and the identification of a novel taxon K grouping (see 3.3.4). This could have implications for the preservative tolerance potential of an organism, given the unique grouping of these organisms amongst related organisms from preserved industrial products (Leong *et al.* 2018). For *Enterobacteriaceae*, this lack of resolution in identification is evidenced by the work presented herein, where sequencing a partial 16S rRNA gene was sufficient to accurately identify *Pluralibacter gergoviae* but provided incorrect identification for other *Enterobacteriaceae* from contamination incidents (see 5.3.2). These findings are also supported in the literature, where difficulty in the identification of *Enterobacteriaceae* is well known (Jesumirhewe *et al.* 2016).

The research presented in this thesis highlights the importance of genomics for the identification of organisms commonly identified as industrial contaminants, which may have implications for the prediction of their preservative tolerance. The presence of megaplasmids in numerous industrial isolates may also play a role in stress and preservative tolerance, but their exact role in fitness remains unconfirmed. Future work should aim to further substantiate these findings using strategies to link the megaplasmid genotypes to tolerance and fitness phenotypes of industrial strains.

Key avenues to follow up from this PhD thesis would be:

1. **Determining the phenotypic antimicrobial and preservative tolerance of isolates from home and personal care product factories which are found elsewhere as industrial contaminants.** This could be achieved simply by using broth dilution MIC assays or agar growth-based minimum bactericidal screens for preservatives of interest (Halla *et al.* 2018) (Weiser *et al.* 2019). Numerous preservation strategies exist, with varying degrees of implementation, owing to both legislation and public perception (Halla *et al.* 2018). Phenoxyethanol and Sodium Benzoate however remain two of the most used preservatives and would provide a good baseline for the biocide characterisation of these isolates. This is especially true for isolates of *P. aeruginosa*, which have been shown to possess higher phenotypic biocide tolerance when derived from industrial products (Weiser *et al.* 2019). Additionally, the findings of Chapter 4 and Weiser *et al.* 2019 highlight that assessing chloramphenicol and trimethoprim tolerance would be warranted in factory *P. aeruginosa* isolates. These assays would serve to elucidate whether *P. aeruginosa* from the factory environment differ phenotypically to industrial contaminants of the same species group.

2. **Determining the role of mutations in efflux pump regulators in preservative resistant *Burkholderia* via knockout of the transcriptional regulators in which point mutations were identified.** Several effective knockout systems exist for this genus, but for this experiment I would propose the approach employed by Flannagan *et al.* 2008. In brief, deletion of the *soxR* region or *marR* family regulators described in Chapter 3 (see 3.3.6) would be induced by introducing sequences flanking the regulators into a suicide plasmid that contains I-SceI, a yeast homing endonuclease. This non-replicating plasmid would then be

introduced into 383-BIT, 383-CMIT, BCC1294, BCC1296 or BCC1297 (see 3.3.6), where it would undergo homologous recombination with the targeted regulator region and introduce a mutant regulator gene allele. A second plasmid would then be introduced, carrying the I-SceI nuclease, causing a double-stranded break at the site of integration. This would stimulate recombination between the mutant and parental alleles, and result in either restoration of the parental allele (non-successful mutation), or regulator deletion (successful mutation). Successful mutants could then be screened for and selected by PCR.

3. **Mapping the exact phenotypic roles of megaplasmids in industrial contaminants, by creating plasmid minus mutants and studying their phenotype.** Megaplasmid curing has been successfully achieved in Bcc previously, where the third replicon of this taxon has been described as a megaplasmid that confers a variety of functions in the Bcc, including secondary metabolism and virulence (Agnoli *et al.* 2012). Curing of this megaplasmid has been achieved using the pMiniC3 vector, a 12.6 kb plasmid with an identical origin of replication to *B. cenocepacia* (*B. cenocepacia*) H111 (Agnoli *et al.* 2012). This vector is introduced via tri-parental mating involving a Bcc strain of interest, donor *Escherichia coli* (*E. coli*) MC1061 and *E. coli* HB101 carrying the helper plasmid pRK2013 (Agnoli *et al.* 2012). The protocol has been successful in curing the megaplasmid from *B. cenocepacia* H111 (Agnoli *et al.* 2012) and *Burkholderia ambifaria* BCC0191 (Mullins *et al.* 2019). To follow up from this thesis (Chapter 4) a similar system could be devised for *P. aeruginosa*, using a vector containing identical origins to those of the industrial *P. aeruginosa* megaplasmid family (see 4.2.3). A similar principle could also be applied to industrial *Enterobacteriaceae* with multireplicon genomes (see 5.1.10).

4. **Determining the extent and prevalence of megaplasmids in industrial *Enterobacteriaceae*.** This could be achieved using long-read technologies such as PacBio (see 4.1.7) or Oxford Nanopore (see 5.1.3) to generate complete genomes for all factory-derived *Enterobacteriaceae* described in Chapter 2, and industrial *Enterobacteriaceae* described in Chapter 5. Completing the genomes of *Enterobacteriaceae* isolates would serve to determine whether the megaplasmids in industrial isolates is ubiquitous or sporadic. This would also serve to validate

or disprove the *in-silico* plasmid prediction conducted using draft genomes from short reads in Chapter 5 (see 5.1.5 and 5.1.9).

6.2. IDENTIFICATION OF INDUSTRIAL BACTERIA

As first highlighted in the aims and objectives (see end of 1.4.3), the studies conducted within this thesis predominantly aimed to understand the genomics of industrial bacteria and identify industrial bacteria accurately using genomic techniques. Therefore, this does not tackle the pertinent issue of identification using other technologies, which may be more widely available and practical to use for routine identification. As such, this section has been synthesised to both broach the topic of non-genomic identification methods and provide identification recommendations to industry. Additionally, Table 6.1 provides several recommendations for the identification of key contaminant organisms.

Methods to accurately identify microorganisms have improved considerably (van Belkum *et al.* 2013) and multiple commercial companies offer both MALDI TOF MS and gene/genome sequence-based analysis for relatively limited costs. Multiple methods to accurately identify the most common causes of microbial contamination in non-food products are available (see Table 6.1). The costs of microbial identification in relation to recall incidents is also small in relation to the overall costs of manufacturers having to deal with removing products, cleaning manufacturing facilities, or develop new formulations to prevent non-food product contamination. A primary recommendation would be that non-food product recall reports require microbial identification to be carried out to a minimum of the genus level for both bacteria and fungi. A secondary recommendation would be that for bacterial species which pose a significant health or AMR risk as opportunistic pathogens (e.g. *P. aeruginosa*, the *Enterobacteriaceae*, Bcc spp., and *Staphylococcus* spp.), that identification to the species level should be reported. In addition, identification should be checked against the current taxonomy of each species, for example by comparison to the List of Prokaryotic names with Standing in Nomenclature (www.bacterio.net) (Parte, 2018), to enable fine grain surveillance and accurate comparative analysis of microbial contamination incidents.

If manufacturers and regulatory bodies do apply whole genome sequencing to characterisation of contaminant and outbreak strains, then they have all the information required for accurate species classification (Bull *et al.* 2012). As such, the

recommendations should not be seen as alternatives to whole genome sequencing but rather, should serve to supplement them. Analysis stemming from whole genome sequencing can also future proof our understanding by archiving this important microbial data within the nucleotide sequence databases to enable retrospective and comparative analysis of contamination incident data. In addition, since the non-food product and cosmetics industry landscapes are changing, methods for surveillance and recall recording need to be updated to ensure they remain relevant. The rise of new market trends such as natural non-food products may result in a shift in the diversity of microorganisms encountered as contaminants. Improved vigilance of non-food microbial contaminations is key to ensure methodologies remain relevant to the products they govern. For example, the non-food global ISO 11930 and European Pharmacopeia testing methods do not currently include the prevalent contaminating bacteria *Enterobacter*, *Burkholderia* or *Klebsiella* (Chapter 1, Figure 1.1), and only recommend using *E. coli*, *P. aeruginosa*, *Staphylococcus aureus* (*S. aureus*), *Candida albicans* and *Aspergillus brasiliensis* (Siegert 2013). A risk-based assessment advocates the inclusion of *Enterobacter*, *Burkholderia* or *Klebsiella* strains to reflect those compromising hygienic integrity of non-foods products (Chapter 1, Figure 1.1).

Table 6.1 – Recommendations for the identification of key contaminants in non-food product recalls		
Microorganism	**Phenotypic analysis: culture and identification**	**Genotypic identification**
P. aeruginosa and other *Pseudomonas* spp.	• Culture on nutrient and selective media for *P. aeruginosa* (Weiser *et al.* 2014) • Biochemical profiling generally accurate for *P. aeruginosa* • MALDI TOF MS accurate for *P. aeruginosa*	• 16S rRNA gene analysis adequate but *oprL* gene sequencing provide additional accuracy • *P. aeruginosa* MLST scheme and database available
Fungus	• Culture on a range of fungal growth media, such as Sabaroud Dextrose or Brain Heart Infusion agar with antibiotics (Hong *et al.* 2017) • MALDI TOF MS emerging as useful method for yeast and mould identification (Cassagne *et al.* 2016)	• Nuclear ribosomal internal transcribed spacer (ITS) and combined molecular methods (Raja *et al.* 2017)
Enterobacteriaceae: • *Enterobacter* spp. • *Klebsiella* spp.	• Culture on MacConkey agar – *Enterobacter* and *Klebsiella* can then be differentiated by features such as pigment, lactose fermentation (*Klebsiella* and *Enterobacter*) and Urease production (*Enterobacter*) • Chromogenic agar and MALDI TOF MS also perform well (Perry 2017)	• 16S rRNA gene sequence insufficient to discriminate all genera • Curated MLST database available for *Enterobacter cloacae, K. aerogenes* and *K. oxytoca* • EnteroBase database provides additional reference sequence comparison (Alikhan *et al.* 2018)
Burkholderia spp.	• Culture on nutrient and selective media for Bcc spp. (Henry *et al.* 1999) • Biochemical profiling and MALDI TOF MS insufficient for accurate identification of certain Bcc spp.	• 16S rRNA gene sequence insufficient for accurate species identification (Mahenthiralingam *et al.* 2000) • *recA* gene sequence comparison enables a provisional species identification (Mahenthiralingam *et al.* 2000) • Curated Bcc MLST database (Baldwin *et al.* 2005)
S. aureus and other *Staphylococcus* spp.	• Culture on Mannitol Salt Agar (*Staphylococcus* spp.) or CHROMagar (*S. aureus*) (Perry 2017) • Biochemical profiling and MALDI TOF MS accurate	• 16S rRNA gene and *rpoB* gene sequence provides accurate species identification (Mellmann *et al.* 2006) • Curated MLST databases for *S. aureus, S. epidermidis, S. hominis and S. haemolyticus*

6.3. APPLICATION OF GENOMICS IN COLLABORATIVE STUDIES

The techniques and skills developed in this PhD, particularly with regard to genomic and bioinformatic analyses, have a wide range of applications. This is evidenced by publications produced as both a part of the PhD project, and as collaborative, external ventures where the skillsets gained were used to generate further publications as detailed in Table 6.2.

Table 6.2 - Application of genomics in collaborative projects during PhD

Publication	PhD or external project?	Skills used
Beaton *et al*. 2018, FEMS Microbiol Lett.	External	• Database curation • ANI analysis • Core-gene analysis
Weiser *et al*. 2019. Microb Genomics	PhD	• Evolutionary and phylogenetic analysis • Comparative genomics
Webster *et al*. 2019a. and Webster *et al*. 2019b. Microbiol Resour Announc.	External	• Webster *et al*. 2019a - Core-gene analysis • Webster *et al*. 2019b – Library preparation, genome assembly and quality assessment
Cunningham-Oakes *et al*. 2020. FEMS Microbiol Lett.	PhD	• Database management and curation
Cunningham-Oakes *et al*. 2020. Microbiol Resour Announc.	PhD	• Comparative genomics • Plasmid evolutionary analysis

References

Abdel Malek, S.M.A. and Badran, Y.R. (2010). *Pseudomonas aeruginosa* PAO1 Adapted to 2-Phenoxyethanol Shows Cross-Resistance to Dissimilar Biocides and Increased Susceptibility to Antibiotics. *Folia Microbiologica* 55(6), pp. 588–592.

Addai, E.K., Gabel, D. and Krause, U. (2016). Experimental investigations of the minimum ignition energy and the minimum ignition temperature of inert and combustible dust cloud mixtures. *Journal of Hazardous Materials* 307(1), pp. 302–311.

Agnoli, K., Frauenknecht, C., Freitag, R., Schwager, S., Jenul, C., Vergunst, A., … Eberl, L. (2014). The third replicon of members of the *Burkholderia cepacia* complex, plasmid pC3, plays a role in stress tolerance. *Applied and Environmental Microbiology* 80(4), pp. 1340-1348.

Agnoli, K., Schwager, S., Uehlinger, S., Vergunst, A., Viteri, D. F., Nguyen, D. T., Sokol, P. A., Carlier, A. and Eberl, L. (2012). "Exposing the Third Chromosome of *Burkholderia cepacia* Complex Strains as a Virulence Plasmid." *Molecular Microbiology* 83(2), pp. 362-378.

Alexandre, H., Mathieu, B. and Charpentier, C. (1996). Alteration in membrane fluidity and lipid composition, and modulation of H+-ATPase activity in *Saccharomyces cerevisiae* caused by decanoic acid. *Microbiology* 142(3), pp. 469-475.

Alikhan, N.F., Zhou, Z., Sergeant, M.J. and Achtman, M. (2018). A genomic overview of the population structure of *Salmonella*. PLoS Genetics. 14(4), e1007261.

Andrews, S., Krueger, F., Seconds-Pichon, A., Biggins, F. and Wingett, S. (2015). FastQC. A quality control tool for high throughput sequence data. *Babraham Bioinformatics*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Last accessed 11 April 2020]

Aruhomukama, D., Sserwadda, I. and Mboowa, G. (2019). Whole-genome sequence analysis of *Vibrio cholerae* from three outbreaks in Uganda, 2014 - 2016. F1000Research 8(1), p. 1340.

Cosmetics Europe, The Personal Care Association (2015). Compliance with regulation 1223/2009 on Cosmetic Products. Available at:

https://cosmeticseurope.eu/files/2214/7945/7676/Updated_Cosmetics_Europe_PIF_Guidelines_-_2015_-_Update.pdf.

Aviv, G., Tsyba, K., Steck, N., Salmon-Divon, M., Cornelius, A., Rahav, G., … Gal-Mor, O. (2014). A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *Salmonella enterica* serovar Infantis strain. *Environmental Microbiology* 16(4), pp. 977-994.

Azevedo, P.A.A., Furlan, J.P.R., Oliveira-Silva, M., Nakamura-Silva, R., Gomes, C.N., Costa, K.R.C., … Pitondo-Silva, A. (2018). Detection of virulence and β-lactamase encoding genes in *Enterobacter aerogenes* and *Enterobacter cloacae* clinical isolates from Brazil. *Brazilian Journal of Microbiology* 49 (Suppl 1), pp. 224-228.

Bagge-Ravn, D., Ng, Y., Hjelm, M., Christiansen, J.N., Johansen, C. and Gram, L. (2003). The microbial ecology of processing equipment in different fish industries - Analysis of the microflora during processing and following cleaning and disinfection. *International Journal of Food Microbiology* 87(3), pp. 239–250.

Baldwin, A., Mahenthiralingam, E., Kathleen, M., Honeybourne, D., Maiden, M.C.J., John, R., … Govan, J.R. (2005). Multilocus Sequence Typing Scheme That Provides Both Species and Strain Differentiation for the *Burkholderia cepacia* complex. *Journal of Clinical Microbiology* 43(9), pp. 4665–4673.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., … Pevzner, P.A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5), pp. 455-477.

Barel, A.O., Paye, M. and Maibach, H.I. (2009). *Handbook of Cosmetic Science and Technology. 3rd edition,* pp. 249-250. CRC Press, Florida.

Barry, A.L., Fay, G.D. and Sauer, R.L. (1972). Efficiency of a transport medium for the recovery of aerobic and anaerobic bacteria from applicator swabs. *Applied Microbiology* 24(1), pp. 31-33.

Bashir, A. and Lambert, P. (2020). Microbiological study of used cosmetic products: highlighting possible impact on consumer health. *Journal of Applied Microbiology* 128(2), pp. 598–605.

Becker, S.L., Berger, F.K., Feldner, S.K., Karliova, I., Haber, M., Mellmann, A., … Gärtner, B. (2018). Outbreak of *Burkholderia cepacia* complex infections associated with contaminated octenidine mouthwash solution, Germany, August to September 2018. *Eurosurveillance* 23(42).

Becks, V.E. and Lorenzoni, N.M. (1995). *Pseudomonas aeruginosa* outbreak in a neonatal intensive care unit: a possible link to contaminated hand lotion. *American Journal of Infection Control* 23(6), pp. 396–398.

Behravesh, C.B., Jones, T.F., Vugia, D.J., Long, C., Marcus, R., Smith, K., … Scallan, E. (2011). Deaths associated with bacterial pathogens transmitted commonly through food: Foodborne Diseases Active Surveillance Network (FoodNet), 1996-2005. *Journal of Infectious Diseases* 204(2), pp. 263–267.

Bergmiller, T., Andersson, A.M.C., Tomasek, K., Balleza, E., Kiviet, D.J., Hauschild, R., … Guet, C.C. (2017). Biased partitioning of the multidrug efflux pump AcrAB-TolC underlies long-lived phenotypic heterogeneity. *Science* 356(6335), pp. 311-315.

Berthele, H., Sella, O., Lavarde, M., Mielcarek, C., Pense-Lheritier, A.M. and Pirnay, S. (2014). Determination of the influence of factors (ethanol, pH and A_w) on the preservation of cosmetics using experimental design. *International Journal of Cosmetic Science* 36(1), pp. 54-61.

Bialek-Davenet, S., Marcon, E., Leflon-Guibout, V., Lavigne, J.P., Bert, F., Moreau, R. and Nicolas-Chanoine, M.H. (2011). In vitro selection of *ramR* and *soxR* mutants overexpressing efflux systems by fluoroquinolones as well as cefoxitin in *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy* 55(6) pp. 2795-2802.

Bom, S., Jorge, J., Ribeiro, H.M. and Marto, J. (2019). A step forward on sustainability in the cosmetics industry: A review. *Journal of Cleaner Production* 225(1), pp. 270-290.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.F., Lió, P., … Fondi, M. (2015). MeDuSa: A multi draft-based scaffolder. *Bioinformatics* 31(15), pp. 2443-51. Available at:

Botelho, J., Grosso, F., Quinteira, S., Mabrouk, A. and Peixe, L. (2017). The complete nucleotide sequence of an IncP-2 megaplasmid unveils a mosaic architecture comprising a putative novel blaVIM-2-harbouring transposon in *Pseudomonas aeruginosa*. *Journal of Antimicrobial Chemotherapy* 72(8), pp. 2225-2229.

Boucher, H.W., Talbot, G.H., Bradley, J.S., Edwards, J.E., Gilbert, D., Rice, L.B., … Bartlett, J. (2009). Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. *Clinical Infectious Diseases* 48(1), pp. 1-12.

Bouillard, J., Vignes, A., Dufaud, O., Perrin, L. and Thomas, D. (2010). Ignition and explosion risks of nanopowders. *Journal of Hazardous Materials*. 181(1), pp. 873–880.

Brady, C., Cleenwerck, I., Venter, S., Coutinho, T. and de Vos, P. (2013). Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA) (A MANO). *Systematic and Applied Microbiology* 36(5), pp. 309–319.

Briasco, B., Capra, P., Cozzi, A.C., Mannucci, B. and Perugini, P. (2016). Packaging evaluation approach to improve cosmetic product safety. *Cosmetics* 213(4), pp. 308-320.

Britton, L.G. (1999a). Evaluating the Hazard of Static Electricity, pp. 47–69. *Avoiding Static Ignition Hazards in Chemical Operations*. John Wiley & Sons, Inc., New Jersey.

Britton, L.G. (1999b). Flammable Liquids, Vapors, and Gases, pp. 83–166. *Avoiding Static Ignition Hazards in Chemical Operations*. New Jersey: John Wiley & Sons, Inc.

Bull, M.J., Marchesi, J.R., Vandamme, P., Plummer, S. and Mahenthiralingam, E. (2012). Minimum taxonomic criteria for bacterial genome sequence depositions and announcements. *Journal of Microbiological Methods* 89(1), pp. 18-21.

Buranasuksombat, U., Kwon, Y.J., Turner, M. and Bhandari, B. (2011). Influence of emulsion droplet size on antimicrobial properties. *Food Science and Biotechnology* 20(1), pp. 793–800.

Caputo, A., Fournier, P.E. and Raoult, D. (2019). Genome and pan-genome analysis to classify emerging bacteria. *Biology Direct* 14(5).

Carattoli, A., Zankari, E., Garciá-Fernández, A., Larsen, M. V., Lund, O., Villa, L., Aarestrup, F. M., & Hasman, H. (2014). In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy* 58(7), pp. 3895–3903.

Cassagne C, Normand AC, L'Ollivier C, Ranque S & Piarroux R (2016) Performance of MALDI-TOF MS platforms for fungal identification. *Mycoses* 59(11), pp. 678-690.

Castro-González, R., Martínez-Aguilar, L., Ramírez-Trujillo, A., los Santos, P. and Caballero-Mellado, J. (2011). High diversity of culturable *Burkholderia* species associated with sugarcane. *Plant and Soil* 345(1), pp. 155–169.

Cazares, A., Moore, M.P., Hall, J.P.J., Wright, L.L., Grimes, M., Emond-Rhéault, J.-G., … Winstanley, C. (2020). A megaplasmid family driving dissemination of multidrug resistance in *Pseudomonas*. *Nature Communications* 11(1), p. 1370.

Centers for Disease Control and Prevention (2011). Vital Signs: Incidence and Trends of Infection with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 1996–2010. *MMWR* 60(22), pp. 749–755.

Chen, N.H., Djoko, K.Y., Veyrier, F.J. and McEwan, A.G. (2016). Formaldehyde stress responses in bacterial pathogens. *Frontiers in Microbiology*. 7(1), p. 257.

Chen, W., Zhang, Ya, Zhang, Yifei, Pi, Y., Gu, T., Song, L., … Ji, Q. (2018). CRISPR/Cas9-based Genome Editing in *Pseudomonas aeruginosa* and Cytidine Deaminase-Mediated Base Editing in *Pseudomonas* Species. *iScience* 31(6), pp. 222-231.

Cherrington, R. (2016). Liz Earle Cleanser Recall: Batch Found To Contain Potentially Dangerous Bacteria. *Huffington Post*. Available at: https://www.huffingtonpost.co.uk/entry/liz-earle-cleanse-and-polish-hot-cloth-cleanser-recall_uk_58206149e4b020461a1ceeef?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAANjGkrECYrtHM2_Wkvxcjow2qI3G5OCVsxBMq-jvVxBE3UBMv0UL [Last accessed: 17 February 2020].

Chindera, K., Mahato, M., Kumar Sharma, A., Horsley, H., Kloc-Muniak, K., Kamaruzzaman, N.F., … Good, L. (2016). The antimicrobial polymer PHMB enters cells and selectively condenses bacterial chromosomes. *Scientific Reports* 6(1), 23121.

CLSI. (2018). Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically. 11th ed. CLSI standards M07. Wayne, PA: Clinical and Laboratory Standards Institute.

European Commission (2014). Commission Regulation (EU) No 1004/ (2014). Commission Regulation (EU) No 1004/2014 of 18 September 2014 amending Annex V to Regulation (EC) No 1223/2009 of the European Parliament and of the Council on cosmetic products. *Official Journal of the European Union*. Available at: https://eur-

lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R1004&from=EN. [Last accessed: 11 April 2020].

Corby-Edwards, A.K. (2013). FDA regulation of cosmetics and personal care products. *Cosmetics and FDA Regulation*. Available at: https://docplayer.net/14948847-Fda-regulation-of-cosmetics-and-personal-care-products.html. [Last accessed: 11 April 2020].

Couto, S.R. and Sanromán, M.Á. (2006). Application of solid-state fermentation to food industry - A review. *Journal of Food Engineering* 76(3), pp. 291-302.

Cundell, T. (2015). The role of water activity in the microbial stability of non-sterile drug products. *European Pharmaceutical Review* 20(1), pp. 58-63.

Cunningham-Oakes, E., Pointon, T., Murphy, B., Connor, T.R. and Mahenthiralingam, E. (2020). Genome Sequence of *Pluralibacter gergoviae* ECO77, a Multireplicon Isolate of Industrial Origin. *Microbiology Resource Announcements* 9(9), e01561-e01619.

Cunningham-Oakes, E., Weiser, R., Pointon, T. and Mahenthiralingam, E. (2020). Understanding the challenges of non-food industrial product contamination. *FEMS Microbiology Letters* 366(23).

D.A., W., G.P., C. and B.P., H. (2017). Current and emerging topical antibacterials and antiseptics: Agents, action, and resistance patterns. *Clinical Microbiology Reviews* 30(3), pp. 827-860.

Davenne, T. and McShane, H. (2016). Why don't we have an effective tuberculosis vaccine yet? *Expert Review of Vaccines* 15(8), pp. 1009–1013.

Davin-Regli, A., Lavigne, J.P. and Pagès, J.M. (2019). *Enterobacter spp.*: update on taxonomy, clinical aspects, and emerging antimicrobial resistance. *Clinical Microbiology Reviews* 32(4), e00002-e00019.

Davin-Regli, A. and Pagès, J.M. (2015). *Enterobacter aerogenes* and *Enterobacter cloacae*; Versatile bacterial pathogens confronting antibiotic treatment. *Frontiers in Microbiology* 6(1), p. 392.

Davis, G.S. and Price, L.B. (2016). Recent Research Examining Links Among *Klebsiella pneumoniae* from Food, Food Animals, and Human Extraintestinal Infections. *Current Environmental Health Reports* 3(2), pp. 128–135.

Deamer, D., Akeson, M. and Branton, D. (2016). Three decades of nanopore sequencing *Nature Biotechnology* 34(1), pp. 518–524.

Depoorter, E., Bull, M. J., Peeters, C., Coenye, T., Vandamme, P., & Mahenthiralingam, E. (2016). *Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. Applied Microbiology and Biotechnology 100(12), pp. 5215-29.

Desmet, S., Nepal, S., van Dijl, J.M., Van Ranst, M., Chlebowicz, M.A., Rossen, J.W., … Bathoorn, E. (2018). Antibiotic resistance plasmids cointegrated into a megaplasmid harboring the bla OXA-427 carbapenemase gene. *Antimicrobial Agents and Chemotherapy* 62(3), e01448-e01517.

Dieckmann, R., Hammerl, J.A., Hahmann, H., Wicke, A., Kleta, S., Dabrowski, P.W., … Lasch, P. (2016). Rapid characterisation of *Klebsiella oxytoca* isolates from contaminated liquid hand soap using mass spectrometry, FTIR and Raman spectroscopy. *Faraday Discussions* 187(1), pp. 353-375.

Dietrich, L.E.P., Teal, T.K., Price-Whelan, A. and Newman, D.K. (2008). Redox-active antibiotics control gene expression and community behavior in divergent bacteria. *Science* 321(5893), pp. 1203-120.

Eckhoff, R.K. (2009). Dust Explosion Prevention and Mitigation, Status and Developments in Basic Knowledge and in Practical Application. *International Journal of Chemical Engineering* 2009.

Eichenberger, E.M. and Thaden, J.T. (2019). Epidemiology and mechanisms of resistance of extensively drug resistant gram-negative bacteria. *Antibiotics* 8(2), p. 37.

Eklund, T. (1980). Inhibition of Growth and Uptake Processes in Bacteria by Some Chemical Food Preservatives. *Journal of Applied Bacteriology* 48, pp. 423-432.

Elliott, A.G., Ganesamoorthy, D., Coin, L., Cooper, M.A. and Cao, M.D. (2016). Complete genome sequence of *Klebsiella quasipneumoniae* subsp. *similipneumoniae* strain ATCC 700603. *Genome Announcements* 4(3), e00438-16.

Erkmen, O. and Bozoglu, T.F. (2016). Food Preservation by Reducing Water Activity, pp. 44-58. *Food Microbiology: Principles into Practice*. John Wiley & Sons, Ltd, New Jersey.

European Centre for Disease Prevention and Control (2018). Carbapenem-Resistant *Enterobacteriaceae*- First Update. Available at:

https://www.ecdc.europa.eu/sites/default/files/documents/RRA-Enterobacteriaceae-Carbapenems-European-Union-countries.pdf [Last accessed: 11 April 2020].

European Commission (2015). RAPEX - Unsafe Products.European Commission, 22. October. Available at: http://ec.europa.eu/consumers/safety/rapex/alerts/main/index.cfm?event=main.listNotifications. [Last accessed: 11 April 2020].

European Commission (2016). ATEX 2014/34/EU Guidelines. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0034&locale=en [Last accessed: 11 April 2020].

European Union (2016). RASFF - Food and Feed Safety Alerts.08/07/2016. Available at: https://ec.europa.eu/food/safety/rasff_en%0Ahttp://ec.europa.eu/food/safety/rasff/index_en.htm. [Last accessed: 11 April 2020].

Evans, H., Bolt, H., Heinsbroek, E., Lloyd, B., English, P., Latif, S., … Puleston, R. (2018). National outbreak of *Pseudomonas aeruginosa* associated with an aftercare solution following piercings, July to September 2016, England. *Eurosurveillance* 23(37).

Farmer, J.J., Davis, B.R., Hickman-Brenner, F.W., McWhorter, A., Huntley-Carter, G.P., Asbury, M.A., … Fanning, G.R. (1985). Biochemical identification of new species and biogroups of *Enterobacteriaceae* isolated from clinical specimens. *Journal of Clinical Microbiology* 21(1), pp. 46-76.

Fei, N. and Zhao, L. (2013). An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *ISME Journal* 7(1), pp. 880–884.

Flannagan, R.S., Linn, T. and Valvano, M.A. (2008). A system for the construction of targeted unmarked gene deletions in the genus *Burkholderia*. *Environmental Microbiology* 10(6), pp. 1652-1660.

Flight, W.G., Smith, A., Paisey, C., Marchesi, J.R., Bull, M.J., Norville, P.J., … Mahenthiralingam, E. (2015). Rapid Detection of Emerging Pathogens and Loss of Microbial Diversity Associated with Severe Lung Disease in Cystic Fibrosis. *Journal of Clinical Microbiology* 53(7), pp. 2022–2029.

Galyov, E.E., Brett, P.J. and DeShazer, D. (2010). Molecular Insights into *Burkholderia pseudomallei* and *Burkholderia mallei*. *Pathogenesis*. *Annual Review of Microbiology* 64(1), pp. 495–517.

Garner, N., Siol, A. and Eilks, I. (2014). Parabens as preservatives in personal care products. *Chem. Action* 103(1), pp. 36–43.

Gfatter, R., Hackl, P. and Braun, F. (1997). Effects of soap and detergents on skin surface pH, stratum corneum hydration and fat content in infants. *Dermatology* 195(3), pp. 258-62.

Giorgio, A., Miele, L., De Bonis, S., Conforti, I., Palmiero, L., Guida, M., … Aliberti, F. (2018). Microbiological stability of cosmetics by using challenge test procedure. *Journal of Pure and Applied Microbiology* 12(1), pp. 23-28.

Glor, M. (2006). Transfer of powders into flammable solvents overview of explosion hazards and preventive measures. *Journal of Loss Prevention in the Process Industries* 19(6), pp. 656–663.

Goodyear, N., Brouillette, N., Tenaglia, K., Gore, R. and Marshall, J. (2015). The effectiveness of three home products in cleaning and disinfection *of Staphylococcus aureus* and *Escherichia coli* on home environmental surfaces. *Journal of Applied Microbiology* 119(5), pp. 1245-1252.

Grant, J.R. and Stothard, P. (2008). The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Research* 36(Web Server issue), W181-4.

De Groot, A.C. and Veenstra, M. (2010). Formaldehyde-releasers in cosmetics in the USA and in Europe. *Contact Dermatitis* 62(4), pp. 221-224.

Gupta, P.L., Rajput, M., Oza, T., Trivedi, U. and Sanghvi, G. (2019). Eminence of Microbial Products in Cosmetic Industry. *Natural Products and Bioprospecting* 9(4), pp. 267-278.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29(8), pp. 1072-1075.

Halla, N., Fernandes, I.P., Heleno, S.A., Costa, P., Boucherit-Otmani, Z., Boucherit, K., … Barreiro, M.F. (2018). Cosmetics preservation: A review on present strategies. *Molecules* 23(7), E1571.

Hamad, M.A., Skeldon, A.M. and Valvano, M.A. (2010). Construction of aminoglycoside-sensitive *Burkholderia cenocepacia* strains for use in studies of intracellular bacteria with the gentamicin protection assay. *Applied and Environmental Microbiology* 76(10), pp. 3170-3176.

Hamouda, T. and Baker, J.R. (2000). Antimicrobial mechanism of action of surfactant lipid preparations in enteric gram-negative bacilli. *Journal of Applied Microbiology* 89(3), pp. 397-403.

Hardoim, P.R., Nazir, R., Sessitsch, A., Elhottová, D., Korenblum, E., Van Overbeek, L.S. and Van Elsas, J.D. (2013). The new species *Enterobacter oryziphilus* sp. nov. and *Enterobacter oryzendophyticus* sp. nov. are key inhabitants of the endosphere of rice. *BMC Microbiology* 13(164).

Hasnan, N.Z.N., Aziz, N.A., Zulkifli, N. and Taip, F.S. (2014). Food Factory Design: Reality and Challenges Faced by Malaysian SMEs. *Agriculture and Agricultural Science Procedia* 2(1), pp. 328–336.

Henry, D., Campbell, M., McGimpsey, C., Clarke, A., Louden, L., Burns, J.L., … Speert, D. (1999). Comparison of isolation media for recovery of *Burkholderia cepacia* complex from respiratory secretions of patients with cystic fibrosis. *Journal of Clinical Microbiology* 37(4), pp. 1004–1007.

Hoffmann, M., Luo, Y., Monday, S.R., Gonzalez-Escalona, N., Ottesen, A.R., Muruvanda, T., … Brown, E.W. (2016). Tracing origins of the *Salmonella bareilly* strain causing a food-borne outbreak in the United States. *Journal of Infectious Diseases* 213(4), pp. 502–508.

Høgsberg, T., Saunte, D.M., Frimodt-Møller, N. and Serup, J. (2013). Microbial status and product labelling of 58 original tattoo inks. *Journal of the European Academy of Dermatology and Venereology* 27(1), pp. 73-80.

Hong G, Miller HB, Allgood S, Lee R, Lechtzin N & Zhang SX (2017) Use of Selective Fungal Culture Media Increases Rates of Detection of Fungi in the Respiratory Tract of Cystic Fibrosis Patients. Journal of Clinical Microbiology 55(4), pp. 1122-1130.

Huang, J., Hitchins, A.D., Tran, T.T. and McCarron, J.E. (2017). Laboratory Methods - BAM: Methods for Cosmetics. U.S Food and Drug Administration. Available at:

https://www.fda.gov/food/laboratory-methods-food/bacteriological-analytical-manual-bam. [Last accessed 11 April 2020]

Ihrmark, K., Bödeker, I.T.M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., … Lindahl, B.D. (2012). New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* 82(3), pp. 666–677.

Ito, S., Yazawa, S., Nakagawa, Y., Sasaki, Y. and Yajima, S. (2015). Effects of alkyl parabens on plant pathogenic fungi. *Bioorganic and Medicinal Chemistry Letters* 25(8), pp. 1774-1777.

Izard, D., Richar, C. and Leclerc, H. (1983). DNA relatedness between *Enterobacter Sakazakii* and other members of the genus *Enterobacter*. *Annales de l'Institut Pasteur Microbiology* 134A(3), pp. 241-245.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T. and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9(1), p. 5114.

Jesumirhewe, C., Ogunlowo, P.O., Olley, M., Springer, B., Allerberger, F. and Ruppitsch, W. (2016). Accuracy of conventional identification methods used for *Enterobacteriaceae* isolates in three Nigerian hospitals. *PeerJ* 4(1), e2511.

Jeukens, J., Kukavica-Ibrulj, I., Emond-Rheault, J.G., Freschi, L. and Levesque, R.C. (2017). Comparative genomics of a drug-resistant *Pseudomonas aeruginosa* panel and the challenges of antimicrobial resistance prediction from genomes. *FEMS Microbiology Letters* 364(18).

Jimenez, L. (2004). *Microbial Contamination Control in the Pharmaceutical Industry*. (1st Edition), CRC Press, Florida.

Jimenez, L. (2007). Microbial diversity in pharmaceutical product recalls and environments. *PDA Journal of Pharmaceutical Science and Technology* 61(5), pp. 383–399.

Jimenez, L. (2011). Molecular Applications to Pharmaceutical Processes and Cleanroom Environments. *PDA Journal of Pharmaceutical Science and Technology* 65(1), pp. 242–253.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., … Wong, G.K.S. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology* 7(1), p. 459.

Kabara, J.J. (1997). *Preservative-Free and Self-Preserving Cosmetics and Drugs: Principles and Practice*. Marcel Dekker, Inc., New York.

Kamaruzzaman, N.F., Chong, S.Q.Y., Edmondson-Brown, K.M., Ntow-Boahene, W., Bardiau, M. and Good, L. (2017). Bactericidal and anti-biofilm effects of polyhexamethylene Biguanide in models of intracellular and biofilm of *Staphylococcus aureus* isolated from bovine mastitis. *Frontiers in Microbiology*, 8(1), p. 1518.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* 44(D1), pp. D457-62.

Kanehisa, M., Sato, Y. and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences *Journal of Molecular Biology* 428(4), pp. 726-731.

Khalifa, A.Y.Z., Alsyeeh, A.M., Almalki, M.A. and Saleh, F.A. (2016). Characterization of the plant growth promoting bacterium, *Enterobacter cloacae* MSR1, isolated from roots of non-nodulating *Medicago sativa*. *Saudi Journal of Biological Sciences* 23(1), pp. 79-86.

Kõljalg, U., Larsson, K.H., Abarenkov, K., Nilsson, R.H., Alexander, I.J., Eberhardt, U., … Ursing, B.M. (2005). UNITE: A database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* 166(3), pp. 1063-1068.

Koutsolioutsou, A., Peña-Llopis, S. and Demple, B. (2005). Constitutive *soxR* mutations contribute to multiple-antibiotic resistance in clinical *Escherichia coli* isolates. *Antimicrobial Agents and Chemotherapy* 49(7), pp. 2746-2752.

Krueger, F. (2015). Trim Galore!: A Wrapper Tool around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files. *Babraham Institute*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [Last accessed: 11 April 2020]

Kuepper, J., Ruijssenaars, H.J., Blank, L.M., de Winde, J.H. and Wierckx, N. (2015). Complete genome sequence of solvent-tolerant *Pseudomonas putida* S12 including megaplasmid pTTS12. *Journal of Biotechnology* 200(1), pp. 17-18.

Ladunga, I. (2017). Finding homologs in amino acid sequences using network blast searches. *Current Protocols in Bioinformatics* 59(1), pp. 3.4.1-3.4.24.

Lambert, P.A. (2013a). Mechanisms of Action of Microbiocides, pp. 95-107. *Russell, Hugo & Ayliffe's Principles and Practice of Disinfection, Preservation & Sterilization*. Blackwell Publishing Ltd, New Jersey.

Lambert, P.A. (2013b). Mechanisms of Action of Biocides, pp. 139–153. *Russell, Hugo & Ayliffe's Principles and Practice of Disinfection, Preservation & Sterilization*. Blackwell Publishing Ltd, New Jersey.

Lane, D. J. (1991). 16S/23S rRNA sequencing, pp. 115-175. *In* E. Stackebrandt and M. Goodfellow (ed.), *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, Inc, New Jersey.

Leong, L.E.X., Lagana, D., Carter, G.P., Wang, Q., Smith, K., Stinear, T.P., … Rogers, G.B. (2018). *Burkholderia lata* infections from intrinsically contaminated chlorhexidine Mouthwash, Australia, 2016. Emerging Infectious Diseases 24(11), pp. 2109–2111.

Li, H. (2018). Minimap2: fast pairwise alignment for long DNA sequences. *Bioinformatics* 34(18), pp. 3094-3100.

Li, L., Yu, T., Ma, Y., Yang, Z., Wang, W., Song, X., … Xu, H. (2019). The genetic structures of an extensively drug resistant (XDR) *Klebsiella pneumoniae* and its plasmids. *Frontiers in Cellular and Infection Microbiology* 8(1), p. 446.

Liapis, E., Bour, M., Triponney, P., Jové, T., Zahar, J.R., Valot, B., … Plésiat, P. (2019). Identification of diverse integron and plasmid structures carrying a novel carbapenemase among *Pseudomonas* species. *Frontiers in Microbiology* 10(1), p. 404.

Lou, Y., Zhang, Z.H., Xiao, L.L., Guo, D.F., Liu, H.Q., Pan, Y.J. and Zhao, Y. (2015). Growth kinetic parameters of multi-drug and single-drug resistant *Vibrio parahaemolyticus* strains in pure culture and in *Penaeus vannamei*. *Modern Food Science and Technology* 31(5), pp. 181-186.

Lü, Y., Zhao, S., Liang, H., Zhang, W., Liu, J. and Hu, H. (2019). The first report of a novel incHI1B blaSIM-1-carrying megaplasmid pSIM-1-BJ01 from a clinical *Klebsiella pneumoniae* isolate. *Infection and Drug Resistance* 12(1), pp. 2103–2112.

Lucchini, J.J., Corre, J. and Cremieux, A. (1990). Antibacterial activity of phenolic compounds and aromatic alcohols. *Research in Microbiology* 141(4), pp. 499-510.

Lundov, M.D. and Zachariae, C. (2008). Recalls of microbiologically contaminated cosmetics in EU from 2005 to May 2008. I*nternational Journal of Cosmetic Science* 30(6), pp. 471–474.

Macià, M.D., Rojo-Molinero, E. and Oliver, A. (2014). Antimicrobial susceptibility testing in biofilm-growing bacteria. *Clinical Microbiology and Infection* 20(10), pp. 981-990.

Macneil, A., Glaziou, P., Sismanidis, C., Maloney, S. and Floyd, K. (2019). Global epidemiology of tuberculosis and progress toward achieving global targets — 2017. *Morbidity and Mortality Weekly Report* 68(11), pp. 263-266.

Mahenthiralingam, E., Bischof, J., Byrne, S.K., Radomski, C., Davies, J.E., Av-gay, Y. and Vandamme, P. (2000). DNA-Based Diagnostic Approaches for Identification of *Burkholderia cepacia* complex, *Burkholderia vietnamiensis*, *Burkholderia multivorans*, *Burkholderia stabilis*, and *Burkholderia cepacia* Genomovars I and III. *Journal of Clinical Microbiology* 38(9), pp. 3165–3173.

Mahenthiralingam, E., Campbell, M.E., Foster, J., Lam, J.S. and Speert, D.P. (1996). Random amplified polymorphic DNA typing of *Pseudomonas aeruginosa* isolates recovered from patients with cystic fibrosis. *Journal of Clinical Microbiology* 34(5), pp. 1129-1135.

Mahenthiralingam, E., Campbell, M.E., Henry, D.A. and Speert, D.P. (1996). Epidemiology of *Burkholderia cepacia* infection in patients with cystic fibrosis: Analysis by randomly amplified polymorphic DNA fingerprinting. *Journal of Clinical Microbiology* 34(12), pp. 2914-20.

Maiden, M.C.J., Jansen van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S. a, Jolley, K. a and McCarthy, N.D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology* 11(10), pp. 728–36.

Maillard, J.Y. (2002). Bacterial target sites for biocide action. *Journal of Applied Microbiology*, 92(1), pp. 16S-27S.

Malek, A., McGlynn, K., Taffner, S., Fine, L., Tesini, B., Wang, J., … Pecora, N. (2019). Next-generation-sequencing-based hospital outbreak investigation yields insight into *Klebsiella aerogenes* population structure and determinants of carbapenem resistance and pathogenicity. *Antimicrobial Agents and Chemotherapy* 63(6), e02577-e02618.

Marchesi, J.R. and Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*. 3(31).

Marquez, L., Jones, K.N., Whaley, E.M., Koy, T.H., Revell, P.A., Taylor, R.S., … Campbell, J.R. (2017). An Outbreak of *Burkholderia cepacia* Complex Infections Associated with Contaminated Liquid Docusate. *Infection Control Hospital Epidemiology* 38(5), pp. 567-573.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17 (1), pp. 10-12.

Maukonen, J., Mättö, J., Wirtanen, G., Raaska, L., Mattila-Sandholm, T. and Saarela, M. (2003). Methodologies for the characterization of microbes in industrial environments: A review. *Journal of Industrial Microbiology and Biotechnology* 30(6), pp. 327–356.

Mayr, A., Hinterberger, G., Lorenz, I.H., Kreidl, P., Mutschlechner, W. and Lass-Flörl, C. (2017). Nosocomial outbreak of extensively drug-resistant *Pseudomonas aeruginosa* associated with aromatherapy. *American Journal of Infection Control* 45(4), pp. 453-455.

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., … Wright, G.D. (2013). The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy* 57(7), pp. 3348-3357.

Mellmann A, Becker K, von Eiff C, Keckevoet U, Schumann P & Harmsen D (2006) Sequencing and *Staphylococci* identification. *Emerging Infectious Diseases* 12(2), pp. 333-336.

Mokarizadeh, M., Kafil, H., Ghanbarzadeh, S., Alizadeh, A. and Hamishehkar, H. (2017). Improvement of citral antimicrobial activity by incorporation into nanostructured lipid carriers: A potential application in food stuffs as a natural preservative. *Research in Pharmaceutical Sciences* 12(5), pp. 409-415.

Moore, I. (2009). Manufacturing Cosmetic Ingredients According to Good Manufacturing Practice Principles, pp. 79–92. *Global Regulatory Issues for the Cosmetics Industry*. William Andrew, New York.

Møretrø, T. and Langsrud, S. (2017). Residential Bacteria on Surfaces in the Food Industry and Their Implications for Food Safety and Quality. *Comprehensive Reviews in Food Science and Food Safety* 16(5), pp. 1022–1041.

Mulani, M.S., Kamble, E.E., Kumkar, S.N., Tawre, M.S. and Pardesi, K.R. (2019). Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: A review. *Frontiers in Microbiology* 10(1), p. 539.

Mullane, N.R., Whyte, P., Wall, P.G., Quinn, T. and Fanning, S. (2007). Application of pulsed-field gel electrophoresis to characterise and trace the prevalence of *Enterobacter sakazakii* in an infant formula processing facility. *International Journal of Food Microbiology* 116(1), pp. 73-81.

Mullins, A. J., Murray, J., Bull, M. J., Jenner, M., Jones, C., Webster, G., Green, A. E., Neill, D. R., Connor, T. R., Parkhill, J., Challis, G. L., & Mahenthiralingam, E. (2019). Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nature Microbiology* 4(6) pp. 996–1005.

Nes, I.F. and Eklund, T. (1983). The effect of parabens on DNA, RNA and protein synthesis in *Escherichia coli* and *Bacillus subtilis*. *Journal of Applied Bacteriology* 54(2), pp. 237-242.

Neza, E. and Centini, M. (2016). Microbiologically Contaminated and Over-Preserved Cosmetic Products According RAPEX 2008–2014. *Cosmetics* 3(1), p. 3.

Nicolaou, K.C. (2014). Organic synthesis: The art and science of replicating the molecules of living nature and creating others like them in the laboratory. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 470(2163), 20130690.

Nicoloff, H., Hjort, K., Levin, B.R. and Andersson, D.I. (2019). The high prevalence of antibiotic heteroresistance in pathogenic bacteria is mainly caused by gene amplification. *Nature Microbiology* 4(1), pp. 504–514.

Noël, A., Vastrade, C., Dupont, S., de Barsy, M., Huang, T.D., Van Maerken, T., … Glupczynski, Y. (2019). Nosocomial outbreak of extended-spectrum β-lactamase-producing *Enterobacter cloacae* among cardiothoracic surgical patients: causes and consequences. *Journal of Hospital Infection* 102 (1), pp. 54-60.

Oliphant, C.M. and Eroschenko, K. (2015). Antibiotic resistance, Part 1: Gram-positive pathogens. *Journal for Nurse Practitioners* 11(1), pp. 70–78.

Orth, D.S., Kabara, J.J., Denyer, S.P. and Tan, S.K. (2006). *Cosmetic and Drug Microbiology*. Informa Healthcare, New York.

Orus, P. and Leranoz, S. (2005). Current trends in cosmetic microbiology. *International Microbiology* 8(2), pp. 77–79.

Page, A.J., Alikhan, N.F., Carleton, H.A., Seemann, T., Keane, J.A. and Katz, L.S. (2017). Comparison of classical multi-locus sequence typing software for next-generation sequencing data. *Microbial Genomics* 3(8), e000124.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., … Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22), pp. 3691–3693.

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E. and Larsson, D.G.J. (2014). BacMet: Antibacterial biocide and metal resistance genes database. *Nucleic Acids Research* 42(Database issue), pp. D737–D743.

Pareek, C.S., Smoczynski, R. and Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics* 52 (4), pp. 413-435.

Parte AC (2018) LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology* 68(6), pp. 1825-1829.

Pendleton, J.N., Gorman, S.P. and Gilmore, B.F. (2013). Clinical relevance of the ESKAPE pathogens. *Expert Review of Anti-infective Therapy* 11(3), pp. 297–308.

Penna, V.T.C., Martins, S.A.M. and Mazzola, P.G. (2002). Identification of bacteria in drinking and purified water during the monitoring of a typical water purification system. *BMC Public Health* 2(13).

Périamé, M., Pagès, J.M. and Davin-Regli, A. (2014). *Enterobacter gergoviae* adaptation to preservatives commonly used in cosmetic industry. *International Journal of Cosmetic Science* 36(4), pp. 386–395.

Périamé, M., Philippe, N., Condell, O., Fanning, S., Pagès, J.M. and Davin-Regli, A. (2015). Phenotypic changes contributing to *Enterobacter gergoviae* biocide resistance. *Letters in Applied Microbiology* 61(2), pp. 121–129.

Perry JD (2017) A Decade of Development of Chromogenic Culture Media for Clinical Microbiology in an Era of Molecular Diagnostics. *Clinical Microbiology Reviews* 30(2), pp. 449-479.

Petersen, J. (2011). Phylogeny and compatibility: Plasmid classification in the genomics era. *Archives of Microbiology* 193(5), pp. 313–321. Available at: https://doi.org10.1007/s00203-011-0686-9

Pirttijärvi, T.S.M., Graeffe, T.H. and Salkinoja-Salonen, M.S. (1996). Bacterial contaminants in liquid packaging boards: Assessment of potential for food spoilage. *Journal of Applied Bacteriology* 81(4), pp. 445–458.

Potter, R.F., D'Souza, A.W., Wallace, M.A., Shupe, A., Patel, S., Gul, D., … Dantas, G. (2018). *Superficieibacter electus* gen. nov., sp. nov., an extended-spectrum β-Lactamase possessing member of the *Enterobacteriaceae* family, isolated from intensive care unit surfaces. *Frontiers in Microbiology* 9(1629).

Promega (2012). Maxwell(R) 16 DNA Purification Kits Technical Manual. Available at: https://worldwide.promega.com/-/media/files/resources/protocols/technical-manuals/0/maxwell-16-dna-purification-kits-protocol.pdf?la=en [Last accessed: 11 April 2020]

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., … Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research* 41(Database issue), pp. D590–D596.

Rabaey, K., Boon, N., Siciliano, S.D., Verhaege, M. and Verstraete, W. (2004). Biofuel cells select for microbial consortia that self-mediate electron transfer. *Applied and Environmental Microbiology* 70(9), pp. 5373-5382.

Raja HA, Miller AN, Pearce CJ & Oberlies NH (2017) Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community. *Journal of Natural Products* 80(3), pp. 756-770.

Rajamohan, G., Srinivasan, V.B. and Gebreyes, W.A. (2009). Novel role of *Acinetobacter baumannii* RND efflux transporters in mediating decreased susceptibility to biocides. *Journal of Antimicrobial Chemotherapy* 65(2), pp. 228–232.

Ramamurthy, T., Ghosh, A., Pazhani, G.P. and Shinoda, S. (2014). Current Perspectives on Viable but Non-Culturable (VBNC) Pathogenic Bacteria. *Frontiers in Public Health* 2(1), p. 103.

Rhoads, A. and Au, K.F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics* 13(5), pp. 278-89.

Richter, M. and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* 106 (45), pp. 19126-19131.

Ricke, S.C. (2003). Perspectives on the use of organic acids and short chain fatty acids as antimicrobials. *Poultry Science* 82 (4), pp. 632-639.

Rushton, L., Sass, A., Baldwin, A., Dowson, C.G., Donoghue, D. and Mahenthiralingam, E. (2013). Key role for efflux in the preservative susceptibility and adaptive resistance of *Burkholderia cepacia* complex bacteria. *Antimicrobial Agents and Chemotherapy* 57(7), pp. 2972–2980.

Russell, A.D. (2003). Challenge testing: principles and practice. International Journal of Cosmetic Science 25(3), pp. 147–53.

Saker, L., Lee, K., Cannito, B., Gilmore, A., Campbell-Lendrum, D.H. (2004). Globalization and infectious diseases: a review of the linkages. Available at: https://www.who.int/tdr/publications/documents/seb_topic3.pdf. [Last accessed: 11 April 2020].

Salo, S., Ehavald, H., Raaska, L., Vokk, R. and Wirtanen, G. (2006). Microbial surveys in Estonian dairies. *LWT - Food Science and Technology* 39(5), pp. 460–471.

Sandle, T. (2015). Characterizing the Microbiota of a Pharmaceutical Water System - A Metadata Study. *SOJ Microbiology & Infectious Diseases* 3(1), pp. 1-8.

Saxena, S. and Nidhi, S.N. (2015). Antioxidant and Antibacterial Studies on Essential Oils Used as Alternatives for Chemical Preservatives. *Journal of Natural Product and Plant Resources* 5(1), pp. 9-14.

Scheinpflug, K., Wenzel, M., Krylova, O., Bandow, J.E., Dathe, M. and Strahl, H. (2017). Antimicrobial peptide cWFW kills by combining lipid phase separation with autolysis. *Scientific Reports* 7(44332).

Schmaus, G., Pfeiffer, A., Lange, S., Trunet, A. and Pillai, R. (2014). 1, 2-Alkanedoils for Cosmetic Preservation. *Cosmetics & Toiletries* 123(10), pp. 53-64.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30(14), pp. 2068–2069.

Sevastsyanovich, Y.R., Krasowiak, R., Bingle, L.E.H., Haines, A.S., Sokolov, S.L., Kosheleva, I.A., … Thomas, C.M. (2008). Diversity of IncP-9 plasmids of *Pseudomonas*. *Microbiology* 154(1), pp. 2929-2941.

Siegert, W. (2012). ISO 11930 - A Comparison to other Methods to Evaluate the Efficacy of Antimicrobial Preservation. Available at: https://www.teknoscienze.com/Contents/Riviste/PDF/HPC2_2013_RGB_34-41.pdf [Last accessed: 11 April 2020]

Siegert W (2013) Comparison of microbial challenge testing methods for cosmetics. *Household and Personal Care Today* 8(1), pp. 32-39.

Siegert, W. (2014). Boosting the Antimicrobial Efficiency of Multifunctional Additives by Chelating Agents. Available at: https://www.researchgate.net/profile/Wolfgang_Siegert2/publication/260177758_Boosting_the_Antimicrobial_Efficiency_of_Multifunctional_Additives_by_Chelating_Agents/links/02e7e52fe3e5b79f06000000.pdf. [Last accessed: 11 April 2020]

Sila-On, W., Vardhanabhuti, N., Ongpipattanakul, B. and Kulvanich, P. (2006). The influence of physicochemical properties of preservative compounds on their distribution into various phases of oil in water submicron emulsion. *PDA Journal of Pharmaceutical Science and Technology* 60(3), pp. 172-181.

Smart, R. and Spooner, D.F. (1972). Microbiological spoilage in pharmaceuticals and cosmetics. *Journal of the Society of Cosmetic Chemists* 23(1), pp. 721–737.

Smits, T.H.M. (2019). The importance of genome sequence quality to microbial comparative genomics. *BMC Genomics* 20(662).

Sommerstein, R., Führer, U., Priore, E. Lo, Casanova, C., Meinel, D.M., Seth-Smith, H.M.B., … Widmer, A. (2017). *Burkholderia stabilis* outbreak associated with contaminated commercially-available washing gloves, Switzerland, May 2015 to August 2016. *Eurosurveillance* 22(49).

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9), pp. 1312–1313.

Stewart, S.E., Parker, M.D., Amézquita, A. and Pitt, T.L. (2016). Microbiological risk assessment for personal care products. *International Journal of Cosmetic Science* 38(6), pp. 634-645.

Stratford, M. and Eklund, T. (2003). Organic acids and esters, pp. 48-84. In: Russell N.J., Gould G.W. (eds) *Food Preservatives*. Springer, Boston.

Sutton, S. (2006). Cosmetic Microbiology: A Practical Approach. CRC Press, Florida.

Sutton, S. and Jimenez, L. (2012). A review of reported recalls involving microbiological control 2004-2011 with emphasis on FDA considerations of 'objectionable organisms'. *American Pharmaceutical Review* 15(1), pp. 42-57.

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D.L., … Zorzet, A. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases* 18(3), pp. 318-327.

Terjung, N., Löffler, M., Gibis, M., Hinrichs, J. and Weiss, J. (2012). Influence of droplet size on the efficacy of oil-in-water emulsions loaded with phenolic antimicrobials. Food and Function 3(3), pp. 290-301.

Thiel, S., Bunk, B., Spröer, C., Overmann, J., Jahn, D. and Biedendieck, R. (2019). Complete Genome Sequence of *Raoultella electrica* 1GB (DSM 102253^T), Isolated from Anodic Biofilms of a Glucose-Fed Microbial Fuel Cell. *Microbiology Resource announcements* 8(43), pp. e00800-e00819.

Thiemann, A. and Jänichen, J. (2014). The formulator's guide to safe cosmetic preservation. *Personal Care Europe*. Available at:

https://pdfslide.net/documents/preservatives-the-formulators-guide-to-safe-cosmetic-the-formulators.html [Last accessed: 11 April 2020].

Toliver, J. and Narasimhan, S. (2018). Caprylyl Glycol: A Versatile Material to Boost Preservatives. *Cosmetics & Toiletries* 133(7), pp. 18–23.

Torbeck, L., Raccasi, D., Guilfoyle, D.E., Friedman, R.L. and Hussong, D. (2011). *Burkholderia cepacia*: This Decision Is Overdue. PDA Journal of Pharmaceutical Science and Technology 65(5), pp. 535–543.

Trinetta, V., Magossi, G., Allard, M.W., Tallent, S.M., Brown, E.W. and Lomonaco, S. (2019). Characterization of *Salmonella enterica* Isolates from Selected U.S. Swine Feed Mills by Whole-Genome Sequencing. *Foodborne Pathogens and Disease* 17(2), pp. 126-136.

Turton, Jane, Davies, F., Turton, Jack, Perry, C., Payne, Z. and Pike, R. (2019). Hybrid resistance and virulence plasmids in "high-risk" clones of *Klebsiella pneumoniae*, including those carrying blaNDM-5. *Microorganisms* 7(9).

van Belkum A, Durand G, Peyret M, Chatellier S, Zambardi G, Schrenzel J, Shortridge D, Engelhardt A & Dunne WM (2013) Rapid Clinical Bacteriology and Its Future Impact. *Annals of Laboratory Medicine* 33(1), pp. 14-27.

Vandamme, P., Holmes, B., Vancanneyt, M., Coenye, T., Hoste, B., Coopman, R., … Govan, J.R.W. (1997). Occurrence of Multiple Genomovars of *Burkholderia cepacia* in Cystic Fibrosis Patients and Proposal of *Burkholderia multivorans sp. nov. International Journal of Systematic Bacteriology* 47(4), pp. 1188–1200.

Vanlaere, E., Baldwin, A., Gevers, D., Henry, D., De Brandt, E., LiPuma, J.J., … Vandamme, P. (2009). Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans sp. nov*. and *Burkholderia lata sp. nov*. *International Journal of Systematic and Evolutionary Microbiology* 59(Pt 1), pp. 102-11.

Varvaresou, A., Papageorgiou, S., Tsirivas, E., Protopapa, E., Kintziou, H., Kefala, V. and Demetzos, C. (2009). Self-preserving cosmetics. *International Journal of Cosmetic Science* 31(3), pp. 163-175.

Vincze, S., Al Dahouk, S. and Dieckmann, R. (2019). Microbiological safety of non-food products: What can we learn from the RAPEX database? *International Journal of Environmental Research and Public Health* 16(9).

Vives-Flórez, M. and Garnica, D. (2006). Comparison of virulence between clinical and environmental *Pseudomonas aeruginosa* isolates. *International Microbiology* 9(4), pp. 247-252.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., … Earl, A.M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9(11), p. e112963.

Wallner, A., King, E., Ngonkeu, E.L.M., Moulin, L. and Béna, G. (2019). Genomic analyses of *Burkholderia cenocepacia* reveal multiple species with differential host-Adaptation to plants and humans. *BMC Genomics* 20(1), p. 803.

Wang, F., Yang, Q., Kase, J. a, Meng, J., Clotilde, L.M., Lin, A. and Ge, B. (2013). Current Trends in Detecting non-O157 Shiga Toxin-Producing *Escherichia coli* in Food. *Foodborne Pathogens and Disease*, 10(8), pp. 665–677.

Warner, D.M., Yang, Q., Duval, V., Chen, M., Xu, Y. and Levy, S.B. (2013). Involvement MarR and YedS in carbapenem resistance in a clinical isolate of *Escherichia coli* from China. *Antimicrobial Agents and Chemotherapy* 57(4), pp. 1935–1937.

Weiser, R., Donoghue, D., Weightman, A. and Mahenthiralingam, E. (2014). Evaluation of five selective media for the detection of *Pseudomonas aeruginosa* using a strain panel from clinical, environmental and industrial sources. *Journal of Microbiological Methods* 99(1), pp. 8-14.

Weiser, R., Green, A.E., Bull, M.J., Cunningham-Oakes, E., Jolley, K.A., Maiden, M.C.J., … Mahenthiralingam, E. (2019). Not all *Pseudomonas aeruginosa* are equal: strains from industrial sources possess uniquely large multireplicon genomes. *Microbial Genomics* 5 (7).

White, T.J., Bruns, T., Lee, S. and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA Genes for phylogenetics. In book: PCR - Protocols and Applications - A Laboratory Manual. Academic Press, Cambridge.

WHO (2017). Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. Available at: https://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1. [Last accessed: 11 April 2020]

Wick, R.R., Judd, L.M., Gorrie, C.L. and Holt, K.E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology* 13(6), p. e1005595.

Wick, R.R., Schultz, M.B., Zobel, J. and Holt, K.E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* 31 (20), pp. 3350-3352.

Wilkinson, K.M., Winstanley, T.G., Lanyon, C., Cummings, S.P., Raza, M.W. and Perry, J.D. (2012). Comparison of four chromogenic culture media for carbapenemase-producing *Enterobacteriaceae*. *Journal of Clinical Microbiology* 50 (9), pp. 3102-3104.

Williams, T.R. and Marco, M.L. (2014). Phyllosphere microbiota composition and microbial community transplantation on lettuce plants grown indoors. *mBio* 5(4).

Wood, D.E. and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15(3), p. R46.

Yoon, S.H., Ha, S.M, Lim, J., Kwon, S. and Chun, J. (2017). A Large-Scale Evaluation of Algorithms to Calculate Average Nucleotide Identity. *Antonie van Leeuwenhoek*, 112 (10), pp. 1409-1423.

Yuan, M., Chen, H., Zhu, X., Feng, J., Zhan, Z., Zhang, D., … Li, J. (2017). pSY153-MDR, a p12969-DIM-related mega plasmid carrying blaIMP-45 and *armA*, from clinical *Pseudomonas putida*. *Oncotarget* 8(40), pp. 68439–68447.

Zhao, X., Zhong, J., Wei, C., Lin, C.-W. and Ding, T. (2017). Current Perspectives on Viable but Non-culturable State in Foodborne Pathogens. *Frontiers in Microbiology* 8(1), p. 580.

Zheng, D., Wang, X., Wang, P., Peng, W., Ji, N. and Liang, R. (2016). Genome sequence of *Pseudomonas citronellolis* SJTE-3, an estrogen and polycyclic aromatic hydrocarbon-degrading bacterium. *Genome Announcements* 4(6), pp. e01373-e01416.