Metagenomic analyses of the scalp microbiome and insights from *Staphylococcus capitis* genomics

Thesis submitted in accordance with the requirements of

the University of Liverpool for the degree of Doctor in Philosophy by

Charlotte Elizabeth Chong

February 2021

UNIVERSITY OF LIVERPOOL

ABSTRACT

Human skin is a complex ecosystem that supports a diverse population of microorganisms. Variation in the physiological and topological conditions of the skin gives rise to diverse niches dependent upon the distinct local environment. Staphylococcus species are common residents of the human skin and in some locations form a major part of the microbiota, including the scalp. The human scalp boasts a unique microbial community primarily arising from the host physical conditions, including follicular density, sebum content, moisture, and pH. Several scalp studies from different human populations have revealed an association of dandruff with bacterial and fungal dysbiosis. The factors that select for the frequency of certain species in different niches are poorly understood and the mechanisms that structure each community remain to be determined.

Metagenomic sequencing of scalp samples from UK volunteers exhibiting high and low dandruff profiles were used to explore the role of the scalp microbiome in dandruff presentation. The study identified the presence of known key scalp-associated taxa, such as *Malassezia, Staphylococcus* and *Cutibacterium* spp.. No genera displayed altered frequency correlated with the presentation of dandruff. Bioinformatic analysis revealed the enrichment of KEGG orthologs linked to the bacterial cell wall and regulatory processes such as biofilm formation, nitrogen metabolism and quorum sensing on dandruff and dandruff healthy scalps.

Comparative genome analyses of *S. capitis* were undertaken to study the key scalp resident species and to expand current knowledge of the species. A highly recombinogenic population structure was revealed, with genomes including the presence of a range of described staphylococcal virulence factors, cell wall-associated proteins, mobile genetic elements, and two-component systems. Genomic differences between *S. capitis* subspecies were explored and reveal determinants associated exclusively with each. The subspecies *ureolyticus* could be distinguished from subspecies *capitis* by their gene differences in antimicrobial resistance genes, gene clusters linked to survival on the skin, and phage content.

Investigation of Pda, an *S. aureus* putative carbohydrate deacetylase revealed that the target of its activity is not peptidoglycan but may act together with described bacterial cell wall modification activities such as *oatA* and *tagO*. The regulation of *pda* appeared linked to genes implicated in biofilm formation, lipoteichoic acid synthesis and extracellular protease production, and the function enhances staphylococcal species competition.

The study utilises significant advances in DNA sequencing technologies to provide new insight into *S. capitis*, dandruff linked metagenomes and a staphylococcal putative surface modification enzyme. The data therein increases our understanding of staphylococcal mechanisms that enable them to colonise, compete and persist are essential for informing pathogenesis, epidemiology, and therapeutics.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr Malcolm Horsburgh for his continued expertise and advice. I will always be grateful for your constant support, for always pushing me (even when I complained), giving me endless opportunities to learn and for having confidence in me when my own was lacking. Thank you for providing me with such a nurturing environment to grow both personally and academically.

I would also like to thank my secondary supervisor Dr Alistair Darby as well as my many assessor's Dr Alan McCarthy, Dr Kate Baker, Dr Meriel Jones, Dr Jay Hinton and Dr Mark Caddick for their interest and advice during my time at the University of Liverpool. Additionally, to my supervisory team at Unilever thank you for providing me with many invaluable experiences. A special thanks to Dr Barry Murphy, Sally Grimshaw and Jo Hunt for all your expertise, continued support, and answers to my many emails. I am also fortunate to have been a part of an engaging and friendly BBSRC DTP cohort. I am very grateful for the opportunities they afforded me and the many friends I have made.

To the members of Lab H past and present, my PhD wouldn't have been the same without you! Thank you all for the chats, laughs at lunch, answers to my endless list of questions, trips to the AJ and many nights out! All of which have kept me sane! A special shout out to Dr Jo Moran for all your help when I was finding my feet and to Dr Becky Ji Bengtsson for answering all of my bioinformatics questions and sharing my obsession to eat and talk about food.

Lastly, I'd like to thank all my friends and family who have always cheered me on and been there for me throughout all of the ups and downs! A special thank you to Danielle, Ciara, Archie and Jordan, I couldn't have done the last 4 years without you! I am eternally thankful to my family, most especially my mum, without whom none of this would have been possible. Thank you to you, David, Nana and our angels for your unconditional love, encouragement, guidance and unwavering belief in me. I'm sure you're all as relieved as I am to be finishing!

DECLARATION OF AUTHORSHIP

I declare that except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of other, is indicated as such and acknowledged.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

ABBREVIATIONS

ACME	Arginine Catabolic Mobile Element
AD	Atopic Dermatitis
AFA	Antimicrobial Fatty Acids
AMP	Antimicrobial Peptide
AMR	Antimicrobial Resistance
ANI	Average Nucleotide Identity
ANOVA	Analysis of Variance
API	Analytical Profile Index
BAPS	Bayesian Analysis Population Structure
BGC	Biosynthetic Gene Cluster
BHI	Brain & Heart Infusion
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
BUSCO	Benchmarking Universal Single-Copy Orthologs
CGR	Centre of Genomic Research University of Liverpool
COG	Clusters of Orthologous Groups
CoNS	Coagulase Negative Staphylococci
CWA	Cell Wall Associated
DBG	De Bruijn Graph
DD	Dandruff scalps
DDH	DNA-DNA Hybridisation
DH	Healthy site on dandruff scalp
DNA	Deoxyribonucleic Acid
FDR	False Discovery Rate
GLM	Generalised Linear Model
GMM	Generalised Method of Moments
GOLD	Genome Online Database
GWAS	Genome Wide Association Study
HBD	Human Bacterial Defensins
HGT	Horizontal Gene Transfer
HH	Healthy Scalp
HPLC	High Performance Liquid Chromatography
HSD	Honest Significant Difference
IBD	Inflammatory Bowel Disease
ITS	Internal Transcribed Spaces region
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KO	KEGG Orthology
LCA	Lowest Common Ancestor
LCMS	Liquid Chromatography Mass Spectrometry
LDA	Linear Discriminant Analysis
lfcSE	log fold change Standard Error

LTA	Lipoteichoic Acid
MAG	Metagenome Assembled Genome
MGE	Mobile Genetic Element
MLEE	Multilocus Enzyme Electrophoresis
MLST	Multilocus Sequence Typing
MLVA	Multilocus Variable number tandem repeat
MS	Mass Spectrometry
MSCRAMM	Microbial Surface Component Recognising Adhesive Matrix Molecules
NCBI	National Centre for Biotechnology Information
NICU	Neonatal Intensive Care Unit
NMDS	Non-metric Multidimensional Scaling
NPRS	Nonribosomal Peptide Synthetases
OD	Optical Density
OG	Orthologous Group
OTU	Operational Taxonomic Unit
PCA	Principal Component Analysis
PCoA	Principal Coordinates Analysis
PERMANOVA	Permutational Multivariate Analysis of Variance
PFGE	Pulsed-Field Gel Electrophoresis
PG	Peptidoglycan
PIA	Polysaccharide Intercellular Adhesin
PKS	Polyketide Synthase
PSM	Phenol Soluble Modulin
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
SAG	Single-cell Assembled Genomes
SNP	Single Nucleotide Polymorphism
ST	Sequence Type
TCA	Trichloroacetic Acid
TCS	Two-Component System
TEM	Transmission Electron Microscopy
TNF	Tetra-Nucleotide Frequency
TWHS	Total Weighted Head Score
UV	Ultraviolet
WT	Wild Type
WTA	Wall Teichoic Acid
ZOI	Zone of Inhibition

CHAPTER 1

General introduction

1.1 The physiology of human skin

Human skin is comprised of two layers: an outer stratified squamous epithelium known as the epidermis, and a sub-epidermal layer called the dermis, which are connected via a basement membrane (Venus et al., 2010). The epidermis is the outermost layer of the skin and is divided into four sublayers: stratum corneum (horny layer), stratum granulosum (granular layer), stratum spinosum (spinous cell layer) and stratum basale (basal cell layer) (Fuchs and Raghavan, 2002)(Figure 1.1).

The epidermis is a terminally differentiated stratified epithelium made up of keratinocytes. Each layer of the epidermis is host to cells at different stages of keratinisation, whereby cells that form through mitosis in the basal later move up the strata and change shape and composition, eventually reaching the stratum corneum where terminal differentiation of cells produces corneocytes that are ultimately shed during the process of desquamation (Kolarsick et al., 2011, Blanpain and Fuchs, 2009).

Keratinocytes synthesise keratin, a fibrous filamentous protein arranged in an alpha-helical coil pattern linked by disulphide bonds between adjacent cysteine amino acids (Venus et al., 2010). Corneocytes are composed of keratin filaments surrounded by a cornified envelope (Harding et al., 2000) and are lost from the stratum corneum following the breakdown of corneodesmosomes. Corneodesmosomes are modified desmosomal structures that anchor corneocytes to one another and are proteolytically degraded in the uppermost

layers of the cornified layer to allow desquamation, which counterbalances regeneration to maintain epidermal homeostasis (Candi et al., 2005)(Figure 1.1).

The dermis is tightly connected to the epidermis by a basement membrane and internally by subcutaneous fat (Venus et al., 2010). The dermis is comprised of connective tissue, nerve endings and multiple appendages and accessory structures such as, hair follicles, sweat glands, sebaceous glands and apocrine glands (Kolarsick et al., 2011). The appendages found in this layer of the skin differ depending upon location, for example sebaceous glands are ubiquitous throughout the skin and secrete sebum, a lipid-rich oily fluid that protects skin against water loss, maintaining moisture (Rittié, 2016) (Figure 1.1).

The basement membrane connecting the epidermis and dermis is composed of extracellular matrix proteins, including collagen IV, fibronectin and laminin 5. The extracellular matrix is an important component in the skin structure and contributes to vital processes, such as wound healing. Skin repair is the result of interactions involving soluble factors, blood elements, extracellular matrix components and cells (Rousselle et al., 2019).

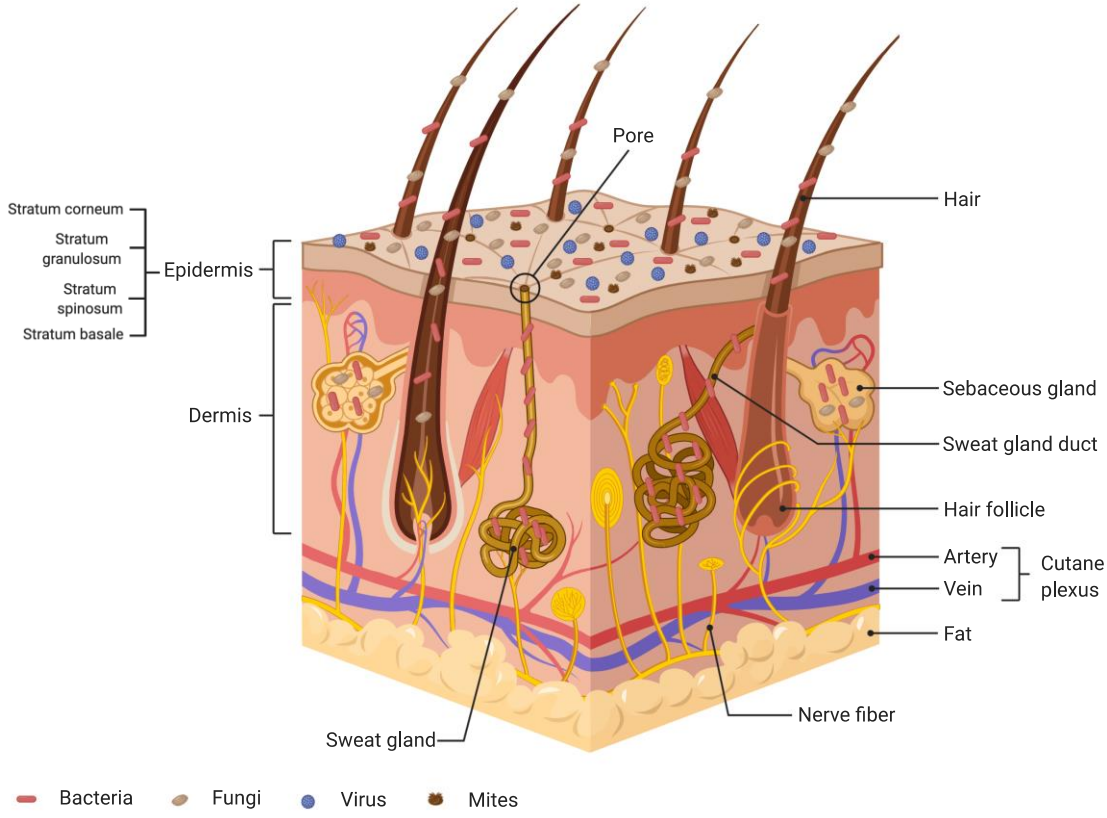


Figure 1.1 Schematic of skin histology viewed in cross section with microorganisms and skin appendages. The human epidermis has a stratified epithelial structure. Each layer of the epidermis is host to cells at different stages of keratinisation, becoming terminally differentiated corneocytes. The dermis is comprised of connective tissue, nerve endings and multiple appendages and accessory structures such as hair follicles, sweat glands, sebaceous glands and apocrine glands. Microorganisms (viruses, bacteria and fungi) and mites cover the surface of the skin and reside deep in skin associated appendages. Figure adapted from Grice & Serge (2001) and "Anatomy of the skin", by BioRender.com (2020). Retrieved from https://app.biorender.com/biorender-templates.

1.2 The human skin microbiome

The skin is both the largest organ of the human body and the primary interface with the external environment (Prescott et al., 2017). Skin is vital to human health, providing protection against dehydration and harmful external physical, chemical and biological agents (Grice and Segre, 2011). Whilst skin forms the first line of defence against pathogens, it is also colonised by a diverse range of microorganisms, including bacteria, fungi and viruses (Oh et al., 2016). The physical and chemical features of skin select for specific species of microorganisms that are adapted to this niche. In general, the skin is cool, acidic and desiccated, although unique niches are determined by skin thickness, folds and the density of hair follicles and glands (Grice and Segre, 2011) (Figure 1.1).

The skin microbiome has been studied in multiple DNA sequencing surveys of healthy adults to understand normal variation and to compare disease states (Oh et al., 2014, Oh et al., 2016, Grice and Segre, 2011, Costello et al., 2009). Past culture-dependent and -independent studies have presented *Staphylococcus epidermidis* and other coagulase-negative staphylococci as being among a few major bacterial inhabitants of the skin microbiome. Other dominant bacterial phyla include Actinobacteria (genera *Corynebacterium*, *Cutibacterium* and *Micrococcus*), Proteobacteria (genus *Pseudomonas*) and Bacteroidetes (genus *Prevotella*) (Grice and Segre, 2011, McGinley et al., 1975).

Eukaryotic microorganisms are also isolated from skin. The fungal genus *Malassezia* is remarkably dominant and commonly isolated from the skin, it was first reported in scientific literature in 1847 (Findley et al., 2013, McGinley et al., 1975). In addition to *Malassezia,* several fungal phyla are reported as skin

commensals, including Basidiomycetes (genera *Rhodotorula* and *Cryptococcus*) and Ascomycetes (genera *Aspergillus*, *Debaromyces* and *Penicillium*) (Findley et al., 2013, Roth and James, 1988).

Variation in the physiological and topological conditions of skin gives rise to diverse niches, dependent upon the distinct local environmental conditions. Changes in the relative abundance of different taxa are associated with moist, dry and sebaceous microenvironments, with sebaceous areas presenting the lowest microbial diversity (Grice and Segre, 2011). The microenvironments of the skin are influenced by appendages such as sweat glands, hair follicles and sebaceous glands.

Sebaceous sites are characterised as oily and include the face, chest and back. These sites often have high follicular density and produce a high rate of sebum production and secretion due to the presence of many pilosebaceous units and sweat glands (Grice and Segre, 2011, Fitz-Gibbon et al., 2013). Pilosebaceous units are comprised of sebaceous glands connected to a hair follicle and secrete sebum. Sebum is a lipid-rich secretion that provides a hydrophobic coating that lubricates the skin and hair shaft to provide an antimicrobial shield (Byrd et al., 2018, Grice and Segre, 2011). Sebaceous sites are dominated by lipophilic *Cutibacterium* species, which are abundant skin commensal bacteria. *C. acnes* is a facultative anaerobe able to survive in sebaceous glands by using proteolytic enzymes that promote both adherence and pathogenesis (Grice and Segre, 2011). Certain *C. acnes* proteases release arginine from skin proteins that are used as a carbon source (Holland et al., 1979). Other key enzymes include lipases that degrade triglyceride lipids of sebum, releasing free fatty acids, promoting

bacterial adherence (Gribbon et al., 1993, Brüggemann et al., 2004). Other key bacterial colonisers of pilosebaceous units include coagulase-negative staphylococci, more specifically *Staphylococcus capitis* (Byrd et al., 2018, Grimshaw et al., 2019).

Lipids of sebum secreted by pilosebaceous units together with those on the stratum corneum are metabolised by auxotrophic *Malassezia* and *Corynebacterium* species, which are unable to produce their own lipids, to allow colonisation and survival. For example, *Corynebacterium* spp. utilise skin lipid compounds to generate corynemycolic acids that coat their cell surface to aid pathogenicity and survival (Burkovski, 2018, Scharschmidt and Fischbach, 2013). *Malassezia* spp. genomes reveal the enrichment of lipase genes that allow these fungi to dominate the skin mycobiome (Wu et al., 2015).

Further microbes routinely isolated from healthy skin include the *Demodex* mite. First described in the hair follicles of humans in 1841 by Henle and Berger the mites were later classified into two species, *Demodex folliculorum* and *Demodex brevis* (Aylesworth and Vance, 1982, Elston and Elston, 2014). These microscopic arthropods reside within the pilosebaceous unit and are acquired shortly after birth then become more abundant in sebaceous glands during puberty while feeding on sebum (Elston and Elston, 2014).

Staphylococcus and *Corynebacterium* species were revealed by microbiome sequencing to be abundant in moist areas, including the bends of the elbows, feet, nares and groin (Kong et al., 2012, Kong and Segre, 2012). These regions of the skin are partially occluded, which results in a higher temperature and humidity which encourages growth of bacteria that thrive in these conditions, such as

Gram-negative bacilli, coryneforms and staphylococci (Kong et al., 2012, Kong and Segre, 2012).

From microbiome sequencing data, the most diverse skin sites were dry skin areas such as the volar forearm, buttock and palm, with the four main phyla associated with the skin microbiome represented, including Actinobacteria, Proteobacteria, Firmicutes and Bacteriodetes. Interestingly these sites harbour great phylogenetic diversity, including multiple Gram-negative species. It has been hypothesised that this high level of diversity results from the frequent exposure and contact of palms and forearms with microbially colonised material (Grice and Segre, 2011, Grice et al., 2008, Costello et al., 2009).

The microbiome populations between individuals varies greatly. Host factors such as age, sex and geographical location are known to contribute to variability in the microbiome of the skin (Grice and Segre, 2011). In particular, age has a significant effect on the microflora of the skin with certain skin conditions being more common across particular age ranges. For example, *Staphylococcus*-associated atopic dermatitis is common in prepubescent children (Merriman et al., 2016). Adults are thought to maintain the composition of their skin microbial communities (Byrd et al., 2018, Oh et al., 2016).

In utero, foetal skin is sterile, but colonisation occurs in newborn babies immediately after birth and is dependent on delivery mode. Vaginally born neonates are colonised by the bacteria that reside in the vagina, whereas neonates born via caesarean section acquire skin associated microorganisms (Mueller et al., 2015, Dominguez-Bello et al., 2010).

The stability of the human skin microbiome as we age differs compared with those of other areas, such as the gut, which stabilises around 3 years of age. Microflora of the skin changes during puberty, due to hormone levels causing an increase in sebum production that supports the growth of lipophilic bacteria, such as *Cutibacterium* spp., *Corynebacterium* spp. and the fungal *Malassezia spp.* Microbial differences between males and females also occur at this time due to anatomical differences in the cutaneous environment arising from sweat, sebum and hormone production. Other studies have linked ageing skin with greater microbial diversity, however little is known about the ageing skin microbiome and how it potentially predisposes the elderly to skin infections (Byrd et al., 2018).

Environmental factors specific to individuals, such as occupation, clothing choice and cosmetic use, modulates colonisation of the skin. Each factor can cause variation in the skin microbiome by altering the conditions of the skin barrier. Geographical location also effects the skin microbiome, with high temperatures and humidity being associated with increased bacterial colonisation on the back, axillae and feet (McBride et al., 1977, Grice and Segre, 2011). Geographic variability is also linked with ultraviolet (UV) light exposure, which is often used as a bactericidal treatment affecting the skin microbiome. Ultraviolet radiation modulates the host immune system and triggers the production and release of antimicrobial peptides that can affect the skin microflora; however, the skin microbiome is proposed to contribute towards diminishing UV-induced immune responses (Patra et al., 2016, Patra et al., 2019).

The functional potential of the human microbiome is beginning to be understood through metagenomic sequencing efforts, although this can often be technically difficult due to the low biomass of organisms on skin (Byrd et al., 2018). For example, gene expression profiles of *C. acnes* differ between individuals with and without acne in relation to vitamin B_{12} (Kang et al., 2015a).

The importance of the human skin microbiome to both health and disease is increasingly being recognised and has undergone extensive characterisation since the advent of high-throughput sequencing. Advances linking functionality with particular organisms residing on the skin could be used to inform the development of therapeutic techniques to counteract pathogens and ameliorate dysbiosis associated with skin conditions (Byrd et al., 2018).

1.3 The human skin microbiome and disease

Many common skin disorders are postulated to be associated with dysbiosis of the microflora present on skin. Dysbiosis can be driven by commensal bacteria that are ordinarily beneficial to their hosts but become pathogenic (Iebba et al., 2016).

1.3.1 Atopic Dermatitis (AD)

AD or eczema is a chronic, relapsing inflammatory disease of the skin, causing patients with AD to suffer from intensely itchy and inflamed skin lesions (Byrd et al., 2017). The disease has a high prevalence affecting around 10-30% of children in industrialised countries (Eyerich et al., 2015). AD typically manifests in the antecubital fossa and the popliteal fossa, areas which are similarly colonised (Grice et al., 2009).

Atopic dermatitis is treated with emollients that promote barrier integrity, with steroids, and in cases where there is disease persistence antimicrobial approaches are taken, including topical and systemic antibiotics and dilute bleach baths (Huang et al., 2009). Treatments that decrease *S. aureus* generally correlate with an improvement in symptoms (Kong et al., 2012).

Many studies have shown that more than 90% of AD patients are colonised with *S. aureus* on both lesional and non-lesional skin compared with <5% of healthy individuals (Leyden et al., 1974, Hanifin and Rogge, 1977). Other staphylococci such as *S. epidermidis* are also present in patients with mild disease flares (Byrd et al., 2017).

Proliferation of *S. aureus* during AD flare ups is encouraged by reduced competition within the skin microbiota and favourable conditions, such as a higher pH (Proksch et al., 2008). *S. aureus* also expresses several virulence factors that have roles in AD pathogenesis. *S. aureus* cell wall, associated proteins, including ClfB and fibronectin-binding proteins FnBP A and B are expressed on the cell surface and promote strong adhesion to fibronectin and cornified envelope proteins loricrin and cytokeratin 10 of AD skin (Foster et al., 2014, Mulcahy et al., 2012, Cho et al., 2001). *S. aureus* also expresses superantigens, cytolytic α- and δ- toxins, phenol-soluble modulins (PSMs), protein A and several proteases, and each have roles in AD pathogenesis by causing inflammation by immune dysregulation (Geoghegan et al., 2018). Superantigens bind to major histocompatibility class II molecules on the surface of antigen-presenting cells and T cell receptors resulting in overproduction of T cell cytokines that cause toxicity and generate an IgE immune response (Herz et al., 1998, Geoghegan et

al., 2018). Bacterial proteases act in concert with PSMs to protect *S. aureus* from antimicrobial peptides of AD skin, promoting bacterial penetration. These proteases act together with those expressed by the host to perturb skin barrier integrity (Syed et al., 2015). Additionally, toxins produced by *S. aureus* isolates from AD lesional sites can cause the degranulation of the stratrum granulosum to produce a further loss of barrier function, inflammation and increased penetration of irritants, allergens and bacteria (Nakamura et al., 2013, Berube and Wardenburg, 2013)

The pathogenesis of AD is multifactorial involving elements of barrier dysfunction, altered immune responses and environmental factors. Host susceptibility to AD was associated with over 30 gene loci, including genes important in cell envelope formation and linked to the immune system (Palmer et al., 2006, Byrd et al., 2018). Genetic predispositions for disruption of the epidermal barrier of the skin have been found in AD patients. Genome-wide scans have identified polymorphisms in several genes involved in controlling skin barrier integrity (Boguniewicz and Leung, 2011). These gene polymorphisms are associated with loss of function in filaggrin, with up to 60% of Europeans affected by AD possessing the mutation (Palmer et al., 2006). Mutations causing a loss of function in filaggrin genes produce disruption of the stratum granulosum, influencing the differentiation and growth of the stratum corneum, changes lead to an increase in trans-epidermal water loss, pH alterations and dehydration, all of which promote bacterial colonisation of the skin (Irvine et al., 2011).

Other genetic changes are associated with altered skin barrier function. The imbalance of Th2 to Th1 cytokines observed in atopic dermatitis causes altered

cell mediated immune responses. High levels of Th2 cytokines cause alterations in the immune response that promote-IgE mediated sensitivity. Overproduction of IgE antibodies can result in asthma, AD and other allergic diseases in the early years of life (Furue et al., 2017). IgE together with IL-4 contribute to the release of antimicrobial compounds from mast cells through degranulation. Decreased degranulation caused by *S. aureus* δ-toxins enables its persistence (Nakamura et al., 2013). Dysregulation of the innate immune system therefore results in decreased production of antimicrobial peptides, reduced recruitment of neutrophils to the skin and epidermal barrier, and in turn loss of barrier function and increased penetration of irritants and allergens (Boothe et al., 2017).

The role of the environment must also be considered in the causation of AD. The use of harsh alkaline detergents and chemicals in skin care may unfavourably alter skin pH, which can in turn cause inflammation, bacterial colonisation and a change in enzyme activity (Elias, 2007, Törmä et al., 2008).

1.3.2 Acne vulagris

Acne vulgaris is a chronic inflammatory skin condition that affects up to 90% of teenagers (Lee et al., 2019). Acne is a disease of the pilosebaceous unit resulting from increase sebum production, altered follicular keratinisation and bacterial colonisation (Williams et al., 2012). The clinical features of acne include seborrhoea, inflammatory and non-inflammatory lesions and varying degrees of scarring on the skin where there is a high density of pilosebaceous units (face, neck, upper chest, shoulders and back) (Williams et al., 2012, Lee et al., 2019).

Sebaceous areas of the skin are prone to colonisation by lipophilic microorganisms such as *Cutibacterium* spp., *Malassezia* spp. and *Corynebacterium*

spp. (Byrd et al., 2018, Grice et al., 2009, Costello et al., 2009). Sebaceous areas tend to be low in skin microbiome diversity and relatively stable compared to moist and dry areas, however a puberty-associated shift in the microflora leads to an enrichment of Actinobacteria (Oh et al., 2016, Capone et al., 2011).

C. acnes is a major commensal of the normal skin microflora and is ubiquitous in the sebaceous areas of healthy skin but it also contributes to acne pathogenesis (Fitz-Gibbon et al., 2013). Studies have associated *C. acnes* over-proliferation with the disease (Omer et al., 2017, Miura et al., 2010), although other metagenomic studies have reported the relative abundance of *C. acnes* as similar in healthy individuals and acne patients. Instead, studies linked a loss of diversity in *C. acnes* phylotypes to the cause of acne (Fitz-Gibbon et al., 2013, Spittaels et al., 2020, Pécastaings et al., 2018). The loss of *C. acnes* phylotype diversity was shown to trigger the innate immune system and inflammation in acne (Dréno et al., 2020).

C. acnes types IA and IB have been consistently associated with acne across studies and were found to produce significantly higher levels of porphyrin pro-inflammatory metabolites, which generate reactive oxygen species and induce inflammation in keratinocytes (Johnson et al., 2016). *C. acnes* has been shown to cause an increase in the proliferation of keratinocytes and dysregulation of the expression of epidermal markers, including filaggrin and desmocollin 1 (DSC1) (Choi et al., 2018). These specific phylotypes were also found to induce production of human antimicrobial peptide (AMP) and β-defensin 2 (hBD2) (Borrel et al., 2019).

Although *C. acnes* is commonly thought to be the main cause of acne, recent studies proposed that acne could result from an imbalance between *C. acnes* and

S. epidermidis frequency. *S. epidermidis* controls growth of *C. acnes* by competing for niche colonisation and by favouring the fermentation of glycerol produced by skin it releases succinic acid. Consequent inflammatory effects of *S. epidermidis* suppress *C. acnes*-induced IL-6 and tumour necrosis factor TNF-alpha production by keratinocytes (Claudel et al., 2019). Through competition, *C. acnes* can inhibit *S. epidermidis* through the secretion of propionic acid as a result of hydrolysing sebum triglycerides that maintain the acidic pH of skin unfavourable to staphylococci (Grice et al., 2009, Wang et al., 2014). In addition, *Malassezia* were considered to have a potential role in acne pathophysiology due to, abundance in pilosebaceous areas, lipase activity and pro-inflammatory potential (Lee et al., 2019).

The gut microbiome was linked to acne. Studies have revealed correlation between increased abundance of Proteobacteria and an increased ratio of Bacteroidetes to Firmicutes in patients with acne. These alterations in the gut microbiome were reported to be the enterotype of the Western diet (Deng et al., 2018). The connection between gut microbiota and acne development could be related to the fact that bacterial dysbiosis in the gut causes increased intestinal permeability and release of inflammatory mediators into circulation (Dréno et al., 2020).

1.3.3 Psoriasis

Dry areas of the skin are the most diversely populated by resident microflora compared to sebaceous and moist sites, and with less of a selective force no individual microbe dominates (Grice et al., 2009). Greater amounts of

Proteobacteria and Bacteriodetes are present in drier areas compared to moist and sebaceous sites (Grice, 2014).

Psoriasis vulgaris is a chronic inflammatory skin condition commonly associated with drier skin sites, such as the elbows, knees and trunk (Grice, 2014). It is currently thought to affect approximately 100 million individuals worldwide and is characterised by the hyperproliferation of keratinocytes and inflammation (Visser et al., 2019).

It has been hypothesised that psoriasis results from failure in immune tolerance to microbiota, similar to Crohn's disease (Fry et al., 2013), with the disease causing not only localised inflammation of the skin but also systemic inflammation (Visser et al., 2019). The pathogenesis of psoriasis vulgaris involves abnormal crosstalk between the innate and adaptive immune systems, including the dysregulation of the IL-23/IL-17 axis and tumour necrosis factor-alpha causing hyper proliferation of keratinocytes as well as the activation and recruitment of a number of other immune factors (Boehncke and Schön, 2015, Ogawa et al., 2018). In addition to these immune aspects, gut microbiome dysbiosis also co-presents with psoriasis through an association with inflammatory bowel disease (IBD) (Cohen et al., 2009, Whitlock et al., 2018).

Dysbiosis of the skin microbiome was implicated in the pathogenesis of psoriasis. Studies comparing the composition of the skin microflora found on lesional skin of psoriasis patients and non-lesional skin are mostly based on 16S rDNA sequencing. Two earlier studies found that *S. epidermidis* and *C. acnes* were underrepresented in psoriatic lesions compared to healthy skin, with a relative enrichment of *S. aureus* associated on both lesional and non-lesional psoriatic

skin (Gao et al., 2008, Fahlén et al., 2012). Although the opposite was found in a larger study, where the psoriatic plaque microbiota were determined to contain an increased combined relative abundance of *Cutibacterium, Staphylococcus, Corynebacterium* and *Streptococcus* (Alekseyenko et al., 2013).

Although such studies provide an insight into the microbiome that is associated with psoriasis, there are discrepancies in their results: they are based on a small number of patients; each is based on 16S rDNA sequencing that does not account for unculturable and eukaryotic organisms. Consequently, it is unclear whether these microbiome changes are a cause or consequence of alteration of the skin barrier and the ecosystem. Longitudinal studies and more extensive sequencing of the psoriatic microbiome will be needed to provide insight into the role of microbiota during triggering, propagation and maintenance of psoriatic plaques.

The first high-resolution shotgun metagenomics study to characterise the microbiome of psoriatic and unaffected skin was carried out by Tett, *et al.* (2017). This study revealed that psoriatic ear sites have a decreased diversity and are associated with an increase in *Staphylococcus* but overall, the microbiome at all sites did not reveal discriminative features at the species level (Tett et al., 2017). These results may highlight that strain-level variations could be key determinants of the psoriatic microbiome and the study of the role of the microbiome in psoriasis requires further high-resolution study, with a larger subject size.

1.4 Staphylococci and the skin

Members of the genus *Staphylococcus* are among the most abundant skin colonising bacteria and important causes of a wide spectrum of diseases. The

dominant *Staphylococcus* species on the skin is *S. epidermidis*, which is considered to be a universal coloniser (Byrd et al., 2018). Other coagulase-negative staphylococci (CoNS) make up a substantial proportion of the skin microbiome, including *S. homins, S. haemolyticus, S. capitis, S. warneri, S. simulans* and *S. cohnii* (Coates et al., 2014). While most CoNS are considered commensal, they are now emerging as common causes of nosocomial infection (Heilmann et al., 2019). Whereas coagulase-positive *S. aureus,* although not frequently found as part of the natural skin flora, is still an important host-associated staphylococcal species due to its role in disease (Otto, 2010). Across the species, staphylococcal diseases are diverse and range from common skin ailments, to post-surgical infections and systemic infection (Coates et al., 2014, Li et al., 2014).

Skin colonising bacteria require the ability to adhere to the flattened, cornified keratinocytes (Figure 1.1). Therefore, staphylococcal molecular determinants of skin colonisation include surface polymers and proteins that promote adhesion and aggregation and a wide variety of mechanisms which aid in the evasion of the host immune defences (Otto, 2010).

Staphylococci are decorated with a variable number of proteins that are covalently linked to the cell wall peptidoglycan. Many of these proteins are classified as a microbial surface component recognising adhesive matrix molecules (MSCRAMMs) (Foster et al., 2014). Interspecies variation between adhesion proteins within staphylococci may explain their different host and target cell range. *S. aureus* has a larger repertoire of encoded MSCRAMMs compared to CoNS. Nevertheless, several CoNS such as *S. epidermidis*, has a similar repertoire to *S. aureus* for cell wall associated proteins that promote

adhesion to components of the extracellular matrix including collagen, fibrinogen and fibronectin (Foster, 2020). Following attachment to surfaces, bacteria multiply and form biofilms. Intercellular adhesion is also promoted by surface proteins and adhesins (PIA), again aiding persistence and immune evasion (Vuong et al., 2004). MSCRAMMs and cell wall-anchored teichoic acids also contribute to antimicrobial fatty acids resistance via alterations in the capsule and peptidoglycan of staphylococci to reduce hydrophobicity and in turn AFA interaction (Hachem et al., 2003). AFAs together with lactic acid and amino acids in sweat contribute to maintaining the acidic pH of the skin, in an effort to protect the skin against invasion by pathogenic bacteria (Parsons et al., 2012, Kohler et al., 2009, Kenny et al., 2009, Clarke et al., 2007).

In addition to promoting adhesion to biotic and abiotic surfaces, staphylococci must also maintain colonisation in the face of host antimicrobial defence mechanisms, such as antimicrobial peptides (AMPs) (Otto, 2010). Epithelial tissues produce cationic AMPs as part of the host's innate immune system: defensins, cathelicidins and dermcidin (Coates et al., 2014). Defensins and cathelicidins are amphipathic peptides that are associated with human sweat glands and thought to bind directly to anionic microbial cell walls and membranes. Human cathelicidins and defensins such as LL-37 and Human β-defensins (HBDs) 1 and 2 have multiple functions including wound healing and innate and adaptive immunity (Kościuczuk et al., 2012, Yamasaki and Gallo, 2008). Dermcidin is an AMP, constitutively produced in eccrine sweat glands and transported to the epidermal surface, that possesses staphylocidal activity (Rieg et al., 2014). Amphipathic peptides bind to the negatively charged cell membrane of bacteria and in response to these, staphylococci alter their cell surface charge

to repel the peptides by cell surface electrostatic interactions. Such processes are regulated by several two component systems (TCS) and global regulators (Yang et al., 2012, Foster, 2020, Joo and Otto, 2015).

Staphylococci live within complex microbial communities in different niches of the body, so they must also compete for resources such as nutrients and space. Competitive exclusion within skin niches occurs for example between *S. epidermidis* and *S. aureus*, whereby *S. epidermidis* excludes *S. aureus* by stimulating the production of AMPs by the host innate immune system, thus inhibiting *S. aureus* colonisation of the epithelium (Iwase et al., 2010, Lai et al., 2010). In addition to this, staphylococci secrete proteins, such as phenol-soluble modulins (PSMs) that are inhibitory to a number of species that reside on the skin, such as *C. acnes* (O'Neill et al., 2020, Kumar et al., 2017).

1.5 Molecular analysis of microbial community composition

The human body is host to a vast diversity of microorganisms, with the number of microbial cells outnumbering human cells by a factor of 10 (Grice, 2014). Resident microbes are increasingly viewed as an integral part of human health, interacting directly with host tissues and immune networks (Prescott et al., 2017). The human microbiome is a dynamic and complex entity, contributing to genetic diversity, metabolic processes, and health and disease (Grice, 2014). The microbiome is a prime target for manipulation and investigation of health and disease, and with skin it is more accessible for both testing and intervention (Grice, 2014). Multiple studies by microbiologists and dermatologists over the years have employed both culture-based and DNA methodology to uncover the

species composition of the skin and scalp microbiome and how microbial diversity may contribute to dermatological conditions (Dréno et al., 2016).

Microbiome research was once limited to culture-based techniques including colony morphology, Gram staining, biochemical characteristics, i.e. coagulase tests, sugar fermentation and antibiotic resistance profiles (Findley and Grice, 2014). Historically, these techniques were the standard, but they can impart bias through sampling and simplification (Findley et al., 2013). It is now evident they do not account for microbe-microbe interactions and only select for the unfastidious organisms that are able to flourish in the standard nutritional and physiological conditions provided in a laboratory (Grice and Segre, 2011).

The advent of high-throughput sequencing technologies and the development of bioinformatics tools and reference databases has accelerated all aspects of biological research, but in particular it has supported the investigation of human-associated microbial communities (Proctor et al., 2019). DNA sequencing now allows us to characterise microbial communities more accurately and with less bias.

Metagenomics refers to the study of all genetic material recovered from a community within an environmental sample. Early metagenomic studies focused on bacterial and fungal identification marker gene sequencing. Small subunit ribosomal RNA (16S rRNA) sequences for bacterial classification are well accepted. Hypervariable regions within this gene allow for species-specific identification (Lane et al., 1985). Similarly, for fungal identification, sequencing the internal transcribed spaces region (ITS) that lies between the fungal 18S and 5.8S rRNA genes, can give species resolution (Findley et al., 2013). Advances in

high-throughput sequencing approaches have expanded metagenomic studies to whole genome sequencing. Metagenomic WGS allows us to both profile selected organisms through marker gene sequencing and taxonomic composition to interrogate the functional potential of microbial communities and recover whole genome sequences with shotgun metagenomics (Quince et al., 2017, Thomas et al., 2012).

Microbial genomes currently deposited in public repositories are heavily biased towards organisms with medical importance and those that can be cultured in laboratory conditions but ultimately do not represent the naturally occurring diversity in the environment (Solden et al., 2016). In recent times, scientists have addressed missing information from uncultured organisms through advances in genomic sequencing technologies, with microbial genomes being directly sequenced from the environment using metagenomics and single cell genomics (Parks et al., 2017, Solden et al., 2016). Consequently, the number of single-cell assembled genomes (SAGs) and metagenome assembled genomes (MAGs) registered in the Genome Online Database (GOLD) has risen significantly (Reddy et al., 2015).

Although short-read NGS technologies have greatly improved our understanding of microbial function and composition of many ecosystems, short-read sequencing is limited by fragmented assemblies, which impedes the analysis of genomic content and the detection of certain microorganisms such as viruses (Yahara et al., 2021). Long reads are crucial for deciphering genomic regions that remain inaccessible to short-read sequences due to repeat sequences (Bharti and Grimm, 2021).

The use of third generation, long-read sequencing derived from Oxford Nanopore Technologies and Pacific Biosciences is currently the only method which allows strain resolution via reconstructing *de novo*, highly accurate, closed, MAGs from assembled and binned long sequence reads without mapping to reference genomes (Byrd et al., 2018, Overholt et al., 2020, Tsai et al., 2016). Moreover, long-read sequencing also offers full length gene profiling of ITS and 16s rRNA regions (Earl et al., 2018). Additionally, the portability of the Oxford Nanopore MinION DNA sequencer has resolved several issues regarding sample collection and preservation in clinical and field settings (Jain et al., 2018b).

Correctly describing genomes present in different environments, although improves with the resolution given by long-read sequencing, is not without inherent error. Therefore, recent studies have suggested a hybrid approach of long and short read sequencing resolves challenges in metagenomics using more accurate reads (Bharti and Grimm, 2021, Wongsurawat et al., 2019, Slaby et al., 2017). Current sequencing and computational developments will continue to improve the preciseness of genomic findings, with additional metatranscriptomic and metaproteomic technologies, helping to interrogate biological information from microbiomes (Bharti and Grimm, 2021).

1.6 Bioinformatics methods in metagenomic analysis

The analysis of high throughput metagenomic datasets presents a considerable computational challenge due to the complexity of the data. Metagenomic sequence data contains a complex mixture of numerous reads from a wide variety of microorganisms that vary in abundance which presents unique challenges for sequence assembly and analysis (Quince et al., 2017). However, the recent rise in

metagenomic sequencing has led to the development of a suite of bioinformatic approaches and tools that can cope with large data volumes and decipher datasets to answer the fundamental questions regarding microbial community structure and function.

A typical shotgun metagenomics study comprises of, but it not limited to, four steps after initial study design: (i) the collection, processing and sequencing of the samples; (ii) pre-processing of the sequencing reads; (iii) sequence analysis to profile taxonomic, functional and genomic features of the microbiome; and (iv) statistical and biological post-processing analysis and validation (Figure 1.2)

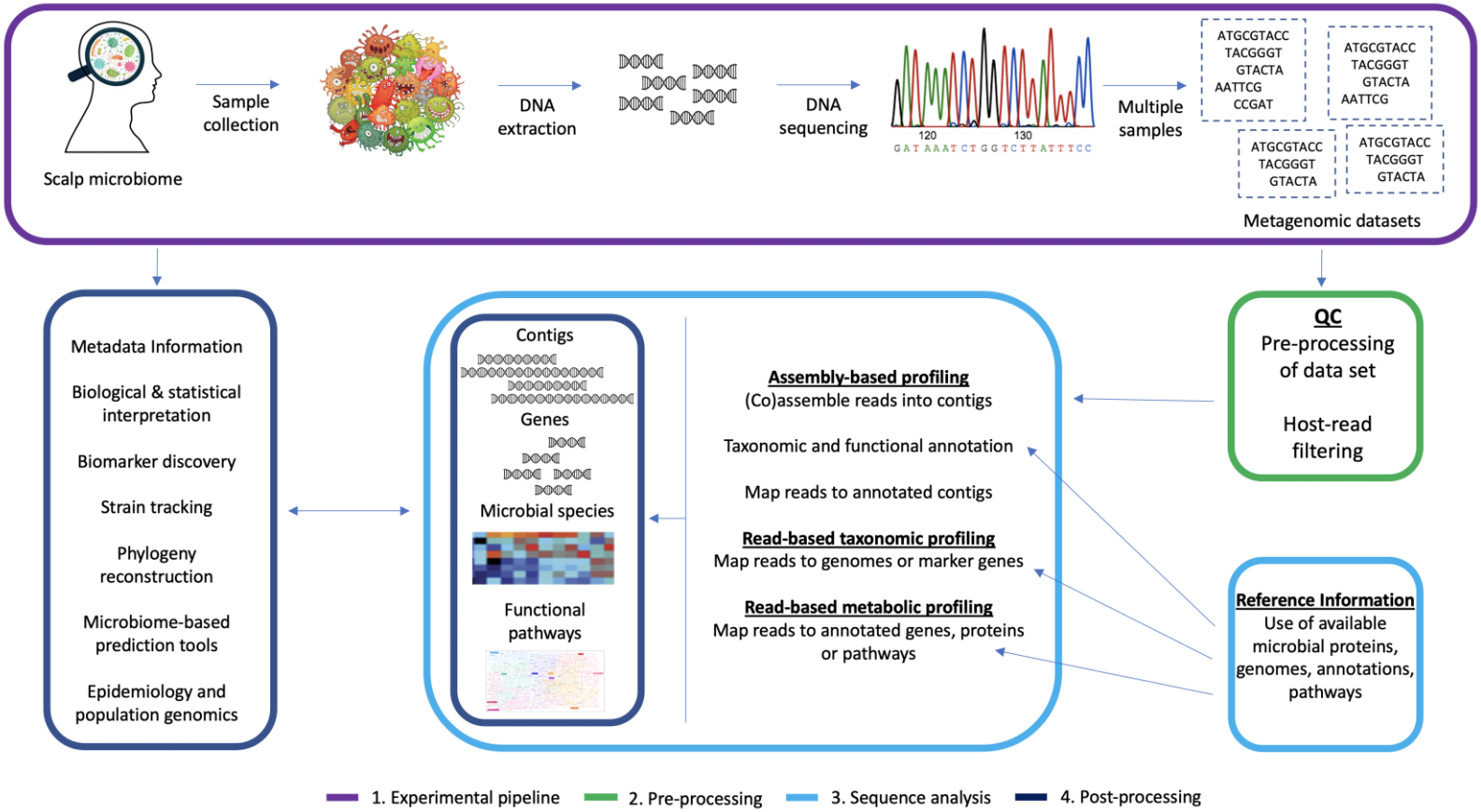


Figure 1.2 Summary of a shotgun metagenomics workflow. 1: Study design and experimental protocol, 2: Pre-processing- computational quality control steps such as quality control, removal of sequencing duplicates and adaptors and non-target DNA sequences, 3: Sequence analysis- revealing the taxonomic and genetic composition of metagenomic samples using a combination of read based and assembly-based approaches, 4: post-processing- various multivariate statistical techniques can be used to interpret the data, to allow conclusions to be made. Figure adapted from Quince, et al (2017).

1.6.1 Pre-processing of metagenomic data

Pre-processing and quality control filtering prior to downstream analysis is imperative to generate accurate conclusions from metagenomic datasets. Errors can originate from a variety of sources including the sequencing process and environmental contamination. Computational quality control steps minimise fundamental sequence biases or artifacts such as the removal of sequence adaptors, quality trimming and removal of sequencing duplicates. These quality control steps are performed using tools, such as FASTQC (Andrews, 2010), PRINSEQ (Schmieder and Edwards, 2011), BBDuK (Bushnell) and Trimmomatic (Bolger et al., 2014).

Sample-associated contamination from soil particles, faecal matter or human tissue is often filtered prior to DNA sequencing, however not all contaminants can be removed without imposing biases upon the microbial cells within the sample. Therefore, there is often a considerable portion of the sample genetic material that is made up of foreign or non-target DNA sequences. This genetic contamination is filtered prior to downstream analyses, normally using alignment-based tools (Quince et al., 2017, Breitwieser et al., 2019).

1.6.2 Assembly-free metagenomic profiling

Taxonomic profiling of metagenomes identifies which microbial species are present in a metagenome and estimates their abundance. Profiling can be carried out without assembly using publicly available reference genomes. This approach can mitigate assembly problems, speed up computation time and enable profiling of low abundance organisms that cannot be assembled *de novo*. The main limitation is that previously uncharacterised microbes are difficult to profile,

although the number of reference genomes available is increasing rapidly, including those from metagenomic assembly and single cell sequencing approaches, and new cultivation methods (Rinke et al., 2013, Stewart, 2012). The diversity of reference genomes available for some sample types, such as the human gastrointestinal tract, oral cavity and the skin are now extensive enough for assembly-free profiling whilst the analysis of some more diverse environments such as soil and oceans may be hindered by the lack of representative reference genomes (Human Microbiome Jumpstart Reference Strains Consortium, 2010).

Assembly-free taxonomic profilers commonly follow two main approaches: utilising information available in reference genomes as whole or selective strain representative gene markers from available reference sequences. Mapping of reads to genomes can result in profiles with many false positives, although usually tools couple this with lowest common ancestor (LCA) strategies or Markov models. This approach does not always improve the computational run-times compared to assembly-based methods, though programs like Kraken exploits LCA and speeds up computation by substituting read mapping with *k*-mer matching (Wood and Salzberg, 2014, Huson et al., 2007, Brady and Salzberg, 2009). Marker gene-based approaches can be used to taxonomically profile metagenomic datasets by looking for the presence of discriminative gene markers from preassembled gene catalogues (Sunagawa et al., 2013, Quince et al., 2017). For example, MetaPhlAn adopts several thousand clade-specific markers with high discriminatory power and is able to effectively profile several areas of the human microbiome (Truong et al., 2015, Huttenhower et al., 2012). The

accuracy of these methods again will improve as more reference genomes and metagenomic assembled genomes become available.

1.6.3 Metagenome assembly

Metagenome *de novo* assembly is conceptually similar to whole-genome assembly, with most existing assemblers for metagenomic shotgun sequencing using the De Bruijn graph (DBG) approach. A DBG represents all connections (edges) between all sequence reads, which are split into sub-sequences of defined length *k* (k-mers) (nodes) and organised into a DBG of overlapping *k*-mers with *k*-1 matching bases. Contiguous sequences are then assembled from the DBG by traversing through a path across the graph to satisfy the requirement that every edge is visited once (Vollmers et al., 2017, Pevzner et al., 2001) (Figure 1.3).

Selection of *k*-mer length is an important parameter for DBG construction, if too short then multiple *k*-mers will overlap with each other and introduce branches into the graph, which results in very short contigs. Whereas selecting longer *k*-mer lengths will usually improve an assembly, but they can result in no overlaps being found between *k*-mers, particularly if the reads have high error rate resulting in gaps in the assembly. Longer *k*-mer lengths can also distinguish smaller repeats, which often introduce branches into graphs (Quince et al., 2017, Pevzner et al., 2001). The presence of repeats, sequencing errors and chimeric reads may complicate DBG construction by introducing multiple path ambiguities in the graph which results in genome miss-assemblies and fragmentation (Quince et al., 2017).

Metagenome assembly presents unique challenges. Sequencing coverage of genomes is highly variable within metagenomic data, since the coverage of each

constituent genome depends on the abundance of each genome in the community (Quince et al., 2017). Varying levels of different species within sampled communities will result in a highly non-uniform sequencing coverage of genomes. This can cause low abundance genomes to be fragmented if overall sequencing depth is insufficient for connections in the DBG (Vollmers et al., 2017). Using short *k*-mer lengths in DBG formation can aid recovery of low abundance genomes but this can obscure the reconstruction of the genomes due to an increased chance of mis-assemblies. However, the use of longer *k*-mer lengths will provide more accurate contigs but bias towards highly abundant genomes (Quince et al., 2017). Metagenome assemblers must therefore satisfy a trade-off between recovering low-abundance genomes and obtaining long, accurate contigs for high-abundance genomes. Therefore, assemblers developed for metagenome assembly address this issue by applying a multi *k*-mer approach which iterates from short to long *k*-mer length during assembly to account for reconstruction of low and high abundant species (Vollmers et al., 2017).

Strain resolution is another challenge for metagenome assembly, since microbial community samples can contain different strains of the same bacterial species. These closely related genomes in varying abundance can cause problems for DBG construction, similar to those presented by sequencing errors (Olson et al., 2019). Closely related genomes can cause branches in the DBG where they may differ by a single nucleotide variant or by the presence or absence of an entire gene or operon, causing the assembly to stop at these branch points, which leads to fragmented reconstructions. However, these events normally occur at lower frequencies and can be filtered using a defined cut-off and removed from

sequence data based on low coverage (Quince et al., 2017, Vollmers et al., 2017) (Figure 1.4).

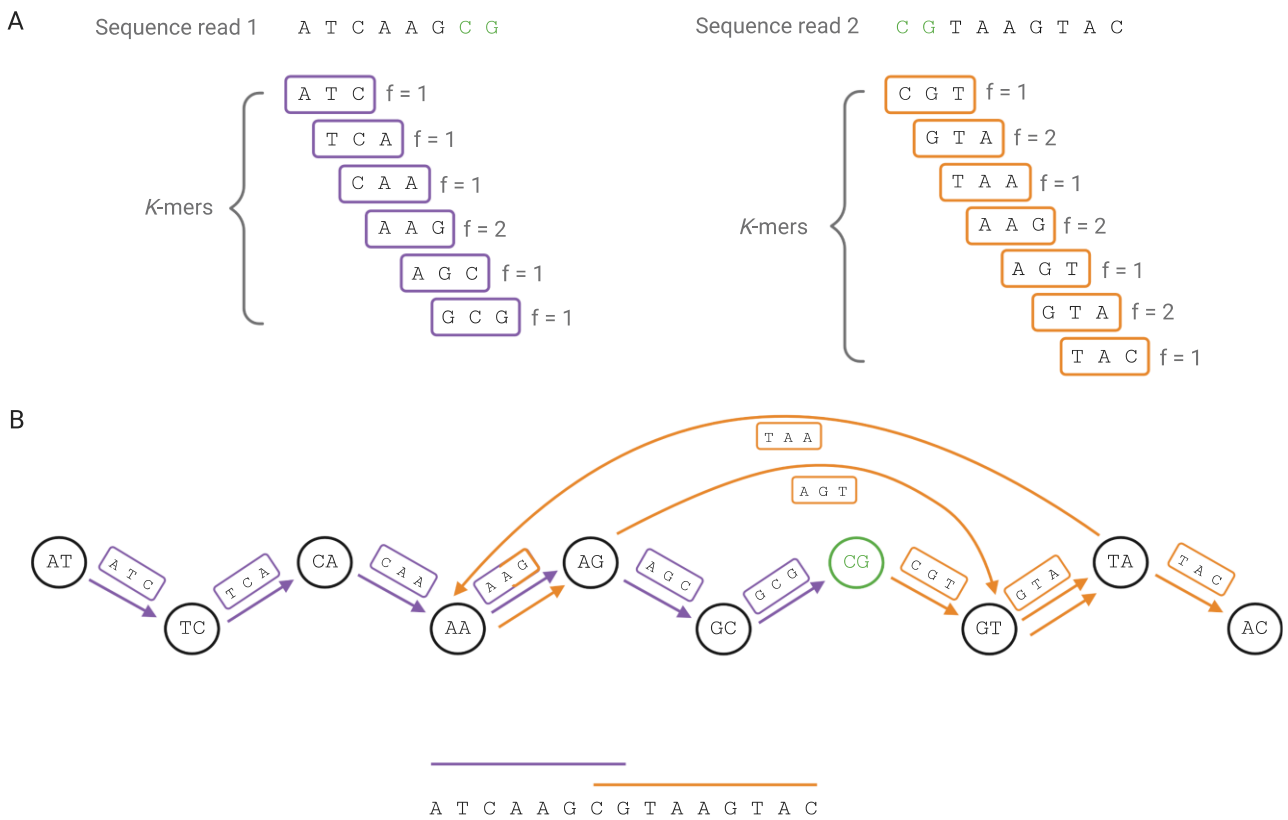


Figure 1.3 Schematic overview of the de bruijn graph concept. (A) Sequence data, represented here by two reads, overlap by two base pairs (blue). These reads are initially broken into unique sub-sequences of length *k* (*k*= 3 is used here) and are subsequently used to build a de Bruijn graph comprising of edges and nodes. All overlapping *k*-mers are derived from the sequence reads and counted (f = frequency) and adjacent *k*-mers share *k*-1 base pairs. (B) Constructed and combined representative de Bruijn graph for sequence read 1 (purple) and read 2 (orange). An additional branch is introduced to the graph by the repetitive *k*-mer AAG, as it occurs in each of the sequence reads. Figure adapted from Vollmers *et al* (2017) and created in BioRender.com (2020).

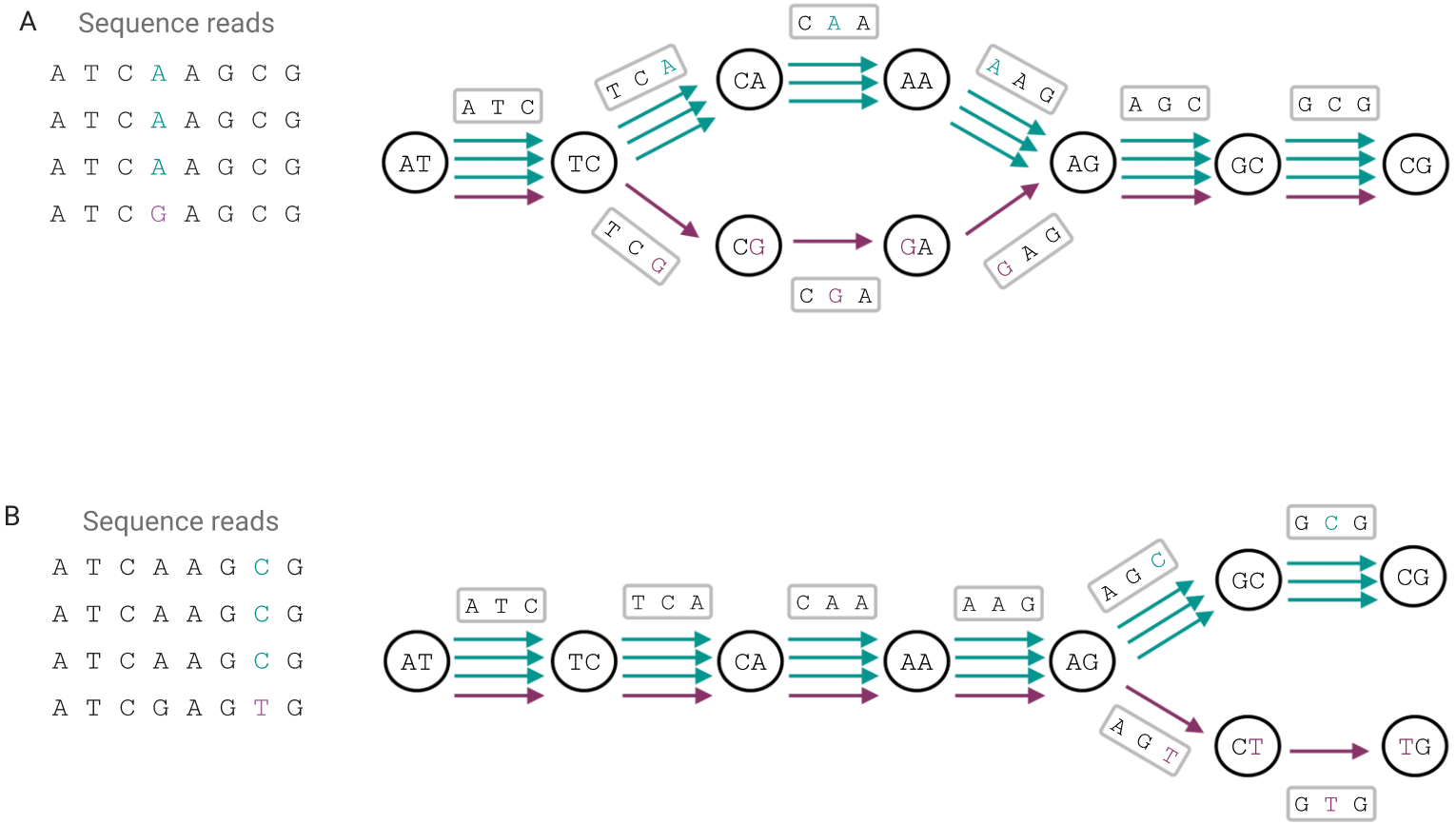


Figure 1.4 Impact of single nucleotide substitution on de bruijn graph construction. (A) A single nucleotide substitution error that occurs near the middle of a sequence read results in a bubble-like structure. (B) A single nucleotide substitution error that occurs near the beginning or end of a sequence read results in an unconnected tip-like structure. Correct nucleotide positions and edges are coloured in teal whereas incorrect nucleotide positions are coloured in purple. Erroneous reads containing a single nucleotide substitution error (purple) introduce false *k*-mers to the DBG, resulting in additional branches. Figure adapted from Vollmers *et al* (2017) and created in BioRender.com (2020).

1.6.4 Binning contigs

Metagenomic assemblies are highly fragmented, comprising of many contigs derived from genomes of hundreds to thousands of species within a community. The aim of the 'binning' process is to group contigs which are derived from the same genome together in an imaginary bin (Tyson et al., 2004).

Binning can be performed either supervised or unsupervised. Supervised binning methods use databases of sequences to taxonomically classify or group contigs based on homology with known reference genomes. However, the taxonomic-binning approach is limited by incomplete reference databases (Yue et al., 2020, Quince et al., 2017).

Unsupervised binning methods cluster contigs from the same genome together based on various shared properties (Quince et al., 2017). Clustering algorithms used in unsupervised genome binning can be divided into three types: (i) sequence composition based; (ii) differential abundance based; (iii) hybrid methods that combine both sequence composition and differential abundance (Yue et al., 2020).

Sequence composition-based binning strategies presume the sequence features from different genomes are distinct whereas the sequence features of a genomes are similar. Different sequence composition features are observed depending on the clustering algorithm used, including genome GC content, essential single copy genes, tetranucleotide frequencies (Sangwan et al., 2016). Differential abundance-based binning strategies presume that sequences belonging to the same genome have parallel abundance in the same metagenomic sample and the

sequences belonging to the same species have similar abundance patterns across multiple metagenomic samples (Yue et al., 2020).

Clustering of contigs via unsupervised binning has the advantage of not requiring reference genomes. However, it may lack the discriminatory power to resolve strain-level heterogeneity within a complex community, although with current NGS providing increased sequence depth microbial population coverage information is more reliable to obtain high quality microbial genomes from metagenomic datasets (Doughty et al., 2014, Yue et al., 2020).

Methods for reconstructing metagenomic assembled genomes (MAGs) are essential to the discovery of unculturable microbes. Binning was first applied to shotgun metagenomic sequencing data from a natural acidophilic biofilm from which two near-complete microbial genomes were recovered (Tyson et al., 2004). Since then, this approach has been adapted in large-scale recovery of thousands of genomes from public metagenomes (Parks et al., 2017).

Taxonomic classification is commonly performed using a database of marker genes that are clade-specific single copy core genes and found across bacterial and archaeal genomes to assign the taxon of binned draft genomes (Parks et al., 2015). This approach is often used to evaluate the quality of MAGs, by examination of the presence and absence of marker genes such as tRNA synthetases or ribosomal proteins to assess the level of completeness and contamination in each genome bin. A pure MAG will have these genes present in single copies. MAGs can provide a rich dataset for further analysis and comparisons (Quince et al., 2017).

1.6.5 Functional annotation of metagenomic datasets

Analysis of metagenomic sequencing data can reveal the taxonomic composition of a microbial community and has accelerated the accumulation of genomic sequences of microbial species, some of which had not been characterised previously. These same data can also be used to identify genes and pathways to generate hypotheses regarding the roles played by communities or species within specific environments and novel microbial functions (Quince et al., 2017, Sharifi and Ye, 2017).

The genetic repertoire of microbial communities can be identified using the tools created for single-genome characterisation. These tools can usually be adapted for use with metagenome assemblies and generally involve a gene identification step, followed by homology-based annotation pipelines (Quince et al., 2017).

Reconstruction of functional profiles of metagenomic datasets can also be interpreted by comparing translated sequences against databases of functionally characterised orthologous protein families, pathways and subsystems or protein signatures, such as KEGG and UniProt (Kanehisa and Goto, 2000, The UniProt Consortium, 2019). Sequences are queried normally by either a variant of the BLAST algorithm or the Hidden Markov Model (Altschul et al., 1990, Eddy, 2011). The main limiting factor in profiling metagenomic datasets is the lack of annotations for some accessory genes in microbial species, especially those from newly identified organisms. This leads to conserved genes and pathways being more consistently detected and quantified which may explain why functional traits are often consistent across different samples despite differences in taxonomic composition (Abubucker et al., 2012).

Experimental characterisation of coding genes and other genomic features such as tRNAs and non-coding RNAs is necessary to expand the current knowledge contained in reference databases. The most effective way to achieve this remains to carry out experimental characterisation of isolates, their metabolic activities and interactions with other microbes. However, due to the nature of metagenomic samples this has a crucial impact on the ability to functionally profile metagenomes (Quince et al., 2017, Walker et al., 2014, Prakash and Taylor, 2012). Functional profiling of metagenomes can also involve characterisation of specific functions of interest, such as antimicrobial resistance related genes, which can be searched for using specially curated databases (Pehrsson et al., 2016).

1.7 Bacterial genomics

Bacterial population structure is shaped by various evolutionary processes and forces. Understanding the population biology of bacteria allows us to interpret the response of a bacterial population to various selection pressures. Such knowledge improves our understanding of pathogen biology and informs on antibiotic resistance, emerging infectious diseases and interventions for control of pathogenesis. With improvements in high-throughput sequencing technologies, the availability of bacterial whole genome sequencing and the number of publicly available published and complete bacterial genome sequences has greatly increased. These data are invaluable for understanding population structures, although this does require characterisation of a large representation of sample isolates (Feil and Enright, 2004).

1.7.1 Bacterial population structure

The population structure of bacterial pathogens is defined by genetic variation and is driven by the combined effects of different evolutionary processes, including recombination, mutation, genetic drift, demographic history and natural selection (Andam et al., 2017).

Bacteria are primarily clonal, as they reproduce asexually through binary fission, whereby a single cell divides into two identical daughter cells. Each of the daughter cells contain a chromosome identical to the original mother cell (Spratt and Maiden, 1999). Evolution via mutation occurs when genetic variation generated by errors made during DNA replication are vertically inherited by all progeny bacteria. Consequently, the distribution of genetic variation among clonal, vertically descended bacterial populations will exhibit linkage disequilibrium (Shapiro, 2016). In the absence of sexual reproduction, given mutations in clonal bacterial populations will be associated with other mutations that have occurred and accumulated in the chromosome during the history of the lineage in which they arose (Spratt and Maiden, 1999).

The evolutionary forces of natural selection and recombination determine the clonality of a bacterial population. Clonality of bacterial lineages can be disrupted by bacterial sexual reproduction, which involves inheritance of exogenous genetic material through recombination (Spratt and Maiden, 1999, Spratt, 2004). In the absence of recombination, genetic variation arises through mutation leading to a clonal population, however the more recombination found within genomes, the less clonality.

Natural selection favours clonal expansion of adaptive mutants that confer competitive advantage within a specific ecological niche, thus removing genetic diversity in the population that in turn results in a genome wide sweep (Shapiro, 2016, Cohan and Perry, 2007). This process is referred to as periodic selection, which can cause evolving lineages to present sets of mutations associated with each other in higher frequencies that expected by chance. Periodic selection generates patterns of linkage disequilibrium that are indicative of clonal evolution (Cohan, 2005). Such selection can lead to an occasional epidemic clonal population structure, in which the emergence of a successful clonal lineage consequently increased its capacity to cause disease and dominate the population (Spratt and Maiden, 1999, Slatkin, 2008).

Linkage disequilibrium observed in clonal bacterial populations can also be attributed to genetic drift. Genetic drift is the random change of gene frequency between one generation and the next and can alone create linkage disequilibrium between closely linked loci (Slatkin, 2008). The consequence is reduction of sequence diversity as linkage disequilibrium caused by genetic drift reduces the response of loci to selection (Felsenstein, 1974, McVean and Charlesworth, 2000). In turn this causes selection to be inefficient in purging deleterious mutations in a species with low recombination rate. The effects of random elimination of lineages is greater in smaller populations compared to larger populations (Shapiro, 2016).

Some bacterial species are highly clonal, such as *Staphylococcus aureus* and *Mycobacterium tuberculosis*, which are genetically uniform and emerged as independent lineages from the accumulation of mutations from other species or

due to bottlenecks or selective sweeps, rather than large scale recombination events (Andam et al., 2017).

1.7.2 Recombination and horizontal gene transfer

The opposite of clonality is panmixis, whereby the rate of horizontal transfer is higher than the rate of vertical cell division, resulting in random association of loci in the genome (Shapiro, 2016). Sexual reproduction events in bacteria leads to the emergence of variation and usually involves the horizontal transfer of DNA fragments between donor and recipient bacteria by mechanisms of transformation (absorption of DNA from the environment), transduction (mediated by bacteriophages) and conjugation (plasmid exchange) (Narra and Ochman, 2006).

Horizontal gene transfer (HGT) is the principal mechanism of genetic exchange for bacteria. HGT can occur between distant taxa or closely related organisms, allowing the incorporation of exogenous genetic material into genomes (Andam et al., 2017). Recombination may lead to the acquisition of adaptive mutations and novel combinations of alleles that provide evolutionary benefits that facilitate rapid adaptation within an ecological niche and in some cases enable the microorganism to occupy a new niche (Cohan and Perry, 2007, Narra and Ochman, 2006).

The rate of horizontal transfer incorporation varies widely across the genome, and a population can be mostly clonal except for a few loci (Shapiro, 2016). These defined loci are often termed genome islands, and those that contain virulence factors are pathogenicity islands, others found in non-pathogenic organisms confer adaptation to different ecological niches (Hacker et al., 1997). An example

of the latter is a *Prochlorococcus* genomic island that confers adaptation to light and nutrient conditions such as phosphate availability (Coleman and Chisholm, 2010). In addition to this, the transfer of antibiotic resistance genes through HGT is one of the most prominent examples of recombination facilitating epidemic emergence of pathogens (von Wintersdorff et al., 2016). For example, multiple resistances including tetracycline and chloramphenicol resistance were transferred between strains of the human pathogen *Streptococcus pyogenes* (Ubukata et al., 1975).

HGT is one of the major evolutionary processes shaping bacterial population structure. Information regarding the level of recombination of a given species is essential to understanding its evolutionary dynamics. Low levels of recombination results in highly clonal populations, with well-defined lineages, whereas highly recombinogenic pathogens are associated with diversifying species (Spratt and Maiden, 1999).

1.7.3 Tools for investigating bacterial genetic diversity and population structure

Differentiating and classifying bacterial isolates into sub-groups or types is an important step in understanding their evolution and the epidemiology of any pathogenesis. This information combined with metadata allows the inference of pathogen and disease association (Bentley and Parkhill, 2015).

Traditionally, typing methods of bacteria relied on phenotypic characterisation of distinguishing properties such as, serotype, phage-type or biotype (Sabat et al., 2013). Subsequent methods that examine the relatedness of isolates at the molecular level have provided valuable insights for understanding the diversity of bacterial populations. Traditional molecular typing methods for investigating

bacterial population structure and phylogeny includes multilocus enzyme electrophoresis (MLEE) (Selander et al., 1986), multilocus variable number tandem repeat (MLVA) (Schouls et al., 2009), multilocus sequence typing (MLST) (Maiden et al., 1998) and pulsed-field gel electrophoresis (PFGE) (Sharma-Kuinkel et al., 2016).

MLEE, MLVA and MLST rely on sequence analysis from a few genes or loci whereas PFGE provides a measurement of differences spanning the entire genome and was long considered the gold standard approach for molecular epidemiological investigation (Sabat et al., 2013). Although these methods provided valuable insights for understanding the diversity in bacterial populations, they are limited in terms of reproducibility achieved and, in many cases, did not provide enough resolution to discriminate between bacterial lineages (Singh et al., 2006).

MLST is considered a more robust approach, which aims to capture the sequence variation at multiple loci, typically six to eight, distributed around the chromosome (Bentley and Parkhill, 2015, Singh et al., 2006). These loci encode housekeeping functions, so are not likely to be under strong selective pressures to capture ancestral relationships between isolates, providing a better resolution. MLST allows the unambiguous identification of clones within pathogen species and resolve ancestral relationships among isolates that have particular associations with diseases (Bentley and Parkhill, 2015). In addition, the information can be uploaded into accessible databases that can be expanded by the research community (https://pubmlst.org).

MLST has proven to be effective for studying pathogen evolution, including *S. aureus*. MLST was used in genotyping of *S. aureus* including MRSA strains, using seven housekeeping genes that led to the identification of over 1000 sequence types grouped into clonal complexes (Enright et al., 2000). MLST analysis of *S. aureus* enabled the study of clonality, evolution, virulence and antibiotic resistance (Feil et al., 2003). Although this methodology has been essential for investigating multiple aspects of bacterial populations, it is still limited to fully comprehend the evolution and population structure (Achtman, 2008). By using only, a small portion of the genome this method lacks sufficient discriminatory power compared with sequencing methods for distinguishing among highly clonal pathogens and closely related outbreak strains (Achtman, 2008, Bentley and Parkhill, 2015).

1.8 Population genomics of Coagulase Negative Staphylococci

Coagulase-negative staphylococci (CoNS) are a defined group of microorganisms that were initially considered to be apathogenic constituents of the healthy human microbiome, but have now emerged as common causes of nosocomial infections (Heilmann et al., 2019). Despite their growing clinical impact, CoNS are under characterised compared with *S. aureus*, although studies suggest that clonal diversity among CoNS species varies much more than *S. aureus*, which has evolved via point mutation (Becker et al., 2014).

S. epidermidis is the most widely characterised CoNS and has pronounced genomic diversity, greater than that exhibited by other CoNS species such as *S. haemolyticus*, *S. lugdunensis* and *S. schleriferi* when analysed with PFGE (Lang et al., 1999, Lina et al., 1992). PFGE analysis targeting the entire bacterial genome

was regarded as an appropriate short-term surveillance tool for CoNS outbreaks. However, more recent studies using MLST based on the selected conserved loci are more suitable for long term evolutionary analysis and surveillance of more virulent strains (Schoenfelder et al., 2010).

MLST studies revealed a *S. epidermidis* population with an epidemic structure of emerging well-adapted clonal lineages evolving rapidly through genetic recombination (Becker et al., 2014). Several MLST schemes were introduced to assess the distribution of frequently transferred mobile genetic elements (MGE) such as the staphylococcal cassette chromosome mec (SCC*mec*) element and virulence factors such as biofilm formation related genes (Wisplinghoff et al., 2003, Kozitskaya et al., 2005, Miragaia et al., 2007). Studies revealed that the majority of disease-associated isolates belong to sequence type, ST27. This sequence type was detected in hospitals across Europe and the US and were found to harbour the biofilm related *ica* operon, MGE SCC*mec* and the insertion sequence element IS*256* (Kozitskaya et al., 2005).

The widespread use of whole genome sequencing of CoNS species is improving resolution of their genetic organisation. Several CoNS possess a large repertoire of non-aggressive virulence factor genes including those linked to adhesion and biofilm formation; affording them both commensal and pathogenic traits. CoNS were proposed to act are a reservoir for uptake of mobile genetic elements because they are commensals and exist closely with organisms and so via recombination events and frequent acquisition of mobile genetic elements, they increase their genetic diversity (Otto, 2013, Becker et al., 2014, Argemi et al., 2019).

In addition to *S. epidermidis,* other CoNS were found to be genetically diverse. For example, *S. haemolyticus* genomes are reported to contain an abundance of insertion sequence elements, while *S. capitis* has exhibited close evolutionary links to *S. epidermidis* and large areas of recombination within its genomes (Cameron et al., 2015, Takeuchi et al., 2005, Carter et al., 2018, Bouchami et al., 2016). CoNS genetic diversity often leads species to display open pangenomes even though they have a very high number of existing genes sometimes shared by a limited number of clones (Argemi et al., 2019). However, not all CoNS exhibit the same patterns, for example *S. lugdunensis* has a closed pangenome, possibly due to the presence of barriers to horizontal gene transfer that reduce recombination rates (Argemi et al., 2018).

1.9 Thesis Aims

The research aim of this project was to utilise genomics to characterise *S. capitis* and its associated scalp microbiome to help inform industry of possible therapeutic targets to combat dandruff. Specific aims include:

1. Study the composition of the scalp microbiota using metagenomic approaches. Shotgun metagenomic profiles for both healthy and dandruff subjects were to be compared to understand in more detail the microbes that colonise a dandruff scalp.
2. Study and compare the genomes of *S. capitis* strains isolated from the same healthy scalp skin controls and dandruff scalps, to investigate the hypothesis that *S. capitis* undergoes selection of the scalp skin of dandruff sufferers. Additionally, this study aims to probe the poorly described

subspecies genotypic classification of *S. capitis* (subsp. *capitis* and subsp. *ureolyticus*).

3. Skin survival and competition are essential to staphylococci, distinct and less connected aim was to study a putative surface modification enzyme and competition determinant with respect to resistance of *S. aureus* and its interactions with other species. The focus was to determine the regulation of the *pda* gene, encoding a polysaccharide deacetylase family enzyme; and how regulation of cell wall charge mediated by Pda, in relation to other described mechanisms, alters the interactions of *S. aureus* with other niche inhabitants.

CHAPTER 2

Metagenomic profiling of healthy and dandruff scalps

2.1 Introduction

2.1.1 Structure and topography of the scalp

The scalp surface provides a distinct skin microenvironment to its resident microbes, primarily arising from the host physical conditions. These conditions include sebum content, moisture, pH and topography (Oh et al., 2016, Saxena et al., 2018). The scalp is unique among skin areas in humans, with a high follicular density which produce a high rate of sebum production and secretion due to the many pilosebaceous units and sweat glands (Grimalt, 2007, Xu et al., 2016).

Sebaceous glands are important endocrine bodies, which have a close association with hair follicles, forming the pilosebaceous units of the skin that secrete sebum onto the skin surface (Xu et al., 2016). Hair follicles are among the multiple appendages and accessory structures found embedded in the dermis, extending to the skin surface. They are in direct contact with the sebaceous gland and create a lipid-rich hydrophobic niche (Polak-Witka et al., 2020). These cutaneous appendages comprise sub-habitats of the skin (Brinkac et al., 2018) and collectively contribute to a significant increase of the surface area; the infundibulum, which extend into deeper scalp skin layers are typically filled with sebum and debris that promotes microbial colonisation (Buffoli et al., 2014, Polak-Witka et al., 2020).

Human sebum is a mixture of lipids, consisting of triglycerides, wax esters, squalene, free fatty acids and cholesterol; it coats the hair and the skin surface following secretion at the hair follicle (Pappas et al., 2009). Sebum is an important

energy source for the growth of bacteria and fungi (Sparber and LeibundGut-Landmann, 2017) and therefore plays a key role in the metabolic exchanges between the scalp surface and the microbiome. It is these metabolic interactions that support the formation of symbiotic, commensal and pathogenic microbial biofilms (Brandwein et al., 2016, Saxena et al., 2018).

In addition to pilosebaceous units, eccrine sweat glands are distributed throughout the entire skin surface, whilst apocrine sweat glands are restricted to body regions associated with hair, such as the axillae, anogenital and inguinal regions, and the periareolar regions, although rarely the face and scalp (Gilaberte et al., 2016). Both secrete watery-like fluids (Biedermann et al., 2010), and secretions from sweat and sebaceous glands promote vital processes such as thermoregulation (Kolarsick et al., 2011) and maintain the scalp microbiota (Brinkac et al., 2018).

2.1.2 Microbial composition of the scalp

The scalp has a high density of pilosebaceous units that secrete sebum. Sebum is secreted by sebaceous glands, passing along the sebaceous duct into the lumen of the pilosebaceous follicle and from there onto the skin surface (Gribbon et al., 1993). The substance provides an anoxic, antimicrobial, hydrophobic coating that protects and lubricates both the skin and hair. The resulting environment supports the growth of lipophilic microorganisms such as *Cutibacterium* spp. and *Malassezia* spp. compared with other, more dry or moist skin sites (Leeming et al., 1984).

Both *Malassezia* spp. and *Cutibacterium* spp. thrive on and within the sebaceous areas of the body due to their ability to metabolise the lipids. Both genera produce

non-specific lipases that hydrolyse available triglycerides on the skin surface (Brüggemann et al., 2004, Xu et al., 2016, Sparber and LeibundGut-Landmann, 2017) to produce free fatty acids on the skin surface, which are thought to assist bacterial adherence and colonisation of the pilosebaceous unit (Brüggemann et al., 2004). Released free fatty acids also contribute to the acidic pH (~5) together with lactic acid on of the skin surface. An acidic pH limits many common pathogens such as *S. aureus* and *Streptococcus pyogenes* but selects growth and colonisation of commensal organisms, including coagulase-negative staphylococci and corynebacteria (Grice and Segre, 2011, Elias, 2007). In addition, *Cutibacterium* spp. secrete bacteriocins to compete with other species (Gribbon et al., 1993, Xu et al., 2016).

Further microbes routinely isolated from healthy skin include the *Demodex* mite. First described in the hair follicles of humans in 1841 by Henle and Berger the mites were later classified into two species, *Demodex folliculorum* and *Demodex brevis* (Aylesworth and Vance, 1982, Elston and Elston, 2014). These microscopic arthropods reside within the pilosebaceous unit and are acquired shortly after birth then become more abundant in sebaceous glands during puberty while feeding on sebum (Elston and Elston, 2014).

Taxonomic diversity in the normal microbiota of sebaceous areas such as the scalp is lowest compared with dry and moist sites, comprising few stable inhabitants: *Cutibacteria* spp.; *Staphylococcus* spp.; and *Malassezia* spp. Transiently colonising species found within the skin microbiome of these areas is proposed to account for the majority of interpersonal variation that is affected by host factors, including pathophysiology, immune status, age (e.g. puberty

increases sebum production), sex and external environmental factors (e.g. climate and personal activities) (Grice and Segre, 2011).

2.1.3 Scalp microbiome link to dandruff

Dandruff is a persistent, inflammatory scalp disorder that affects around half of the world's adult population and is characterised by abnormal flaking of the scalp (Xu et al., 2016). The condition is accompanied with redness on the scalp and itchiness which can extend to the neck. The flaking of the scalp is the result of disrupted cohesion between corneocytes causing them to be shed from the upper layer of the epidermis, thereby disrupting barrier function and decreasing hydration of the scalp (Wang et al., 2015, Meray et al., 2018, Coates et al., 2014). Seborrheic dermatitis is a more severe disorder that in many cases is still classed as dandruff. The skin presents as patches of red, yellow flaking, greasy skin and can include increased desquamation and inflammation of the scalp together with other areas, such as the face, ears, eyebrows and chest (Dawson, 2007, Schwartz et al., 2012).

Both dandruff and seborrheic dermatitis were shown to be more than just superficial disorders of the stratum corneum, with Transmission Electron Microscopy (TEM) images of scalp tape strips from dandruff scalps presenting abnormalities in the stratum corneum ultrastructure in flaking and non-flaking sites. The epidermis is altered with hyperproliferation, parakeratosis, interdigitation of the corneal envelope and a disorganised excess of inter- and intracellular lipids (Warner et al., 2001).

Despite dandruff not being classed as a serious disease, the presence of distinguishable skin flakes on the scalp, in the hair and on clothing is considered

abnormal and its persistence can negatively affect many people socially and mentally (Warner et al., 2001, Schwartz et al., 2012).

The development of dandruff has been studied for many years, frequently associated with a reported cause of dysbiosis in the cutaneous microbiome (Park et al., 2012, Clavaud et al., 2013, Tanaka et al., 2016). The most widely accepted microbial aetiopathology is that an increased presence of *Malassezia* spp. is found on dandruff scalps, being around 1.5 - 2 times above their normal level on healthy scalps (Park et al., 2012). In addition to fungal colonisation, dandruff pathobiology is linked to abnormal sebum production, environment (seasonality) and individual susceptibility, including host inflammatory response, host predisposition and the integrity of the scalp stratum corneum (Grimshaw et al., 2019).

Sebaceous gland activity in neonates results in the microbial colonisation of the scalp and can often cause cradle cap. Following initial colonisation there is little sebaceous gland activity until puberty, which actively increases the amount of sebum found on the skin surface. Ultimately an energy source for the microbiome, that in turn is often responsible for the changes in the skin surface lipid composition, which can lead to barrier disruption and flaking of the epidermis (Wikramanayake et al., 2019).

Skin *Malassezia* spp. are found within and on the stratum corneum, the uppermost layers of the human epidermis and the follicular infundibulum. Comparatively with other fungi, *Malassezia* spp. have very small genomes of around 9 Mb and are the only free-living fungi to lack a fatty acid synthase gene (Xu et al., 2007). Although these fungi have limited genomic content, they are

highly adapted to survival on the sebaceous skin of warm-blooded vertebrates as they encode a plethora of lipase genes to satisfy lipid requirements through the hydrolysis of triglycerides found in sebum (Saunders et al., 2012).

On the skin surface *Malassezia* interact with enucleated keratinocytes but in the follicle infundibulum they are exposed to sebum lipids, a mixture of triglycerides, fatty acids, wax esters sterol esters, cholesterol, cholesterol esters and squalene, which are metabolised as an energy source. Secreted sebum, once exposed to the skin microenvironment, is broken down into its constituent di- and mono-glycerides and both saturated and unsaturated fatty acids (Grice and Dawson, 2017).

Malassezias are in part responsible for sebum break-down using their multiple secreted lipases with broad spectrum activity, as they fulfil their specific nutritional requirements. However, Malassezias can only metabolise saturated fatty acids, leaving unsaturated fatty acids, such as oleic acid on the skin surface (DeAngelis et al., 2007). Unsaturated fatty acids that remain on the skin surface can penetrate the upper layers of the skin inducing hyperproliferation, inflammation and flaking of the skin in both individuals and models, which show underlying skin barrier permeability dysfunction that causes susceptibility to dandruff and seborrheic dermatitis (Grice and Dawson, 2017, Gupta et al., 2004, DeAngelis et al., 2005).

Inflammation on the scalp is associated with more severe forms of dandruff. Dandruff sufferers exhibit increased expression of non-specific proinflammatory genes relative to healthy individuals on skin with and without lesions. This difference suggests the existence of predisposing factors related to inflammation,

all of which could have detrimental effects on epidermal barrier integrity (Mills et al., 2012, Wikramanayake et al., 2019).

The vast majority of data in published literature supports a link between increased *Malassezia* species colonisation and dandruff. The link emerged from effective treatment of the condition using antifungal compounds such as zinc pyrithione, ketoconazole, piroctone olamine and climbazole. (Schmidt-Rose et al., 2011, Piérard-Franchimont et al., 2002, Turner et al., 2012, Grice and Dawson, 2017). Improvements in dandruff symptoms were also associated with reduced numbers of *Malassezia spp.*, however, this finding does not explain why steroids are an effective treatment of dandruff and seborrheic dermatitis symptoms (Adalsteinsson et al., 2020). The presence of *Malassezia* spp. on healthy skin is not likely to be sufficient alone to cause dandruff (Grice and Dawson, 2017).

In addition to Malassezias*,* recent studies have pointed to relationships with other scalp microflora in the causation of dandruff. Species of staphylococci and cutibacteria comprise the dominant bacteria of the scalp and the relative abundance of these two genera could be key to maintaining a healthy scalp (Wikramanayake et al., 2019). Park *et al.*, showed that *S. aureus* colonisation was significantly more common in patients with dandruff/seborrheic dermatitis compared with healthy controls. Similarly, a recent metataxonomic study also associated *Staphylococcus* spp. frequencies with dandruff, despite treatment with antimicrobials not leading to improvement of symptoms (Grimshaw et al., 2019, Grice and Segre, 2011, Leyden et al., 1976).

2.1.4 Aims of chapter

Shotgun metagenomic profiles for both healthy and dandruff subjects will be compared to understand in more detail the composition of the scalp microbiota and determine the frequency of the taxa present. These data could be exploited to enable novel personal care products to be produced to counteract undesirable properties associated with increased frequency of a particular taxon. Preliminary data from a study conducted by Unilever Plc revealed raised *S. capitis* frequency that could indicate a staphylococcal associated skin condition, whereby increased levels of *S. capitis* colonisation correlate with higher levels of skin shedding.

2.2 Methods and materials

2.2.1 Subject recruitment

Ethical clearance for the experimental portion of the study was obtained by Unilever Research and Development, Strategic Science Group, Port Sunlight (Unilever R&D, Port Sunlight, HAI-SSG-2937). Scalp samples were jointly isolated by me working with members of the Unilever Strategic Science Group, Port Sunlight.

Volunteers were split into two groups for sample collection and enrolled to the study for up to one week, with two visits required. Generally, healthy male and female subjects aged 18-45 were recruited. The study set out to recruit ten subjects with visibly healthy scalps and ten subjects with visible dandruff flakes. All participants were required to provide information regarding health status and general medical history. Volunteers were not receiving any medical treatment for any active skin conditions, were not pregnant or breast-feeding, not

using any antibiotic or antifungal medication or shampoo in the 6 weeks prior to sampling and did not have visible head lice (Table 2.1).

Each subject visited the study centre twice: firstly, for consent and screening, and second for sample collection. Screening and sampling were a maximum of seven days apart and subjects were not permitted to wash their hair or use any hair products for 48 h prior to sampling to avoid major changes in scalp conditions. One scalp sample was taken from each healthy volunteer (HH). Two scalp samples from subjects with dandruff were taken from the least (DH) and most flaking sites (DD).

Table 2.1 Description of inclusion and exclusion criteria of suitable subjects for study.

Inclusion Criteria	Exclusion Criteria
Male or female	Females with a current pregnancy, pregnancy within the previous 6 months, or currently breast feeding
Aged 18-45 years	Current treatment by a doctor for an active skin condition
In good health on the basis of medical history and physical appearance	Have ever had psoriasis on any part of the body
Healthy group: TWHS [a] <8, at least one grade 0 site to sample from	Have had eczema on any part of the body in the last 5 years
Dandruff group: TWHS [a] >32, at least one grade c site and one grade 0 site to sample from	Have any other active skin conditions on any part of the body
Fitzpatrick I-II [b]	History of serious illness that may affect the outcome of the study
	Regular use of medication that may affect the study
	Current use of systemic antibiotic or antifungal medication
	Visible headlice
	Use of anti-dandruff shampoo within the last 6 weeks
	Use of shampoo containing selenium sulphide or Ketaconazole within the last 6 weeks
	Extremely curly or thick hair which makes it difficult to collect suitable samples
	Currently participating in any other study involving the hair and scalp

[a] Scalp condition was evaluated using a visual assessment method by splitting the head into four quadrants and scoring the amount of the scalp affected by dryness and flakes. The resulting numerical score, Total Weighted Head Score (TWHS) was recorded (Rogers et al., 2003). [b] The Fitzpatrick scalp for the numerical classification for human skin colour (Type I (scores 0-6) always burns, never tans (palest; freckles) & Type II (scores 7-13) usually burns, tans minimally.

2.2.2 Dandruff grading and clinical evaluation of physiology of the scalp

Subjects in the healthy scalp group were required to have visibly healthy scalps, with no surface skin flakes and uniform texture (TWHS <8 and dandruff grade 0). Subjects in the second group classed as dandruff sufferers were also required to have a healthy site of grade 0 and at least one dandruff site suitable for sampling (TWHS >32 and dandruff grade > C). Dandruff sites were described as having loosely adhered skin flakes with an irregular skin surface (Table 2.2) (Rogers et al., 2003).

Table 2.2 Description of scalp assessment grades.

Grade	Description
0	Perfect scalp. Uniform texture, no surface flakes
A	Minimal powdery flakes on the scalp
B	Small flakes at least partially adhered to scalp
C	Moderate flakes loosely adhered to scalp. Scalp surface irregular and white
D	Large pronounced dry and crusting flakes, adhering to scalp
E	Very large crusting flakes. Often congealed together into yellow plates

2.2.3 Sampling of the scalp microbiome

Scalp samples were taken following the standard scalp surface wash procedure used by Unilever Research and Development, Port Sunlight. This method is used to collect surface material from the scalp, including cells from the stratum corneum of the skin plus materials that are typical of the skin surface, such as sebum and skin microflora.

The collection sample site was located, and an appropriate clear and straight parting in the hair ~10 cm long was secured to maximise exposure of the scalp. A Teflon cup (18 mm internal diameter and 6 cm high) was placed onto the parting and held firmly in place with sufficient down force to prevent any liquid escaping during the sampling procedure. A volume of 2.0 mL sampling collection medium (phosphate buffered saline + 0.1% Triton-X-100) was then applied to the collection site and the skin agitated in the liquid for one minute, by moving a Teflon rod across the surface in a back-and-forth motion, varying in direction to ensure the entire sampling area was covered evenly. The sample collection buffer with superficial skin cells and other surface material was removed and stored in sealed container, and the process repeated, giving a total volume of 4.0 mL. The scalp washes (~4.0 mL) were stored and processed within two hours of collection once 100 μL of each sample was removed and stored on cryobeads at -80 °C.

2.2.4 Extraction of whole genomic DNA

DNA was extracted from all scalp samples. Cells were lysed by adding 3 μL Ready-Lyse lysozyme (Epicentre) to each sample in pathogen lysis tubes (Qiagen) and incubating for 18 h at 37 °C with gentle shaking (300 rpm). After incubation, bead beating was carried out and samples were transferred to a TissueLyser (Qiagen)

beating at 20 Hz, for 3 min. A second lysis step was then carried out before loading samples onto the QIAsymphony (Qiagen). 640 µL of lysis buffer (Proteinase K, Carrier RNA, ATL and ACL buffers) was added to 400 µL of sample and transferred to a microtitre plate. The plate was then incubated at 68 °C for 15 min (Eppendorf Thermomixer) with gentle agitation at 300 rpm for the first 10 s. DNA was extracted from samples using the QIAsymphony Virus/Pathogen Midi kit (937055) and the Complex Fix 400 protocol and quantified with the Qubit dsDNA high sensitivity (HS) kit using a Qubit spectrophotometer (Invitrogen).

2.2.5 Library Prep/Sequencing

DNA extracted from each scalp sample was the source material to prepare Illumina fragment libraries using a Nextera XT kit. Whole genome, shotgun metagenomic sequencing was performed by loading onto two lanes of an Illumina HiSeq 4000 (paired end, 2*150 bp sequencing, generating data from >280 M clusters per lane). Using supplied DNA, libraries were constructed and sequenced by the Centre for Genomic Research (CGR), University of Liverpool according to their established manufacturer's procedures.

2.2.6 Bioinformatics analysis

2.2.6.1 Pre-processing of raw data

All sequencing datasets underwent an in-house curation with a filtering protocol to remove low quality, human contaminating, and artificially duplicated reads. The CGR has a standard read-filtering pipeline on all sequenced datasets, which comprised: a) removal of Illumina adaptor sequences using Cutadapt v1.2.1 (Martin, 2011); b) trimming of low-quality bases, which removed read segments without a minimum phred quality score of 20, using Sickle v.1.2 (Joshi and Fass,

2011). Following filtering, artificial and real duplicates that may have been introduced during the PCR step in the library preparation protocol were removed using PRINSEQ v0.20.4 (Schmieder and Edwards, 2011).

After quality filtering of raw reads, human contaminant reads were removed using two different methods to compare their relative stringency. The first method used the bioinformatics tool suite BBTools (Bushnell) and the second combined SAMtools v1.9 (Li et al., 2009), Bowtie2 v2.3.4.2 (Langmead and Salzberg, 2012) and BEDTools v2.16.2 (Quinlan and Hall, 2010). Both methods identify contaminating human sequences after firstly creating a custom database of the human reference genome (*Homo sapiens*, NCBI Genome accession GCA_000001405.28). Filtered reads were mapped against this database and reads classified as host were discarded while unclassified reads were used for downstream analyses.

2.2.6.2 Taxonomic profiling

Following quality filtering of raw reads the taxonomic composition of each sample was characterised using both Kraken v2.0 and Bracken (Wood and Salzberg, 2014, Lu et al., 2017) and MetaPhlAn 3.0 (Segata et al., 2012). Taxonomic profiling was carried out for each scalp sample individually as well as per scalp sample state i.e. dandruff, healthy and dandruff healthy, whereby FASTQ files were concatenated for these samples.

MetaPhlAn 3.0 uses an alignment-based approach to map unassembled reads against a pre-computed set of clade-specific marker genes using Bowtie2 (Langmead and Salzberg, 2012). Clade-specific marker genes are coding sequences that are strongly conserved within genomes of a certain clade and

have substantial dissimilarity to sequences outside that clade. The clade-specific marker gene database includes information from bacteria, archaea, eukaryotes and viruses and was selected for the ability to distinguish microbial clades at a high taxonomic resolution. MetaPhlAn v3.0 method infers the presence and read coverage of clade-specific markers to detect the taxonomic clades present in the microbiome sample and estimates their relative abundance. Relative abundance of each clade is calculated by normalising the total length of all read hits by the length of the total number of reads, compares the community composition of datasets of different sizes without the need for normalisation and with less bias (Truong et al., 2015, Segata et al., 2012). MetaPhlAN v3.0 was run with default parameters.

Kraken v2.0 was used to assign taxonomic labels to metagenomic DNA sequences from each dataset. A database was created consisting of RefSeq (O'Leary et al., 2016) genomic information for archaea, bacteria, viruses and fungi. Kraken v2.0 identifies the *k*-mers within query sequences and maps them to the lowest common ancestor (LCA) of all genomes known to contain a given *k*-mer in a chosen database. The use of the *k*-mer strategy rather than an exact sequence match to the reference database achieves classification accuracy comparable to the fastest BLAST programs. Kraken v2.0 was used with default parameters, building the database with $k = 31$ (Wood and Salzberg, 2014, Wood et al., 2019).

Braken (Bayseian Reestimation of Abundance with KrakEN) was used to estimate species abundance of metagenomic samples taxonomically classified by Kraken v2.0. Bracken uses the taxonomic assignments made by Kraken v2.0 that had previously classified reads to the best matching location in the taxonomic tree

and probabilistically re-distributes the reads within the taxonomic tree. Reads assigned to nodes above the species level were distributed down to the species nodes, while reads assigned at the strain level were re-distributed upward to their parent species (Lu et al., 2017). Bracken was run with default parameters.

2.2.6.3 De novo metagenome assembly and quality assessment

For metagenomic *de novo* assembly, the processed sequence raw reads were assembled using three De Bruijn graph assemblers: MEGAHIT (Li et al., 2015); metaSPAdes (Nurk et al., 2017); and IDBA-UD (Peng et al., 2012). Each assembler was run with default recommended settings and assembly statistics were compared. The quality and statistics of each metagenome assembly was evaluated with QUAST v4.4.0 without a reference genome (Gurevich et al., 2013).

2.2.6.4 Binning contigs

Unsupervised binning of the assembled contigs was carried out using MetaBAT2 (Kang et al., 2019) and CONCOCT (Alneberg et al., 2014). Prior to binning with MetaBAT2, BWA MEM (Li, 2013) was used to map processed sequence reads onto scalp state co-assembled contigs. Contig coverage from each assembly was calculated using the script jgi_summarize_bam_contig_depths from the MetaBAT2 package, followed by binning of contigs using the options –minContigLength 2000.

Binning contigs with CONCOCT (Alneberg et al., 2014) was done within the Anvi'o v6.2 platform (Eren et al., 2015). Anvi'o first profiles contigs using Prodigal v2.3.6 (Hyatt et al., 2010) with default parameters to identify genes, as well as HMMER v3.3 (Eddy, 2011) to identify genes matching to archaeal and bacterial single-copy core gene collections. Kaiju v1.7.2 (Menzel et al., 2016) infers the taxonomy

of genes with the NCBI BLAST *nr* database containing proteins belonging to archaea, bacteria, viruses, fungi and microbial eukaryotes.

Processed metagenomic reads were mapped to Anvi'o profiled co-assembled contigs using Bowtie2 v2.4.1 (Langmead and Salzberg, 2012). Mapped reads were processed using SAMTools (Li et al., 2009) and input into CONCOCT via the Anvi'o program anvi-cluster-contigs using the option –length-threshold 2000. Anvi'o was also used to infer taxonomy, completion and redundancy of genome bins.

2.2.6.5 Functional annotation of the scalp microbiome

Metagenomic annotation was carried out within the Anvi'o v6.2 platform (Eren et al., 2015). Protein sequences were extracted from each Anvi'o profile created from assembled metagenome contigs as described in section 2.2.6.1 and were input into GhostKOALA.

GhostKOALA is an automatic annotation server that was used to assign KO (KEGG Orthology) to metagenome sequences, to characterise individual gene functions and reconstruct KEGG pathways, BRITE hierarchies and KEGG modules to infer high-level functions of the organisms (Kanehisa and Goto, 2000, Kanehisa et al., 2016).

GhostKOALA utilises GHOSTX for a more rapid database search and is suitable for metagenome annotation, including those containing eukaryotes (Suzuki et al., 2014). The annotation of metagenome sequences assigns both functional and taxonomic information. The non-redundant database of sequences is supplemented by sequence similarity to groups of genes without K assigned numbers to better characterise taxonomic compositions of metagenomes.

2.2.7 Understanding variation in scalp microbiome

2.2.7.1 Correlation of microbial species with scalp physiology

2.2.7.1.1 LEfSe analysis of enriched microbial clades

Taxonomic abundance profiles created in MetaPhlAn3 for each scalp sample were compared using a linear discriminant analysis (LDA) effect size (LEfSe) algorithm to identify clades most likely to account for the presence of dandruff (Segata et al., 2011).

LEfSe initially analyses all features in each comparison class, identifying features that are differentially distributed between each scalp state using a non-parametric Kruksal-Wallis sum-rank test with a p-value cut-off of 0.05. Microbial clades that violated the null hypothesis of equal distributions were analysed further using a pair wise Wilcoxon test, and microbial clades which did not were removed.

The differentially abundant set of features from LEfSe were subjected to LDA, and a model was created using the relative differences in scalp state to rank the taxonomic features. The output consists of a list of features that are discriminative with respect to each scalp state, ranked according to the effect size which estimates the degree of responsibility associated with each microbial clade with respect to scalp state (Segata et al., 2011).

2.2.7.1.2 Alpha and beta diversity analysis

Taxonomic classifications made by Kraken v2.0 analysis were used to create a count table based on the number of reads estimated to belong to different taxa. These counts were used to analyse alpha and beta diversity of scalp samples. All

analysis was done using the statistical programming language R (Team, 2010), as follows. Firstly, rarefaction curves were generated using the "rarecurve" function in the community ecology package "vegan" (Jari Oksanen, 2019). Alpha diversity for each scalp sample was calculated using the "diversity" function in the package "vegan", comparing the observed OTU numbers, Shannon index and Simpson index. The Shannon index is used to describe the disorder and uncertainty of individual species, considering the number or species and the average individual distribution in the species. In this case the diversity index is larger when each individual in a sample belongs to a different species (Shannon, 1948). The Simpson index is based upon the probability that two individuals within a sample belongs to a different species, so the higher the Simpson index the higher the diversity (Simpson, 1949). The alpha diversity values calculated by each of the three indices for each scalp sample were plotted using the function "plot_richness" in the package "phyloseq" (McMurdie and Holmes, 2013) together with the package "ggplot2" (Wickham, 2016). Diversity mean values were statistically compared using analysis of variance (ANOVA) and a Tukey's Honest Significance Differences (HSD) test in the package "agricolae" (Mendiburu, 2020).

Permutational Multivariate Analysis of Variance (PERMANOVA) was applied to verify differences between scalp sample sites using the function "adonis" in the vegan package, using Bray-Curtis distances calculated in vegan (Jari Oksanen, 2019). Nonmetric Multidimensional Scaling (NMDS) and Principal Coordinates Analysis (PCoA) plots of all samples based on the Bray-Curtis distance matrix were then created using the functions "metaMDS" and "pcoa" within the vegan

package to highlight clustering according to scalp sample type. For all statistical analyses, *p*-values <0.05 were considered significant.

2.2.7.1.3 Statistical analysis to identify taxa significantly enriched between scalp samples

The R package DESeq2 was used to calculate differences in read counts assigned to taxa by Kraken v2.0 using a database consisting of RefSeq genomic information for archaea, bacteria, viruses and fungi (Wood et al., 2019, Love et al., 2014). The "DESeqDataSetFromMatrix" and "DESeq" functions were both used to test for differentially abundant taxa. Data were normalised to sample-specific size factors determined by the median ratio of gene counts relative to a geometric mean per gene. Normalised read counts were analysed with DESeq2 default settings, which fits a negative binomial generalised linear model (GLM) for each gene and uses the Wald test for significance testing. Significance testing was done in a pairwise manner, comparing each of the three scalp states against each other. The correction of *p*-values relating to taxonomic profiles were performed using the Benjamini-Hochberg false discovery rate (FDR). An FDR-adjusted *p*-value <0.05 was considered significant. The R package ggplot2 was used to plot read counts per sample, before and after DESeq2 normalisation (Wickham, 2016).

2.2.8 Functional Analysis of the scalp microbiome

2.2.8.1 Statistical analysis of global differences in the functional profiles of the scalp microbiome

Functional annotation of each metagenome was curated within the Anvi'o v6.2 platform (Eren et al., 2015). Protein sequences were extracted from each Anvi'o profile and created from assembled metagenome contigs as described in section

2.2.6.1; data were input into GhostKOALA. GhostKOALA, then assigned KO (KEGG Orthology) to metagenome sequences to characterise individual gene functions and reconstruct KEGG pathways (Kanehisa et al., 2016). These functional annotations were used to create a count table based on the number of reads estimated to map to each KEGG functional ortholog. These counts enabled analysis of statistical differences in the functional potential of healthy, dandruff healthy and dandruff scalp states.

The overall similarity and variation in functional profiles between the microbiomes from the three scalp sites were examined using the Bray Curtis distance matrix. Permutational Multivariate Analysis of Variance (PERMANOVA) was applied to verify differences between scalp sample sites and Nonmetric Multidimensional Scaling (NMDS) and Principal Coordinates Analysis (PCoA) plots of all samples based on the Bray-Curtis distance matrix were also created as described in Section 2.2.8.1.2.

2.2.8.2 Identifying functional differences between dandruff, dandruff healthy and healthy scalps

The R package DESeq2 was used to calculate coverage differences for individual KEGG orthologs assigned to metagenome sequences using GhostKOALA (Love et al., 2014, Kanehisa et al., 2016). These functional annotations were used to create a count table based on the number of reads estimated to map to each KEGG functional ortholog. These counts enabled analysis of statistical differences in the functional potential of healthy, dandruff healthy and dandruff scalp states.

The “DESeqDataSetFromMatrix” and “DESeq” functions were both used to test for differentially abundant KEGG orthologs as described in Section 2.2.8.1.3. The

R package ggplot2 was used to plot read counts per sample before and after DESeq2 normalisation as well as the normalised read counts of the top 30 differentially abundant KEGG orthologs found on dandruff and dandruff healthy scalp sites compared to healthy scalps (Wickham, 2016).

2.3 Results

2.3.1 Scalp sampling and dandruff assessment

Scalp samples were obtained from sixteen subjects. Prior to sampling, all subjects underwent a scalp assessment by trained assessors at Unilever R&D, Port Sunlight. Scalps were visually assessed and graded according to the criteria laid out in Table 2.2. A scalp grade of 0 was characterised as a healthy scalp, whilst a grade of >C was characterised as a dandruff scalp. The scalps of subjects A1-A5 and E1-E5 were characterised as healthy scalps (HH), presenting with a uniform scalp and no surface flakes, whereas the scalps of subjects B1, B3, D1, D3, D3 and D7 were categorised as dandruff scalps with grades C or D, presenting with skin flakes and scalp surface irregularities (DD). Dandruff categorised subjects were also sampled at a grade 0 site, or most healthy site on their scalp and were given separate subject IDs: B2, B4, D2, D6 and D8 (DH) (Table 2.2 & Table 2.3).

2.3.2 Generation of metagenomic sequencing libraries

DNA extraction methodology can affect the composition of downstream sequence data. The DNA extraction method must be effective to ensure the detection of a diverse range of microbial taxa. Therefore, a robust extraction protocol was followed, combining both chemical and mechanical lysis. The resulting DNA concentrations ranged from 2.16-12.2 ng μl^{-1} (Table 2.3) and Illumina libraries were successfully generated from all samples using Nextera XT preparation kit.

Large high-quality datasets were generated for all samples with an average of 57 ± 4 million raw reads generated across all samples (Table 2.3).

Table 2.3 Scalp study subject information. Scalp assessment grade and sample information for the 22 scalp samples isolated from 16 female subjects enrolled in the study. HH: Healthy scalp group (TWHS <8 and dandruff grade 0). DH: Dandruff group with a healthy site (TWHS <8 and dandruff grade 0). DD: Dandruff group with a dandruff site (TWHS >32 and dandruff grade >C).

Sample number	Subject ID	Scalp state [α]	TWHS Score	Scalp assessment grade	Concentration of DNA extracted ($ng\mu l^{-1}$)	Total Quantity of DNA (ng)	Number of raw reads generated
1	002	HH	4	0	5.26	26.30	58,669,400
2	019	HH	0	0	2.16	10.80	49,022,644
3	023	HH	0	0	5.96	29.80	42,604,436
4	007	HH	4	0	3.08	15.40	40,719,724
5	008	HH	4	0	3.64	18.20	66,618,484
6	010	DD	40	C	8.60	43.00	60,720,864
7	010	DH	40	0	4.44	22.20	62,500,154
8	021	DD	48	D	9.70	48.50	64,310,044
9	021	DH	48	A-B	4.38	21.90	70,346,556
10	009	DD	32	C-D	6.72	33.60	73,788,904
11	009	DH	32	A	2.42	12.10	44,509,048
12	020	DD	40	C	9.46	47.30	46,680,554
13	020	DH	40	0	5.06	25.30	58,804,714
14	028	DD	48	D	12.2	61.00	66,550,904
15	028	DH	48	A	11.9	59.50	60,911,302
16	031	DD	31	D	11.7	58.50	64,777,538
17	031	DH	31	A	9.38	46.90	57,335,560
18	015	HH	4	0	5.56	27.80	49,936,274
19	025	HH	4	0	3.80	19.00	57,295,432
20	026	HH	4	0	3.40	17.00	51,138,426
21	043	HH	4	0	2.66	13.30	74,071,560
22	040	HH	4	0	5.36	26.80	41,640,322

2.3.3 Removal of human host sequences

The scalp sampling process involved agitation of the scalp surface to collect surface material, including cells from the stratum corneum of the skin and resident skin microflora. The mechanical disruption meant it was likely that each sequenced dataset would contain considerable contamination from human host reads.

Two different methods were used to remove human reads from each dataset. The first, BBMap is a splice aware global aligner for DNA and RNA sequencing reads from the BBTools suite (Bushnell) and the second, a combination of the alignment tool Bowtie2 (Langmead and Salzberg, 2012), SAMtools (Li et al., 2009) and BEDTools (Quinlan and Hall, 2010). BBMap quickly maps with very high precision but with an option to lower sensitivity by allowing the user to set the minimum percentage identity, so reducing the risk of false positives. Using Bowtie2 to align reads to the human genome (*Homo sapiens*, NCBI Genome accession GCA_000001405.28) provides higher sensitivity with default options. Bowtie2 removed more contaminating reads from each sample compared with BBMap. Across all datasets, the average percentage of reads remaining post contaminant filtering was 22.94% and 21.01% for BBMap and Bowtie2 respectively. With the average percentage of reads remaining post contaminant filtering for both methods being very similar each method was deemed suitable for host read removal, but BBMap filtered reads were taken forward for further analysis.

The proportion of human contaminating reads varied considerably across samples, ranging from 44% to 97% (BBMap) and 45% to 99% (Bowtie2) (Figure

2.1). On average, 77% (BBMap) and 79% (Bowtie2) of reads within each sample were classified as being of human origin. Samples obtained from dandruff (DD) graded sites contained more human reads compared with samples from healthy (HH & DH) graded sites (Table 2.3), presumably due to the relative nature of the skin. A point-biserial correlation was run to determine the relationship between the percentage of reads remaining in each dataset post removal of human associated reads, and the disease state dataset. A statistically significant, positive correlation was observed between the percentage of reads remaining, post host filtering and scalp state ($rpb = 0.631$, $t = 3.6406$, $df = 20$, p-value = 0.001627). This meant it is likely that samples from dandruff graded sites on the scalp contain more human contaminated reads as more human material from the scalp was collected during sampling. Dandruff graded sites have by definition, flaking and excess sebum on the surface of the scalp (Table 2.2), which would be expected and was confirmed to lead to a significant amount of human-originating sequence in the final samples.

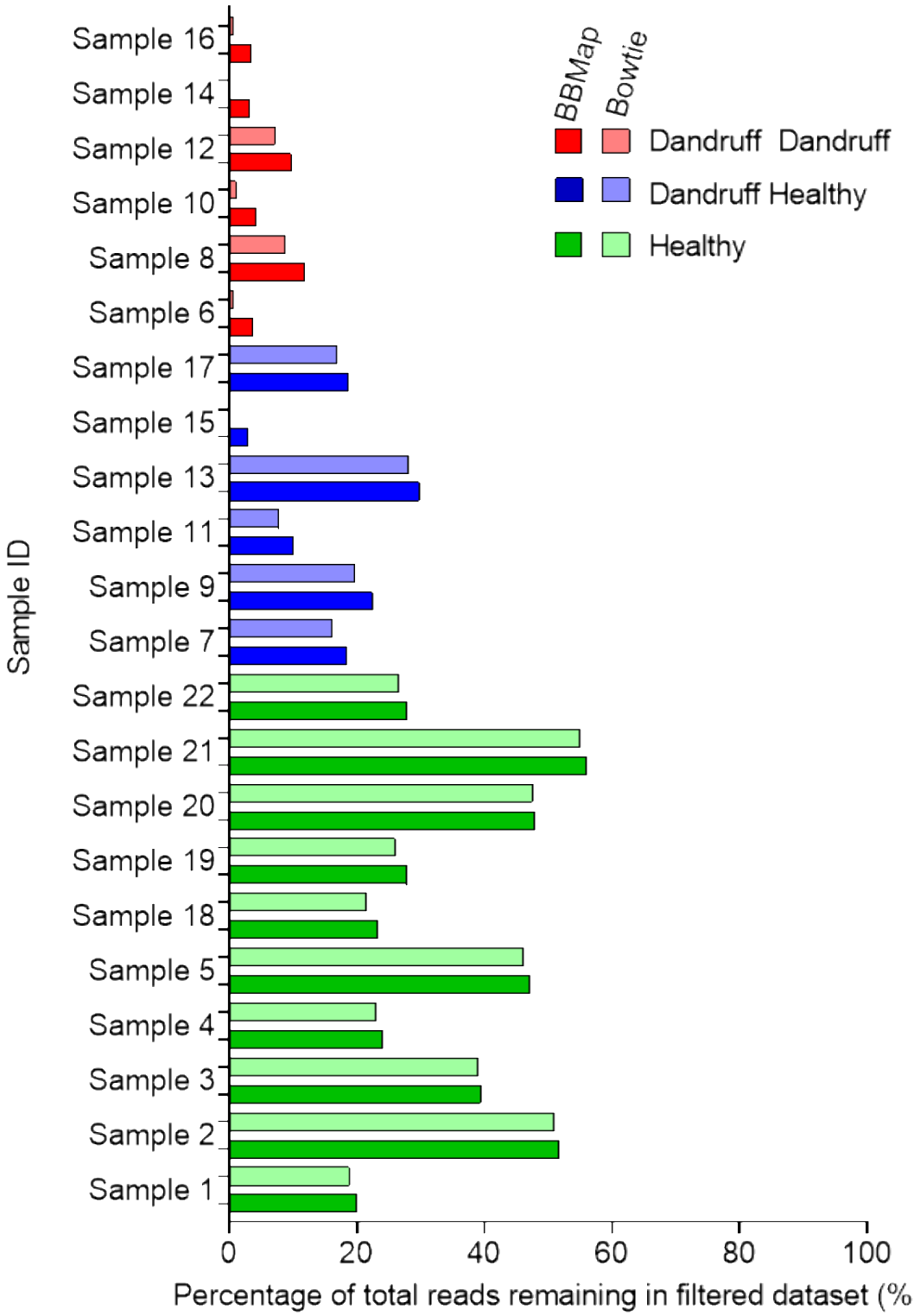


Figure 2.1 Proportion of non-human reads from metagenomes. The proportion of original reads remaining in the final dataset following quality control steps removing human host originating, and artificially duplicated sequences using PRINSEQ and either BBMap (dark shading) or Bowtie2 and SAMtools (light shading).

2.3.4 Microbial composition of the scalp microbiome

The microbial composition of all filtered and quality-controlled scalp sample reads was predicted firstly using two different profiling programs: MetaPhlAn3 that maps reads to a clade-specific marker gene database (Segata et al., 2012), or Kraken v2.0 that maps DNA sequences to a database of genomes (Wood et al., 2019). To further analyse the compositional features of scalp samples, contigs from assembled reads were grouped and taxonomically classified using MetaBAT/CONCOCT.

2.3.4.1 Kraken v2.0 prediction of taxonomic composition

Kraken v2.0 is a taxonomic sequence classifier that assigns taxonomic labels to short DNA reads, by aligning *k*-mers of query sequences to databases made up of in this case all NCBI taxonomic information and complete genomes in the RefSeq database (O'Leary et al., 2016). Every *k*-mer that is aligned to the lowest common ancestor is then analysed to give a single taxonomic label for the read (Wood and Salzberg, 2014). Bracken (Bayesian Reestimation of Abundance with KrakEN) was then used to compute the abundance of species, using the Kraken v2.0 taxonomy-assigned labels to estimate the number of reads originating from each species present in a sample (Lu et al., 2017).

To understand the most prevalent microbial clades within the scalp microbiome, the top 25 abundant species were identified in each sample and also in each scalp state combined. *M. restricta* was found to be in all but one sample as one of the 25 most abundant species, accounting for 80 %, 23 % and 12 % across dandruff, dandruff healthy and healthy samples, respectively (Table 2.3, Figure 2.2, Figure 2.3, Table S 2 & Table S 3).

In addition, *C. acnes* and *S. capitis* were only abundant in 18 and 15 samples out of the 22 samples across all scalp states, although both were present in every healthy sample. *C. acnes* was most abundant on healthy scalps making up 15 % of the microbial abundance, compared with 2.5 % and 7 % for dandruff and dandruff healthy samples, respectively. Comparatively, *S. capitis* was found to make up only 0.4 % of the microbial abundance in dandruff and healthy samples, compared with 0.2 % in dandruff healthy samples.

Overall, the data obtained using Kraken v2.0 to taxonomically profile all 22 samples supports the current understanding of key genera present on the scalp: *Malassezia, Staphylococcus* and *Cutibacterium*, with an increased abundance of *Malassezia* spp. on dandruff scalps.

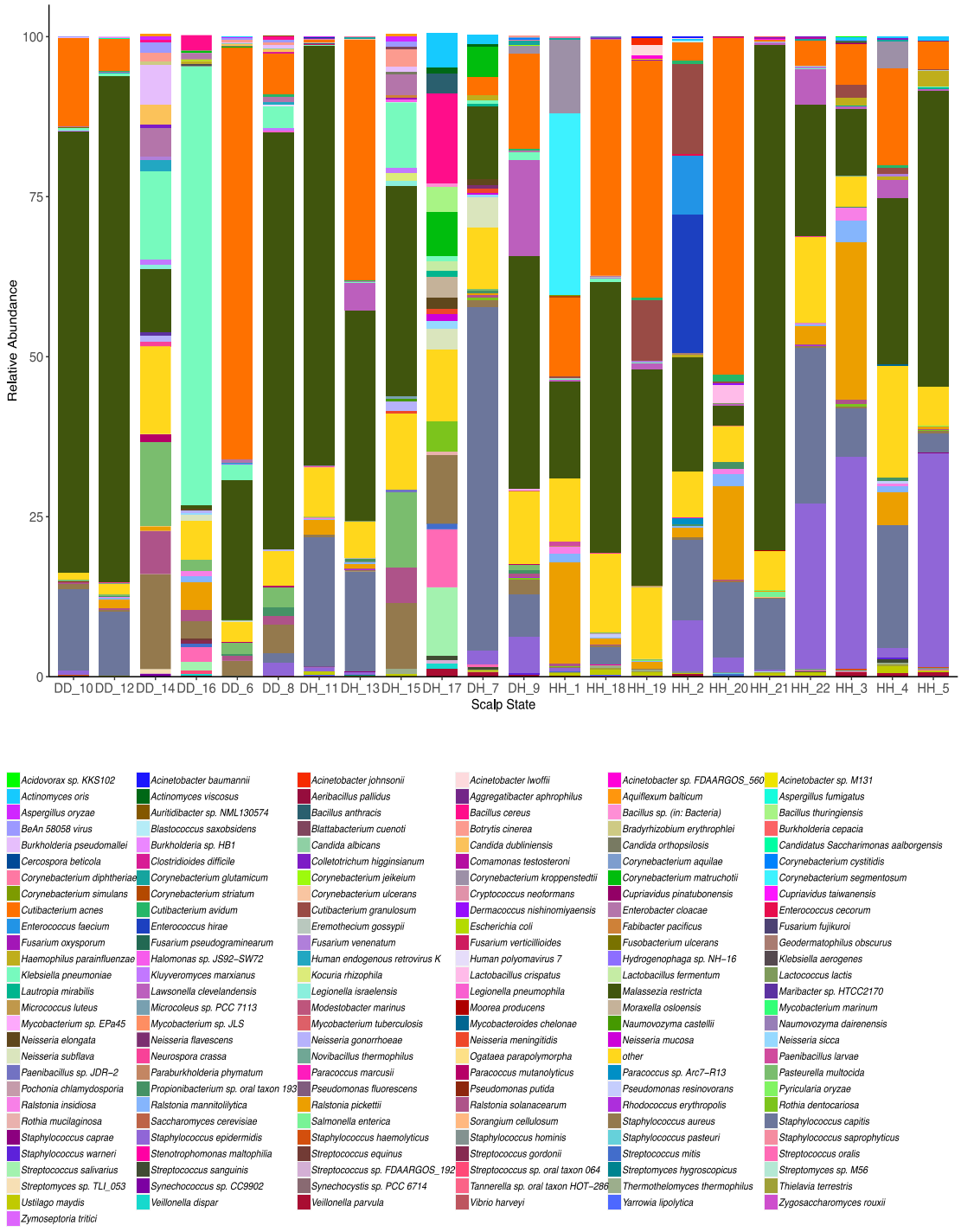


Figure 2.2 Taxonomic composition of the scalp microbiome for each sample. The relative abundance of the top 25 most abundant taxa are shown for each scalp sample. Taxonomic profiles were generated using Kraken v2.0 and relative abundance estimations were generated using Bracken.

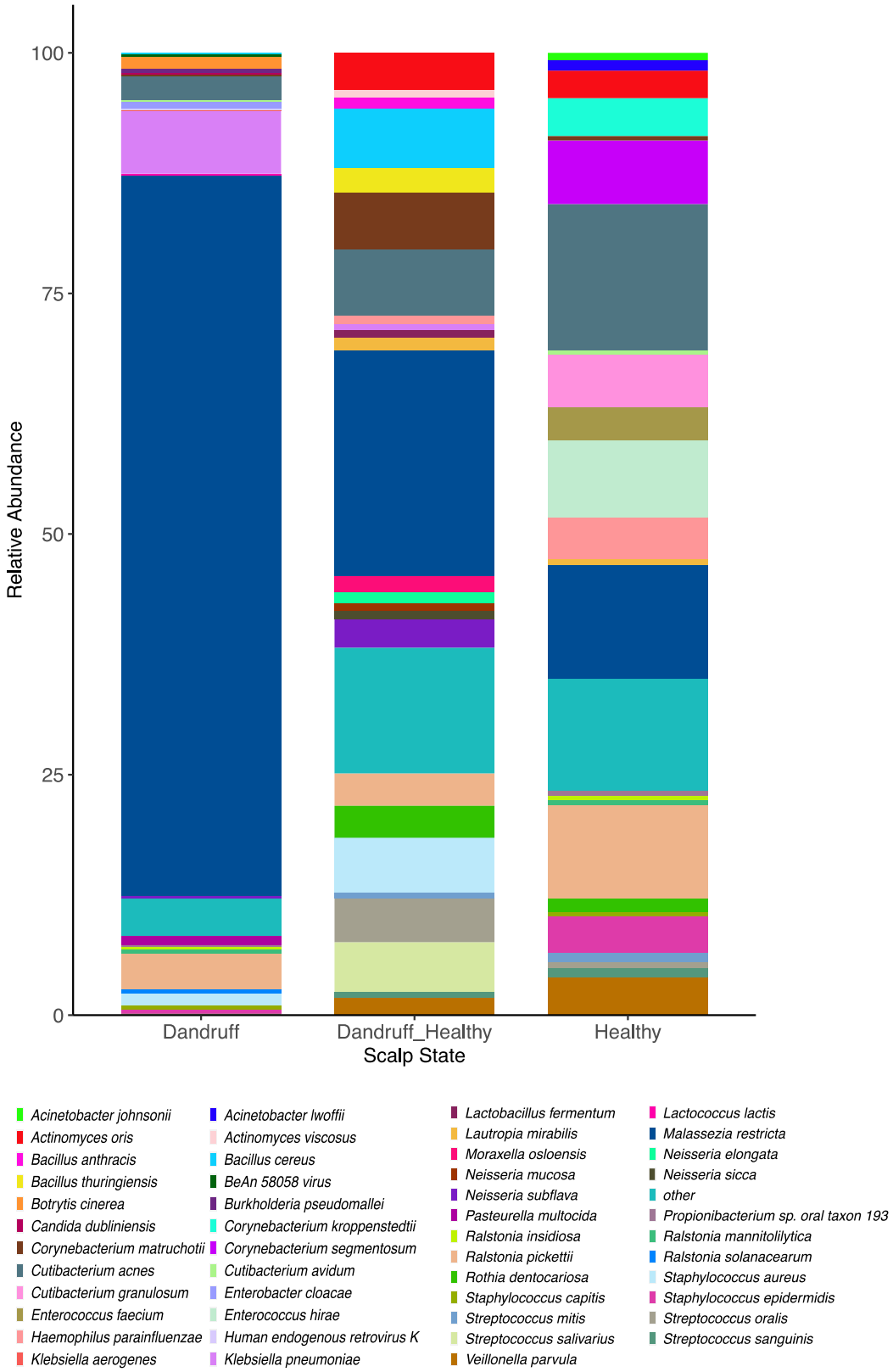


Figure 2.3 Taxonomic composition of the scalp microbiome for each scalp state. The relative abundance of the top 25 most abundant taxa are shown for each scalp state. Taxonomic profiles were generated using Kraken2 v2.0 and relative abundance estimations were generated using Bracken.

2.3.4.2 MetaPhlAn3 prediction of taxonomic composition

MetaPhlAn v3.0 uses an alignment-based approach to map unassembled reads against a pre-computed set of clade-specific marker genes from Bowtie2 (Langmead and Salzberg, 2012), allowing for the accurate assignment of taxonomic classification at species level. This method infers the presence and read coverage of clade-specific markers to detect the taxonomic clades present in the microbiome sample and estimate their relative abundance. MetaPhlAn3 v3.0 estimates relative abundance, therefore it is possible to directly compare the community composition of datasets of different sizes, meaning there is no need for normalisation and there are less biases (Truong et al., 2015).

The top 25 most abundant clades found by MetaPhlAn3 v3.0 in all samples (Figure 2.4) and for each scalp sample state (Figure 2.5) are presented. The most abundant species in the datasets based on reads across all samples and scalp states was *C. acnes*. Meanwhile, both *S. capitis* and *M. restricta* were identified in every sample and scalp state, with lesser abundance than *C. acnes*, but were more abundant in dandruff compared with healthy samples. The frequency of *M. restricta* was overall higher in dandruff samples compared with the levels of *S. capitis.* The opposite was true with *M. globosa*, *M. sympodialis*, *C. acnes*, *S. epidermidis* and *Propionibacterium namnetense*, all of which were more abundant in healthy samples compared with dandruff. In addition to organisms commonly found on the scalp, multiple species were found sporadically and in low levels throughout all samples, most commonly associated with the oral bacterial community, which have been documented as being frequent on human skin (Shibagaki et al., 2017).

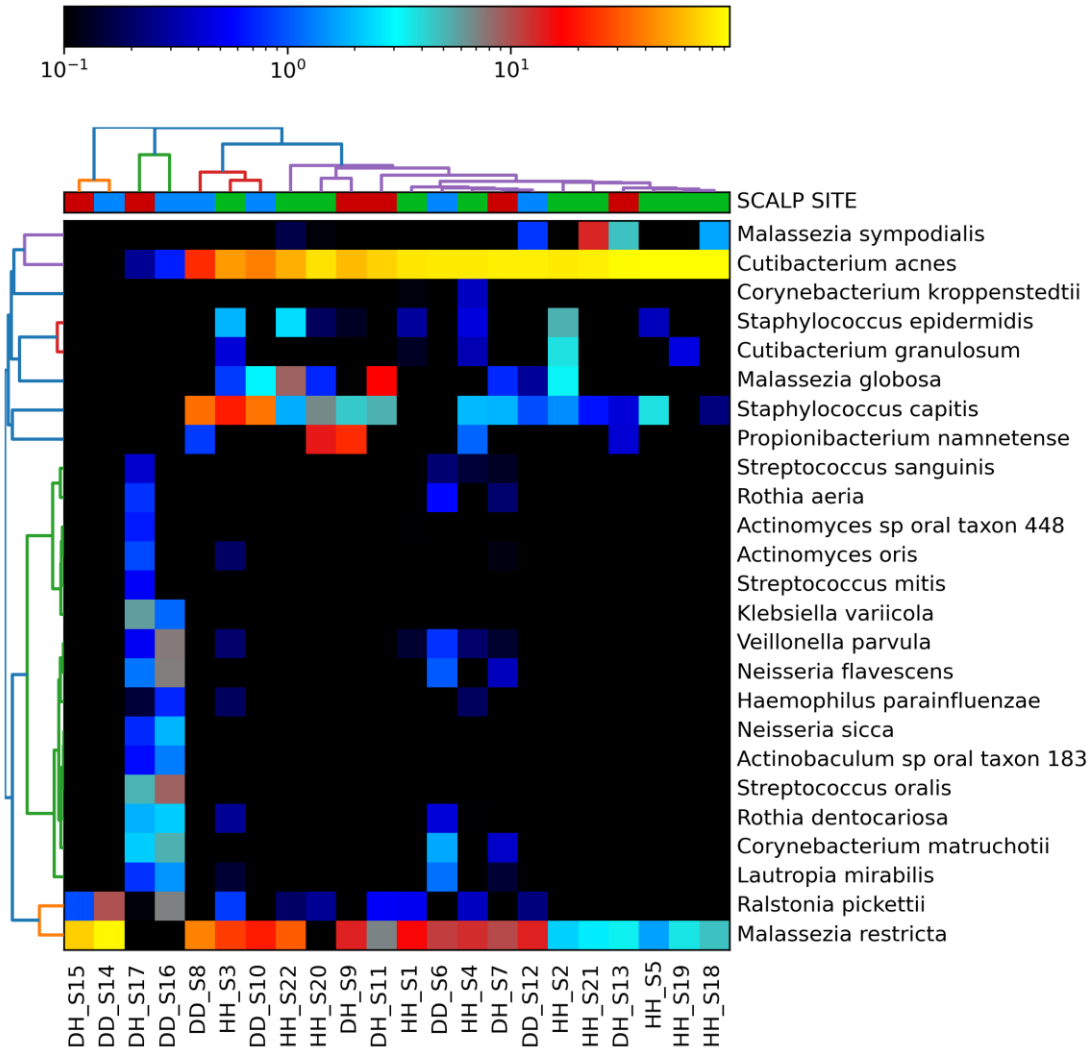


Figure 2.4 The top 25 most abundant clades in each individual scalp sample. Hierarchal clustering of samples was performed with the Bray-Curtis method using MetaPhlan3. The black to yellow scale represents the relative abundance of each microbial clade in each dataset. Image was generated using hclust2.

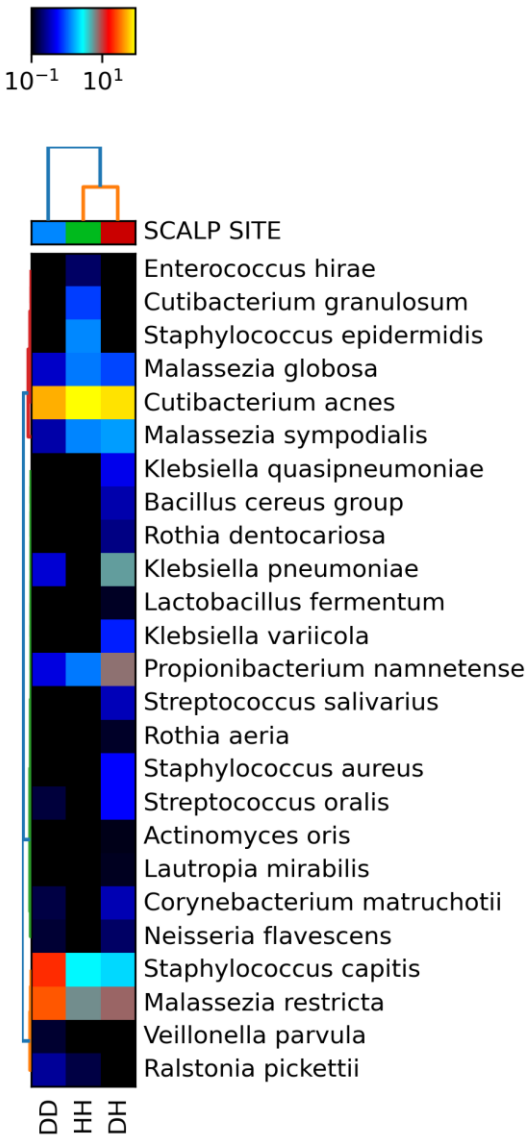


Figure 2.5 The top 25 most abundant clades clustered by scalp state. Hierarchal clustering of samples was performed with the Bray-Curtis method using MetaPhlan3. The black to yellow scale represents the relative abundance of each microbial clade in each dataset. Image was generated using hclust2.

2.3.5 Metagenome assembly and contig binning

2.3.5.1 Metagenome Assembly

Three of the most widely used assemblers for metagenomics, metaSPAdes, MEGAHIT and IDBA-UD were used and compared to investigate which assembler was the most appropriate for the study dataset. Each are de Bruijn graph-based assemblers and were developed for the assembly of sequencing data with uneven read depth of coverage (Vollmers et al., 2017). The assembly quality was assessed for each metagenomic sample individually and combined via scalp sample state, since both groups of assemblies were used in further analysis.

The quality of the metagenome assemblies from the three assemblers was assessed using QUAST, a quality assessment tool that is capable of evaluating assemblies, with and without a reference genome. In this instance QUAST was used to evaluate metagenome assemblies, so this was done without a single reference genome. QUAST provides several metrics, including: i) the total number of contigs in the assembly; ii) the length of the largest contig in the assembly; iii) the total number of bases in the assembly; and iv) the minimum contig length that comprises over half of the entire assembly (N_{50}) (Gurevich et al., 2013).

Metagenome assemblies with metaSPAdes produced better contiguity, with a larger N_{50} value, compared with assemblies from IDBA-UD and MEGAHIT (Figure 2.6, Figure 2.7, Table S 4 & Table S 5). However, metaSPAdes also produced the greatest number of contigs (Figure 2.6, Figure 2.7, Table S 4 & Table S 5). IDBA-UD and MEGAHIT performed similarly, with MEGAHIT presenting assemblies with the lowest N_{50} values on average across all assemblies and the lowest

number of contigs. The average N_{50} and total number of contigs of assemblies produced by IDBA-UD fell between those of metaSPAdes and MEGAHIT.

Overall, because metaSPAdes produced the most fragmented assembly it was not used to assemble each metagenome. Both MEGAHIT and IDBA-UD performed very similarly with negligible differences in average N_{50} values and total number of contigs, so either assembler could be used. For all subsequent analysis MEGAHIT was used to assemble metagenomes.

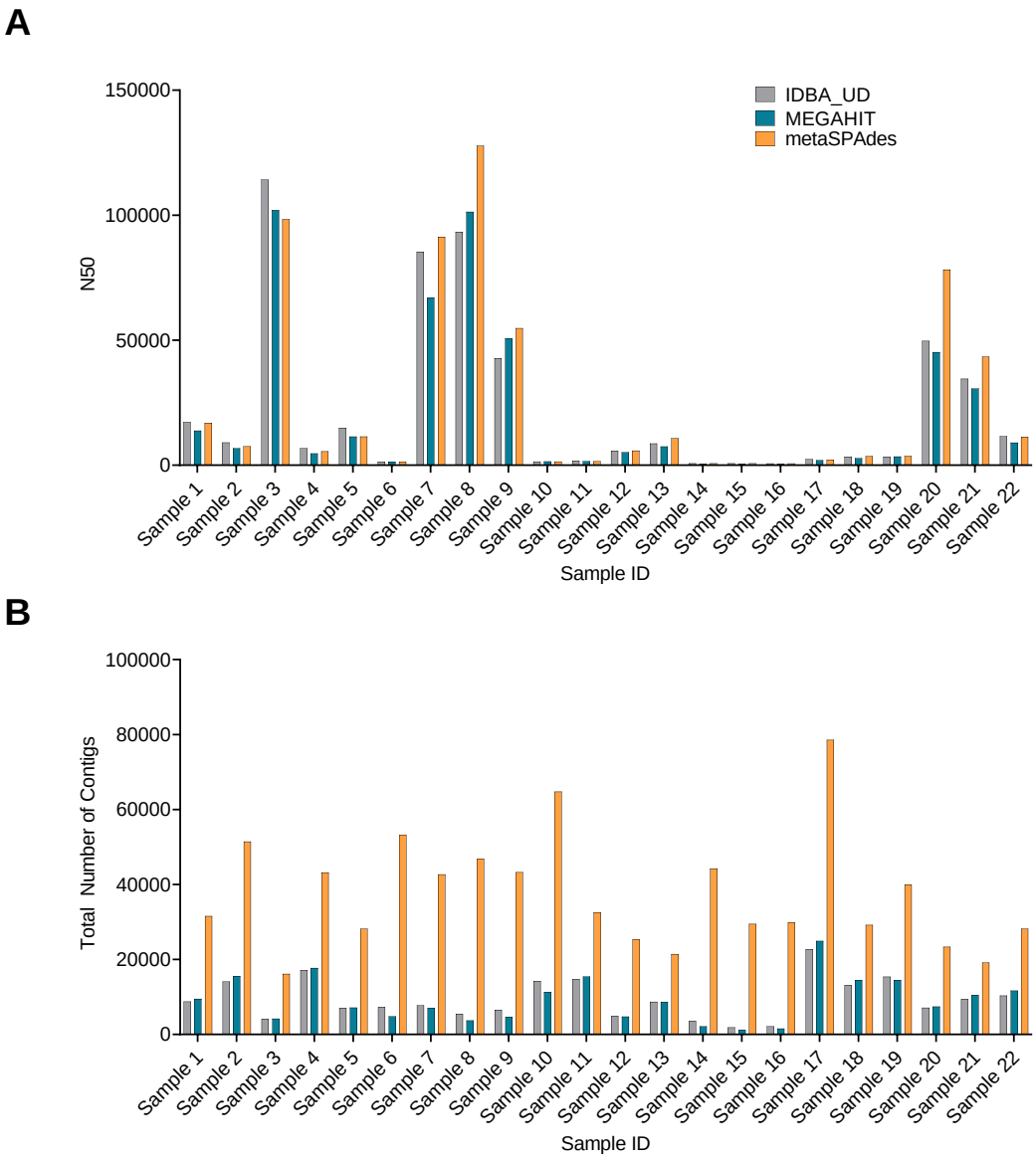


Figure 2.6 Comparison of metagenomic assembler performance on each scalp sample. Comparison of assembly quality (A: N50 and B: Total number of contigs) of each metagenomic sample using three different metagenomic assemblers: IDBA-UD, MEGAHIT and metaSPAdes.

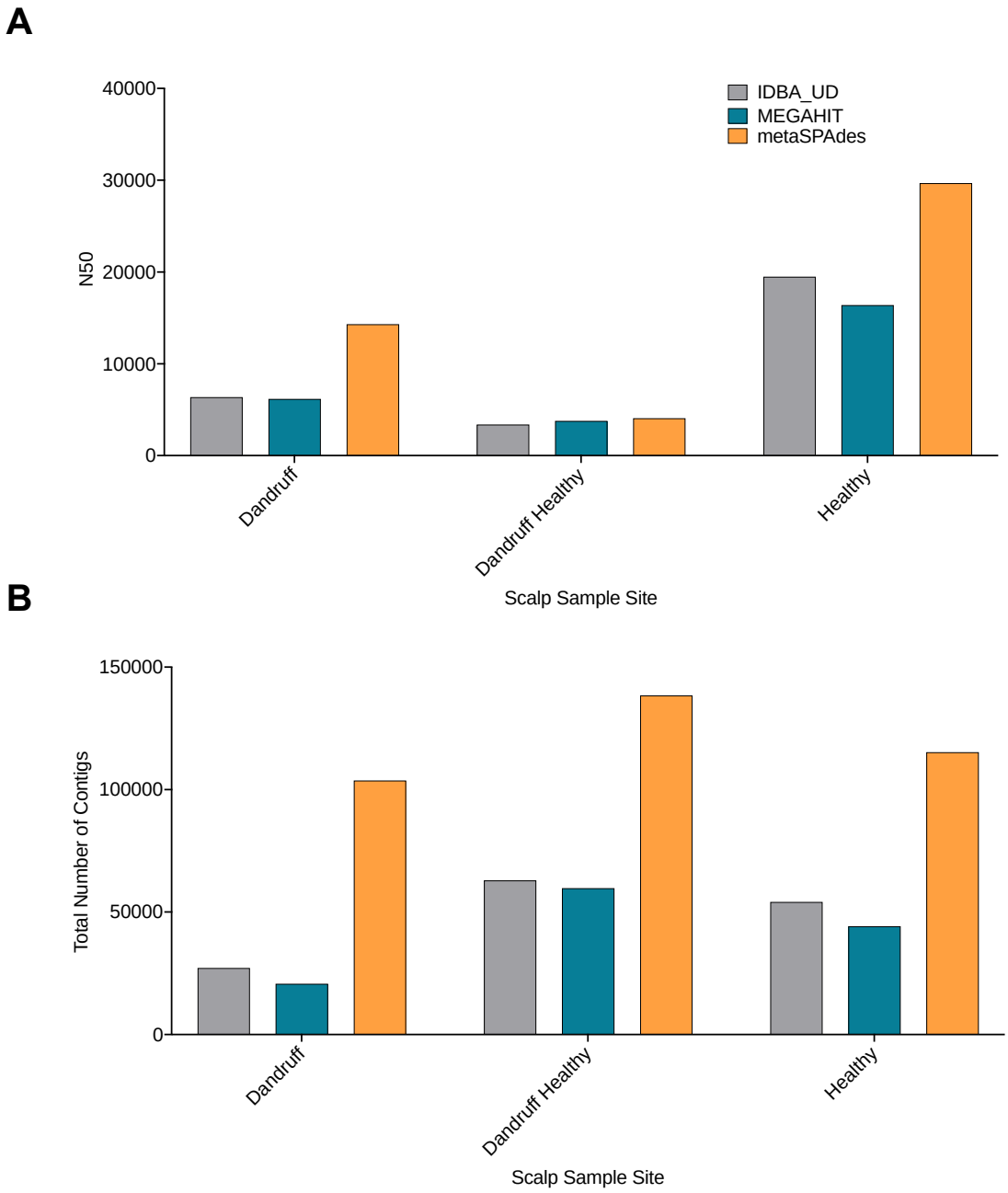


Figure 2.7 Comparison of metagenomic assembler performance on clustered scalp samples. Comparison of assembly quality (A: N50 and B: Total number of contigs) of metagenomic samples assembled via scalp sample state using three different metagenomic assemblers: IDBA-UD, MEGAHIT and metaSPAdes.

2.3.5.2 Binning contigs

Metagenomic shotgun sequencing enables the reconstruction of genomes from complex microbial communities. However, assembling metagenomes does not reconstruct entire genomes but rather produces fragments as contigs derived from the many species sampled in these communities. The process of 'binning' groups those contigs derived from the same genome together in a bin (Tyson et al., 2004, Alneberg et al., 2014). Here, binning was performed unsupervised, whereby contigs from the same genome are clustered together based upon their shared properties. The programs MetaBAT2 and CONCOCT were used to bin contigs.

CONCOCT uses a variational Bayesian approach combined with the Gaussian mixture models (GMM) to cluster contigs into genomes based on their tetranucleotide frequencies and their average coverage over multiple samples. Starting from a large number of clusters, the program selects only those necessary to explain the data (Alneberg et al., 2014).

MetaBAT2 software measures the probabilistic distance of tetra-nucleotide frequency (TNF) between a contig pair, then determines whether the contig pair originates from the same genome or not using a precomputed posterior probability distribution model (Kang et al., 2015b). MetaBAT2 relies on the assumption that contigs originating from the same genome should share similar TNF and sequence coverage (Kang et al., 2019).

Although both CONCOCT and metaBAT2 produced genome bins with ≥70 % completeness and ≤10 % redundancy for all scalp state datasets, it was determined that CONCOCT clustered contigs into bins with greater completeness

and lower redundancy compared with metaBAT2 (Table.2.4.). Both binning programs were able to resolve species variation for multiple genomic bins, such as *M. restricta, Lawsonella clevelandensis* and *C. acnes*, whilst other bins were harder to differentiate for species.

Table 2.4 Genome bin completeness, redundancy and taxon associations. Genomic bins were created with CONCOCT and MetaBAT2 for each scalp sample state. Bins were filtered for ≥70 % completeness and ≤10 % redundancy.

Binning algorithm	Scalp sample state	Bin	Taxon	Number of Contigs	N50	Completeness (%)	Redundancy (%)
CONCOCT	Dandruff	C_DD_Bin_1	*Malassezia restricta*	1050	6581	74.70	6.02
		C_DD_Bin_2	Unknown *Actinobacteria*	18	247034	98.59	1.41
		C_DD_Bin_3	Unknown *Staphylococcus*	16	290613	100	1.41
	Dandruff Healthy	C_DH_Bin_1	Unknown *Enterobacteriaceae*	73	146534	100	0
		C_DH_Bin_2	Unknown *Staphylococcus*	82	63601	100	1.41
		C_DH_Bin_3	*Lawsonella clevelandensis*	107	21759	100	2.82
		C_DH_Bin_4	Unknown *Actinobacteria*	450	6029	90.14	9.86
		C_DH_Bin_5	*Malassezia sympodialis*	324	24729	77.11	3.61
		C_DH_Bin_6	*Malassezia restricta*	1350	4535	72.29	4.82
	Healthy	C_HH_Bin_1	*Lawsonella clevelandensis*	25	196808	100	0
		C_HH_Bin_2	Unknown *Staphylococcus*	71	84107	97.18	2.82
		C_HH_Bin_3	*Malassezia restricta*	814	11825	85.54	3.61
		C_HH_Bin_4	*Malassezia globosa*	682	19911	85.54	2.41
		C_HH_Bin_5	*Malassezia sympodialis*	53	363903	83.13	1.20

		C_HH_Bin_6	*Malassezia sympodialis*	204	71916	81.93	6.02
		C_HH_Bin_7	*Cutibacterium acnes*	567	4163	80.28	9.86
		C_HH_Bin_8	*Malassezia sympodialis*	313	46577	79.52	3.61
		C_HH_Bin_9	*Cutibacterium granulosum*	428	4213	77.46	4.23
MetaBAT2	Dandruff	M_DD_Bin_1	Unknown *Staphylococcus*	11	324698	91.55	1.41
		M_DD_Bin_2	*Malassezia restricta*	1207	6408	79.52	7.23
		M_DD_Bin_3	Unknown *Actinobacteria*	18	247034	98.59	1.41
	Dandruff Healthy	M_DH_Bin_1	Unknown *Actinobacteria*	421	6259	88.73	9.86
		M_DH_Bin_2	*Lawsonella clevelandensis*	44	26944	88.73	0
		M_DH_Bin_3	Unknown *Staphylococcus*	62	63601	100	1.41
	Healthy	M_HH_Bin_1	*Cutibacterium acnes*	407	4458	70.42	4.23
		M_HH_Bin_2	Unknown *Staphylococcus*	63	84107	71.83	2.82
		M_HH_Bin_3	*Malassezia globosa*	601	20876	83.13	6.02
		M_HH_Bin_4	Unknown *Staphylococcus*	63	84107	71.83	2.82
		M_HH_Bin_5	*Cutibacterium acnes*	407	4458	70.42	4.23

2.3.6 Functional annotation of the scalp microbiome

Protein sequences extracted from dandruff, dandruff healthy and healthy metagenomes were functionally annotated using GhostKOALA (Kanehisa et al., 2016). Amino acid sequences were assigned KEGG Orthology numbers, based upon the gene functions and KEGG pathways they mapped to. The relative abundance of each KEGG pathway module for each scalp sample was compared across each scalp state. Carbohydrate metabolism was found to the be the most abundant KEGG pathway module found within all scalp samples, but overall, there is no over or under representation of any one KEGG pathway module in relation to dandruff and healthy scalp states.

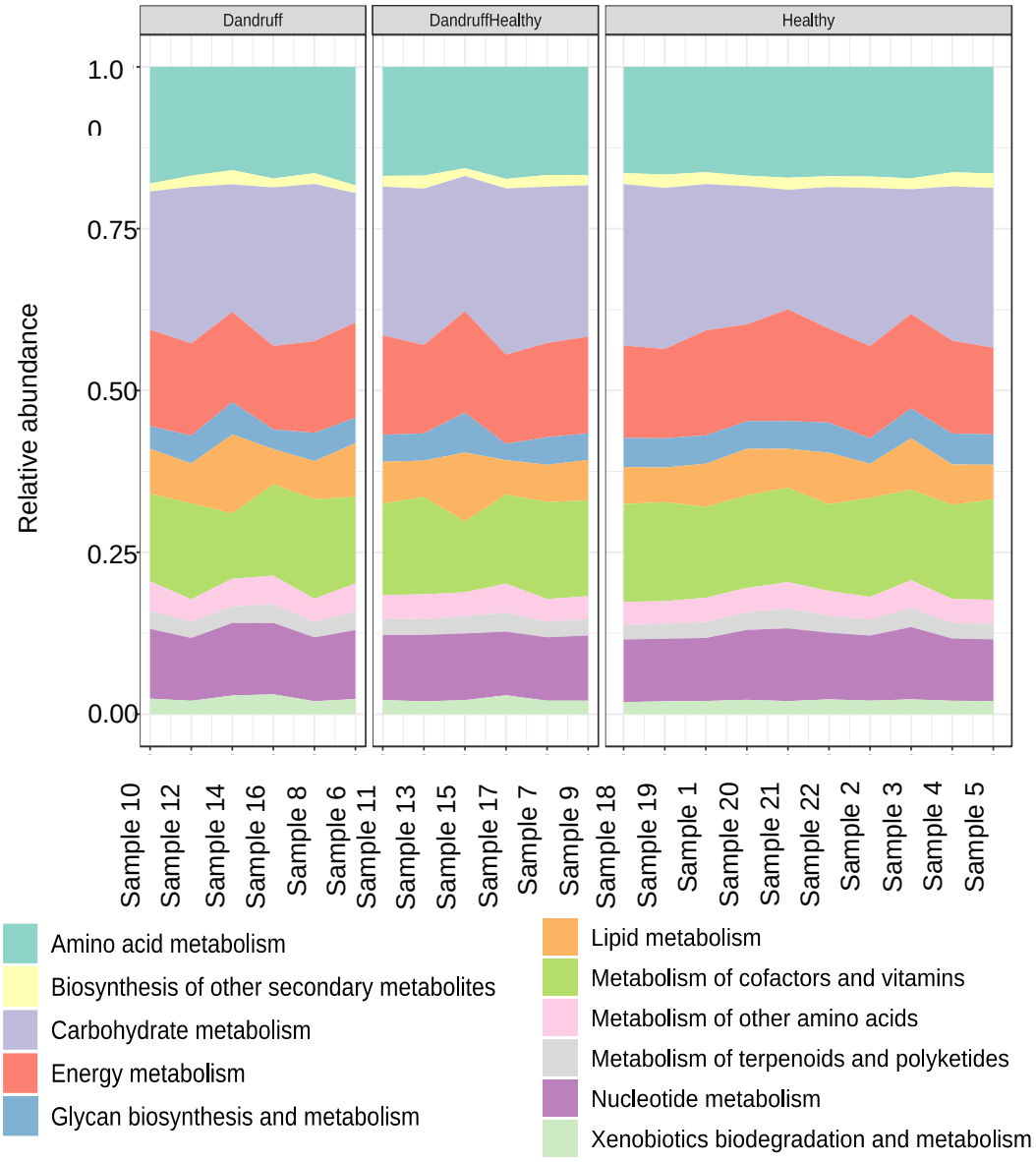


Figure 2.8 Functional composition of the scalp microbiome. GhostKOALA assigned KEGG Orthology of metagenomes from dandruff, dandruff healthy and healthy scalp states, shown per KEGG pathway module.

2.3.7 Understanding variation in scalp microbiome

2.3.7.1 Correlation of microbial species with scalp physiology

2.3.7.1.1 LEfSe analysis of enriched microbial clades

To understand which taxonomic groups were most likely to account for the causation of dandruff, relative abundance profiles of all samples created in MetaPhlAn3 (see section 2.3.4.2) were analysed by LEfSe, a linear discriminant analysis (LDA) effect size tool. LEfSe uses both a Kruksal-Wallis sum-rank test and a Wilcoxon test to identify microbial clades significantly associated with dandruff (DD), dandruff healthy (DH) and healthy (HH) samples. LDA effect size then estimates the degree of responsibility associated with each microbial clade with respect to scalp state (Segata et al., 2011).

LEfSe identified 7 microbial clades that were significantly enriched within microbial communities of healthy scalps, with an LDA score of 4.26 or above, indicating these groups were significantly associated with a healthy scalp (Figure 2.9). Only 1 microbial clade was found to be significantly associated with microbial communities present on healthy sections of the scalp of subjects with dandruff sites, while no microbial clades were significantly associated with microbial communities found on dandruff scalps (Figure 2.9).

Seven clades were associated with healthy scalps, though only three clades were classified to genus-level or below, all but one significant clades were associated with the *Actinobateria* phylum highlighting its dominant contribution to the skin microbiota of a healthy scalp. In addition, the major human commensal *Staphylococcus epidermidis* was also associated with healthy scalp sites of non-dandruff scalps. Healthy sites on dandruff scalps were significantly associated

with a single taxon compared with healthy scalps and dandruff sites, the nasopharynx and oral commensal bacterium, *Streptococcus mitis*.

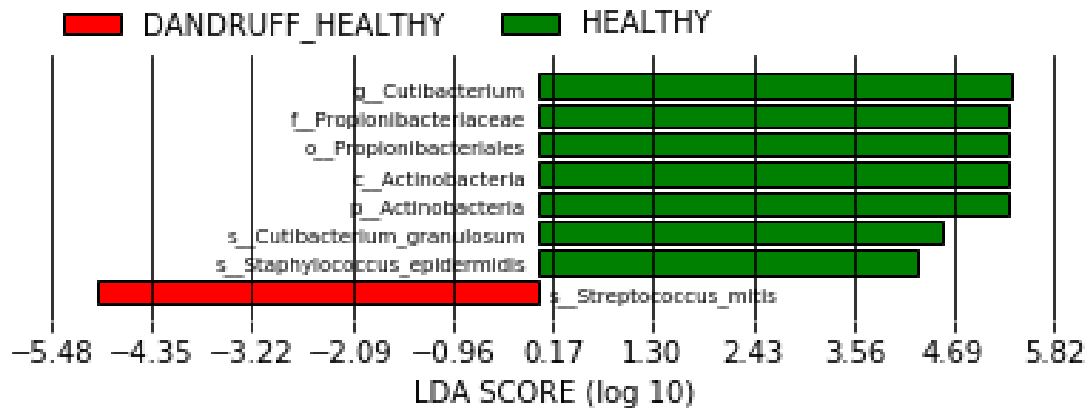


Figure 2.9 LEfSe analysis of enriched microbial clades. Significantly enriched microbial clades within either dandruff (DD), dandruff healthy (DH) or healthy (HH) datasets. Analysis and image were generated using LEfSe. P, C, O, F, G and S refer to Phylum, Class, Order, Family, Genus and Species.

2.3.7.1.2 Alpha and beta diversity analysis

The scalp microbiome from 10 subjects with a healthy (HH) scalp site and 6 subjects with both a healthy (DH) scalp site and a dandruff scalp site (DD) was established by classifying metagenomic shotgun sequence reads using Kraken v2.0 (section 2.3.4). Taxonomic classifications made by Kraken v2.0 were used to create a count table based on the number of reads estimated to belong to different taxa. These counts were then used to analyse alpha and beta diversity of scalp samples.

Analysis of alpha diversity (species richness) was made for samples of each scalp state (HH, DH, DD), based on their observed OTU number, Shannon index and Simpson index. Species richness represents the number of distinct species within a microbial community, not taking into account the abundance of each species. To generate meaningful comparisons of species richness between each scalp sample, rarefaction analysis was carried out to understand how comprehensively species diversity had been represented within each community. Rarefaction analysis calculates the average species richness represented by increasing numbers of individual microbes. Due to the nature of the samples, filtering of low quality and human contaminant reads has led to a reduction in the number of reads per dataset (Figure 2.1). Therefore, curves generated for all datasets did not level off, indicating that additional reads would lead to the identification of new species-level classifications such that estimates of species richness could be comprehensively represented (Figure 2.10).

The healthy scalp microbiome had the lowest alpha diversity within the Simpson index, observed OTU number and Shannon diversity metrics. The dandruff scalp microbiome had a greater alpha diversity compared with healthy scalp microbiota in all indices. The healthy site on a dandruff scalp microbiome was found to have the highest alpha diversity in the Simpson index and observed OTU number. Whereas DH alpha diversity was found to be between that of dandruff and healthy with the Shannon diversity metrics (Figure 2.11). Detailed median values for each scalp sample site (DD, HH & DH) are reported in Figure 2.11.

To further understand the differences between the microbiota found on each of the scalp sites, statistical analyses were carried out on the alpha diversity values based on the three different matrices. There were no statistical differences between the species richness of each scalp state, within either the observed OTU (range of medians between 4200-5300), Shannon (range of medians between 3-3.15) or Simpson (range of medians between 0.92-0.94) indices, as determined by one-way ANOVA (Observed OTU, p= 0.08; Shannon index, p= 0.971; Simpson index, p= 0.794). Tukey post hoc tests also revealed no significant differences between the three scalp states (Figure 2.11).

The overall structural similarity and variation (beta diversity between the samples) between the microbiomes from the three scalp sites were then examined using the Bray Curtis distance matrix. Assessment of these differences by a Permutational Multivariate Analysis of Variance (PERMANOVA), using the "adonis" function in R revealed no evidence of differences between samples being based on scalp sample site ($p > 0.05$, $R_2 = 0.05$ PERMANOVA). The resulting NMDS

and PCoA plots of all scalp samples based on the Bray-Curtis distance matrix showed a strong overlap of samples at each of the three scalp sites (Figure 2.12). An NMDS stress value <0.05 (stress = 0.03) indicates that the difference between the distances in the reduced dimension compared to the complete multidimensional space of samples is optimised (Figure 2.12).

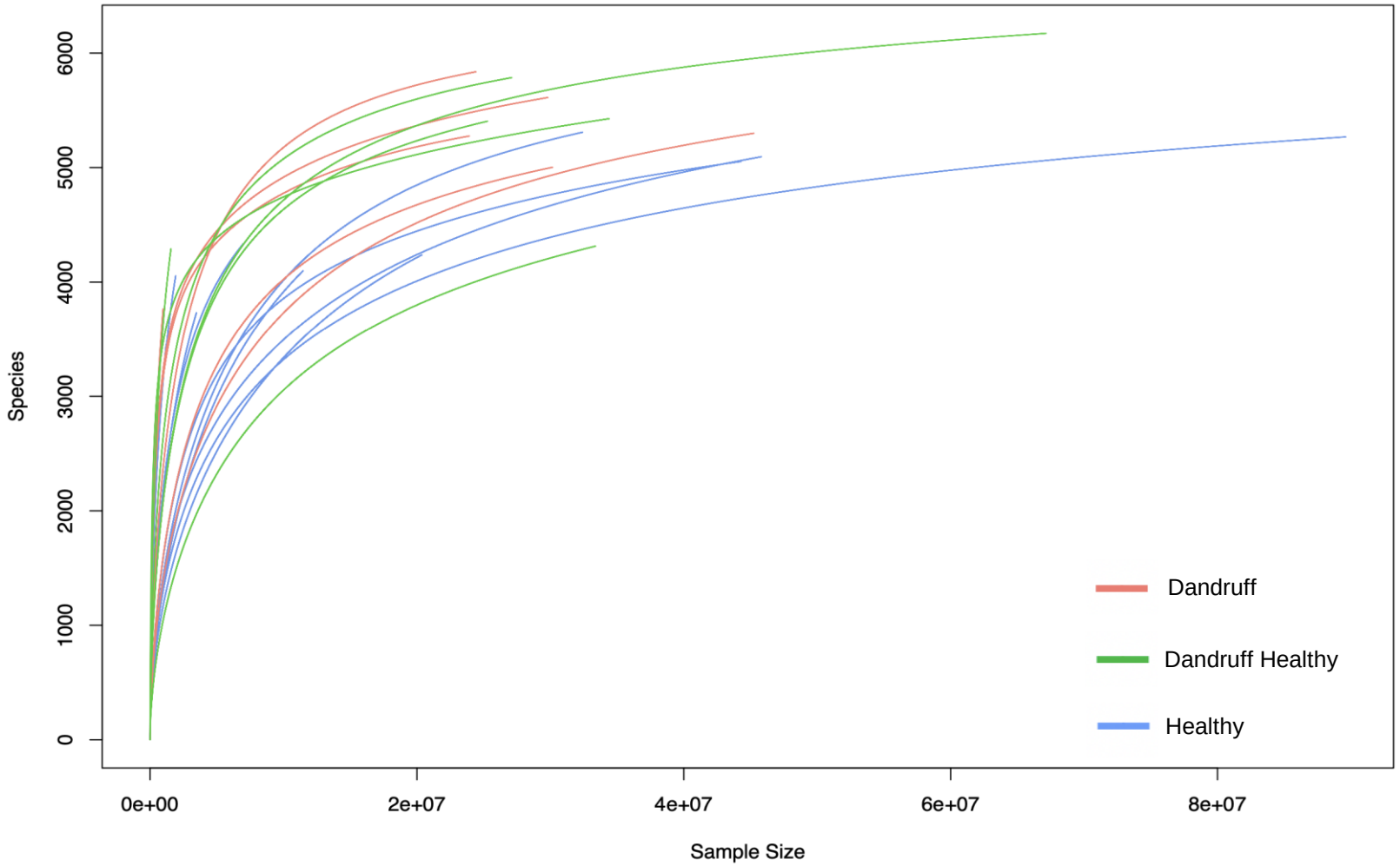


Figure 2.10 Rarefaction analysis of 22 scalp samples. Species richness is estimated as a function of the number of samples. Rarefaction analysis was calculated using the "vegan" package in R.

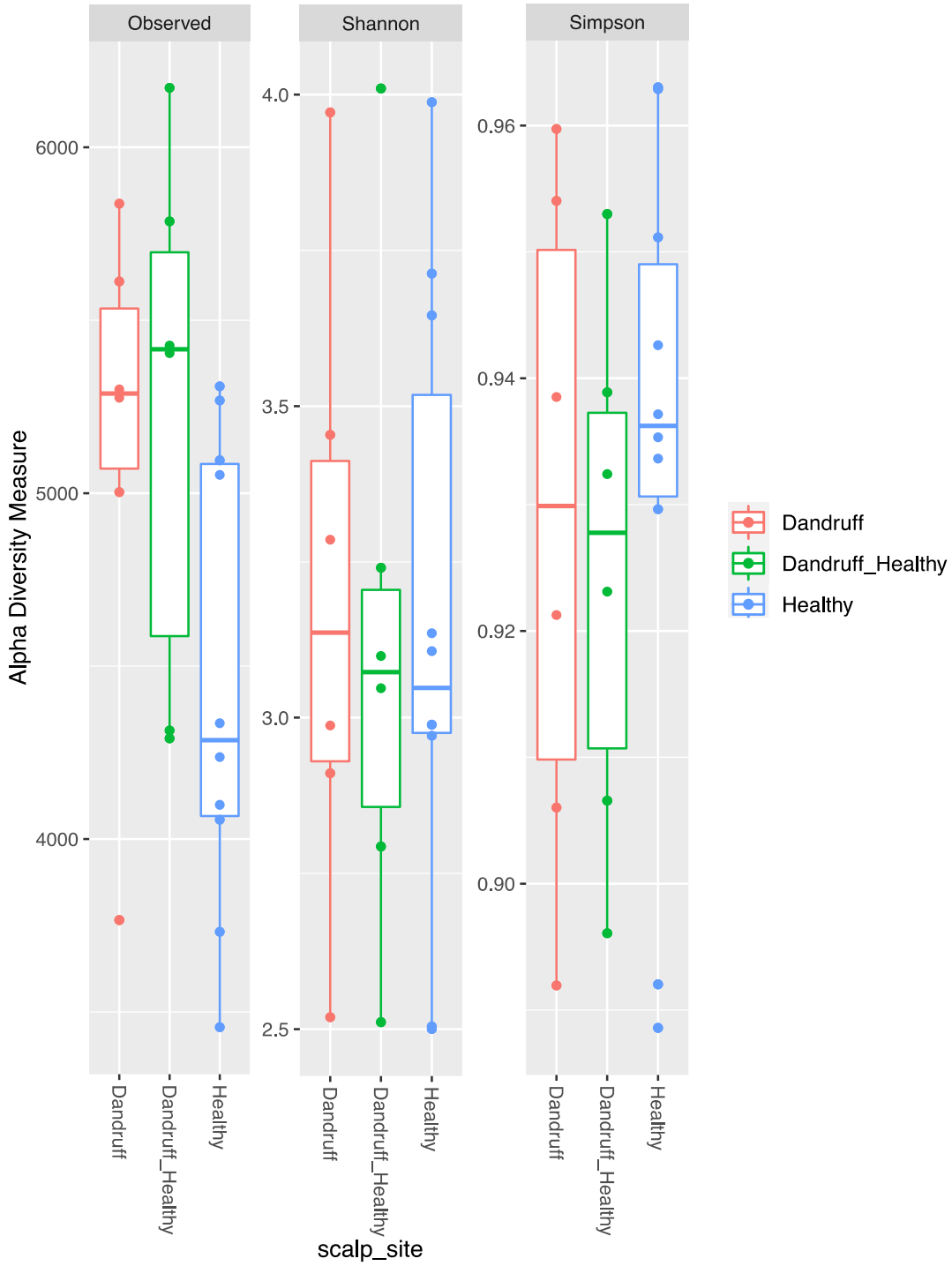


Figure 2.11 Comparison of alpha diversity of scalp microbiomes from different scalp states. Differences in alpha diversity in each skin microbiome of dandruff, dandruff healthy and healthy scalps (n=6, dandruff; n=6, dandruff healthy; n=10, healthy) are shown by observed OTU number, Shannon index and Simpson index. Line in box is a median of index scores, boxes represent interquartile range, whiskers represent lowest and highest values and dots represent samples and outliers.

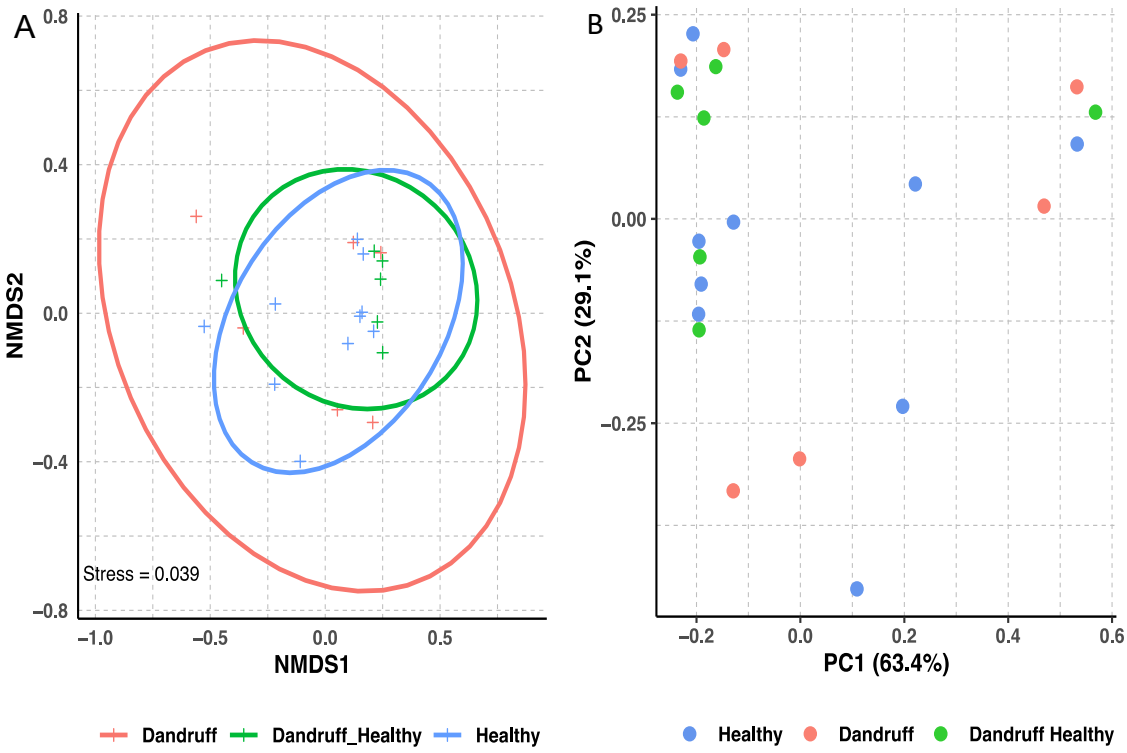


Figure 2.12 Ordination plots of scalp sample beta diversity. A: Non-metric multidimensional scaling plot of all scalp samples based on the Bray-Curtis distance matrix. The plot shows the two main dimensions of variation, with points colour coded according to scalp sample site. The model stress value was 0.03. B: Principal coordinates analysis of all samples using Bray-Curtis distance matrix, calculated using R package ape PCOA function, with the first 2 principal components plotted against each other. Points represent scalp samples. Clustering according to scalp sample type was not significant ($p > 0.05$, $R_2 = 0.05$ PERMANOVA).

2.3.7.1.3 Statistical analysis to identify taxa significantly enriched between scalp samples

The DESeq2 package was used to identify taxa that were significantly different between dandruff, dandruff healthy and healthy scalp samples, (Love et al., 2014). Read count files created from Kraken v2.0 taxonomic classifications were first input into the R package "DESeq2" and read counts were normalised using the DESeq2 median of ratios method, which divides read counts by sample-specific size factors determined by median ratio of gene counts and relative to a geometric mean per OTU (Figure 2.13).

Normalised read counts were then analysed with the DESeq function. DESeq2 fits a negative binomial generalised linear model for each OTU and uses the Wald test for significance testing. Significance testing was done in a pairwise manner, comparing each of the three scalp states against another. While there were no significant differences found in the microbiome composition in terms of overall diversity (Section 2.3.7.1.2), there were also no specific taxa that exhibited significant differences in abundance between the three scalp states (p >0.05). To investigate the overall effect of scalp sample state on samples, a PCA plot was used to visualise the samples. The resulting plot shows no clustering of samples by scalp state and no obvious outliers (Figure 2.14).

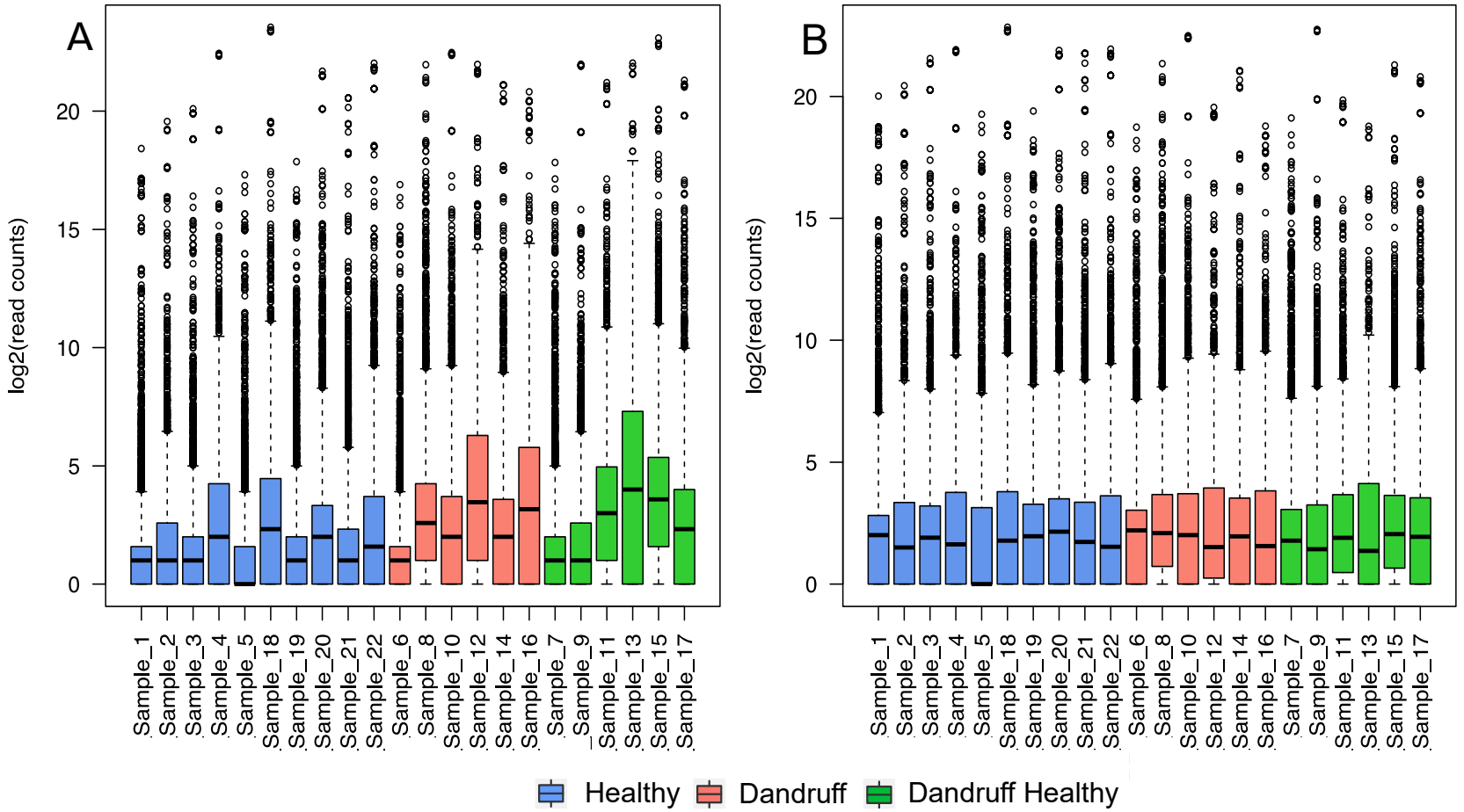


Figure 2.13 DESeq2 normalisation of read counts from Kraken v2.0 based taxonomic classifications. A: Log_2 read counts per sample prior to DESeq2's median of ratios normalisation. B: Log_2 read counts divided by sample-specific size factors determined by median ratio of gene counts relative to geometric mean per OTU, following normalisation.

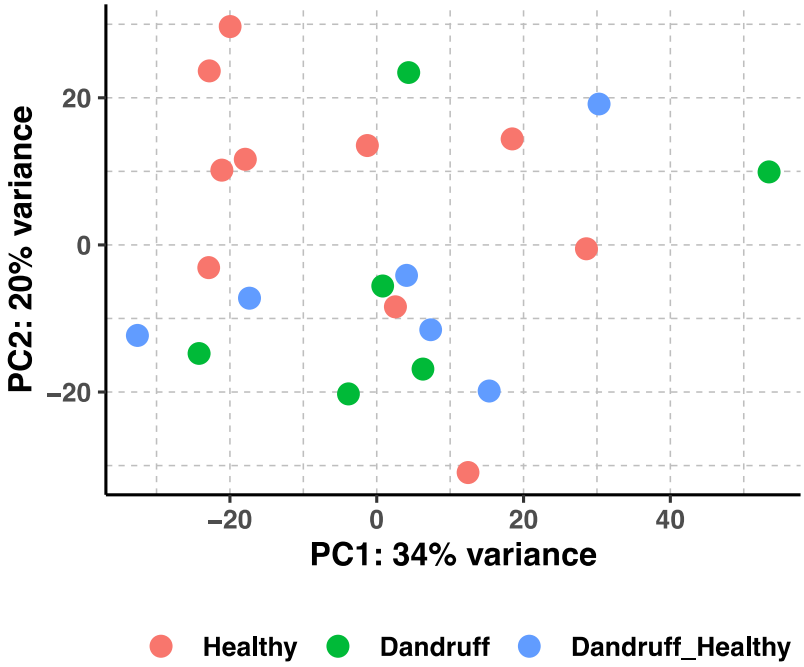


Figure 2.14 Clustering of sample-sample diversity in DESeq2. Principal Coordinates Analysis (PCA) plot of scalp samples in DESeq2, with the first 2 principal components plotted against each other. Points represent scalp samples, showing clustering of samples via scalp sample state.

2.3.7.2 Functional Analysis of the scalp microbiome

2.3.7.2.1 Statistical analysis of global differences in the functional profiles of the scalp microbiome

Functional annotation of each metagenome was achieved using the Anvi'o v6.2 platform (Eren et al., 2015). Protein sequences were extracted from Anvi'o profiles, each of which were created from assembled metagenome contigs as described in section 2.2.6.1. After input into GhostKOALA, KEGG Orthology (KO) was assigned to metagenome sequences to characterise individual gene functions and reconstruct KEGG pathways (Kanehisa et al., 2016). These functional annotations were used to create a count table based on the number of reads estimated to map to each KEGG functional ortholog. Counts were used to analyse statistical differences in the functional potential of each scalp state.

The overall similarity and variation in functional profiles between the microbiomes from each scalp state were examined using the Bray Curtis distance matrix. Assessment of these distances by a Permutational Multivariate Analysis of Variance (PERMANOVA) using the "adonis" function in R revealed no significant functional differences between the microbiomes of dandruff (DD), dandruff healthy (DH) and healthy (HH) scalp sites ($p > 0.05$, $R_2 = 0.10$ PERMANOVA).

Subsequently NMDS (Non-metric multidimensional scaling) analysis and PCOA (Principal Coordinate Analysis) was done using the Vegan R package (Jari Oksanen, 2019) (Figure 2.15). The resulting analysis plots based on the Bray-Curtis distance matrix showed a strong overlap of samples from each scalp state,

with an absence of clustering based on scalp state. An NMDS stress value <0.05 (stress = 0.04) indicates that the difference between the distances in the reduced dimension compared with the complete multidimensional space of samples is optimised (Figure 2.15).

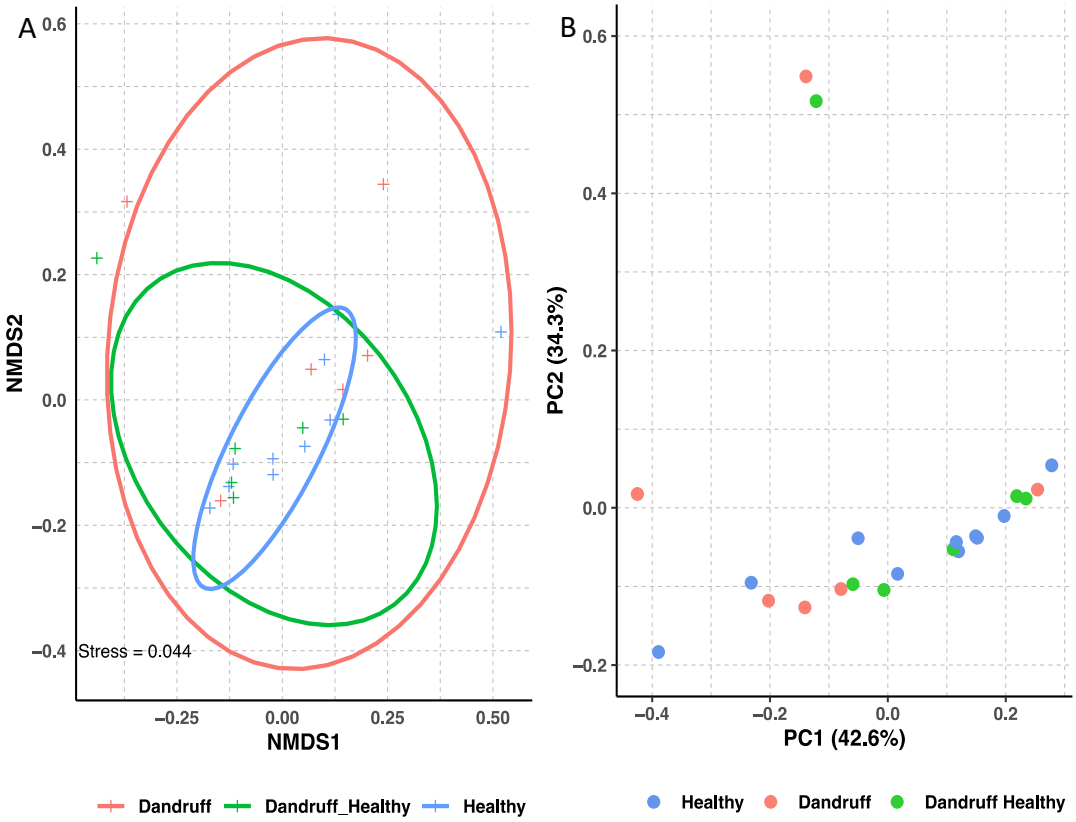


Figure 2.15 Ordination plots of scalp sample functional diversity. A: Non-metric multidimensional scaling plot of all scalp samples based on the Bray-Curtis distance matrix. The plot shows the two main dimensions of variation, with points colour coded according to scalp sample site. The model stress value was 0.04. B: Principal coordinates analysis of all samples using Bray-Curtis distance matrix, calculated using R package ape PCOA function, with the first 2 principal components plotted against each other. Points represent scalp samples. Clustering according to scalp sample type was not significant (p > 0.05, PERMANOVA).

2.3.7.2.2 Identifying functional differences between dandruff, dandruff healthy and healthy scalps

The DESeq2 package was used to investigate whether KEGG Orthology modules differed between the three scalp sample sets (Love et al., 2014). Read count data were normalised using DESeq2's median of ratios method, which divides read counts by sample-specific size factors determined by median ratio of gene counts, relative to a geometric mean per gene (Figure 2.16).

Normalised read counts were analysed using DESeq2 default settings that fit a negative binomial generalised model for each gene and uses the Wald test. Significance testing was done in a pairwise manner, comparing each of the three scalp states against one another. No significant differences were found in the functional profiles globally by Permutational Multivariate Analysis (PERMANOVA) based on Bray-Curtis dissimilarity matrices (Section 2.3.7.2.1). However, significant differences were identified between the functional profiles of samples from dandruff and dandruff healthy sites when compared with healthy scalp sites.

There were 46 KEGG orthologs identified that were differentially abundant in the microbiota found on either dandruff or dandruff healthy scalp sites, compared with healthy scalp sites (Adjusted *p*-value <0.05). There were no differentially abundant KEGG orthologs between dandruff and dandruff healthy microbiomes (Adjusted *p*-value >0.05) (Figure 2.17 & Table 2.5). Six KEGG orthologs were more abundant in species found on both dandruff healthy (DH) and dandruff (DD) when compared with healthy (HH) scalp sites (Adjusted *p*-value <0.05)

(K03566, K04080, K08194, K23107, K01993 and K06915). The functions of these six orthologs include biofilm formation, metabolism, transcription and cellular processes. Thirty additional orthologs were differentially abundant in dandruff (DD) versus healthy (HH) state samples and ten additional orthologs were differentially abundant in dandruff healthy versus healthy state samples. These forty orthologs were ascribed to pathways linked to metabolism, biosynthesis of secondary metabolites, quorum sensing, fatty acid biosynthesis, peptidoglycan biosynthesis, antimicrobial resistance and bacterial motility. Human disease related pathway associations are likely due to incomplete curation of microbiomes (Table 2.5).

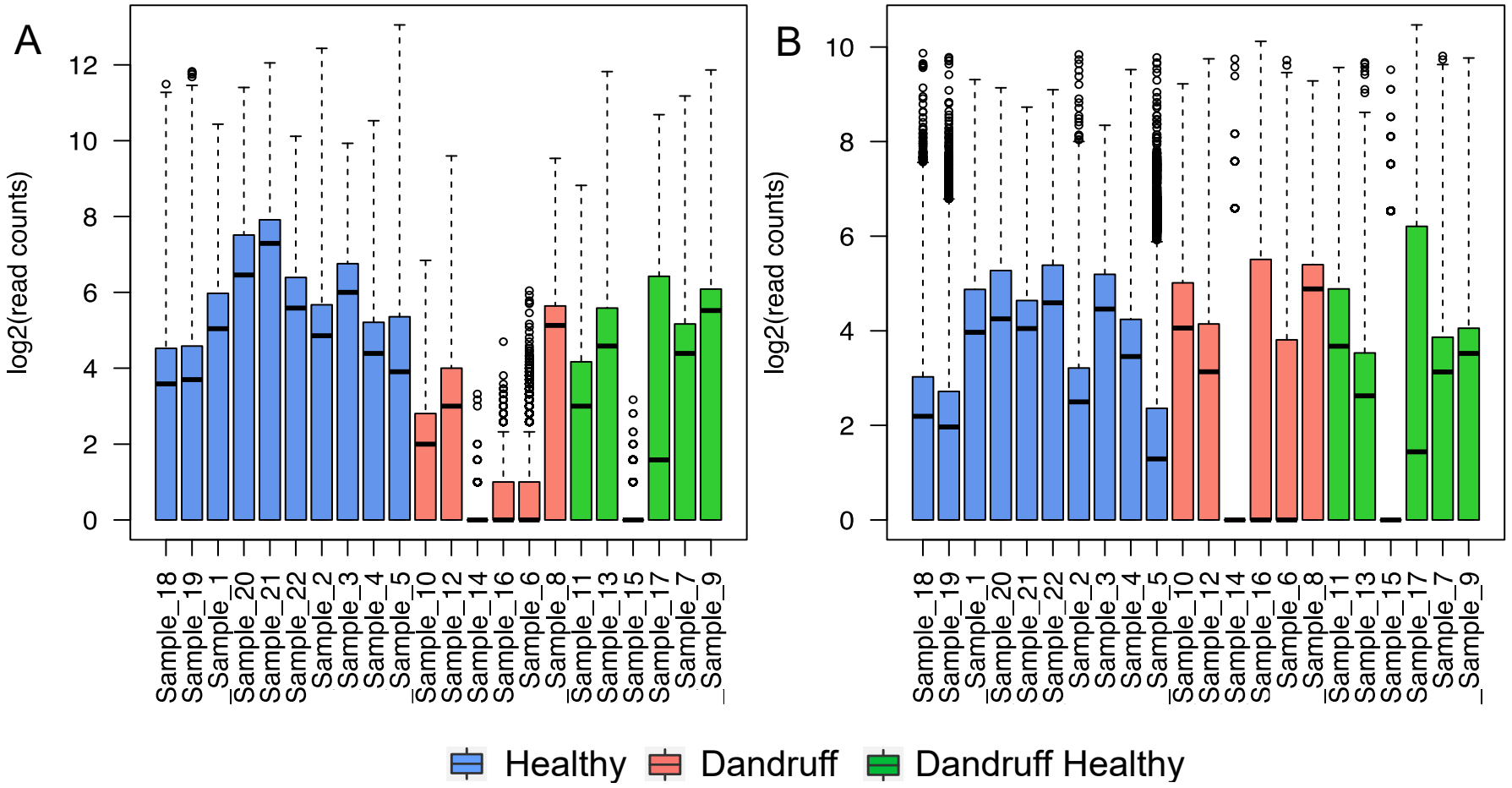


Figure 2.16 DESeq2 normalisation of read counts from functional profiles of scalp samples. A: Log_2 read counts per sample prior to DESeq2's median of ratios normalisation. B: Log_2 read counts divided by sample-specific size factors determined by median ratio of gene counts relative to geometric mean per gene, following normalisation.

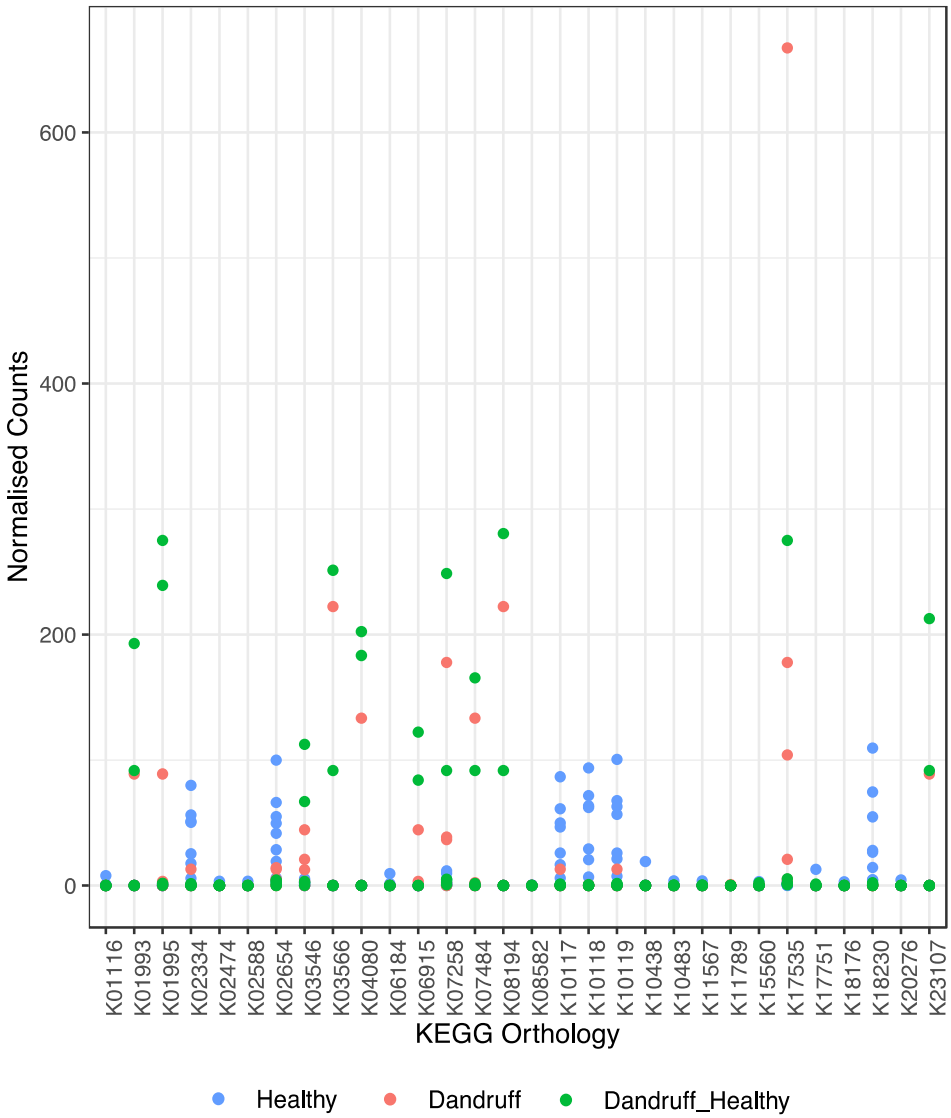


Figure 2.17 Top 30 differentially abundant KEGG Orthologs. Differentially abundant KEGG orthologs found on dandruff and dandruff healthy scalps compared to healthy scalps. Counts were normalized with DESeq2-size factors.

Table 2.5 Significant differentially abundant KEGG functional orthologs. KEGG functional orthologs that are significantly differentially abundant on dandruff (DD) and dandruff healthy (DH) scalp sites compared to a healthy (HH) scalp site. Log_2 fold change highlights the abundance increase or decrease between scalp states and lfcSE values show the standard error of each estimate. Wald test p-values were adjusted with the Benjamin-Hochberg method.

	KO Number	log_2 Fold Change	lfcSE	Adjusted P value	KO Title and Description
Dandruff vs Healthy	K03566	24.06	3.58	4.52E-08	LysR family transcriptional regulator, glycine cleavage system transcriptional activator
	K08194	24.05	3.58	4.52E-08	MFS transporter, ACS family, D-galactonate transporter
	K04080	23.30	3.59	1.38E-07	Molecular chaperone IbpA
	K01993	22.71	3.60	2.61E-07	HlyD family secretion protein
	K23107	22.70	3.60	2.61E-07	1-deoxyxylulose-5-phosphate synthase
	K17535	7.15	1.16	5.81E-07	TNNI3K; serine/threonine-protein kinase TNNI3K
	K06915	21.78	3.61	1.05E-06	K06915; uncharacterized protein
	K11789	5.05	0.93	3.85E-05	DDB1- and CUL4-associated factor 1
	K17751	4.78	0.97	0.0005	Myosin heavy chain 6/7
	K02474	4.46	0.95	0.0010	UDP-N-acetyl-D-glucosamine/UDP-N-acetyl-D-galactosamine dehydrogenase
	K02588	4.45	0.95	0.0010	Nitrogenase iron protein NifH
	K10438	4.45	0.95	0.0010	3-hydroxyphenylacetate 6-hydroxylase

	K20276	4.45	0.95	0.0010	bapA; large repetitive protein
	K18176	4.18	0.94	0.0031	Cytochrome c oxidase assembly factor 3, fungi type
	K08582	3.91	0.93	0.0079	CAPN15; calpain-15
	K15560	5.20	1.24	0.0079	NRD1; protein NRD1
	K10483	3.94	0.95	0.0092	BTB/POZ domain-containing protein 11
	K01116	3.96	0.96	0.0097	Phosphatidylinositol phospholipase C, gamma-1
	K06184	3.97	0.97	0.0097	ATP-binding cassette, subfamily F, member 1
	K11567	3.97	0.97	0.0097	DASH complex subunit DAD2
	K22653	4.89	1.21	0.0124	NPTN; neuroplastin
	K13295	4.06	1.01	0.0126	APLF; aprataxin and PNK-like factor
	K12994	3.67	0.93	0.0166	Mannosyl-N-acetyl-alpha-D-glucosaminyl-diphospho-ditrans, octacis-undecaprenol 3-alpha-mannosyltransferase / alpha-1,3-rhamnosyltransferase
	K16211	3.70	0.95	0.0194	maltose/moltooligosaccharide transporter
	K01447	3.72	0.96	0.0206	N-acetylmuramoyl-L-alanine amidase
	K13240	3.79	1.00	0.0292	NOS1; nitric-oxide synthase
	K02011	3.63	0.98	0.0300	Iron (III) transport system permease protein
	K04850	3.78	1.01	0.0300	Voltage-dependent calcium channel L type alpha-1C

	K10380	5.54	1.47	0.0300	ANK; ankyrin
	K11981	3.78	1.01	0.0300	E3 ubiquitin-protein ligase NRDP1
	K14442	3.78	1.01	0.0300	ATP-dependent RNA helicase DHX36
	K18027	3.79	1.02	0.0315	Tyrosine-protein phosphatase non-receptor type 3
	K22184	4.97	1.36	0.0384	Bromodomain-containing protein 9
	K07512	3.27	0.91	0.0424	Mitochondrial enoyl-[acyl-carrier protein] reductase / trans-2-enoyl-CoA reductase
	K15588	3.53	0.98	0.0457	[histone H3]-lysine36 N-dimethyltransferase NSD1
	K03753	3.76	1.05	0.0469	Molybdopterin-guanine dinucleotide biosynthesis adapter protein
Dandruff Healthy vs Healthy	K03566	24.70	3.57	5.89E-09	LysR family transcriptional regulator, glycine cleavage system transcriptional activator
	K04080	24.86	3.57	5.89E-09	Molecular chaperone IbpA
	K08194	24.82	3.57	5.89E-09	MFS transporter, ACS family, D-galactonate transporter
	K01993	24.46	3.57	5.92E-09	HlyD family secretion protein
	K23107	24.55	3.57	5.92E-09	1-deoxyxylulose-5-phosphate synthase
	K06915	24.07	3.57	1.07E-08	K06915; uncharacterized protein
	K01995	7.75	1.65	0.0015	Branched-chain amino acid transport system ATP-binding protein
	K07484	6.63	1.46	0.0028	Transposase

	K02654	-3.60	0.80	0.0030	N-methyltransferase
	K10118	-6.01	1.36	0.0041	Raffinose/stachyose/melibiose transport system permease protein
	K02334	-5.18	1.31	0.0296	Dpo; DNA polymerase bacteriophage-type
	K07258	4.06	1.06	0.0407	Serine-type D-Ala-D-Ala carboxypeptidase (penicillin-binding protein 5/6)
	K03546	4.17	1.10	0.0428	DNA repair protein SbcC/Rad50
	K18230	-4.50	1.18	0.0428	Macrolide transport system ATP-binding/permease protein
	K10117	-5.15	1.37	0.0481	Raffinose/stachyose/melibiose transport system substrate-binding protein
	K10119	-4.99	1.34	0.0481	Raffinose/stachyose/melibiose transport system permease protein

2.4 Discussion

This study provides an overview of the scalp microbiome and its association with healthy and dandruff scalps using high throughput sequencing approaches (shotgun metagenomic sequencing) in a small human cohort based in the United Kingdom.

The data presented in this investigation agree with previous taxonomic studies that described *Malassezia* species as a major component of the mycobiome of the scalp, irrespective of scalp condition. An identified ratio of *M. globosa* lower than *M. restricta* was determined in this study to be associated with a dandruff scalp, which matches the findings of some studies (Park et al., 2012, Clavaud et al., 2013, Gemmer et al., 2002, Soares et al., 2016). However, the opposite was observed in two separate studies, which could be attributable to multiple differences between studies, including cohort size, sequencing technology and bioinformatic analyses workflow and differences in subject population origin (Saxena et al., 2018, Xu et al., 2016). Aside from *M. restricta*, *Malassezia* species, including *M. sympodialis* as well as *M. globosa* were found to be in very low proportions in all samples, which was similar to other studies related to dandruff and seborrheic dermatitis with Brazilian subjects, (Soares et al., 2015).

Cutibacterium acnes was revealed as the major bacterial coloniser, found in all scalp states but most abundantly on healthy scalps compared with dandruff, and this matched previous reports (Clavaud et al., 2013, Fawley et al., 2007, Saxena et al., 2018, Wang et al., 2015, Xu et al., 2016). Other bacterial colonisers including the *Staphylococcus* species, *S. epidermidis, S. capitis* and *S. aureus* formed part of the scalp microbiome.

S. aureus was the most prevalent *Staphylococcus* species on dandruff scalps, similar to a scalp study in South Korea (Park et al., 2017). *S. capitis* was similarly more abundant on dandruff versus healthy scalps, whereas *S. epidermidis* and *C. acnes* were more prevalent on healthy scalps. This relationship was recently described by (Grimshaw et al., 2019), along with the similar observation that, *S. capitis* was not the most abundant of the staphylococci on any of the scalp sites, with *S. aureus* and *S. epidermidis* presenting increased levels on dandruff and healthy scalps respectively.

Increased prevalence of *S. epidermidis* on healthy compared with dandruff scalps, is a commonly reported finding. In the study, this species was found to be less abundant compared with *Cutibacterium* species across all scalp samples., a finding that does not support suggestions that *Cutibacterium* enhancement and suppression of *S. epidermidis* could be a potential solution to lessen dandruff (Xu et al., 2016). These differences again could be attributable to sample cohort size, differences in subject population origin and sex, and differences in sequence technology. Xu *et al.* used barcoded pyrosequencing with Roche 454 FLX+ to sequence amplicons, whereas this study utilised shotgun sequencing. Metagenomic shotgun sequencing can offer species and strain level classification and is less vulnerable to biases through sample preparation and sequencing errors, which could contribute to observed differences (Xu et al., 2016, Rausch et al., 2019b).

Among the most abundant clades in all scalp samples were species that are members of the core oral bacterial community, such as *Streptococcus, Rothia, Veillonella* and *Haemophilius*. Changes in the composition of the scalp, cheek,

forearm and forehead microbiome have been correlated to an age-related shift, with an increase in oral bacterial flora being found on subjects over the age of 37, with aging skin (Shibagaki et al., 2017). That age range is similar to the subjects enrolled in this study and could account for the increased species richness and diversification of the sampled microbiomes.

In the current study, description was made of the microbial composition of scalp microbiota present on dandruff and healthy scalps based on taxonomy using two different classifiers. Both Kraken v2.0 (Wood et al., 2019) and MetaPhlAn v3.0 (Segata et al., 2012) were used to classify reads to compare results between mapping metagenomic reads to a curated database of clade-specific marker genes verses mapping to complete RefSeq databases. Wood, *et al.* 2019 have argued that MetaPhlAn may classify a smaller proportion of reads due to the limited database compared to Kraken v 2.0. Nevertheless, both methods produced results consistent with known scalp microbiota. However, not all results correlated with those in the literature and no significant differences were found in the frequency of taxa of the scalp microbiome in all cases. While LEfSe analysis identified enriched microbial clades of a healthy scalp microbiome: *S. epidermidis* and members of the phylum *Actinobacteria,* including *C. gransulosum;* plus, *S. mitis* as an enriched clade of a dandruff scalp microbiome; these results were not mirrored in the statistical analysis using R (Team, 2010, Segata et al., 2011).

Microbial content can vary across samples from the same environment and this can complicate the detection of statistically significant and biologically meaningful differences among small sets of samples (Quince et al., 2017). A major

limitation of this study is the small study size and uneven number of dandruff versus healthy samples that affects the statistical power of conclusions drawn from the analyses. In addition, results from this study are skewed due to the lack of diversity within the subjects sampled as all subjects were women, between the ages 20-45 years and residents of a small area of the UK. Human microbiome data is influenced heavily by host genotype, age, diet, and environment. A more diverse study over multiple time points could have given more power to the comparisons made in this study and made comparison with other studies more enlightening.

A further drawback of the study was the loss of significant amounts of sequence data in dandruff samples. Due to the nature of human metagenomics, filtering steps were undertaken to remove human host contaminating reads, which left dandruff and dandruff healthy site samples with fewer sequence reads compared with healthy samples. This difference was to be expected because dandruff samples contain more human matter as a result of the flaking on the scalp surface, but additional sequence reads would likely have led to the identification of new species-level classifications as seen in rarefaction analysis (Figure 2.10).

This study would also have benefitted from sequencing the control sample extracted from collection buffer (Phosphate buffered saline + 0.1 % Triton-X-100), which could have highlighted any possible contamination from the DNA extraction processes used as well as the sequencing itself. Previous studies have shown that such contamination can critically impact the results obtained from samples containing low microbial biomass (Salter et al., 2014).

The two main methodologies for analysing microbiome data include read classification, and metagenomic assembly. Currently, there is a lack of consensus on the performance of different metagenomic assembly software (Sczyrba et al., 2017). Specifically, the variation in metagenomic samples, such as the microbial community structure and complexity of datasets, as well the chosen sequencing technology used for sequencing, which all influence the performance of different software (Vollmers et al., 2017, Quince et al., 2017). Here, with the scalp sample datasets we compared the performance and assembly results from three De Bruijn graph assemblers: metaSPAdes, IDBA-UD and MEGAHIT.

The De Bruijn graph assembler approach is currently the popular metagenome assembly method; it involves breaking reads into overlapping sub-sequences of fixed length, known as *k*-mers and organising them into a De Bruijn graph of overlapping *k*-mers (Pevzner et al., 2001). Contigs are then assembled to reconstruct genomes by traversing a path across the graph. This task is complicated by sequencing errors and repeats which can introduce genome mis-assemblies and fragmentation. Sequencing coverage of genomes is highly variable within metagenomic data, since genome coverage corresponds to genome abundance, meaning varying levels of different species within metagenomic samples can result in highly non-uniform sequencing coverage of genomes (Quince et al., 2017, Vollmers et al., 2017). Metagenome-specific assemblers try to overcome these issues using a multiple *k*-mer approach that will work well for both low and high abundance species.

Comparatively, quality assessment of metagenome assemblies constructed with metaSPAdes produced better contiguity, with a larger N_{50} value, but the greatest

number of contigs, indicating a more fragmented assembly (Figure 2.6 & Figure 2.7). IDBA-UD and MEGAHIT generally performed similarly, with MEGAHIT presenting assemblies with the lowest N_{50} values on average across all assemblies but the lowest number of contigs. The average N_{50} and total number of contigs of assemblies produced by IDBA-UD fell between those of metaSPAdes and MEGAHIT. These results were consistent with findings from a study by Quince, *et al.* that showed MetaSPAdes generated longer contigs, which appeared to be less accurate, while MEGAHIT succeeded in reconstructing more of the true genomes that could be functionally annotated (Quince et al., 2017).

Although metagenome assemblers are optimised for metagenome reconstruction, assemblies are still highly fragmented, so in this study I chose to 'bin' contigs into species groups. Here, unsupervised binning was carried out using both CONCOCT and MetaBAT2, two unsupervised binning methods that cluster contigs based on tetranucleotide frequencies and their average coverage across samples (Alneberg et al., 2014, Kang et al., 2019).

The Genomic Standards Consortium (GSC) developed specific criteria for assessing MAG quality and these state that a genome with more than 90 % completeness and less than 5 % redundancy or contamination is regarded as a 'high-quality draft' (Bowers et al., 2017). A MAG with completeness of more or equal to 50 % with less than 10 % redundancy should be classified as a 'medium-quality draft' and a MAG that falls below both of these values is a 'low-quality draft' (Bowers et al., 2017).

Table 2.4 contains high- and medium quality draft MAGs constructed by either MetaBAT2 or CONCOCT for each scalp state. Overall CONCOCT was found to

produce 8 genome bins with more than 90% completeness compared to only 3 bins with MetaBAT2. Currently metagenome assembly quality is based on the presence and absence of single copy marker genes, which can pose problems as this is based on core genes that can be contaminated with accessory genome in final assemblies, so can be missed and contamination can be underestimated (Parks et al., 2017).

An additional problem posed in this study was the assessment of MAGs originating from fungal species, as many MAG quality assessment tools, such as checkM are based upon only bacterial and archaeal single core copy genes (Parks et al., 2015). This issue was overcome by using Anvi'o, which combines the use of HMM profiles to identify bacterial and archaeal single-copy genes (Eddy, 2011) as well as a collection of single-copy core genes curated from BUSCO (Benchmarking Universal Single-Copy Orthologs) for identification of eukaryotic genome bins (Simão et al., 2015, Eren et al., 2015).

Finally, this study provides supportive insights of the functional potential of the scalp microbiome in healthy and dandruff conditions. The fungal fatty acid biosynthesis in mitochondria pathway is enriched in the functional profile of the dandruff microbiome compared with the healthy microbiome, specifically the MECR; mitochondrial enoyl-CoA reductase (K07512). This pathway was implicated in the regulation of bioactive lipid levels such as lysophospholipids and sphingolipids (Clay et al., 2016). In addition, this fungal mitochondrial function was linked to survival of *M. restricta* on the scalp, whereby the common antifungal zinc pyrithione used to treat dandruff was proposed to inhibit mitochondrial function, increase cellular zinc levels and decrease lipase

expression in *M. restricta* (Park et al., 2018). Zinc pyrithione contributes to growth inhibition of fungi through the copper-mediated inactivation of the Fe-S cluster assembly in mitochondria (Park et al., 2018).

A KEGG functional ortholog linked to nitrogen fixation (nifH -K02588), was also significantly enriched within the dandruff microbiome. This gene is part of the microbial metabolism in diverse environments module and may be involved in enhancing symbiotic relationships between bacterial and fungal species on the scalp. Therefore, the enrichment of this gene may be linked to the enhancement of the colonisation of the scalp that includes environmental species capable of this restricted metabolism. Distinctly, staphylococci are hypothesised to occupy aerobic niches on the skin and use skin secretions as a nitrogen source (Grice and Segre, 2011).

Further bacterial genes that were found to be significantly enriched on dandruff and dandruff healthy scalps included those linked to peptidoglycan and capsular biosynthesis (K07258 & K02474). The KEGG ortholog K07258, a penicillin-binding protein, PBP4, has transpeptidase activity that contributes a high level of cross linking in the peptidoglycan structure, which may affect staphylococcal colonisation and adherence to the skin (Maya-Martinez et al., 2019). In addition to this K02474, a lipopolysaccharide biosynthesis protein with orthology to staphylococcal transferase protein CapL, is also enriched on dandruff scalps. This gene has functionality linked to the cascade of reactions in capsule biosynthesis and expression may be upregulated as a protective response to the host environment increasing pathogenesis (O'Riordan and Lee, 2004) (Rausch et al., 2019a).

Additionally, bacterial protein genes known to contribute to biofilm formation were significantly enriched on dandruff and dandruff healthy scalps (K20276, K01447 & K03566). K20276 is a LPXTG cell wall anchor domain-containing protein. This KEGG ortholog, has links to the Gram-negative protein BapA, a biofilm determinant that is also found in staphylococcal strains linked with persistent infection (Latasa et al., 2005). The LPXTG motif is conserved in many bacterial adhesin proteins that decorate peptidoglycan enhancing colonisation and biofilm production (Foster, 2019). Another enriched protein gene linked to biofilm formation and colonisation is K01447, the Atl autolysin precursor protein. Atl is a bifunctional surface protein with amidase and glucosaminidase activities that promote bacterial attachment and mediate cell wall degradation and separation during cell division (Porayath et al., 2018).

Regulatory processes such as biofilm formation and quorum sensing are key to the survival and formation of microbiomes on cutaneous sites and so the enrichment of genes linked this function on dandruff scalps is unsurprising, especially as these genes are often linked to pathogenesis (Brandwein et al., 2016, Saxena et al., 2018).

In addition to the microbial orthologous proteins that were differentially abundant in the scalp functional profiles, there were also multiple hits to human inflammation and disease proteins, despite the filtering of human reads before metagenome assembly and analysis. This is a potential downfall to using GhostKOALA to functionally annotate metagenomes since this program does not allow for very specific customisation of mapping databases and so human proteins were included. Due to the fragmented nature of metagenome assemblies

small contigs may have been wrongly assigned human function or human filtering should have been more stringent.

Overall, this study provides a comprehensive overview of the scalp microbiome and its association with healthy and dandruff scalps, specifically comparing the taxonomic and functional profiles of metagenomic samples isolated from a small human cohort based in the United Kingdom. The taxonomic composition of the scalp microbiome of both healthy and dandruff scalp sites mostly correlates with previously published literature, despite statistical testing not supporting this. Nevertheless, this study did find enriched KEGG functional orthologs that may be important to bacterial and fungal survival on the scalp, the formation of dandruff. To gain more powerful insight into the differences between the flora found on dandruff and healthy scalp a larger, more diverse study would be needed applying the extensive metagenomic and interrogative approaches used here.

CHAPTER 3

Comparative genomics of *Staphylococcus capitis* isolated from healthy and dandruff scalps

3.1 Introduction

Staphylococcus capitis was first isolated from healthy human skin in 1975 and classified as a coagulase negative *Staphylococcus* species (CoNS) (Kloos and Schleifer, 1975). Later biochemical studies divided the species into two subspecies, *capitis* and *ureolyticus* due to differences in their urease production and maltose fermentation (Bannerman and Kloos, 1991). *S. capitis* is frequently found on the human scalp and the forehead, as it thrives in lipid-rich areas where sebaceous glands are abundant (Maggs and Pennington, 1989, Gras-Le Guen et al., 2007). Although *S. capitis* is often isolated from healthy human scalps it was also associated with dandruff presenting scalps (Grimshaw et al., 2019).

Phylogenetic analysis of multi-locus DNA sequence data (16S rRNA gene, *dnaJ*, *rpoB* and *tuf* gene fragments) by Lamers *et al.,* estimated that *S. capitis* belongs to the 'Epidermidis cluster group', which comprises *S. epidermidis*, *S. saccharolyticus* and *S. caprae*. This group of staphylococci were considered skin commensal bacteria and apathogenic constituents of the healthy human skin and mucosal microbiome but are now emerging as important opportunistic pathogens (Cameron et al., 2015, Heilmann et al., 2019).

Multiple studies have linked *S. capitis* with a range of human diseases, and frequently being isolated from bloodstream infections (Cameron et al., 2015, Carter et al., 2018). *S. capitis*-associated sepsis is most frequently reported from neonatal intensive care units, causing up to 20 % of cases of neonatal sepsis

(Carter et al., 2018, Rasigade et al., 2012, Stenmark et al., 2019, Wirth et al., 2020). *Staphylococcus* spp. are known as the most common cause of prosthetic join infections, *S. capitis* are identified as the predominant CoNS isolated from these sites (Tevell et al., 2017, Flurin et al., 2019); lastly, *S. capitis* is increasingly implicated in prosthetic valve endocarditis (Nalmas et al., 2008, Cone et al., 2005, Al Hennawi et al., 2020).

The role of *S. capitis* in these infections has been studied with reference to the well-described and clinically important species within the Epidermidis cluster group, i.e., *S. epidermidis* and *S. caprae* (Cameron et al., 2015, Lamers et al., 2012, Watanabe et al., 2018). *S. capitis* encodes important virulence factors important for biofilm production, persistence and immune evasion (Cameron et al., 2015). Relatively, *S. capitis* together with other CoNS species, such as *S. epidermidis,* have a lower virulence potential compared with *S. aureus*, the most virulent species, since they do not code for such an extensive suite of exotoxins associated with invasive infection (Otto, 2009, Massey et al., 2006). Whilst, potentially not as virulent as other staphylococci, species of the CoNS possess a sufficient number of virulence factors to be classed as an opportunistic pathogen, contributing significantly to nosocomial infection (Heilmann et al., 2019).

Biofilm formation is highly important for CoNS pathogenicity as the exopolysaccharide matrix allows bacteria to colonise abiotic surfaces of indwelling medical devices and host biotic surfaces, helping to establish infections (Cameron et al., 2015). Adherence of staphylococcal cells is mediated by hydrophobic interactions, surface charge, surface associated proteins, in addition to van der Waal's forces (Heilmann et al., 2019).

Both cell surface charge and hydrophobicity have been attributed to surface proteins, which can be associated with or covalently linked to staphylococcal peptidoglycan (Heilmann et al., 2019). Covalently linked cell wall polymers such as teichoic acids are implicated in biofilm development, maintaining cation homeostasis and anchoring cell surface proteins (Holland et al., 2011). Teichoic acids can be modified by the addition of positively charged D-alanine residues by the *dltABCD* operon-encoded enzymes, helping to repel Cationic Antimicrobial Peptides (CAMPs), such as lysozyme, an important factor in host immune evasion (Bera et al., 2007, Peschel et al., 1999).

The surface associated *S. epidermidis* autolysin AtlE promotes adherence to both abiotic and biotic surfaces and is present in other staphylococci (Allignet et al., 2002, Hussain et al., 2015, Biswas et al., 2006, Heilmann et al., 1997). This bifunctional autolysin functions as a peptidoglycan hydrolase, important in cell wall degradation and separation during cell division (Porayath et al., 2018). The autolytic activity of AtlE and its homolog leads to the release of extracellular DNA, which is important for bacterial attachment to surfaces in the early stages of biofilm production (Qin et al., 2007).

The latter stages of biofilm production include cell-cell aggregation and the formation of a multi-layered structure (Heilmann et al., 2019). The main molecule responsible for this is the polysaccharide intercellular adhesin (PIA) poly-N-acetylglucosamine (PNAG) (Cramton et al., 2001). PIA is regulated, synthesised and modified by the *icaADBCR* operon (Heilmann et al., 1996). The deacetylation of PIA introduces a positively charged element to PIA by liberating free amino acid groups at neutral and acidic pH, similar to that of human skin (Vuong et al.,

2004). As the staphylococcal cell surface is negatively charged, PIA works to adhere cells together via electrostatic attraction (Rohde et al., 2010, Otto, 2008, Heilmann et al., 1996).

Staphylococcal MSCRAMMs, which are covalently linked to peptidoglycan, contribute to CoNS colonisation and pathogenicity (Foster, 2019, Heilmann et al., 2019). MSCRAMMs have a common structural organisation, including an N-terminal signal peptide, ligand binding domain, various repeated sequences and a C-terminal LPTXG motif responsible for cell wall anchorage (Foster, 2019). Collectively, cell wall-anchored proteins associated with CoNS biofilm include: the biofilm associated protein, Bap; accumulation associated protein, Aap; SdrG; and extracellular matrix binding protein, Embp (Arora et al., 2016). All participate in biofilm formation by mediating bacterial attachment or intercellular aggregation, through fibronectin binding, proteolytic modification of PIA, and mediation of both bacterial-bacterial and bacterial-host interactions (Taglialegna et al., 2016, Hartford et al., 2001, Rohde et al., 2005, Cucarella et al., 2001, Christner et al., 2010).

Phenol Soluble Modulins (PSMs) are secreted amphipathic peptides that contribute to staphylococcal pathogenicity and are regulated by Agr (Peschel and Otto, 2013, Wang et al., 2011). PSMs are proinflammatory, cytolytic and contribute to biofilm production, and have anti-microbial activity (Yao et al., 2005, Wang et al., 2007b, Cogen et al., 2010). These peptides recruit, activate and lyse human neutrophils, eliminating the main immune defence against bacterial infection (Wang et al., 2007b). In addition, the amphipathic properties of PSMs afford them the capacity to facilitate biofilm structuring, detachment and

spreading on surfaces, particularly in the β subclass of PSMs, commonly expressed in CoNS (Wang et al., 2011, Otto, 2008). PSMs are also an important competition determinant of CoNS. *S. epidermidis* γ and δ PSMs directly induce lipid vesicle leakage and exert selective antimicrobial action against other skin colonisers, such as *S. aureus* (Cogen et al., 2010).

Studies on *S. capitis* virulence have revealed outbreaks of vancomycin resistance in neonatal care units, associated with sepsis in preterm infants (Wirth et al., 2020). Certain *S. capitis* strains also express the lysostaphin-like glycylglycine endopeptidase, Ale-1, which cleaves peptidoglycan interpeptide bridges (Sugai et al., 1997). The *epr* immunity gene that renders *S. capitis* resistant to glycylglycine endopeptidase is present upstream of *ale-1* and shares amino acid sequence similarity to, *femAB*, methicillin resistance in *S. aureus* (Sugai et al., 1997).

The pathogenesis of inflammatory skin conditions such as atopic dermatitis are linked to the microbiome and in particular *S. aureus* (Grice and Segre, 2011). Dermatitis was associated with immune dysregulation, decreased diversity of the skin microbiome and impaired epithelial barrier integrity (Geoghegan et al., 2018). Staphylococcal virulence factors, including PSMs, biofilm production and exogenous proteases, expressed by *S. aureus* exacerbate atopic dermatitis (Blicharz et al., 2019). These same virulence factors are found throughout CoNS and are specifically described in *S. capitis* genomes. This study therefore hypothesises that this set of factors may also contribute to inflammatory skin disorders of the scalp, such as dandruff and seborrheic dermatitis, whereby the microbiome can be perturbed by *S. capitis* and CoNS virulence factors cause skin lesions, inflammation and immune dysregulation.

3.1.1 Aims of chapter

This study uses comparative approaches to study the genomes of *S. capitis* strains isolated from healthy scalp skin controls and dandruff scalps to investigate the hypothesis that *S. capitis* express determinants that could lead to dandruff. To examine the expressed determinants used by *S. capitis* to colonise the scalp, a particular focus will be surface adhesin and secreted protease genes, since these are known factors of *S. aureus* associated with its successful colonisation that leads to atopic dermatitis and other skin inflammation disorders (Geoghegan et al., 2018). Additionally, this study aims to probe the poorly described genomic subspecies classification of *S. capitis* ssp. *capitis* and ssp. *ureolyticus*.

3.2 Methods and materials

3.2.1 *Staphylococcus capitis* strain isolation from the scalp

The majority of *S. capitis* isolates used in this study were isolated from scalp samples, as described in section 2.2.3. For each scalp sample taken from dandruff, dandruff healthy and healthy scalp sites (Table 2.3), 100 µL of wash was plated onto the following staphylococcal selective media and if possible four distinct colonies were isolated: (1% (w/v) Tryptone (Oxoid), 0.5% (w/v) lab lemco powder (Oxoid), 0.3% (w/v) yeast extract, 1.3% (w/v) agar (Lab M), 0.1% (w/v) sodium pyruvate (JT Baker Chemicals), 0.05% (w/v) glycine (JT Baker Chemicals), 2.25% (w/v) potassium thiocyanate (JT Baker Chemicals), 0.12% (w/v) NaH2 PO4.2H20 (JT Baker Chemicals), 0.2% (w/v) lithium chloride (JT Baker Chemicals), 0.5 % (v/v) glycerol (JT Baker Chemicals), 1.0% (v/v) sodium azide (Sigma Aldrich), 3.0% (v/v) sterile egg yolk emulsion (Lab M), pH 7.2).

3.2.2 Whole-genome sequencing of *S. capitis* isolates

All isolates selectively obtained in this study, together with 7 further strains comprising type strains and skin isolates (Table 3.1), were grown in 10 mL BHI broth (Lab M) overnight, shaking at 37 °C. Subsequently, 1 mL of each overnight culture was centrifuged for 1 min at 5000 rpm and resuspended in 180 µl lysis buffer (20 mM Tris- HCl pH8, 2 mM EDTA, 1.2% Triton X-100); the cells of each clone were extracted to obtain high-quality genomic DNA using the QIAGEN DNeasy Blood & Tissue Kit and eluted in 10 mM Tris-HCl (pH 8.5).

DNA concentration was measured using a Thermo Scientific Nanodrop, a Qubit plus visualisation after gel electrophoresis on 1% (w/v) agarose gels (at 90 mV for 40 min with a 1 kb ladder). For sequencing, the extracted DNA was required in a final volume of 60 µl with a concentration between 1-30 ng μl^{-1}. Illumina DNA sequencing was performed by MicrobesNG (http://www.microbesng.uk, which is supported by the BBSRC (grant number BB/L024209/1)) using the Nextera XT library preparation kit protocol on a MiSeq platform, using 2*250 bp paired end reads (Illumina, San Diego, CA, USA).

3.2.3 Obtaining contextual references

From the 86 staphylococcal strains isolated in this study, 58 were classified as *S. capitis*, had acceptable sequence quality, and used as a part of this study. In order to represent the breadth of clonal complexes, host isolation site diversity and multiple geographical locations, we also included an additional 30 strains in this study. Of these, 7 were skin isolates provided by Unilever Plc or isolates available from type culture collections and 23 were reference genomes publicly available via NCBI (Table 3.1).

3.2.4 Pre-processing of raw data

All sequencing datasets underwent a standard analysis pipeline by MicrobesNG (http://www.microbesng.uk). Sequence reads were trimmed using Trimmomatic (Bolger et al., 2014) and their quality assessed using in-house scripts. Following this, artificial and read duplicates that may have been introduced during the PCR step in the library preparation protocol were removed using PRINSEQ v0.20.4 (Schmieder and Edwards, 2011). Further trimming of low-quality bases, which removed read segments that were below a minimum phred quality score of 20; the removal of phiX and sequencing adapters was carried out using the bioinformatics tool suite BBTools (Bushnell). Sequence read quality was assessed using FastQC v0.11.7 (Andrews, 2010) and MultiQC v1.0 (Ewels et al., 2016).

Only high-quality reads were used in downstream analysis and were assessed on passing the following criteria: basic quality statistics, per base sequence quality, per base N content, adapter content and an average GC content of between 30–40%. The remaining DNA sequence reads were mapped to the RefSeq bacterial database using Kraken v2.0 to ensure reads were free from contamination and to taxonomically classify each staphylococcal scalp sample (Wood et al., 2019).

3.2.5 Genome assembly and annotation

For *de novo* genome assembly, the processed sequence raw reads were assembled using the *De Bruijn* graph assembler Unicycler (Wick et al., 2017). Unicycler builds initial assembly *De Bruijn* graphs from short reads using the *de novo* assembler SPAdes (Nurk et al., 2013) and then simplifies the graph using information from short reads (Wick et al., 2017). Unicycler was run with default

recommended settings and assembly statistics evaluated with QUAST v4.40 without a reference genome (Gurevich et al., 2013) (Table S 6). Following *de novo* assembly, genomes were annotated using Prokka v 1.14.6 (Seemann, 2014).

3.2.6 Reference mapping to *S. capitis* AYP1020

Following pre-processing, trimmed sequencing reads were mapped against the *S. capitis* reference genome AYP1020 (NCBI Genome accession GCA_001028645.1). This reference genome consists of a single circular chromosome of 2.44 Mb encoding 2304 predicted proteins (RefSeq accession NZ_CP007601.1), including a prophage ΦAYP1020 (48.5 Kb) and a plasmid pAYP1020 (59.6 Kb) (RefSeq accession NZ_CP007602.1), (Cameron et al., 2015).

Sequence reads were aligned to the reference genome using BWA MEM (Li, 2013). Removal of optical or PCR duplicates, unmapped reads, supplementary aligned reads and poor quality reads was completed using SAMtools (Li et al., 2009). QualiMAP v2.2.2 was then used to assess mapping quality (Okonechnikov et al., 2016).

3.2.7 Phylogenetic reconstruction of *S. capitis* dataset

Sequence reads from the 88 *S. capitis* isolates from this study were mapped to the reference genome *S. capitis* AYP1020 (NCBI Genome accession GCA_001028645.1) using Snippy v.4.4.3 with minimum coverage of 4 (Seemann, 2015). QualiMap v2.2.2 was used to assess mapping quality and only samples with greater than 10X genome coverage were included in downstream analysis. Next, snippy-core v.4.4.3 was used to generate core genome SNP alignment files, which were then passed to snippy-clean v.4.4.3 to replace all non-ATCG characters with the letter N (Seemann, 2015).

Gubbins v2.3.4 was used to remove the effect of potential recombination events (Croucher et al., 2014). Maximum-likelihood (ML) phylogenetic inference was performed on the core genome alignments using IQTREE v1.6.12 (Chernomor et al., 2016). ModelFinder implemented in the IQTREE software package was used to find the best nucleotide substitution model that fitted the dataset (Kalyaanamoorthy et al., 2017). A phylogenetic tree inferred based on the core genome alignment of all 88 *S. capitis* isolates (scalp study *n* = 58, contextual *n* = 30), was built using the TVMe+ASC+R2 model, selected by Modelfinder, with 1000 bootstrap replicates. Phylogenetic tree inferred based on the core genome alignment of all 88 *S. capitis* isolates and rooted using *S. caprae* strain 26D (NCBI Genome accession GCA_007814385.1) as an outgroup, was built using the GTR+F+ASC+R2 model selected by ModelFinder with 1000 bootstrap replicates (Chernomor et al., 2016).

To determine the population structure of the *S. capitis* dataset and identify any potential smaller subpopulations of for example the two *S. capitis* ssp. *capitis* and ssp. *ureolyticus*, the rhierBAPS algorithm was used to carry out BAPS (Bayesian Analysis of Population Structure), specifying two cluster levels, 20 initial clusters and infinite extra rounds (Tonkin-Hill et al., 2018). Visualisations were performed using iTOL v4.2 (Letunic and Bork, 2016).

3.2.7.1 Rooting the phylogenetic tree

Identifying the root of a phylogenetic tree allows evolutionary and taxonomic inferences to be made and gives an indication of the directionality of evolutionary change although unrooted trees can be beneficial when depicting clusters of related sequences (Kinene et al., 2016). There are several methods to root

phylogenetic trees, but in this study both outgroup and midpoint rooting were explored. The most widely used is the outgroup method, whereby one or more taxa in the phylogenetic tree are assumed to lie outside of the rest of the taxa (ingroup) and the root of the ingroup phylogeny. The branch connecting the ingroup and outgroup is defined as the root and determines all subsequent evolutionary events within the tree (Maddison et al., 1984, Wheeler, 1990, Boykin et al., 2010). The outgroups of phylogenetic trees are often phylogenetically closely related (Kinene et al., 2016).

To root phylogenetic trees inferred based on the core genome alignment of all 88 *S. capitis* isolates with an outgroup a number of different staphylococcal related species were considered. The first strain considered as a potential outgroup was *Macrococcus caseolyticus* (strain JCSC5402 NCBI Genome accession GCA_000010585.1) because *Macrococcus* is the most closely related genus to *Staphylococcus* (Baba et al., 2009). *Macrococcus* was included in the *Staphylococcus* family until it was reassigned to its own genus due to its small chromosome size (2.1 MB) and lack of many sugar and amino acid pathways and virulence genes that are present in *S. aureus* (Baba et al., 2009, Kloos et al., 1998).

However, when attempting phylogenetic reconstruction as described above in section 3.2.7 it was found that when *M. caseolyticus* was included in the snippy core genome alignment, the alignment was only 110 bp long with less than 2% of the genome aligning to the *S. capitis* AYP1020 reference genome. Therefore, alternative *Staphylococcus* species were included in the phylogenetic reconstruction to investigate if any would produce an adequate core genome alignment (Table 3.2).

3.2.8 Assembly of the *S. capitis* pan-genome

Pan-genome analyses of all 88 isolates were performed using the Panaroo software package (Tonkin-Hill et al., 2020). The general feature format files produced by PROKKA genome annotation were used as an input for Panaroo pan-genome analysis with default parameters, MAFFT alignment and a core gene threshold of 90%. The interactive visualisation tool Phandango was used to visualise the resulting pangenome, metadata and associated phylogeny (Hadfield et al., 2017). Using IQTREE core genome sequences were aligned and maximum likelihood trees produced for viewing using ITOL (Letunic and Bork, 2016).

Panaroo was used as an alternative approach to inferring the pan-genome with software such as Roary (Page et al., 2015). Panaroo, makes use of a graph-based algorithm, to share information between genomes. This approach corrects for errors introduced during genome annotation and improves the clustering of orthologs and paralogs within the pangenome (Tonkin-Hill et al., 2020). Panaroo then generates gene clusters using CD-HIT to cluster the collection of all gene sequences in all samples (Tonkin-Hill et al., 2020, Li and Godzik, 2006). When comparing the output of Panaroo with Roary in this study, there was a reduction in the size of the accessory genome and an increase in size of the core genome. These results are similar to those found by Tonkin-Hill *et al.,* suggesting that Panaroo produced a more polished pangenome with better ortholog, clusters.

Maximum-likelihood (ML) phylogenetic inference was performed on the core genome alignments using IQTREE v1.6.12 (Chernomor et al., 2016). ModelFinder implemented in the IQTREE software package was used to find the best nucleotide substitution model that fitted to the dataset (Kalyaanamoorthy et al.,

2017). The phylogenetic tree inferred was built using the GTR+F+R5 model, selected by Modelfinder, with 1000 bootstrap replicates, based on the core genome alignment of all 88 *S. capitis* isolates (scalp study *n* = 58, contextual *n* = 30). The roary2svg.pl script was used to plot and visualise the ML phylogenetic tree and assembled pangenome (Page et al., 2015).

3.2.9 Functional annotation of the *S. capitis* pan-genome

A custom python script was used to extract coding gene sequences from the core and accessory genomes created by Panaroo. Predicted coding genes in all of the isolates, to COG (clusters of orthologous genes) categories using eggNOG-mapper (evolutionary genealogy of genes: Non-supervised Orthologous Groups (OGs)) (Huerta-Cepas et al., 2019).

Using each query sequence, eggNOG-mapper works with a precomputed collection of Hidden Markov Models (HMM) available from the eggNOG bacterial database using HMMER 3 and phmmer (Eddy, 2011). The best matching query is then stored and used to retrieve other orthologs that are then assigned orthology from a database of eggNOG phylogenetic trees. Functional descriptors are then transferred to query proteins of the retrieved orthologs to give functional protein annotations based on curated Gene Ontology (GO) terms (Gene Ontology Consortium, 2015), KEGG pathways (Kanehisa et al., 2014) and COG functional categories (Galperin et al., 2015) (Huerta-Cepas et al., 2017).

3.2.10 Understanding the differences between *S. capitis* ssp. *capitis* and ssp. *ureolyticus*

3.2.10.1 Identification of biochemical differences between *S. capitis* ssp. *capitis* and ssp. *ureolyticus*

To biochemically assess the differences between the two subspecies *S. capitis* ssp. *capitis* and *S. capitis* ssp. *ureolyticus* based upon observations by Kloos *et al.* the API *Staph-Ident* Strip system was used (BioMérieux, Marcy l'Etoile, France)(Kloos and Schleifer, 1975, Bannerman and Kloos, 1991). *S. capitis* ssp. *ureolyticus* was characterised differently to *S. capitis* ssp. *capitis* due to its urease activity and maltose fermentation (Bannerman and Kloos, 1991). Therefore, to probe the phenotypic basis for separation of the two subspecies within the phylogenetic reconstruction (Figure 3.1, Figure 3.2 & Figure 3.3) the API *Staph-Ident* Strip system was used analyse the biochemical profiles of multiple isolates included in this study (BioMérieux, Marcy l'Etoile, France). The API *Staph-Ident* Strip system was used according to the manufacturer's instructions (BioMérieux, Marcy l'Etoile, France).

3.2.10.2 Identification of genomic differences between *S. capitis* ssp. *capitis* and ssp. *ureolyticus*

Average nucleotide identity (ANI) indices were used to quantify genetic relatedness of the two *S. capitis* subspecies. ANI estimates the genetic relatedness between two genomes to assess species boundaries and closely reflects traditional microbiological methods such as DNA-DNA hybridisation (DDH) (Jain et al., 2018a, Goris et al., 2007). Studies suggest that there is a close relationship between DDH and ANI values, with the recommended cut-off of 70% DDH for

species delineation corresponding to 85% ANI and 69% conserved DNA (Goris et al., 2007).

DDH was used to initially compare the DNA relationship of *S. capitis* ssp. *ureolyticus* at the time of subspecies identification (Bannerman and Kloos, 1991). This study compared DSM 6717 (included in this study NCBI Genome accession GCA_002901925.1) to other staphylococci, including *S. capitis* ssp. *ureolyticus* and ssp. *capitis*. It concluded that *S. capitis* ssp. *ureolyticus* exhibited high degrees of similarity to other ssp. *ureolyticus* strains, but relatedness was significantly lower when compared to *S. capitis* ssp. *capitis*, at around 89%, similar to levels expected for a separate subspecies (Bannerman and Kloos, 1991).

To compare the relatedness of the *S. capitis* subspecies in this study, FastANI was used with default settings to compare all isolates of potential ssp. *ureolyticus* to ssp. *capitis* isolates, using the recommended cut-off score of >95% that would indicate that isolates belong to the same species (Jain et al., 2018a). *S. capitis* ssp. *ureolyticus* culture collection isolates were also compared to other ssp. *ureolyticus* isolates, similarly to Bannerman and Kloos (1991).

3.2.11 Investigation of cell wall-associated proteins

Studies have described CoNS to be opportunistic pathogens, with a significant virulence potential, albeit being lower than *S. aureus* (Cameron et al., 2015, Massey et al., 2006). Cell wall associated proteins play an essential role in colonisation and virulence of staphylococci, therefore the presence of surface proteins, MSCRAMMs and biofilm-related proteins of *S. capitis* isolates in this study were investigated to gain insight into *S. capitis* virulence. More specifically, this study sought to identify if particular factors were linked to *S. capitis* isolates

from dandruff scalps to investigate potential *S. capitis* subspecies differences in virulence.

Cameron, *et al.,* describe the presence and absence of orthologous virulence factors of *S. capitis* AYP1020 compared with those of *S. epidermidis* RP62a (NCBI Genome accession GCA_000011925.1) (2015). In the study here, BLASTp (Altschul et al., 1990) was used to locally align cell wall associated proteins found in *S. capitis* AYP1020 and *S. epidermidis* RP62a to the isolates from the dataset. Gene presence and absence was then investigated using Panaroo pangenome gene presence and absence output files (Altschul et al., 1990, Tonkin-Hill et al., 2020).

3.2.12 Investigation of *S. capitis* Two-Component Systems

Bacterial genomes encode multiple two-component sensor-response regulator pairs that are triggered by environmental stimuli. The number of Two-Component Systems (TCS) is proportional to genome size, diversity of environment and the complexity of bacterial cellular processes (Rapun-Araiz et al., 2020). TCS are vital to the capacity of staphylococcal species to colonise and survive on different body surfaces (Crosby et al., 2020). *S. aureus* encodes 16 TCS and studies have revealed homologs of these in CoNS, such as *S. epidermidis* and *S. saprophyticus* (Rapun-Araiz et al., 2020, Lou et al., 2014).

To identify the presence of TCS within the *S. capitis* genome, known TCS amino acid sequences from staphylococci were locally aligned to isolates from this dataset using BLASTp so that gene presence and absence could be investigated using panaroo pangenome gene presence and absence output files (Altschul et al., 1990, Tonkin-Hill et al., 2020).

3.2.13 Investigation of *S. capitis* virulence profiles

Genetic determinants for AMR were identified using ABRicate v0.9.8 (Seemann, 2020) with, NCBI Bacterial Antimicrobial Resistance Reference Gene Database (Feldgarden et al., 2019).

Phenol-soluble modulins (PSMs) are a novel toxin family contributing to the virulence potential of staphylococci (Peschel and Otto, 2013). To identify PSMs within the *S. capitis* isolate sequences, known PSM amino acid sequences from staphylococci and *S. capitis* specifically were locally aligned to, isolates from this dataset using BLASTp and gene presence and absence was investigated using Panaroo pangenome gene presence and absence output files (Altschul et al., 1990, Tonkin-Hill et al., 2020). Multiple sequence alignment of PSM peptides was carried out using ClustalW (Thompson et al., 1994).

Secondary metabolites produced by bacteria and fungi are important antimicrobials and bioactive compounds. The tool antiSMASH v5.0 was used to identify biosynthetic gene clusters (BGC) within *S. capitis* isolates. To identify BGC, antiSMASH extracts or predicts genes from input FASTA sequences using Glimmer3 and gene clusters are identified with signature gene pHMMs. Downstream analysis including NRPS/PKS domain analysis and annotation, prediction of the core chemical structure of PKSs and NRPSs, ClusterBlast gene cluster comparative analysis and smCOG secondary metabolism protein family analysis was then carried out and visualisations were generated (Blin et al., 2019).

Extracellular proteases produced by staphylococci contribute to infection by facilitating immune evasion, host colonisation and persistence (Otto, 2009). To

identify exoproteases within the *S. capitis* genome sequences, known protease amino acid sequences from staphylococci and *S. capitis* specifically were locally aligned to isolates from this dataset using BLASTp and gene presence and absence was investigated using panaroo pangenome gene presence and absence output files (Altschul et al., 1990, Tonkin-Hill et al., 2020). Virulence gene presence and absence for each *S. capitis* isolate included in this study was visualised using iTol v4.2 (Letunic and Bork, 2016).

3.2.14 Investigation of *S. capitis* MGEs

Mobile Genetic Elements (MGEs) such as prophages and plasmids were identified in the genomes of this study. Prophages were annotated using PHASTER, which takes genomic sequence data and performs BLAST searches against a custom prophage database (Altschul et al., 1990). Phage-like genes were then clustered into prophage regions using DBSCAN. The density of phage-related proteins across these clusters was identified together with the presence of essential phage protein families. Prophage regions were assigned a completeness score based on the proportion of phage genes in the identified region. A score <70 meant a cluster was designated as an incomplete prophage, 70-90 a questionable prophage, and a score of >90 a likely complete prophage (Arndt et al., 2016).

To find plasmids across the study isolates, the tool ABRicate v0.9.8 (Seemann, 2020) was used to search genomes for potential plasmid sequences using the PlasmidFinder database (Carattoli et al., 2014). Cameron *et al*., (2015) revealed the presence of a plasmid within the AYP1020 genome, and the prevalence of this particular plasmid was investigated. All genomes were mapped against the plasmid sequence (RefSeq NZ_CP007602.1) using BWA MEM (Li, 2013). Removal

of unmapped reads and poor quality reads was completed using SAMtools (Li et al., 2009). QualiMAP v2.2.2 was then used to identify mean sample depth and whether the reference genome represents the diversity within the sequence reads (Okonechnikov et al., 2016).

3.3 Results

3.3.1 *Staphylococcus capitis* scalp isolates and contextual references

Staphylococci from this study were isolated from scalp samples that were obtained as described in section 2.2.3. For each scalp sample, taken from dandruff (DD), dandruff healthy (DH) and healthy (HH) scalp sites, at least 1 and when possible 4 colonies were picked, cultured to purity and sequenced. A total of 86 staphylococcal scalp isolate genomes were sequenced and their speciation confirmed by Kraken v2.0; of this total, 58 were *S. capitis*. Contextual reference strains were also included in the study: 7 sequenced in this study were either provided by Unilever Plc, or were isolates obtained from type culture collections; and 23 were reference genomes publicly available via NCBI (Table 3.1).

Table 3.1 *Staphylococcus capitis* isolates used in this study. Isolates were either isolated in this study, type strains or sequence data obtained from the NCBI database.

Isolate ID	Isolation site	This study	Description
Sample_1	Scalp	Yes	Sample 1. Scalp state HH (Table 2.3)
Sample_2	Scalp	Yes	Sample 1. Scalp state HH (Table 2.3)
Sample_3	Scalp	Yes	Sample 1. Scalp state HH (Table 2.3)
Sample_4	Scalp	Yes	Sample 1. Scalp state HH (Table 2.3)
Sample_7	Scalp	Yes	Sample 2. Scalp state HH (Table 2.3)
Sample_9	Scalp	Yes	Sample 3. Scalp state HH (Table 2.3)
Sample_12	Scalp	Yes	Sample 4. Scalp state HH (Table 2.3)
Sample_13	Scalp	Yes	Sample 4. Scalp state HH (Table 2.3)
Sample_15	Scalp	Yes	Sample 4. Scalp state HH (Table 2.3)
Sample_17	Scalp	Yes	Sample 5. Scalp state HH (Table 2.3)
Sample_18	Scalp	Yes	Sample 5. Scalp state HH (Table 2.3)
Sample_19	Scalp	Yes	Sample 5. Scalp state HH (Table 2.3)
Sample_20	Scalp	Yes	Sample 6. Scalp state DD (Table 2.3)
Sample_21	Scalp	Yes	Sample 6. Scalp state DD (Table 2.3)
Sample_22	Scalp	Yes	Sample 6. Scalp state DD (Table 2.3)

Sample_24	Scalp	Yes	Sample 7. Scalp state DH (Table 2.3)
Sample_25	Scalp	Yes	Sample 7. Scalp state DH (Table 2.3)
Sample_26	Scalp	Yes	Sample 7. Scalp state DH (Table 2.3)
Sample_27	Scalp	Yes	Sample 7. Scalp state DH (Table 2.3)
Sample_28	Scalp	Yes	Sample 8. Scalp state DD (Table 2.3)
Sample_29	Scalp	Yes	Sample 8. Scalp state DD (Table 2.3)
Sample_30	Scalp	Yes	Sample 8. Scalp state DD (Table 2.3)
Sample_32	Scalp	Yes	Sample 9. Scalp state DH (Table 2.3)
Sample_33	Scalp	Yes	Sample 9. Scalp state DH (Table 2.3)
Sample_34	Scalp	Yes	Sample 9. Scalp state DH (Table 2.3)
Sample_35	Scalp	Yes	Sample 9. Scalp state DH (Table 2.3)
Sample_36	Scalp	Yes	Sample 10. Scalp state DD (Table 2.3)
Sample_37	Scalp	Yes	Sample 10. Scalp state DD (Table 2.3)
Sample_38	Scalp	Yes	Sample 10. Scalp state DD (Table 2.3)
Sample_39	Scalp	Yes	Sample 10. Scalp state DD (Table 2.3)
Sample_40	Scalp	Yes	Sample 11. Scalp state DH (Table 2.3)
Sample_41	Scalp	Yes	Sample 11. Scalp state DH (Table 2.3)
Sample_42	Scalp	Yes	Sample 11. Scalp state DH (Table 2.3)

Sample_43	Scalp	Yes	Sample 11. Scalp state DH (Table 2.3)
Sample_45	Scalp	Yes	Sample 12. Scalp state DD (Table 2.3)
Sample_47	Scalp	Yes	Sample 12. Scalp state DD (Table 2.3)
Sample_49	Scalp	Yes	Sample 13. Scalp state DH (Table 2.3)
Sample_50	Scalp	Yes	Sample 13. Scalp state DH (Table 2.3)
Sample_52	Scalp	Yes	Sample 13. Scalp state DH (Table 2.3)
Sample_54	Scalp	Yes	Sample 14. Scalp state DD (Table 2.3)
Sample_55	Scalp	Yes	Sample 14. Scalp state DD (Table 2.3)
Sample_56	Scalp	Yes	Sample 14. Scalp state DD (Table 2.3)
Sample_57	Scalp	Yes	Sample 15. Scalp state DH (Table 2.3)
Sample_58	Scalp	Yes	Sample 15. Scalp state DH (Table 2.3)
Sample_60	Scalp	Yes	Sample 15. Scalp state DH (Table 2.3)
Sample_62	Scalp	Yes	Sample 16. Scalp state DD (Table 2.3)
Sample_63	Scalp	Yes	Sample 16. Scalp state DD (Table 2.3)
Sample_65	Scalp	Yes	Sample 17. Scalp state DH (Table 2.3)
Sample_66	Scalp	Yes	Sample 17. Scalp state DH (Table 2.3)
Sample_75	Scalp	Yes	Sample 20. Scalp state HH (Table 2.3)
Sample_78	Scalp	Yes	Sample 20. Scalp state HH (Table 2.3)

Sample_79	Scalp	Yes	Sample 21. Scalp state HH (Table 2.3)
Sample_80	Scalp	Yes	Sample 21. Scalp state HH (Table 2.3)
Sample_81	Scalp	Yes	Sample 21. Scalp state HH (Table 2.3)
Sample_82	Scalp	Yes	Sample 21. Scalp state HH (Table 2.3)
Sample_83	Scalp	Yes	Sample 22. Scalp state HH (Table 2.3)
Sample_84	Scalp	Yes	Sample 22. Scalp state HH (Table 2.3)
Sample_85	Scalp	Yes	Sample 22. Scalp state HH (Table 2.3)
Sample_86	Scalp	Yes	Sample 22. Scalp state HH (Table 2.3)
Sample_87	Skin	No	Isolated from healthy human skin in the USA. *S. capitis ureolyticus* ATCC 49325. (Bannerman and Kloos, 1991)
Sample_88	Scalp	No	Isolated from the scalp in Bangalore. Unilever Plc strain. *S. capitis* subsp*. ureolyticus*
Sample_104	Scalp	No	Isolation site unknown *S. capitis* VCU116. NCBI Genome accession GCA_000221525.2
Sample_105	Foot	No	Isolated from the human foot in the UK. Unilever Plc strain. *S. capitis*
Sample_106	Foot	No	Isolated from the human foot in the UK. Unilever Plc strain. *S. capitis*
Sample_107	Foot	No	Isolated from the human foot in the UK. Unilever Plc strain. *S. capitis*

Sample_108	Skin	No	Isolated from healthy human skin, in the USA. *S. capitis* DSM 20327. (Kloos and Schleifer, 1975).
AYP1020	Blood	No	Isolated from human blood, in Melbourne, Australia. *S. capitis capitis.* NCBI Genome accession GCA_001028645.1 (Cameron et al., 2015).
FDAARGOS_378	Cerebrospinal fluid	No	Isolated from the cerebrospinal fluid of a 4-month-old male with meningitis in the USA. *S. capitis*. NCBI Genome accession GCA_002591175.1.
TW2795	Blood	No	Isolated from blood of a male with Pacemaker-related endocarditis, in Tokyo, Japan. *Staphylococcus capitis* subsp. *ureolyticus*. NCBI Genome accession GCA_002356175.1.
NCTC 11045	Skin	No	Isolated from human skin, in the USA. *S. capitis* subsp. *capitis.* NCBI Genome accession GCF_002902325.1 (Kloos and Schleifer, 1975).
DSM 6717	Skin	No	Isolated from human skin. *Staphylococcus capitis* subsp. u*reolyticus.* NCBI Genome accession GCA_002901925.1.
SK14	Skin	No	Isolated from the skin of a human arm. *S. captis.* NCBI Genome accession GCA_000174135.1.
LNZR-1	Blood	No	Isolated from human blood in China. *Staphylococcus capitis subsp. ureolyticus*. NCBI Genome accession GCA_000712995.1 (Li et al., 2014).
CR02	Neonate	No	Isolated from human skin in France. Pulsotype NRCS-A, Methicillin-resistant *S. capitis.* NCBI Genome accession GCF_001458675.1 (Lemriss et al., 2016).

CR03	Neonate	No	Isolated from human skin in Belgium. Pulsotype NRCS-A, Methicillin-resistant *S. capitis.* NCBI Genome accession GCA_001215085.1
CR05	Neonate	No	Isolated from human skin in the UK. Pulsotype NRCS-A, Methicillin-resistant *S. capitis.* NCBI Genome accession GCA_001220005.1
CR09	Neonate	No	Isolated from human skin in France. Pulsotype NRCS-A, Methicillin-resistant *S. capitis.* NCBI Genome accession GCA_001220605.1
C87	Upper respiratory tract	No	Isolated from the human upper respiratory tract. *S. capitis.* NCBI Genome accession GCA_000183705.1
QNI	Skin	No	Isolated from human skin, in China. *S. capitis.* NCBI Genome accession AJTH00000000.1 (Qin et al., 2012).
DMG1800578	Neonatal Axilla	No	Isolated from human skin, in Dunedin, New Zealand. *S. capitis.* SRA SRX4598283 (Carter et al., 2018)
DMG1800574	Neonatal unit staff Pinna	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598290 (Carter et al., 2018)
DMG1800575	Neonatal unit staff Pinna	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598289 (Carter et al., 2018)
DMG1800577	Neonatal unit staff Pinna	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598297 (Carter et al., 2018)
DMG1800566	Neonatal unit staff Pinna	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598340 (Carter et al., 2018)
DMG1800560	Neonatal Axilla	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598332 (Carter et al., 2018)

DMG1800582	Neonatal Axilla	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598287 (Carter et al., 2018)
DMG1800663	Neonatal Eye	No	Isolated from the human eye, in Dunedin, New Zealand. SRA SRX4598248 (Carter et al., 2018)
DMG1800572	Neonatal unit staff Thumb	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598292 (Carter et al., 2018)
DMG1800569	Neonatal unit staff Thumb	No	Isolated from human skin, in Dunedin, New Zealand. SRA SRX4598295 (Carter et al., 2018)

3.3.2 Population structure of *S. capitis* genome dataset

To investigate the population structure of *S. capitis*, the 88 genome sequences were aligned, comprising 58 isolated from the scalp in this study, 7 Unilever Plc skin isolates and type culture collection strains and 23 high-quality reference genomes from NCBI (Table 3.1). Alignment of genomes by Snippy v.4.4.3 produced a core nucleotide sequence alignment of 2503265 bp and 29095 SNPs were identified. To infer accurate phylogenetic reconstruction, Gubbins v2.3.4 was employed to remove the effect of potential recombination events. The alignment of genomes prior to this yielded a core nucleotide sequence alignment of 37988 bp with 1225 SNPs identified, which was used to construct a Maximum Likelihood (ML) phylogenetic tree.

The ML phylogenetic tree revealed relatively low diversity between isolates, as indicated by short branch lengths. However, two distinct clades were revealed that separate the two *S. capitis* ssp. *capitis* and *ureolyticus*. Both type strains and contextual reference strains described in literature as each subspecies were spread throughout the phylogenetic tree but were distinctly separated into two clades. The *S. capitis* subspecies clustered into two clades by average pairwise distance of 9254 core genome SNPs, which was congruent with BAPS analysis. The *S. capitis* ssp. *capitis* clade divides into two separate subclades, while the *S. capitis* ssp. *ureolyticus* clade divides into four separate subclades which was again congruent with population structure analysis using BAPS (Figure 3.1).

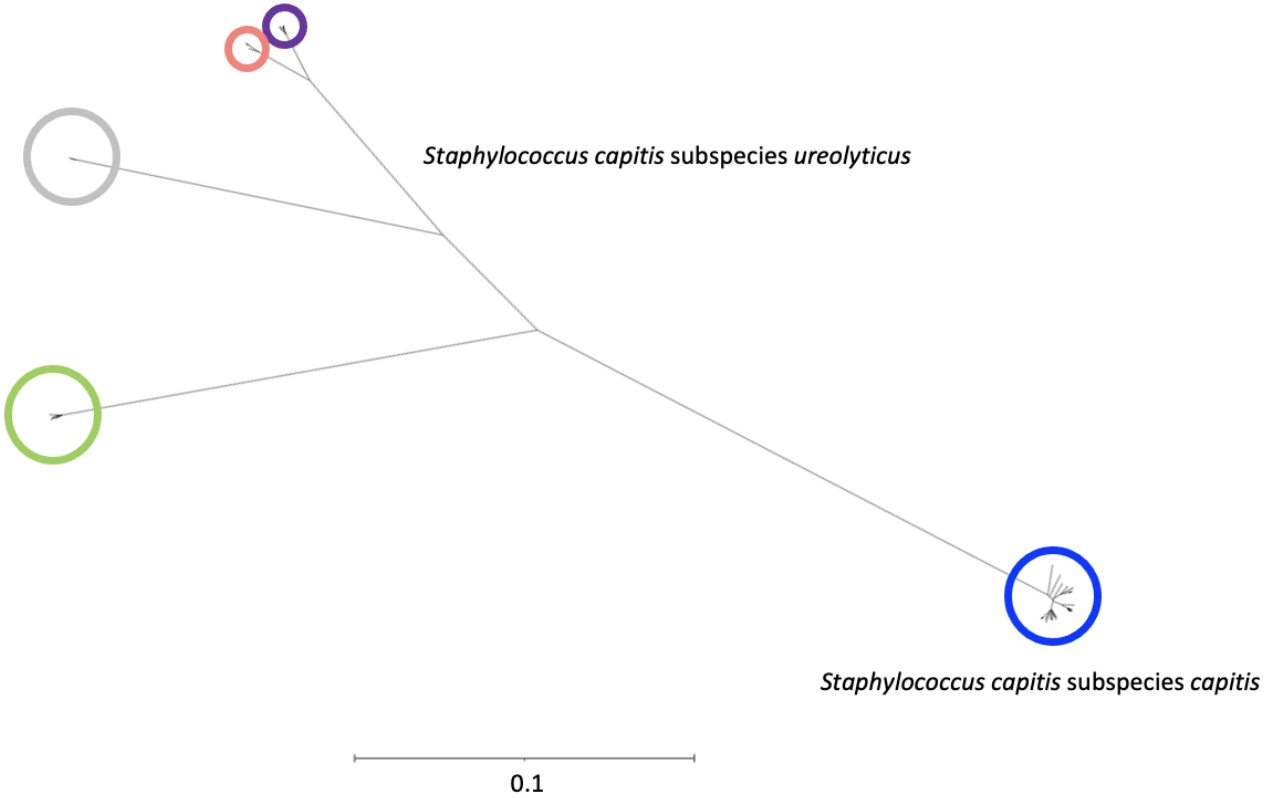


Figure 3.1 Unrooted ML phylogenetic tree generated based on core genome alignments of 88 *S. capitis* isolates. ML tree was generated without recombination regions and was constructed with IQTREE using the best selected substitution model TVMe+ASC+R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates, all nodes displayed contain >80% support. The 61 *S. capitis* ssp. *capitis* isolates are clustered into a clonal group represented in blue, and 27 *S. capitis* ssp. *ureolyticus* are clustered into four groups and represented in green, grey, red and purple. The scale bar represents the number of nucleotide substitutions per site. Different colours represent rhierBAPS clustering. Figure visualised with iTol v4.2 (Letunic and Bork, 2016).

3.3.2.1 Identifying a common ancestral root for the *S. capitis* phylogenetic reconstruction

To root ML phylogenetic trees representing the core genome alignment of all 88 *S. capitis* isolates with an outgroup, several different closely related species were considered (Table 3.2). Identifying an appropriate outgroup to root the ML phylogenetic trees is important to indicate directionality of evolutionary change (Kinene et al., 2016). Of the staphylococcal species tested, *S. caprae* was determined to be the most appropriate outgroup as it produced a sizeable core genome alignment.

Alignment of genomes with Snippy v.4.4.3 produced a core nucleotide sequence alignment of 2503265 bp and 35192 SNPs were identified. To infer accurate phylogenetic reconstruction, Gubbins v2.3.4 was employed to remove the effect of potential recombination events. This yielded a core nucleotide sequence alignment of 70127 bp with 4164 SNPs identified, which was used to construct a ML phylogenetic tree (Figure 3.2). This phylogenetic tree revealed the same subspecies separation seen in the unrooted ML phylogenetic reconstruction of the *S. capitis* dataset, which was congruent with BAPS analysis (Figure 3.1). The topology of the ML tree rooted to *S. caprae* strains 26D and NCTC 12196 revealed a similar topology to the ML tree constructed and midpoint rooted (Figure 3.3). The two *S. caprae* strains 26D and NCTC 12196 were separated into a distinct clade with an average pairwise distance of 25576 core genome SNPs, which along with a longer branch length indicates greater diversity (Figure 3.2).

Table 3.2 Staphylococcal outgroup species. Staphylococcal species and strains that were included in *S. capitis* phylogenetic reconstruction to find a suitable species for outgroup rooting of the tree.

Species	Strain	NCBI Genome accession
Staphylococcus aureus	NCTC 8325	GCA_000013425.1
Staphylococcus simulans	FDAARGOS 124	GCA_001559115.2
Staphylococcus epidermidis	ATCC 12228	GCA_009873455.1
Staphylococcus hyicus	ATCC 11249	GCA_000816085.1
Staphylococcus chromogenes	20B	GCA_007813835.1
Staphylococcus muscae	NCTC 13833	GCA_900187005.1
Staphylococcus auricularis	NCTC 12101	GCA_900478415.1
Staphylococcus hominis	K1	GCA_002850375.1
Staphylococcus saccharolyticus	NCTC 11807	GCA_003970495.1
Staphylococcus pettenkoferi	FDAARGOS 288	GCA_002208805.2
Staphylococcus sciuri	NCTC 12103	GCA_900474615.1
Staphylococcus agentis	908	GCA_001442815.3
Staphylococcus saprophyticus	ATCC 15305	GCA_000010125.1
Staphylococcus cohnii	FDAARGOS 538	GCA_003956025.1
Staphylococcus arlettae	P2	GCA_006741705.1
Staphylococcus intermedius	NCTC 11048	GCA_900458545.1
Staphylococcus caprae	26D	GCA_007814385.1
Staphylococcus caprae	NCTC 12196	GCA_900458385.1

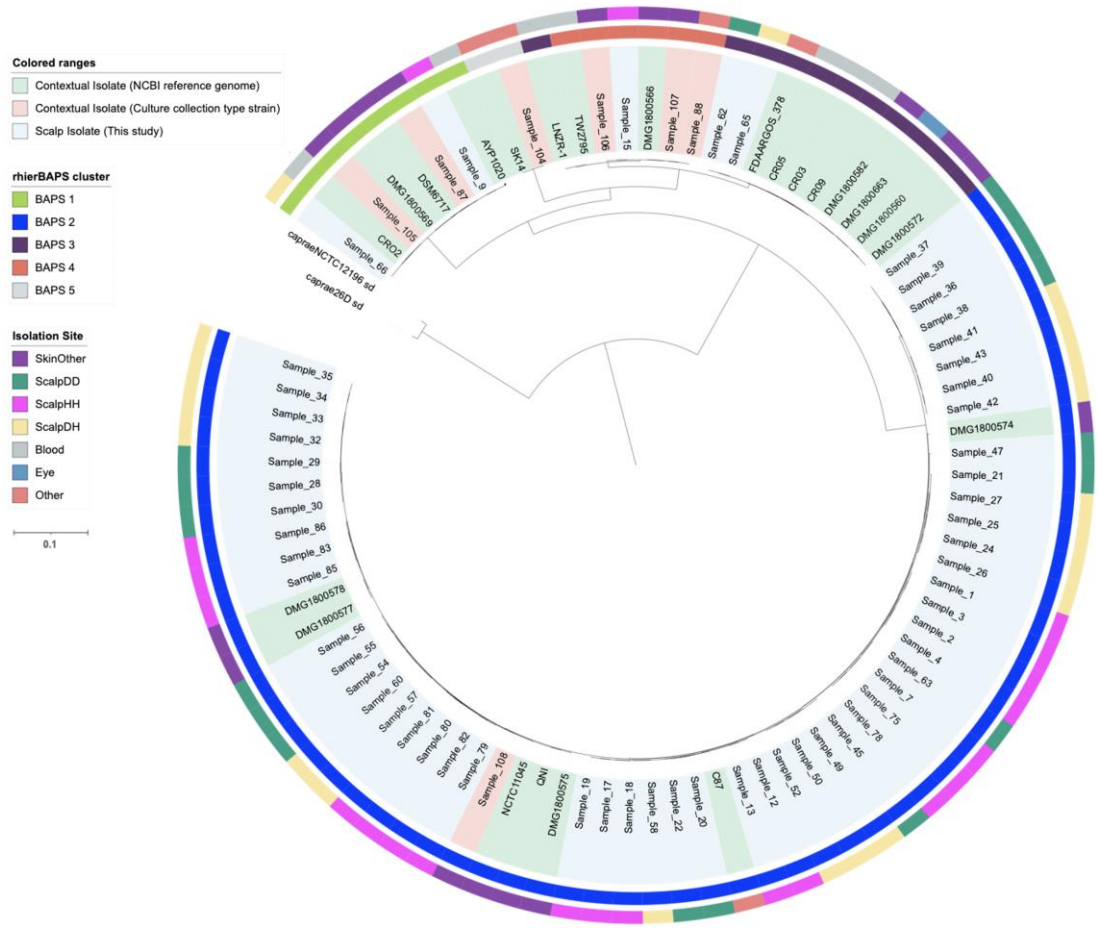


Figure 3.2 ML phylogenetic tree generated based on core genome alignments of 88 S. capitis isolates, rooted to *S. caprae*. ML tree was generated without recombination regions and was constructed with IQTREE using the best selected substitution model GTR+F+ASC+R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates; all nodes displayed contain >80% support. The colour key has identities for each ring. Coloured ranges and the outer ring provide details of the origin of isolates, colours are specified in the key. The 63 *S. capitis* ssp. *capitis* isolates are clustered into a clonal group represented in blue, and 27 *S. capitis* ssp. *ureolyticus* are clustered into four groups and represented in green, grey, red and purple, both based on rhierBAPS clustering (inner ring). The scale bar represents the number of nucleotide substitutions per site. Figure visualised using iTol v 4.2 (Letunic and Bork, 2016).

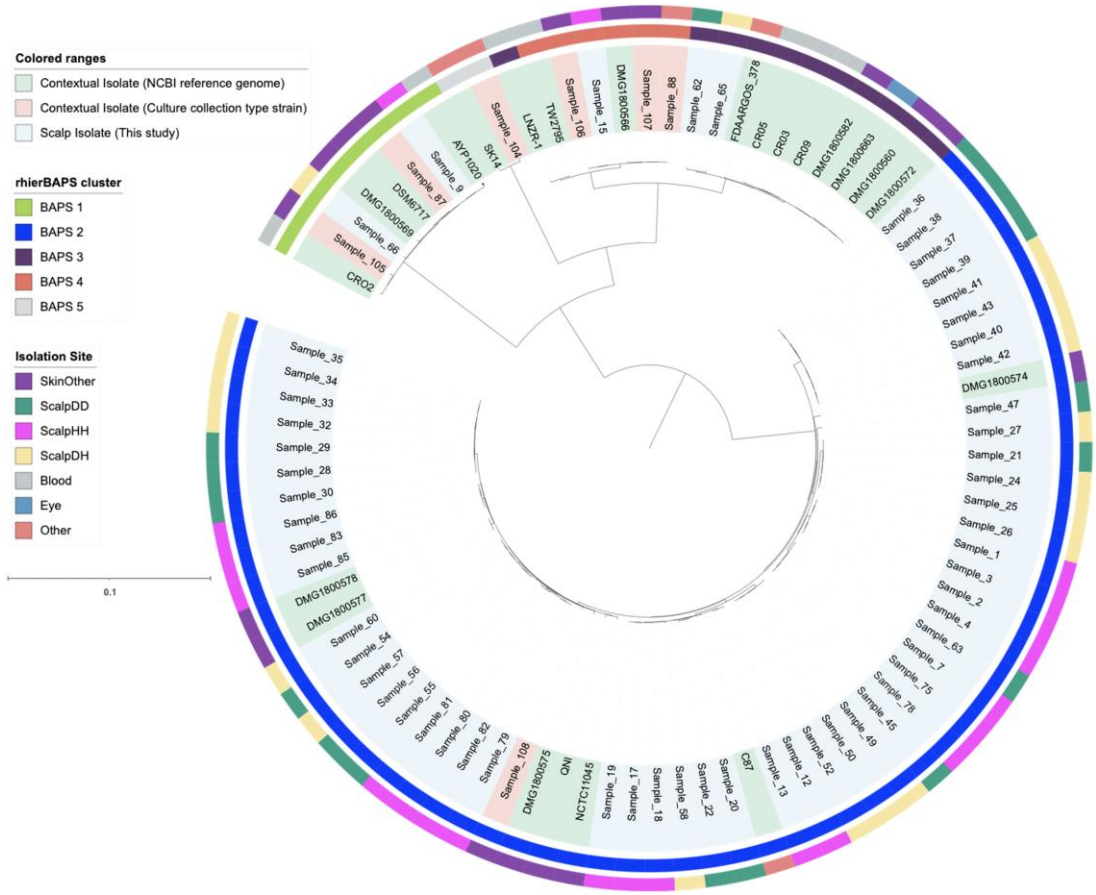


Figure 3.3 Midpoint rooted ML tree generated based on core genome alignments of 88 *S. capitis* isolates. ML tree was generated without recombination regions and were constructed with IQTREE using the best selected substitution model TVMe+ASC+R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates, all nodes displayed contain > 80% support. Coloured ranges and the outer ring provide details of the origin of isolates, colours are specified in the key. The 63 *S. capitis* ssp. *capitis* isolates are clustered into a clonal group represented in blue, and 27 *S. capitis* ssp. *ureolyticus* are clustered into four groups and represented in green, grey, red and purple, both based on rhierBAPS clustering (inner ring). The scale bar represents the number of nucleotide substitutions per site. Figure visualised using iTol v 4.2 (Letunic and Bork, 2016).

3.3.2.2 Genetic diversity of *S. capitis*

S. capitis sequence reads were mapped to reference genome AYP1020 producing a core nucleotide sequence alignment of 2503265 bp, and 29095 SNPs were identified. Filtering of recombination by Gubbins v2.3.4 yielded a core nucleotide sequence alignment of 37988 bp with 1225 SNPs identified, which was used to construct a Maximum Likelihood (ML) phylogenetic tree (Figure 3.1, Figure 3.2 & Figure 3.3).

There was extensive recombination among the study isolates, with recombination most evident in BAPS clusters 2, 3, 4 and 5. BAPS cluster 1 contained less recombination, however this is likely due to the reference genome itself being clustered in this group. Recombination was inferred across large regions of the chromosome, predominantly within the first ~750 kb and the last ~1 Mb of the genome. This suggests that *S. capitis* may have arisen following extensive recombination events (Figure 3.4).

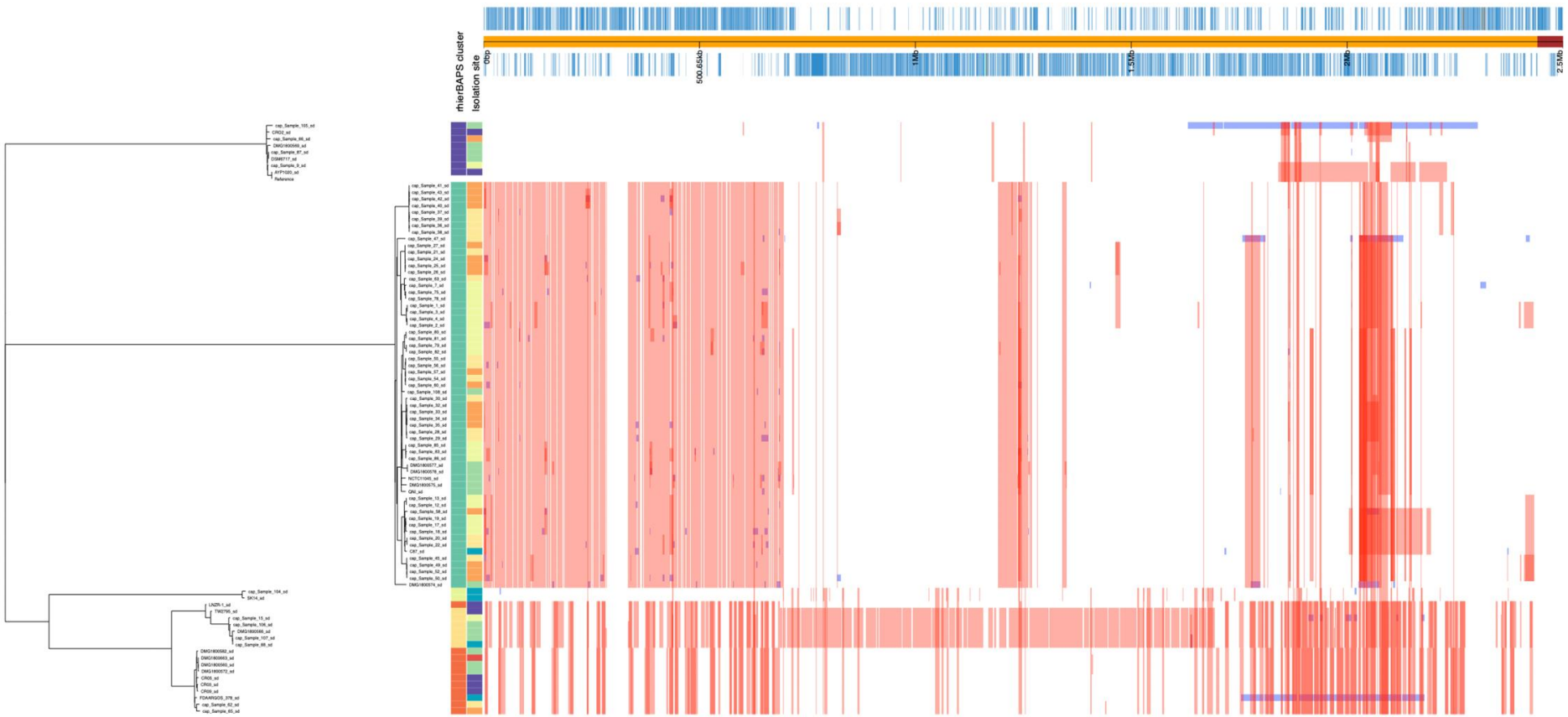


Figure 3.4 Analysis of the *S. capitis* genome alignment with Gubbins. The maximum likelihood phylogenetic reconstruction of *S. capitis* is shown on the left, with coloured bands highlighting rhierBAPS clustering and isolation site of genomes. Homologous recombination events for each *S. capitis* genome ordered based on their position in the AYP1020 reference genome (shown along the top) are shown on the right. Blocks detected in >1 isolate are shown in red, while blocks affecting a single isolate are indicated in blue. Figure visualised using Phandango (Hadfield et al., 2017).

3.3.3 The *S. capitis* pan-genome

To explore the gene content variation among *S. capitis* isolates, pangenome analysis was performed using Panaroo, which identifies core and accessory genes by clustering homologous groups of genes based on CD-HIT sequence identity or derived protein products. A cluster of genes shared by more than 90% of isolates are considered to belong to the core genome.

The pangenome of *S. capitis* comprised of both the core genome, containing many conserved genes required for essential biological and metabolic processes, and the accessory genome representing genes for specific adaptations. Comparative analysis revealed that *S. capitis* is somewhat conserved in terms of gene content, with the total number of predicted genes in each genome ranging from 2,149 to 2,414. The total number of unique gene clusters predicted across all 84 was 3,664, with 1,764 (48%) core genes conserved in all genomes (Figure 3.5). The accessory genome is made up of a total of 1,900 gene clusters (52%) (Figure 3.5). Four genomes (Samples 39, 43, 49, 83) of the original isolates were removed from this analysis as they were duplicate genomes).

The number of new genes added to the pan-genome with the addition of every new genome was investigated. The addition of each genome, at random to the pangenome caused an increase in both the total number of genes and number of unique genes in the pangenome (Figure 3.6). Although this study highlights diversity within the *S. capitis* phylogeny it cannot be conclusively determined from this dataset whether the *S. capitis* pan-genome is open or closed. An open pangenome is described as increasing in size with the addition of new genomes.

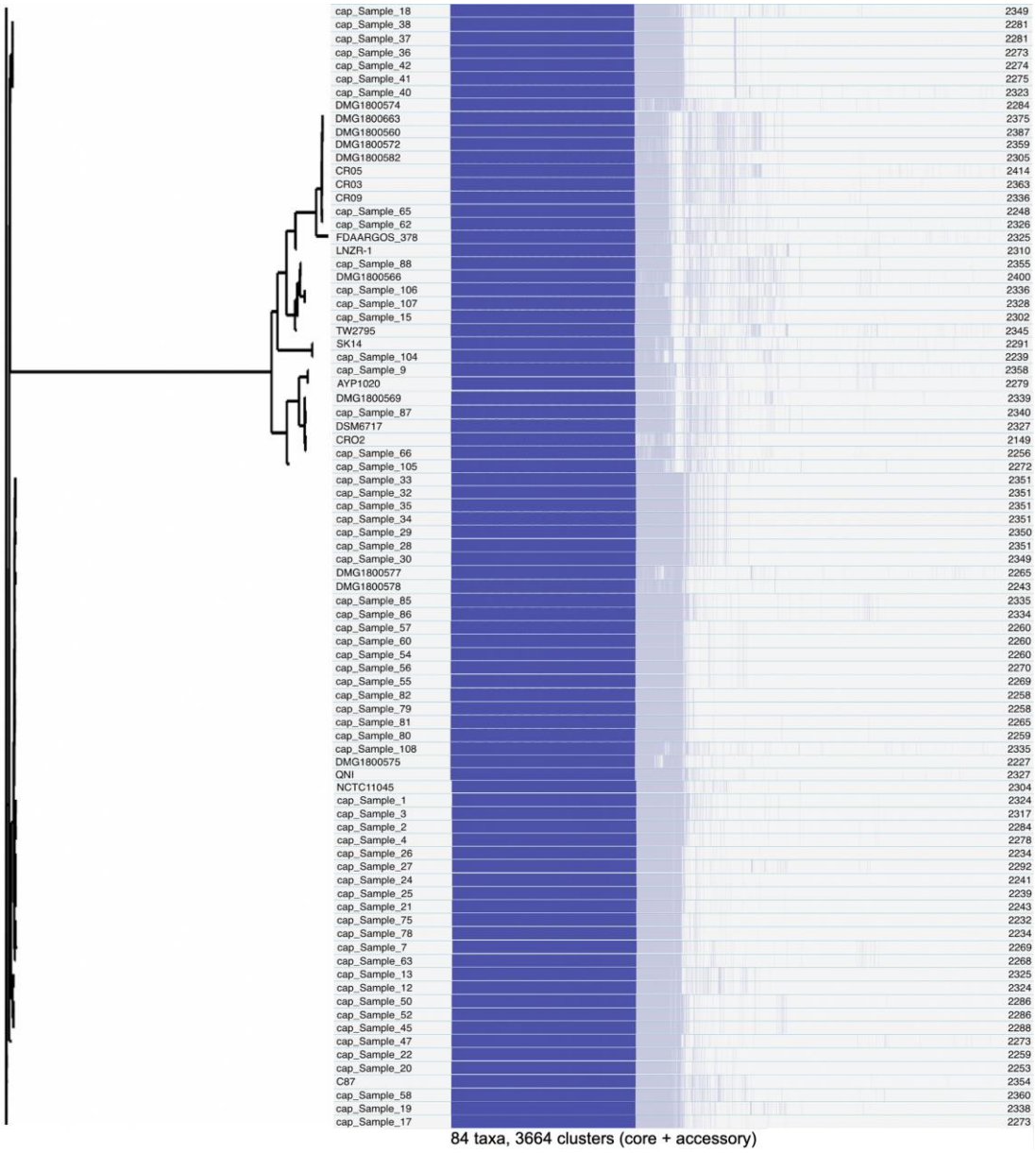
Additional isolates are required to fully represent the diversity within the species and confidently determine the openness of the *S. capitis* pangenome.

To explore the differences in gene content between *S. capitis* ssp. *capitis* and *S. capitis* ssp. *ureolyticus,* the accessory genomes of subspecies were compared to reveal 16 and 33 gene clusters found only in putative *S. capitis* ssp. *ureolyticus* and *capitis* genomes, respectively (Table 3.3). There were 95 gene clusters present in more than half of all putative ssp. *ureolyticus* genomes and absent in less than half of ssp. *capitis* genomes (Table S 7). Whilst 121 gene clusters were present in more than half of all putative ssp. *capitis* genomes and absent in less than half of ssp. *ureolyticus* genomes (Table S 7).

Gene clusters uniquely present in *S. capitis* ssp. *ureolyticus* include a number of phage related proteins, which can be accounted for due to the differences in phage present in each *S. capitis* genomes (Table 3.6). Gene clusters present exclusively in *S. capitis* ssp. *capitis* include gene clusters making up the arginine catabolic mobile element (ACME), encoding arginine deiminase activity found in various species of staphylococci (Lindgren et al., 2014). Other gene clusters differentially found in *S. capitis* ssp. *capitis* include those linked to cellular processes, such as *pyrB* and *crp* (Table 3.3).

There were also multiple gene clusters that were present differentially but not exclusively across the two subspecies (Table S 7). Of note, certain gene clusters with antimicrobial function were more frequent in *S. capitis* ssp. *ureolyticus*, including β-lactam resistance genes BlaI, BlaR and BlaZ, and β-class phenol soluble modulins. The gene cluster AtlE, important to CoNS virulence was also more frequently present in *S. capitis* ssp. *ureolyticus* together with genes

important in metalloid transportation, acquisition and resistance (*arsCB*, *ydhK*, staphylopine biosynthesis and copper transportation (*copB*)) (Table S 7).



<u>Figure 3.5 Pan-genome analysis of 84 *S. capitis* isolates.</u> ML phylogeny based on core genome SNPs identified outside regions of inferred recombination (left) with core and accessory genes identified by Panaroo (right). Total number of genes represented in the column on the right. Dark blue represents the core genome and light blue the accessory genome.

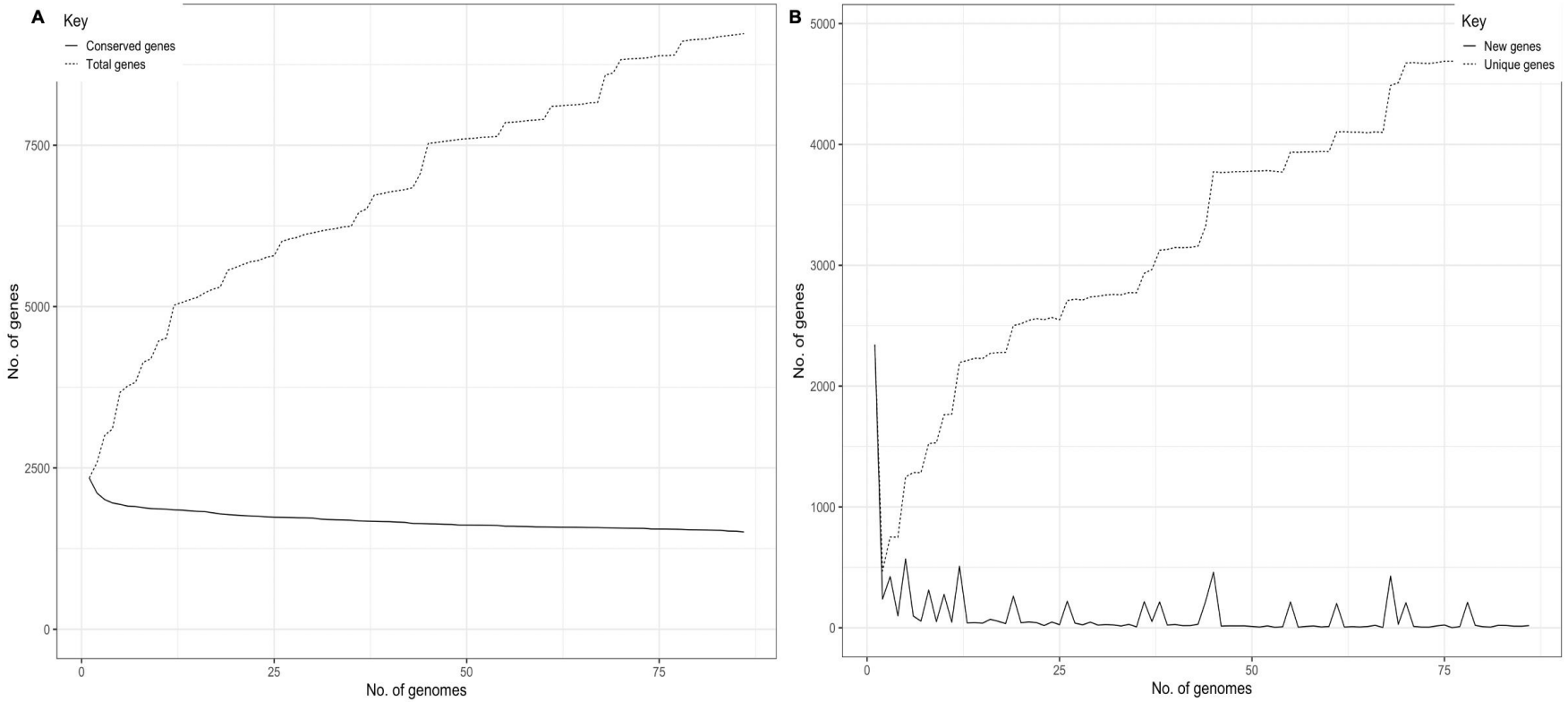


Figure 3.6 Pan-genome plot of total number of genes vs number of genomes. Graphs represent how the pan-genome varies as genomes are added in random order. (A) The number of conserved genes does not increase as genomes are added into the pan-genome, whilst the total number of genes increases. (B) The number of unique genes increases as genomes are added into the pan-genome, whilst the number of new genes, does not exponentially increase.

Table 3.3 Exclusive gene clusters found in *S. capitis* subspecies. Gene clusters found only in *S. capitis* ssp. *capitis* or ssp. *ureolyticus*. Gene clusters, presence and absence, and functional descriptions were obtained from Panaroo pangenome analysis of assembled genomes.

Gene cluster	Function description
Gene functions found only in *S. capitis* ssp. *ureolyticus*	
group_2296	Response regulator transcription factor
group_2011	Hypothetical protein
group_1477	Glycosyltransferase family 2 protein
group_1381	DMT family transporter
group_868	DUF2922 domain-containing protein (unknown function)
group_691	Hypothetical protein
group_401	Hypothetical protein
group_1064	Hypothetical protein
group_1910	Hypothetical protein
group_2080	YSIRK-type signal peptide-containing protein
arsB_2_arsB_4	Inner membrane arsenite extrusion pump, resistance to arsenate, arsenite, and antimonite
group_2735	arsC arsenate reductase (thioredoxin)
group_1779	Phage head protein, minor capsid protein
group_1328	NAD(P)H-binding protein
group_935	Hypothetical protein
Gene functions found only in *S. capitis* ssp. *capitis*	
group_2962	DNA methyltransferase Sau3A
group_2589	Molybdopterin biosynthesis MoeB protein
group_2358	Winged helix-turn-helix transcriptional regulator (ArsR)
group_2065	YSIRK-type signal peptide-containing protein
saeR	Unphosphorylated response regulator SaeR
group_1616	Hypothetical protein
group_1540	Uncharacterised protein
group_1486	YolD-like family protein
group_1448	DNA invertase
group_1431	Uncharacterised protein / molybdopterin biosynthesis MoeB protein
group_1408	FAD-dependent oxidoreductase
pyrB_2	Aspartate carbamoyltransferase
group_1290	Glycosyl transferase, group 1 family/ poly (glycerol-phosphate) alpha-glucosyltransferase

group_1212	Uncharacterised protein / molybdopterin biosynthesis MoeB protein
group_1211	Uncharacterised protein
group_1168	Putative CsbD family general stress response protein
group_558	Hypothetical protein / minor capsid protein
arsR_2	Arsenic/antimonite resistance regulatory protein
group_49	Hypothetical protein
trkG	Na+-transporting ATP synthase (osmoregulator in *E. coli*)
group_1592	Amino acid permease/ arcD_2
group_865	Hypothetical conserved protein / DUF4260 domain-containing protein (unknown function)
hdcA	Histidine decarboxylase proenzyme
group_2150	Cation transporter
group_1767	Crp/Fnr family transcriptional regulator
group_1656	YfcC family protein (Uncharacterized membrane protein YfcC, ion transporter superfamily)
arcC1_1	Utilization of arginine as a source of energy for growth under anaerobic conditions
argF_1	Ornithine carbamoyltransferase, arginine anabolic pathways
arcA_1	Utilization of arginine as a source of energy for growth under anaerobic conditions
arcD_1	Utilization of arginine as a source of energy for growth under anaerobic conditions
group_2499	Hypothetical protein/ restriction endonuclease subunit S
cimH_2_citS	2-hydroxycarboxylate transporter family protein

3.3.4 Functional annotation of the *S. capitis* pan-genome

The functions of genes within the core and accessory genomes of *S. capitis* were investigated by assigning all gene clusters discerned by Panaroo pan-genome analysis to clusters of orthologous groups (COGs) using eggNOG-mapper. The majority of genes in the accessory genomes of both *S. capitis* ssp. *capitis* (39%) and ssp. *ureolyticus* (49%) were poorly characterised, this was also true of core genome, albeit at lower numbers (~32%). This could indicate the presence of novel gene clusters in the *S. capitis* genome and the level of curation (Figure 3.7).

The majority of the 1,764 core genes shared by all 84 *S. capitis* genomes are involved in housekeeping processes such as translation, ribosomal structure and biogenesis (8.1%), and transcription (6.2%). This is in contrast to the accessory genome where gene clusters associated with replication, recombination and repair (7.6%) were enriched. This same pattern was separately observed in the pan-genome analysis of each subspecies (Figure 3.7).

Genes involved in metabolic functions were enriched in the *S. capitis* core genome compared with the accessory genome, although the majority of genes were assigned metabolic functions in both the accessory (29%, 1,004 genes) and core (35%, 617 genes) genomes. Such functions included amino acid transport and metabolism (7.0%), inorganic ion transport and metabolism (5.6%), nucleotide transport and metabolism (4.7%), coenzyme transport and metabolism (3.8%), carbohydrate transport and metabolism (3.4%), lipid transport and metabolism (2.5%), cell wall and membrane biosynthesis (4.8%), energy production and conversion (4.8%).

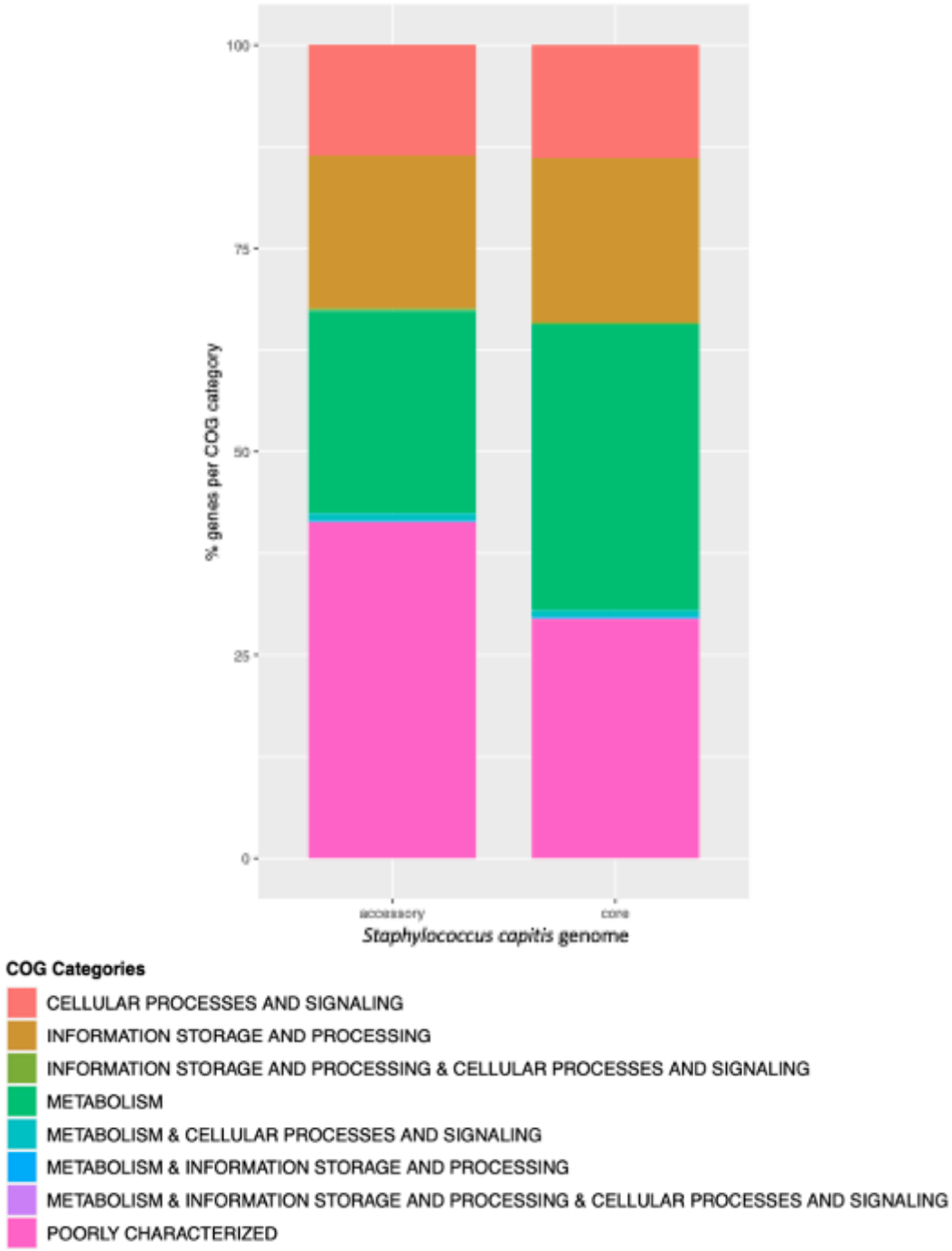


Figure 3.7 Functional annotation of the core and accessory genomes. Percentages of the *S. capitis* core and accessory genomes annotated according to COG categories.

3.3.5 Understanding the differences between *S. capitis* ssp. *capitis* and ssp. *ureolyticus*

3.3.5.1 Identification of biochemical differences between *S. capitis* ssp. *capitis* and ssp. *ureolyticus*

Phylogenetic reconstruction of *S. capitis* isolates included in this study revealed the separation of isolates into two clades by an average pairwise distance of 9254 core genome SNPs, which was concurrent with BAPS analysis. Culture collection type strains and previously described isolates were spread throughout the ML phylogenetic tree, with a clear separation of the *S. capitis* subspecies. Therefore, it was postulated that the separation of the clades represents the separation of *S. capitis* ssp. *capitis* and ssp. *ureolyticus*.

To phenotypically confirm the differences between the subspecies based upon original descriptions, representative isolates from each of the subspecies' clusters were tested with the API *Staph-Ident* Strip system (BioMérieux, Marcy l'Etoile, France), with a particular focus on culture collection type strains (Bannerman and Kloos, 1991, Kloos and Schleifer, 1975). The API *Staph-Ident* strip consists of 20 capsules that contain dehydrated biochemical media for colorimetric tests. Tests were performed and interpreted according to manufacturer instructions.

The API results did not conclusively highlight differences between the subspecies. Bannerman and Kloos (1991) defined *S. capitis* ssp. *ureolyticus* as distinct from ssp. *capitis* as having positive urease and maltose activity, so the results of these tests in particular were noted. A positive urease colorimetric result was identified with a yellow colour and a negative red colour, the opposite

was true for the maltose colorimetric result (Figure 3.8). A positive or negative result is denoted in Table 3.4 by a + or -, and a +/- result represented an orange colorimetric result and limited activity.

The results for *S. capitis* ssp. *capitis* mostly matched with those found previously, with all bar two isolates showing negative results for maltose and urease tests, including culture collection type strain DSM 20327 (Sample 108). However, in contrast to the work of Bannerman and Kloos (1991), tests carried out on *S. capitis* ssp. *ureolyticus* did not confirm positive maltose fermentation or urease activity as discriminating traits of the subspecies. Both culture collection type strains ATCC 49325 (Sample 87) and DSM 6717 exhibited negative colorimetric tests for urease activity and only 3 of the total 11 isolates tested were positive for urease activity. However, 9 out of 11 isolates exhibited maltose fermentation, with 4 isolates showing fully positive results, including both type strains. Although *S. capitis* ssp. *ureolyticus* was previously defined specifically by maltose and urease activity, this study has revealed neither activity is a defining feature.

Table 3.4 *S. capitis* API-Staph test results. Representative *S. capitis* ssp. *capitis* and ssp. *ureolyticus* isolates, including culture collection type strains (*). Tests include substrates (from left to right): NO substrate, D-GLUcose, D-FRUctose, D-ManNosE, D-MALtose, D-LACtose, D-TREhalose, D-MANnitol, XyLiTol, D-MELibiose, potassium NITrate, β-naphthly phosphate, sodium pyruvate, D-RAFfinose, D-XYLose, D-SACcharose, Methy-αD-Glucopyranoside, N-Acetyl-Glucosamine, L-arginine and UREa. Highlighted in blue are tests for maltose fermentation and urease activity, both of which were described as defining features of *S. capitis* ssp. *ureolyticus*. Positive activity is denoted as a +, negative activity as – and +/- for limited activity.

Isolate	Putative *S. capitis* ssp.	API STAPH																			
		O	GLU	FRU	MNE	MAL	LAC	TRE	MAN	XLT	MEL	NIT	PAL	VP	RAF	XYL	SAC	MDG	NAG	ADH	URE
Sample_1	*capitis*	-	+	+	+	-	-	-	+/-	-	-	+	-	+	-	-	+	-	-	-	-
Sample_9	*ureolyticus*	-	+	+	+	+	-	+	+/-	-	-	+	-	+	-	-	+/-	-	-	-	-
Sample_15	*capitis*	-	+	+	+	+/-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	+
Sample_20	*capitis*	-	+	+	+	-	-	-	+/-	-	-	+	-	+	-	-	-	-	-	+	-
Sample_39	*capitis*	-	+	+	+	-	-	-	+/-	-	-	-	-	+	-	-	+	-	-	+	-
Sample_60	*capitis*	-	+	+	+	-	-	-	+/-	-	-	+	-	+	-	-	+/-	-	-	+	-
Sample_62	*ureolyticus*	-	+	+	+	+/-	-	-	-	-	-	+	+	+	-	-	+	-	-	+	-
Sample_65	*ureolyticus*	-	+	+	-	-	-	-	+/-	-	-	-	-	+	-	-	+/-	-	-	+	-
Sample_66	*ureolyticus*	-	+	+	+	+/-	-	-	+/-	-	-	+	-	+	-	-	+/-	-	-	+	-
Sample_87	*ureolyticus**	-	+	+	+/-	+	+	-	+/-	-	-	+	+	+	-	-	+	-	-	+	-
Sample_88	*ureolyticus*	-	+	+	+	+/-	-	-	-	-	-	-	+	+	-	-	+/-	-	-	+	+/-
Sample_104	*ureolyticus*	-	+	+	+	+	+	-	+	-	-	+	-	+	-	-	+	-	-	-	-
Sample_105	*ureolyticus*	-	+	+	+	+	-	-	+	-	-	+	-	+	-	-	-	-	-	-	-
Sample_106	*ureolyticus*	-	+	+	+	+/-	-	-	-	-	-	-	-	+	-	-	+/-	-	-	+	+
Sample_107	*ureolyticus*	-	+	+	+/-	+/-	-	-	-	-	-	-	-	+	-	-	+/-	-	-	+	+
Sample_108	*capitis**	-	+	+	+	-	-	-	+	-	-	+	-	+	-	-	+/-	-	-	+/-	-
DSM 6717	*ureolyticus**	-	+	+	+	+	+	-	+/-	-	-	+	-	+	-	-	+	-	-	+	-

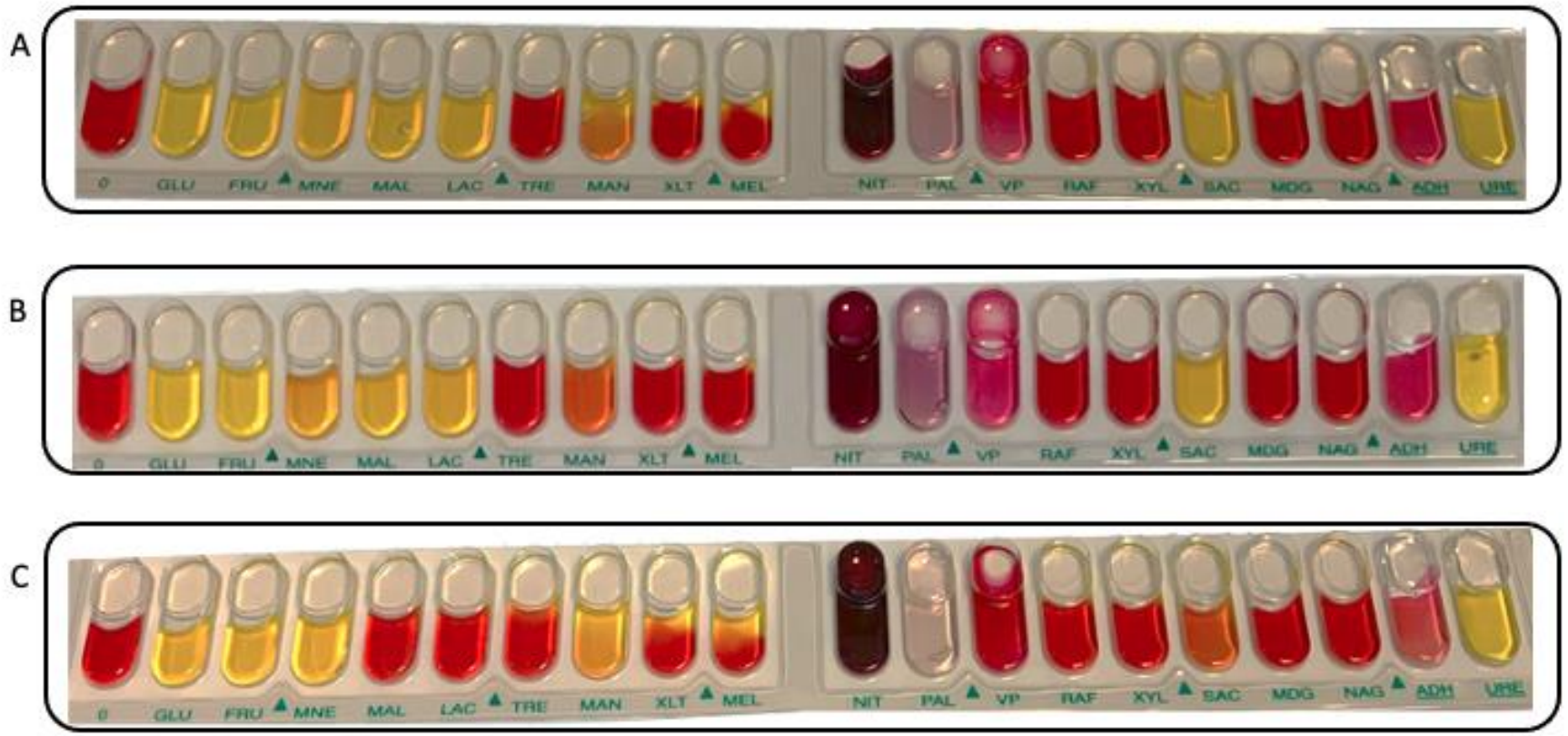


Figure 3.8 *S. capitis* culture collection type strain API-Staph test results. Representative *S. capitis* ssp. *capitis* (C) and ssp. *ureolyticus* isolates (A, B). Isolates DSM 6717 (A), ATCC 49325 (B) and DSM 20327 (C) were tested for substrate activity of a number of species-defining enzymes. Maltose fermentation (MAL) and urease activity (URE) were described as defining phenotypes of *S. capitis* ssp. *ureolyticus*. All isolates show negative urease activity (yellow), positive maltose fermentation in DSM 20327 (C) (red) and negative maltose fermentation in DSM 6717 (A), ATCC 49325 (B) (yellow).

3.3.5.2 Identification of genomic differences between *S. capitis* ssp. *capitis* and ssp. *ureolyticus*

To quantify the genetic relatedness between the two *S. capitis* subspecies this study determined the average nucleotide identity (ANI) indices. Analysis of the average nucleotide identity (ANI) between isolates determined here as *S. capitis* ssp. *ureolyticus* and *S. capitis* ssp. *capitis* revealed that the two shared 96% ANI. An ANI value of >95% indicates test isolates belong to the same species.

Analysis of ANI between culture collection *S. capitis* ssp. *ureolyticus* (DSM 6717 & Sample 87) and the putative *S. capitis* ssp. *ureolyticus* isolates from this study revealed that the ssp. *ureolyticus* isolates share >98% ANI. This study therefore reveals very little genetic difference between the two subspecies, with a high percentage of genetic relatedness.

3.3.6 *S. capitis* cell wall-associated proteins

Cell wall associated (CWA) proteins characterised in *S. capitis* AYP1020 and *S. epidermidis* RP62a were used to characterise the collection of surface proteins, MSCRAMMs and biofilm related proteins in *S. capitis* isolates of this study. CWA proteins investigated included biofilm-associated proteins *icaRADBC*, capsule biosynthesis proteins *capDACB*, surface adhesins *atlE, pls, aap, fbnpA, sesA, sesB, sesC, sesG, ebp* and *bap* and MSCRAMMs *sdrX, sdrZL, sdrH, sdrF* and *sdrG*. Of these genes, 11 were present in *S. capitis* AYP1020 and 14 were present in *S. epidermidis* RP62a (Cameron et al., 2015).

Similar to the study carried out by Cameron *et al,* which characterised the *S. capitis* AYP1020 genome, the genes *sesA, sesB, sesC, sesG, icaRADBC*, *fbnpA, capDACB* and *atlE* were found in all *S. capitis* isolates of this study (2015).

MSCRAMMs *sdrX* and *sdrZL* were also found in 87 of the 88 genomes investigated, and the surface adhesin *pls* was found in more than 50% of *S. capitis* genomes (Figure 3.9).

In addition to this, genes that were absent from the *S. capitis* AYP1020 genome but present in the *S. epidermidis* RP62a including *sdrF, ebp, bap, sdrH, sdrG* and *aap*, were absent, >95% of the genomes included in this study (Figure 3.8)(Cameron et al., 2015).

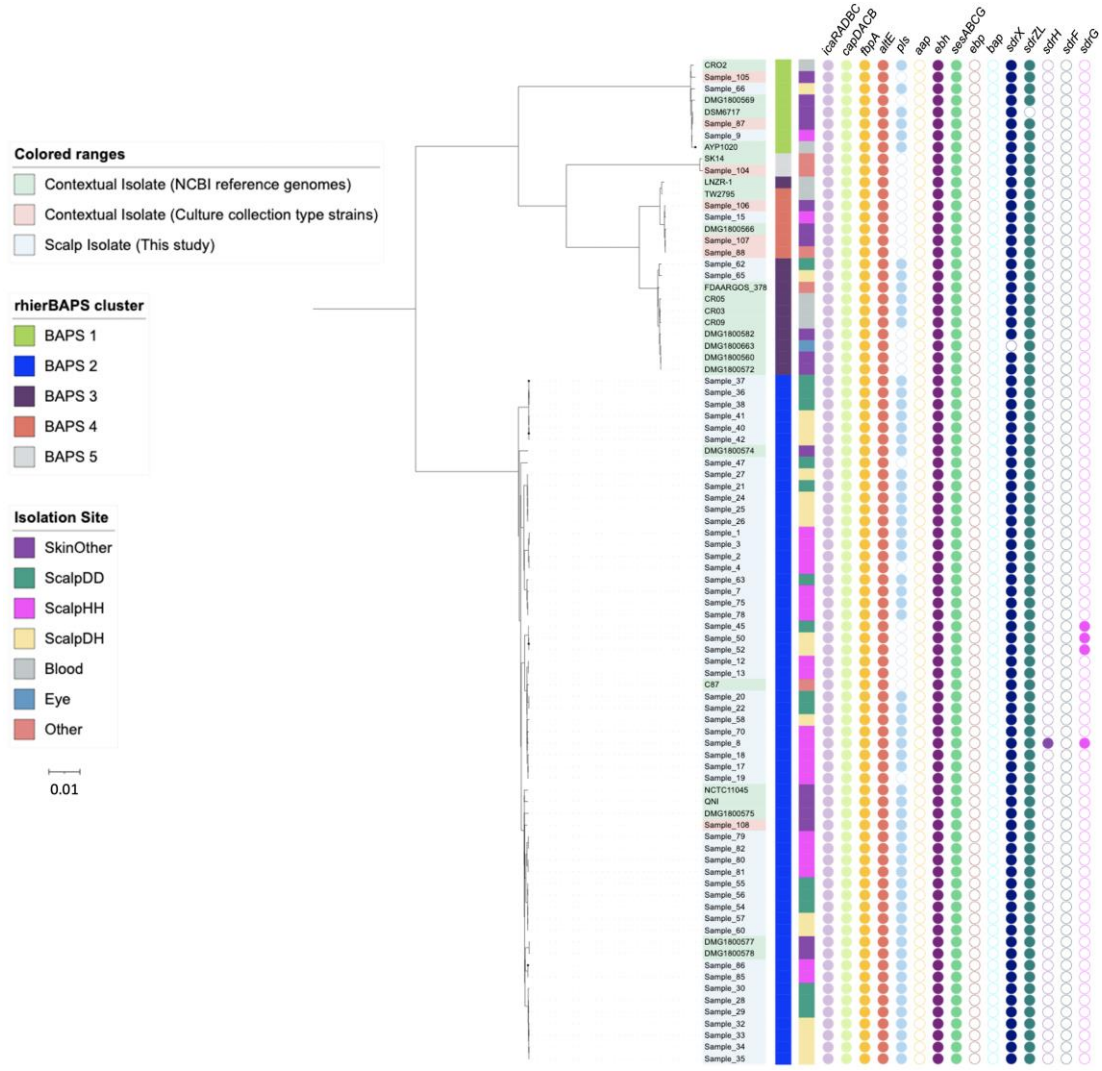


Figure 3.9 ML tree generated based on core genome alignments of *S. capitis* isolates, presenting the presence and absence of staphylococcal cell wall associated proteins. ML tree was generated with IQTREE using the best selected substitution model TVMe_ASC_R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates. ML tree was rooted to *S. caprae* outgroup (isolates not shown). Coloured ranges and the second colour block provide details of the origin of isolates. The first colour block represents rhierBAPS clustering. Presence (filled circle) and absence (empty circle) of cell wall associated proteins is denoted for each isolate. The scale bar represents the number of nucleotide substitutions per site. Figure visualised using iTol v 4.2 (Letunic and Bork, 2016).

3.3.7 *S. capitis* Two-Component Systems

There are 16 known TCSs encoded by *S. aureus* and this study aimed to identify homologs in *S. capitis,* similar to studies of *S. epidermidis* and *S. saprophyticus*. TCS amino acid sequences identified in *S. aureus, S. epidermidis* and *S. saprophyticus* were locally aligned to *S. capitis* genomes included in this study, with a >60% identity threshold used as an indication of TCS presence. *S. capitis* was found to encode 14 of the 16 TCSs found in *S. aureus*. When comparing TCS presence and absence with both *S. epidermidis* and *S. saprophyticus,* this study found that *S. epidermidis* contained all 16 TCS while *S. saprophyticus* contained only 11 TCS. *S. capitis* has most similarity to *S. epidermidis* and *S. aureus* in terms of TCS homology (Rapun-Araiz et al., 2020) (Table 3.5).

Table 3.5 Two-component systems in *S. aureus, S. saprophyticus, S. epidermidis* and *S. capitis*. Presence and absence of the 16 TCS of *S. aureus* (MW2) described in *S. saprophyticus* (ATCC 15305), *S. epidermidis* (ATCC 12228/RP62a) and *S. capitis* (Sample 105, this study) by their corresponding locus IDs.

TCS	*S. aureus*	*S. saprophyticus*	*S. epidermidis*	*S. capitis*	Function
walRK	MW0018	SSP0021	SE0018	KJKAHFGF_01723	Cell wall maintenance, cell viability
	MW0019	SSP0022	SE0019	KJKAHFGF_01724	
hptSR	MW0198		SE0165		Intracellular survival, uptake of hexose phosphate
	MW0199		SE0166		
lytSR	MW0236	SSP0463	SE2011		Autolysis, eDNA release, biofilm
	MW0237	SSP0464	SE2012		
graRS	MW0621	SSP2061	SERP0312	KJKAHFGF_00744	AMPs resistance, growth at low pH
	MW0622	SSP2062	SERP0313	KJKAHFGF_00745	
saeSR	MW0667		SERP0364	KJKAHFGF_01149	Virulence factors regulation
	MW0668		SERP0365	KJKAHFGF_01148	
desKR	MW1208	SSP1146	F1613_RS07815	KJKAHFGF_00026	Uncharacterised function
	MW1209	SSP1547	F1613_RS07810	KJKAHFGF_00025	
arlRS	MW1304	SSP1323	SERP0989	KJKAHFGF_01001	Pathogenesis mechanisms
	MW1305	SSP1324	SERP0990	KJKAHFGF_01002	
srrBA	MW1445	SSP1260	SE1175	KJKAHFGF_00935	

	MW1446	SSP1261	SE1176	KJKAHFGF_00934	Anaerobic respiration, metabolism, growth at low temperatures
phoRP	MW1636	SSP1073	SE1368	KJKAHFGF_00326	Phosphate uptake and homeostasis
	MW1637	SSP1074	SE1369	KJKAHFGF_00325	
airRS	MW1789	SSP0946	SE1529	KJKAHFGF_01868	Oxidative stress response
	MW1790	SSP0947	SE1530	KJKAHFGF_01869	
vraRS	MW1824	SSP0908	SE1569	KJKAHFGF_01778	Cell wall-affecting antibiotic resistance, cell wall biosynthesis
	MW1825	SSP0909	SE1570	KJKAHFGF_01779	
agrCA	MW1962	SSP0839	SE1637	KJKAHFGF_02093	Quorum sensing control of adhesion and virulence factors
	MW1963	SSP0840	SE1638	KJKAHFGF_02094	
kdpDE	MW2002		SERP2489	KJKAHFGF_01712	Potassium homeostasis regulation
	MW2003		SERP2490	KJKAHFGF_01711	
hssRS	MW2282	SSP0540	SERP1953	KJKAHFGF_01263	Haem metabolism regulation
	MW2283	SSP0541	SERP1954	KJKAHFGF_01262	
nreCB	MW2313		SE1969	KJKAHFGF_02138	Response to low oxygen, nitrate reduction
	MW2314		SE1970	KJKAHFGF_02139	
braSR	MW2544		SE2194	KJKAHFGF_00608	Antimicrobial peptide resistance
	MW2545		SE2195	KJKAHFGF_00609	

3.3.8 *S. capitis* virulence profiles

The CoNS are traditionally considered to be non-pathogenic constituents of the healthy human skin and mucosal microbiome but they are now emerging as important opportunistic pathogens (Heilmann et al., 2019). Antimicrobial resistance properties are important factors of nosocomial infection, therefore genetic determinants for AMR were identified using ABRicate v0.9.8 (https://github.com/tseemann/abricate) with NCBI Bacterial Antimicrobial Resistance Reference Gene Database (Feldgarden et al., 2019) and the PlasmidFinder database (Carattoli et al., 2014).

AMR genes responsible for tetracyline, β-lactam, bleomycin, fosfomycin, methicillin resistance, fusidic acid streptogramin A, macrolide, linezolid, trimethoprim and aminoglycoside resistance were found amongst the *S. capitis* genomes included in this study.

The majority of AMR genes were found in isolates associated with neonatal intensive care units and malady afflicted patients, which is expected due to the clinical origin of the isolates. Separately, the healthy scalp and skin isolates Samples 65 and 87, respectively, from this study encoded AMR genes and were classified in this study as *S. capitis* ssp. *ureolyticus*, with the majority of AMR related genes being found in isolates of this cluster. This finding agrees with other studies that found *S. capitis* ssp. *ureolyticus* to be the more virulent of the two subspecies (Cui et al., 2013).

The most resistant strains in terms of AMR gene presence, from this analysis are LNZR-1 (NCBI Genome accession GCA_000712995.1) and TW2795 (NCBI Genome accession GCA_002356175.1), both of which were isolated from blood in

patients with colon cancer or pacemaker related endocarditis (Li et al., 2014). CoNS isolates have been described as serving as reservoirs for antibiotic resistance and virulence traits, especially those isolated from clinical settings (Heilmann et al., 2019).

The AMR gene *dfrC* is present in all *S. capitis* genomes included in this study. This gene is a dihydrofolate reductase linked to trimethoprim resistance in bacteria and is identified in *S. aureus* and CoNS such as *S. epidermidis* (Figure 3.10).

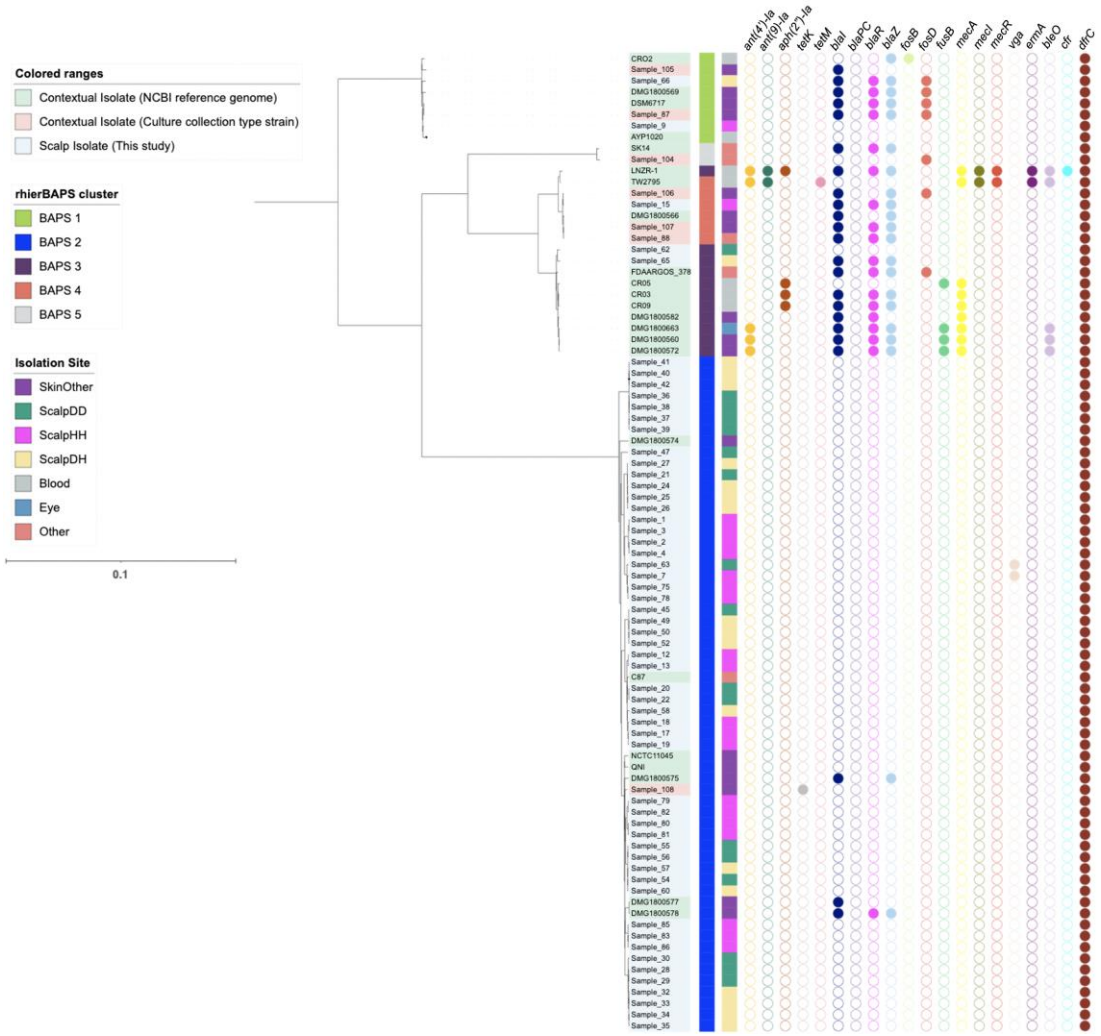


Figure 3.10 ML tree generated based on core genome alignments of *S. capitis* isolates, presenting the presence and absence of antimicrobial resistance genes. ML tree was generated with IQTREE using the best selected substitution model TVMe_ASC_R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates. ML tree was rooted to *S. caprae* outgroup (isolates not shown). Coloured ranges and the second colour block provide details of the origin of isolates. The first colour block represents rhierBAPS clustering. Presence (filled circle) and absence (empty circle) of antimicrobial resistance is denoted for each isolate, based upon ABRicate v0.9.8 (https://github.com/tseemann/abricate). The scale bar represents the number of nucleotide substitutions per site. Figure visualised using iTol v 4.2 (Letunic and Bork, 2016).

In addition to AMR genes present in the *S. capitis* genomes, this study sought to curate PSMs contributing to the virulence potential of *S. capitis*. PSMs are a novel toxin family that have antimicrobial activity and have been attributed to the competitive success of CoNS due to their ability to inhibit the growth of other commensal bacteria such as *Cutibacterium acnes* (O'Neill et al., 2020).

A total of 5 gene clusters encoding β-class PSMs were found within the genomes of this study (Figure 3.11). Gene clusters 2027, 2137 and 2521 were found in over 98% of isolates in this study and shared over 90% similarity when locally aligned to PSMs described and isolated by O'Neill *et al.,* (O'Neill et al., 2020). Similar to AMR gene presence and absence, PSM associated gene clusters were found more abundantly in *S. capitis* ssp. *ureolyticus* compared to *S. capitis* ssp. *capitis*.

Multiple sequence alignment of the 5 gene clusters encoding β-class PSMs revealed the conservation of residues throughout, which have been attributed to the maintenance of the amphipathic nature of the peptides which is essential to antimicrobial activity (Kumar et al., 2017). Kumar *et al*, putatively associate the role of lysine at 3rd and or tryptophan at 20th position in providing antibacterial activity of β-class family peptides (Kumar et al., 2017). These same residues at the same positions are similarly conserved in the peptide sequences of this study (Figure 3.12).

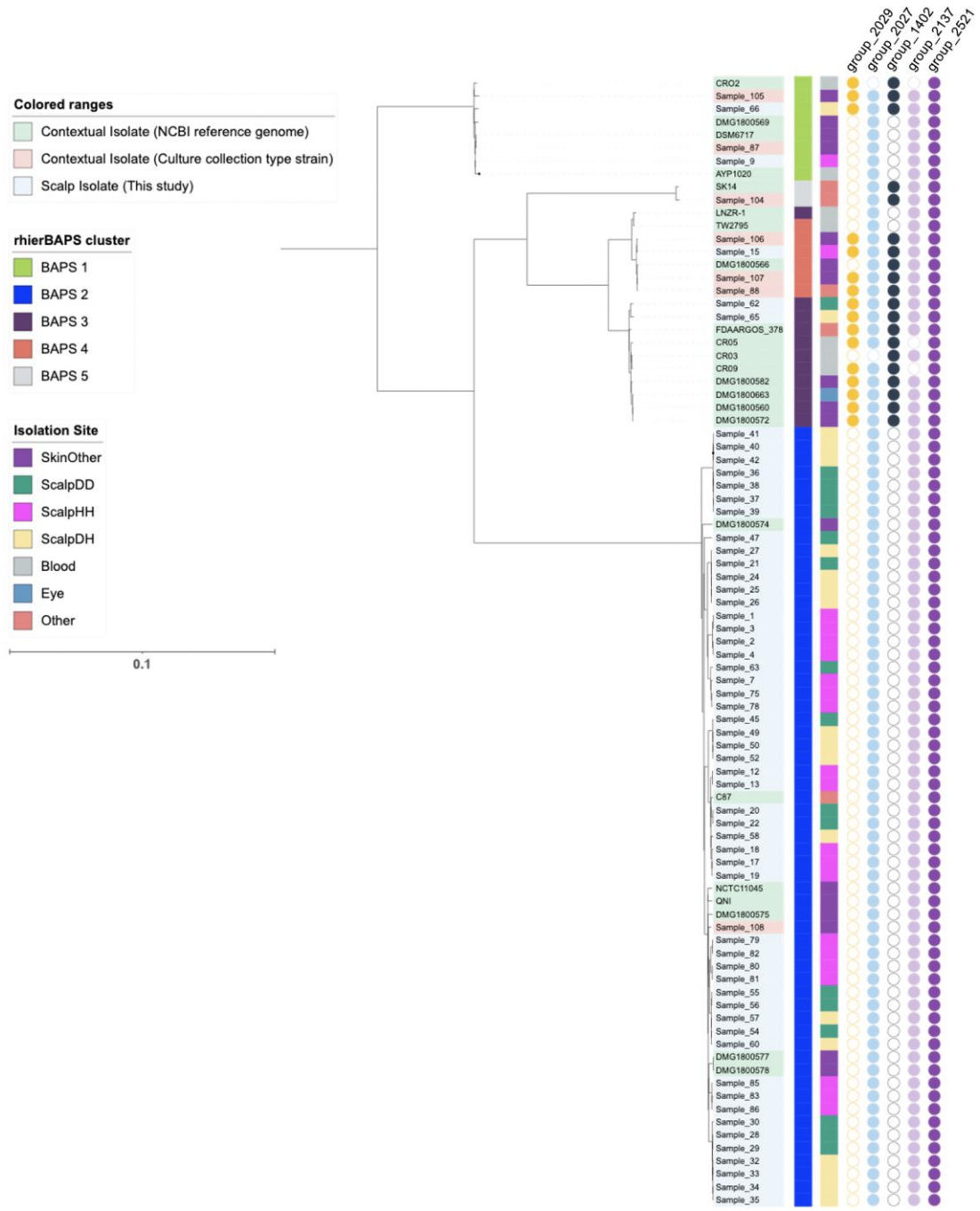


Figure 3.11 ML tree generated based on core genome alignments of *S. capitis* isolates, presenting the presence and absence of gene clusters encoding PSMs. ML tree was generated with IQTREE using the best selected substitution model TVMe_ASC_R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates. ML tree was rooted to *S. caprae* outgroup (isolates not shown). Coloured ranges and the second colour block provide details of the origin of isolates. The first colour block represents rhierBAPS clustering. Presence (filled circle) and absence (empty circle) of gene clusters encoding PSMs is denoted for each isolate. The scale bar represents the number of nucleotide substitutions per site. Figure visualised using iTol v 4.2 (Letunic and Bork, 2016).

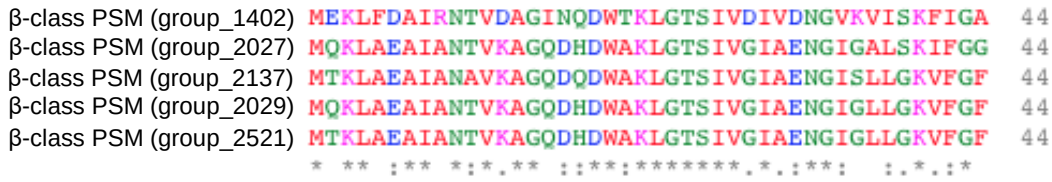


Figure 3.12 Multiple sequence alignment of β-class Phenol Soluble Modulins (PSMs) of *S. capitis* isolates. MSA of β-class PSMs protein sequences found in *S. capitis* genomes , created with ClustalW (Thompson et al., 1994). Residues are coloured based on amino acid property (Red: small and hydrophobic, blue: acidic, magenta: basic, green: hydroxyl, sulfhydryl and amine and grey: unusual), positions that contain fully conserved residues are marked with and asterisk and positions marked with a colon indicate conservation between groups of amino acids with similar properties.

The annotated *S. capitis* genomes included in this study were analysed using antiSMASH to identify genes that encode antimicrobial or bioactive secondary metabolites. The antiSMASH analysis of *S. capitis* sequenced genomes detected secondary metabolite biosynthetic gene clusters (BGCs) in every isolate included in this study. Over 95% of analysed genomes contain BGCs belonging to NRPS, terpene, siderophore and lantipeptide classes, while only 11 genomes contained gene clusters of the bacteriocin biosynthetic class (Figure 3.13). Bacteriocin BGCs were exclusively found in the putatively assigned *S. capitis* ssp. *ureolyticus* cluster, suggestive of a competitive difference between the subspecies.

The two biosynthetic classes, NRPS and siderophore, were found in all genomes with the gene clusters exhibiting 100% similarity to aureusimine and staphyloferrin BGCs respectively. Lantipeptide gene clusters were found to be most similar (~50% similarity) to gallidermin, epicidin and epidermin, whilst BGCs attributed to the bacteriocin and terpene classes were potentially novel as there were no similar clusters found in the antiSMASH database. All BGCs identified by antiSMASH in the isolates of this study have similarity to genes implicated in staphylococcal virulence and competition (Figure 3.13).

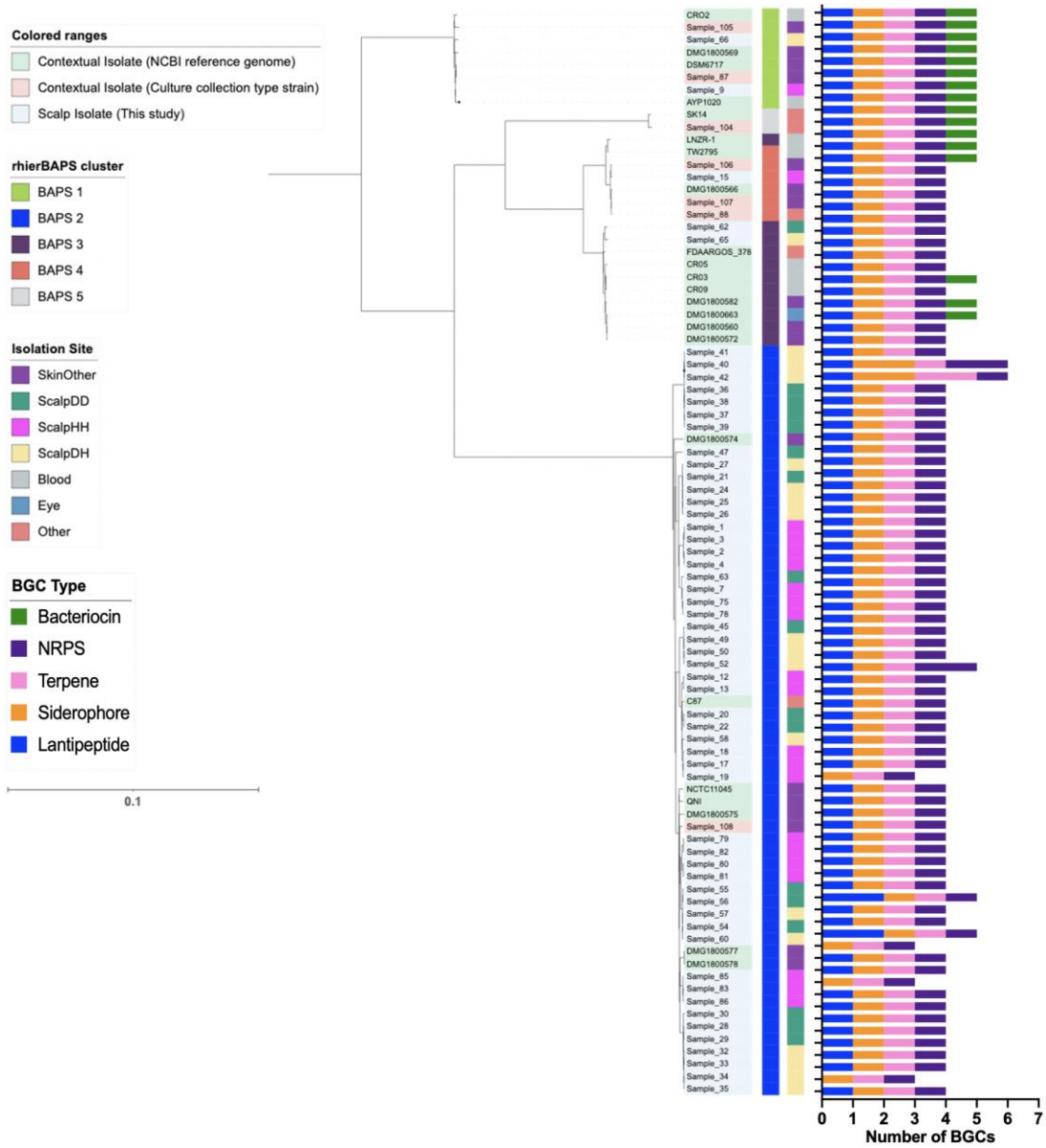


Figure 3.13 ML tree generated based on core genome alignments of *S. capitis* isolates, presenting the number of biosynthetic gene clusters. ML tree was generated with IQTREE using the best selected substitution model TVMe_ASC_R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates. ML tree was rooted to *S. caprae* outgroup (isolates not shown). Coloured ranges and the second colour block provide details of the origin of isolates. The first colour block represents rhierBAPS clustering. The number of secondary metabolite biosynthetic gene clusters predicted by anitSMASH for each *S. capitis* genome is presented (Blin et al., 2019). The scale bar represents the number of nucleotide substitutions per site. Figure visualised using iTol v 4.2 (Letunic and Bork, 2016).

The secreted proteases *hlb, clpP, clpBCX, sepA, htrA, lip geh1*, *geh2* and *lipA* were present in over 95% of all of the 88 isolates included in this study (Figure 3.14). *S. capitis* encodes a suite of exoproteins that are likely to contribute to infection, immune evasion, host colonisation and persistence, important to both pathogenesis and colonisation (Heilmann et al., 2019). Notably, *sspA, sspB* and *sspC* were absent in *S. capitis* genomes, the serine protease SspA promotes invasion in *S. aureus* (Dubin et al., 2001) (Figure 3.14).

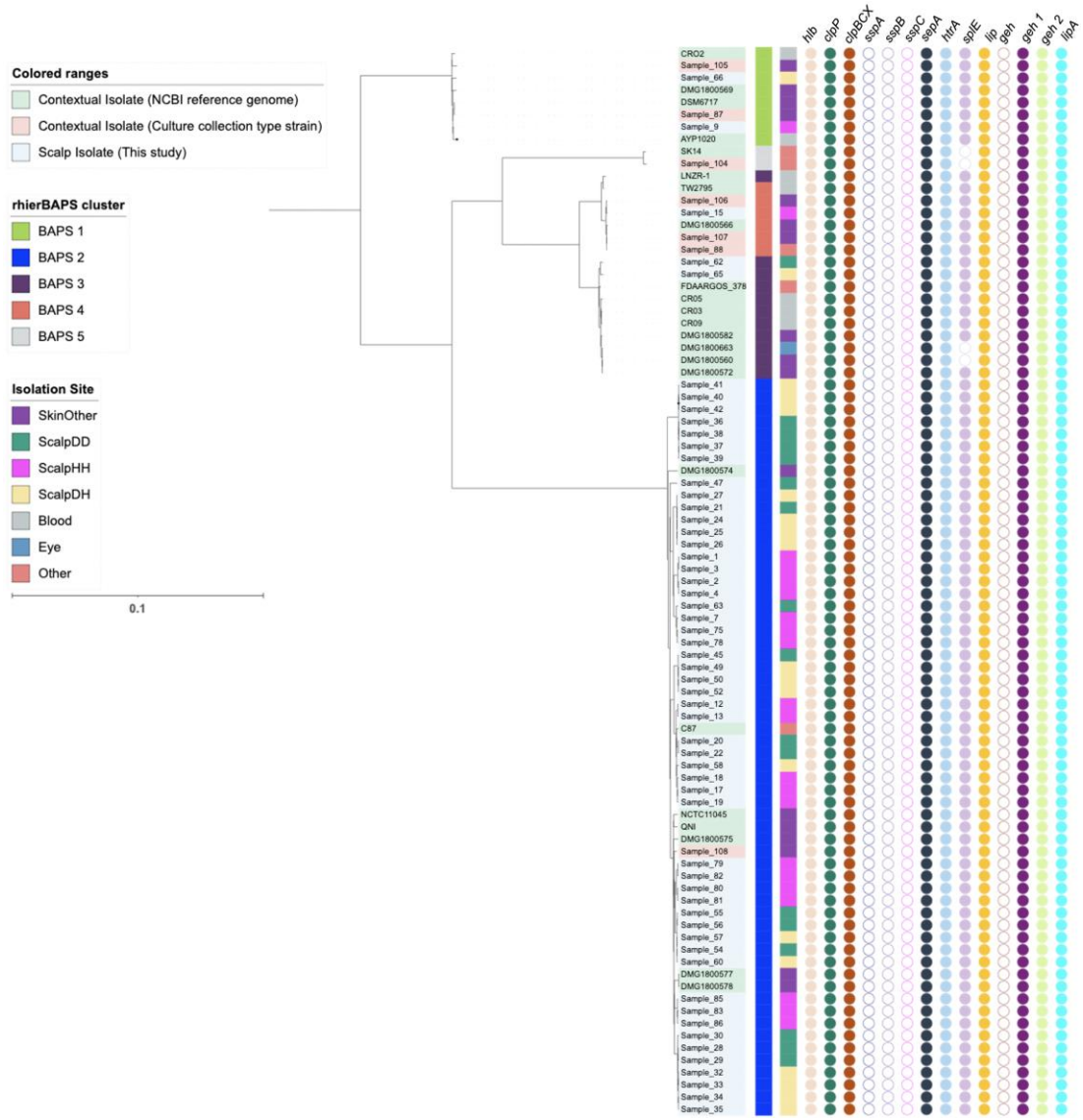


Figure 3.14 ML tree generated based on core genome alignments of *S. capitis* isolates, presenting the presence and absence of gene clusters encoding proteases. ML tree was generated with IQTREE using the best selected substitution model TVMe_ASC_R2 identified by ModelFinder. Bootstrap support value was calculated from 1,000 replicates. ML tree was rooted to *S. caprae* outgroup (isolates not shown). Coloured ranges and the second colour block provide details of the origin of isolates. The first colour block represents rhierBAPS clustering. Presence (filled circle) and absence (empty circle) of gene clusters encoding extracellular proteases is denoted for each isolate. The scale bar represents the number of nucleotide substitutions per site. Figure visualised using iTol v 4.2 (Letunic and Bork, 2016).

3.3.9 Identification of *S. capitis* associated MGEs

Bacteriophages were identified using the PHASTER phage finder program which assigns putative phage regions as intact, questionable or incomplete prophages according to criteria, including the number of coding sequences in the region, and the presence of characteristic phage genes such as tail and capsid proteins.

This study predicted 5 different intact prophages in the isolate collection, including temperate *Siphoviridae*, StB20 and StB12 isolated from *S. capitis* and *S. hominis*; *S. aureus* isolated phage Staus-ST3982-2 and *S. epidermidis* isolated phage Ipla7 (Figure 3.15 & Table 3.6) (Deghorain et al., 2012, Kwan et al., 2005, Gutiérrez et al., 2012, Diene et al., 2017). Many genomes included in this analysis contained incomplete or questionable prophage regions, this could be due to difficulties in MGE assembly causing PHASTER to not appropriately detect them (Table S 8).

Table 3.6 Summary of the prophage present in the *S. capitis* isolates. Complete and intact phage's present *S. capitis* genomes. Complete phages are those with a score above 90.

Isolate	Genome position	Designation	Phage completeness	Possible phage
AYP1020	contig NZ_CP007601.1 :287109-335640	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
C87	contig NZ_GL545272.1: 904359-945507	Intact	150	*Staphylococcus* phage StB20 (NC_019915)
Sample 1	contig 8: 65-50172	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 3	contig 8: 64-50171	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 9	contig 3: 241693-290225	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 12	contig 1: 899413-940560	Intact	150	*Staphylococcus* phage StB20 (NC_019915)
Sample 13	contig 1: 33395-76033	Intact	150	*Staphylococcus* phage StB20 (NC_019915)
Sample 15	contig 1: 578378-601260	Intact	130	*Staphylococcus* phage PT1028 (NC_007045)
Sample 28	contig 1 :33366-77113	Intact	150	*Staphylococcus* phage StB12 (NC_020490)

Sample 29	contig 1 :33367-77114	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 30	contig 1 :33367-77114	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 32	contig 1 :33367-77114	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 33	contig 1 :33366-77113	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 34	contig 1 :33361-77108	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 35	contig 1 :33364-77111	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 57	contig 1 :892102-933248	Intact	150	*Staphylococcus* phage StB20 (NC_019915)
Sample 60	contig 1 :828203-879051	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 62	contig 1 :828203-879051	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 66	contig 1 :730314-789502	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 83	contig 4 :81693-131375	Intact	150	*Staphylococcus* phage Ipla7 (NC_018284)

Sample 85	contig 3 :66468-116803	Intact	150	*Staphylococcus* phage Ipla7 (NC_018284)
Sample 86	contig 4 :20800-71135	Intact	150	*Staphylococcus* phage Ipla7 (NC_018284)
Sample 87	contig 3 :321632-355515	Intact	110	*Staphylococcus* phage StB12 (NC_020490)
	contig 6 :126-27823	Intact	110	*Staphylococcus* phage StB12 (NC_020490)
Sample 104	contig 11 :33360-71846	Intact	120	*Staphylococcus* phage StB20 (NC_019915)
	contig 12 :34896-78353	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 105	contig 3 :608-47688	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 106	contig 14 :113-37599	Intact	120	*Staphylococcus* phage StB20 (NC_019915)
	contig 20 :147-43363	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
Sample 108	contig 13 :14221-68692	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
CR03	contig NZ_CTEB01000001.1 2096596-2146563	Intact	150	*Staphylococcus* phage StB12 (NC_020490)

CR05	contig NZ_CTEO01000014.1 4149-37790	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
	contig NZ_CTEO01000015.1 2581-34828	Intact	110	PHAGE_Staphy_StauST398_2_NC_021323(9)
CR09	contig NZ_CTEL01000003.1:16480-66671	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
CRO2	contig NZ_CZWI01000112.1:1-46866	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
DMG1800560	contig 1 :130716-176448	Intact	130	*Staphylococcus* phage StB20 (NC_019915)
	contig 20 :1663-42759	Intact	150	*Staphylococcus* phage StB20 (NC_019915)
DMG1800566	contig 2 :128582-173012	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
DMG1800569	contig 1 :193412-247705	Intact	130	*Staphylococcus* phage StB20 (NC_019915)
	contig 4 :128599-177130	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
DMG1800572	contig 1 :59927-103907	Intact	130	*Staphylococcus* phage StB20 (NC_019915)

	contig 13 :1936-43032	Intact	150	*Staphylococcus* phage StB20 (NC_019915)
DMG1800582	contig 17 :2406-47222	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
DMG1800663	contig 3 :129430-173658	Intact	130	*Staphylococcus* phage StB20 (NC_019915)
	contig 14 :5288-52700	Intact	150	*Staphylococcus* phage StB20 (NC_019915)
FDAARGOS_378	contig NZ_CP023966.1:861963-907549	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
LNZR-1	contig NZ_JGYJ01000008.1:257-39890	Intact	120	*Staphylococcus* phage StB20 (NC_019915)
QNI	contig NZ_AJTH01000001.1:564481-615373	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
SK14	contig NZ_ACFR01000015.1:159-4494	Intact	150	*Staphylococcus* phage StB12 (NC_020490)
TW2795	contig NZ_AP014956.1:1689028-1732990	Intact	130	*Staphylococcus* phage StB20 (NC_019915)

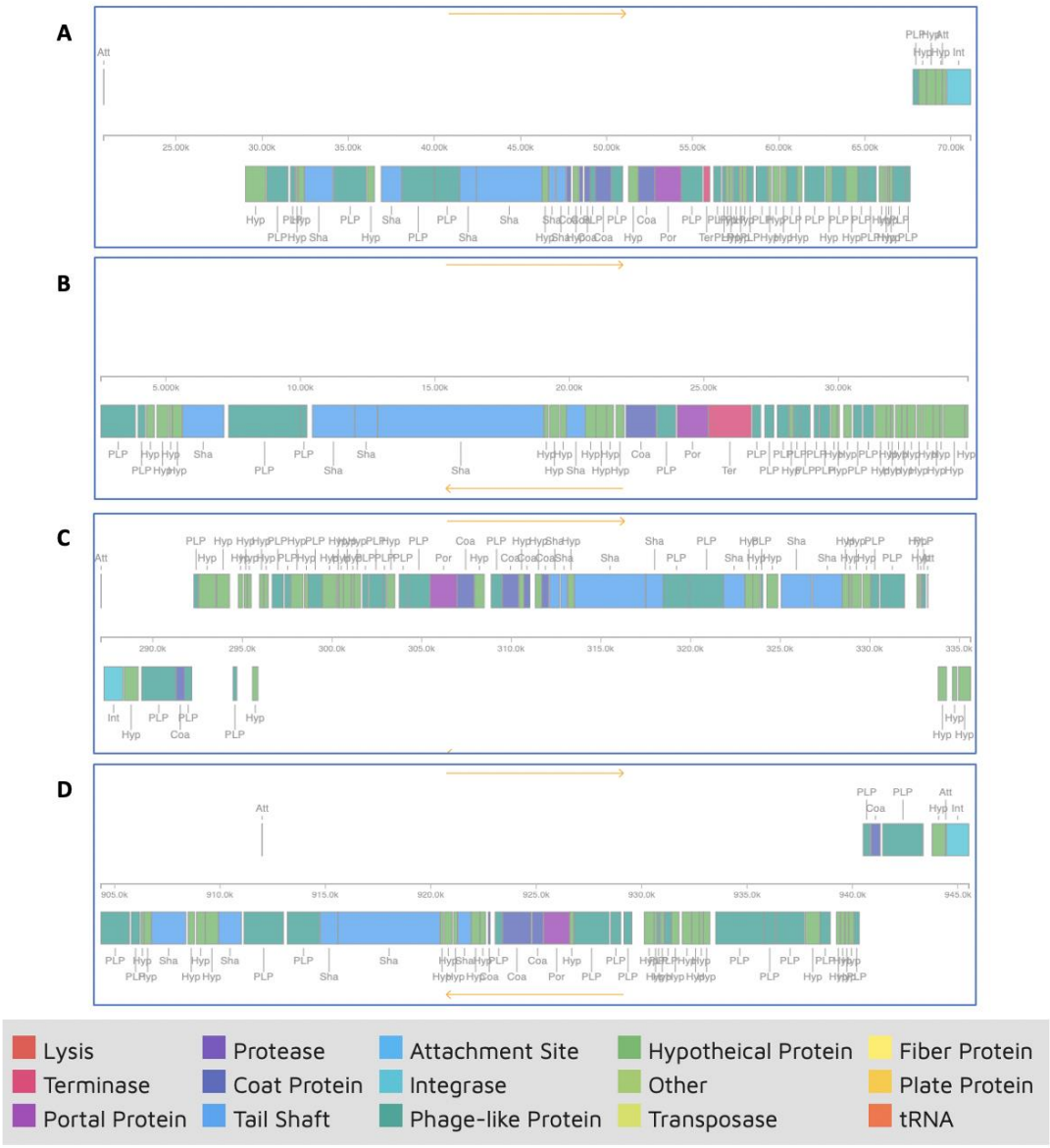


Figure 3.15 Representative intact prophage regions of staphylococcal phages. Graphical representation of intact prophages (A) Ipla7, (B) Staus-ST8982-2, (C) StB12 and (D) StB20, showing CDS regions. Images created in PHASTER (Arndt et al., 2016).

Plasmids were identified using two main approaches, firstly using ABRicate v0.9.8 (Seemann, 2020), which searches genomes for potential plasmid sequences using the PlasmidFinder database (Carattoli et al., 2014) and secondly, investigate the presence of pAYP1020 (RefSeq NZ_CP007602.1) within *S. capitis* isolates.

PlasmidFinder analysis did not reveal the presence of any plasmids in the *S. capitis* genomes included in this study which could be the result of an unexhaustive plasmid database, especially as no plasmids were found in the AYP1020 genome, which is confirmed to contain the pAYP1020 plasmid (RefSeq NZ_CP007602.1).

To investigate whether the pAYP1020 plasmid (RefSeq NZ_CP007602.1) described by Cameron *et al.,* was an ancestral plasmid acquired in a progenitor strain, the plasmid sequence (RefSeq NZ_CP007602.1) was mapped against all genomes and mapping coverage assessed as a means of predicted presence of the plasmid within any given genome. Sample 17 appeared to possess the plasmid pAYP1020 due to 99% mapping coverage, however all other isolates apart from Sample 19, which had 61% mapping coverage had less than 44% coverage. It can therefore be assumed that while not all isolates contain the pAYP1020 plasmid specifically, they do contain some plasmid sequence. As described above, BGCs for bacteriocin were identified in multiple isolates and are typically located on plasmids.

3.4 Discussion

The current work aimed to explore whole genome sequencing of *S. capitis* to understand its population genetic structure, genotypic definition of subspecies and to identify if any competitive traits are linked to *S. capitis* isolated from dandruff afflicted scalps.

A, ML phylogenetic tree based on core genome alignment of 88 isolates derived mostly from the scalp, but also including contextual references provided by Unilever Plc, isolates available from type culture collections and publicly available reference genomes (Table 3.1), revealed two distinct clades. Here, I demonstrated that *S. capitis* ssp. *capitis* and *S. capitis* ssp. *ureolyticus* isolates cluster distinctly based upon their subspecies with isolates described in literature and culture collection type strains for each *S. capitis* subspecies showing a clear separation in the ML tree (Figure 3.1, Figure 3.2 & Figure 3.3).

The most widely used, outgroup and midpoint rooting methods were explored in this study to allow evolutionary and taxonomic inferences to be made (Maddison et al., 1984, Wheeler, 1990, Boykin et al., 2010). Most of the described phylogenetic reconstructions of *S. capitis* are based upon multi-locus datasets of 16s rRNA genes and are rooted with *Micrococcus caseolyticus* as *Micrococcus* is the most closely related genus to *Staphylococcus* (Baba et al., 2009, Li et al., 2014, Naushad et al., 2016). A recent *S. capitis* ML phylogenetic tree presented by Carter et al., uses a minimum ancestor deviation algorithm (Tria et al., 2017). This study found *M. caseolyticus* to be an unsuitable outgroup, due to minimal alignment to the *S. capitis* AYP1020 reference genome, and *S. caprae* was, as an alternative. All phylogenetic reconstructions presented in this study revealed the same

subspecies clustering and similar tree topology, whether midpoint or outgroup rooted.

The rarefaction curve (Figure 3.6) showed that as genomes were sampled, the addition of each genome at random cause an increase in the total number of genes and number of unique genes in the pangenome (Figure 3.6) indicative of an open pan-genome of *S. capitis* (Medini et al., 2005). The results presented here are similar to CoNS pangenome studies that found the pangenomes of *S. epidermidis, S. capitis* and *S. caprae* to be open (Sun et al., 2020). The pangenome status of an organism is partially dependent on its capacity of acquiring exogenous DNA, particularly for species living in bacterial communities (Diene et al., 2013). CoNS inhabiting the skin have high horizontal gene transfer (HGT) rate and are most likely to have an open pangenome (Georgiades and Raoult, 2010). In contrast to other CoNS, *S. lugdunensis* exhibits a closed pangenome, which was attributed to the identification of CRISPR/Cas, RM and T/AT loci that limit HGT (Argemi et al., 2018).

Clonal diversity among CoNS species varies and is less studied than that of *S. aureus* (Becker et al., 2014). *S. epidermidis* studies have revealed well-adapted clones evolving rapidly through genetic recombination by frequent transfer of mobile genetic elements, which contrasts with *S. aureus* that evolves preferentially by point mutation (Miragaia et al., 2007). While *S. epidermidis* is characterised by pronounced genomic diversity, other CoNS species, such as *S. haemolyticus* and *S. schleiferi* exhibit less diversity (Lina et al., 1992). Recent WGS studies on *S. capitis* have also revealed extensive recombination within genomes (Carter et al., 2018), which is consistent with the findings in this study.

Functional classification of the *S. capitis* pangenome according to COG category showed that the majority of core gene clusters were successfully assigned function, compared to those of the accessory genome. While there were still unassigned core gene clusters, this suggests that the core genes of *S. capitis* are largely well-studied and described. In addition, the large number of unassigned gene clusters of the accessory genome supports the presence of novel gene clusters and highlights the diverse biological characteristics of CoNS (Figure 3.7).

The subdivision of *S. capitis* to both ssp. *ureolyticus* and ssp. *capitis* was based upon original descriptions of *S. capitis* ssp. *ureolyticus* urease activity, ability to produce acid from maltose, fatty acid profile, larger colony size and DNA sequence differentiation (Bannerman and Kloos, 1991). Currently, in the literature there is only one previous study that examines the prevalence, phenotypic characteristics and molecular epidemiology (Cui et al., 2013). Cui *et al*., characterised a collection of *S. capitis* isolates associated with neonatal intensive care units (NICUs) in terms of antimicrobial susceptibility, structure of the *ica* operon and biofilm phenotype (2013). They found in a clinical setting that *S. capitis* ssp. *ureolyticus* isolates were characterised by penicillin, erythromycin and oxacillin resistances, and positive biofilm formation, while *S. capitis* ssp. *capitis* isolates were generally antibiotic susceptible and biofilm-negative (Cui et al., 2013).

This study therefore aimed to further characterise the two subspecies, by combining the diseased and healthy scalp-associated isolates with those isolates from NICUs. The API *Staph-Ident* system was used for biochemical analysis of *S. capitis* similar to the methods used in the Bannerman *et al*., study that first

identified *S. capitis* ssp. *ureolyticus* (1991). Biochemical analysis of culture collection type strains confirmed as either of the two *S. capitis* subspecies, plus isolates described in this study as either subspecies, did not conclusively confirm differences between the two subspecies based on urease activity and maltose fermentation. However, later studies have described the API system as only correctly identifying 50-82% of 289 CoNS tested (Renneberg et al., 1995). The inconclusive results presented in this study could be because the API system is not sensitive enough to consistently determine biochemical reactions of isolates. The discrepancy revealed by Rosenberg *et al*., (1995) might indicate that there is greater biochemical diversity in the CoNS than first appreciated by the designers of the API system. Additional experiments using alternative methods for urease and maltose fermentation testing will provide more reliable results but routine identification of isolates by API seems very limited for *S. capitis*.

Analysis of the genetic relatedness between the two subspecies revealed the two shared 96% ANI, 2% less than the relatedness between the isolates of the *ureolyticus* subspecies. These values were compared to the DDH studies originally carried out to compare the DNA relationship of *S. capitis* ssp. *ureolyticus* at the time of subspecies identification (Bannerman and Kloos, 1991), as ANI estimations reflect those from DDH methodology (Jain et al., 2018a). DDH comparing the two subspecies found a relatedness value >70% that was consistent with the test isolates being the same species and scores were lower than those resulting from comparisons between *S. capitis* ssp. *ureolyticus* isolates (Bannerman and Kloos, 1991).

The ANI results agree with those from DDH presented by Bannerman and Kloos, though the difference in genetic relatedness between the two subspecies and within those determined to be *ureolyticus* is not significant (1991). This study therefore reveals very little (4%) overall genetic difference between genomes of the two subspecies, with a high percentage (96%) of genetic relatedness. Although the differences in gene content were few and represented a small percentage of the whole genome, they were mostly concurrent with current literature and highlight differences in virulence between the two subspecies.

Discriminating gene clusters of note that are present in *S. capitis* ssp. *capitis* included genes linked to the arginine catabolic mobile element (ACME), encoding arginine deiminase activity found in various species of staphylococci (Lindgren et al., 2014). One study proposed, this operon confers increased ability to survive and proliferate under acidic conditions, similar to those found on the skin (Thurlow et al., 2013). This would be advantageous to commensal CoNS found on the scalp as the pH is 5.5 and 3.67 at the hair shaft (Gavazzoni Dias et al., 2014). It has been described that ACME was transferred from *S. epidermidis* into *S. aureus* and was associated with the virulent *S. aureus* strain USA300 (Granslo et al., 2010) but does not define community-acquired MRSA (Thurlow et al., 2012). The presence of pathogenicity-linked gene clusters in the ssp. *capitis* disagrees with one study proposing that the ssp. *capitis* is less virulent than ssp. *ureolyticus* (Cui et al., 2013). Granslo, *et al* found that ACME in CoNS *S. epidermidis* was associated with less antibiotic resistance, concurrent with the description of *S. capitis* ssp. *capitis* as having less antibiotic resistance genes (Cui et al., 2013, Granslo et al., 2010).

S. capitis ssp. *ureolyticus* was found in this study to encode multiple gene clusters linked to antimicrobial resistance. Gene clusters with antimicrobial function included β-lactam resistance genes and β-class phenol soluble modulins. In addition to this the gene cluster AtlE, important to CoNS virulence was also differentially present in *S. capitis* ssp. *ureolyticus* along with genes important in metalloid transportation, acquisition and resistance, staphylopine biosynthesis and copper transportation.

These findings agree with published data that also found *S. capitis* ssp. *ureolyticus* to be the more virulent of the two subspecies due to antimicrobial resistance and biofilm formation ability (Cui et al., 2013). However, comparisons made with the study by Cui *et al* (2013) are limited as strains in this study are primarily isolated from clinical samples, which may not be representative of the subspecies as they can be both commensal and pathogen. The presence of more virulent strains of *S. capitis* ssp. *ureolyticus*, in greater frequencies in clinical settings as well as the presence of less virulent strains commensally on the scalp, suggests that HGT events have led to the acquisition of virulence genes as the *S. capitis* genome appears to be highly recombinogenic (Figure 3.4).

To study this further, a larger study with more commensal and clinical representative samples for each subspecies could be used to carry out a genome wide association study (GWAS) to reveal differences between the subspecies linked to a number of discriminative phenotypes.

To further analyse potential virulence differences within *S. capitis*, genetic determinants for AMR for the 88 isolates were identified. A total of 20 AMR genes were found across all *S. capitis* isolates (Figure 3.10). AMR genes were associated

with resistance to tetracycline, β-lactam, bleomycin, fosfomycin, methicillin, fusidic acid streptogramin A, macrolide, linezolid, trimethoprim and aminoglycosides. The majority of AMR genes were found in *S. capitis* ssp. *ureolyticus* and in isolates associated with outbreaks in neonatal intensive care units and other clinical samples. This coincides with the description of clinically isolated CoNS and *S. capitis* ssp. *ureolyticus* specifically, as reservoirs for antibiotic resistance and virulence traits (Heilmann et al., 2019, Cui et al., 2013).

The most prevalent AMR of the *S. capitis* isolates in this study was *dfrC* a dihydrofolate reductase linked to trimethoprim resistance in bacteria and is identified in *S. aureus* and CoNS such as *S. epidermidis*. Trimethoprim resistance in community-acquired methicillin resistant *S. aureus* was reported in regions where trimethoprim is a treatment option for skin and soft tissue infections; this AMR gene may therefore be acquired via *S. aureus* (Harris et al., 2018) (Figure 3.10).

Several studies have described the presence of a similar repertoire of AMR genes in CoNS isolates (Moawad et al., 2019, Sun et al., 2020). This study would benefit from additional phenotypic experiments to evaluate antibiotic susceptibility of *S. capitis* isolates, comparing the AMR profiles of both subspecies, along with biofilm formation, as has been described in other studies.

The emergence of reduced vancomycin susceptibility in CoNS was proposed to be a response to the use of vancomycin as a first-line antimicrobial therapy to treat infections caused by CoNS isolates carrying the *mecA* gene associated with β-lactam antibiotic resistance (Chong et al., 2016, Mello et al., 2008). Although the majority of samples in this study are not clinical isolates, the profiling of

vancomycin resistance could lead to further understanding of AMR across the *S. capitis* population.

This study reveals the presence of additional gene clusters that could contribute to *S. capitis* virulence. Gene clusters encoding β-class PSMs were found within the genomes of this study (Figure 3.11). β-class PSM gene clusters in this study (group_2029, group_2027, group_1402, group_2137 and group_2521) had sequence similarity to PSMs described and isolated by O'Neill et al (2020) (Figure 3.11 & Figure 3.12). The same β-class PSMs showed the ability to inhibit the growth of other commensal bacteria such as *C. acnes*, a property which would be beneficial to *S. capitis* isolates residing on the scalp as *C. acnes* is one of the major bacterial colonisers of the scalp (Xu et al., 2016). β-class PSMs found in this study also revealed the conservation of residues throughout, which have been attributed to the maintenance of the amphipathic nature of the peptides which is essential to antimicrobial activity (Kumar et al., 2017) (Figure 3.11 & Figure 3.12).

Two PSMs found in this study (group_1402 and group_2029) were found only in *S. capitis* ssp. *ureolyticus* isolates (Figure 3.11 & Figure 3.12). These may form part of the more extensive repertoire of virulence factors of *S. capitis* ssp. *ureolyticus*. As these two PSMs are only encoded by a select few isolates, their presence maybe as a result of a HGT event, encoded on a MGE transferred by other CoNS or the more virulent *S. aureus* (Cheung et al., 2014, Otto, 2013). On the other hand, these two PSMs may also be encoded on the *S. capitis* ssp. *ureolyticus* core genome as they are found in virtually all *S. capitis* ssp. *ureolyticus*

isolates similar to other staphylococci and thus these PSMs appear to be a defining characteristic of this subspecies.

Commensal staphylococci are known to produce a number of antimicrobial secondary metabolites. This study sought to investigate the presence of gene clusters in *S. capitis* genomes. Biosynthetic gene clusters were found in all 88 genomes analysed. The biosynthetic classes NRPS and siderophore were found in all genomes with the gene clusters exhibiting 100% similarity to aureusimine and staphyloferrin. Iron availability is of major importance in bacterial pathogenesis as iron is an essential nutrient and staphyloferrin is a siderophore and is involved in iron retrieval, (Beasley et al., 2009). Whereas aureusimine, a dipeptide, has been implicated in the activation of expression of virulence genes in *S. aureus* (Sun et al., 2011), protease inhibition in eukaryotic cells, and interspecies bacterial communication (Secor et al., 2012). The presence of these BGCs may provide, *S. capitis* isolates of this study with virulence and competitive advantages, especially on the scalp in the presence of eukaryotes such as *Malassezia* spp.. Bacteriocins also contribute to virulence and provide a competitive advantage to staphylococci. This biosynthetic class can exert a diverse array of functions, such as colonising peptides that help staphylococci become established in a niche, killing peptides for competitor removal or signal peptides that can interact with bacteria and the immune system (Dobson et al., 2012, O'Sullivan et al., 2019). BGCs encoding bacteriocins were also exclusively found in the genomes of *S. capitis* ssp. *ureolyticus*, adding another determinant to the large virulence repertoire of this subspecies and highlighting the HGT of more virulent traits to this subspecies.

Commensal bacteria such as *S. capitis* are also known to produce lantibiotics which belong to the class of cationic antimicrobial peptides and are active against Gram-positive bacteria, by inducing pores in the cytoplasmic membrane (Becker et al., 2014). Similarity was identified between the lantibiotic BGCs found in the *S. capitis* isolates of this study and known lantibioitcs such as gallidermin, epicidin and epidermin; all of which are competition determinants (Beasley et al., 2009, Sahl and Bierbaum, 1998). The lantibiotic BGCs detected by antiSMASH may be *S. capitis* species specific, similar to those produced by other CoNS such as *S. hominis* and *S. warneri* (Sahl and Bierbaum, 1998).

Both PSMs and exoproteases are strictly regulated by the *agr* locus in CoNS, and they are proposed to be important in biofilm structuring (Otto, 2009, Dubin et al., 2001). The presence of the global regulators *agr* and *sar* loci throughout all of the *S. capitis* isolates of this study and the presence of gene clusters that are regulated by these global regulators, supports that *S. capitis* have inherent virulence potential and are active pathogens (Vuong et al., 2000, Fluckiger et al., 1998).

This study reveals that *S. capitis* genomes encode a plethora of proteases. Functionally, the lipases: *lip, geh1, geh2* and *lipA* are important for staphylococcal persistence in fatty acid secretions (Heilmann et al., 2019). The metalloprotease SepA is important to lipase maturation and antimicrobial peptide maturation, similar to the action of HtrA (Rigoulay et al., 2005, Lai et al., 2007). Exoenzymes Hlb and ClpP, B, C and X found in *S. capitis* are linked to cell membrane destabilisation and biofilm formation (Wang et al., 2007a, Kmieciak et al., 2016).

The genes *sspA, sspB* and *sspC* were absent in, *S. capitis* genomes; the serine protease, *sspA* promotes invasion by degrading fibronectin-binding MSCRAMMs

(Dubin et al., 2001). SspA also stimulates the production of AMPs by keratinocytes, which in turn inhibit the colonisation of the epithelia by *S. aureus* (Iwase et al., 2010). These *S. capitis* observations are similar to those reported by Sun, *et al*. (2020) (Figure 3.14).

Cell wall-associated (CWA) proteins in staphylococci and other Gram-positive bacteria are exposed proteins on the cell surface. CWA proteins therefore promote interactions with target ligands on host cells and the extracellular matrix, such as collagen, fibrinogen and fibronectin (Foster, 2019). CoNS such as *S. epidermidis* have a smaller repertoire of cell wall anchored surface proteins than *S. aureus* but they each promote adherence, biofilm formation, persistence and host immune evasion (Foster et al., 2014).

S. capitis CWA proteins important for intercellular aggregation, including the *ica* locus enzymes involved in PNAG biosynthesis (Otto, 2009) and *capDACB* operon enzymes involved in capsular biosynthesis, are commonly present and may lead to immune evasion (Otto, 2009, Fournier et al., 2013). Genes commonly associated with biofilm formation of *S. epidermidis,* such as *sdrF, ebp, bap, sdrH, sdrG* and *aap*, were not found to be present in this study (Sun et al., 2020, Otto, 2009). Absence could indicate that the *S. capitis* lack of MSCRAMM-related proteins may constrain their ability to colonise particular surfaces and form biofilms. A restricted set of biofilm protein related genes found in *S. capitis,* with all isolates containing *atlE, fbnpA, sdrX, sdrZL*, *sesA, sesB, sesC* and *sesG*, supporting that *S. capitis* possesses multiple biofilm formation effectors that have been described in staphylococci (Heilmann et al., 2019).

Two-component systems (TCS) encoded by staphylococcal genomes control responses to the environmental stimuli and afford them with the capacity to colonise and survive on various body surfaces (Crosby et al., 2020). There are 16 known TCS encoded by *S. aureus* and this study identified 14 homologs in *S. capitis*, compared with 16 described for *S. epidermidis* and 11 in *S. saprophyticus* (Rapun-Araiz et al., 2020).

S. capitis is missing the TCS HptSR and LytSR. HptSR is required to sense extracellular hexose phosphates and to activate the transcription of *uhpT* encoding a glucose 6-phosphate transporter of *S. aureus* that is also involved in fosfomycin uptake (Park et al., 2015). HptSR TCS was found to be important for intracellular survival and multiplication of *S. aureus* within host cells consistent with its more invasive phenotypes (Rapun-Araiz et al., 2020). The LytSR TCS plays a role in *S. aureus* biofilm development as a result of its direct influence on *lrgAB* and *cidABC* operons controlling cell death and lysis during biofilm formation (Sharma-Kuinkel et al., 2009).

The absence of both TCS is consistent with lower virulence of *S. capitis* but could also reflect speciation and diversification of niche colonisation during evolution. Differences in metabolic pathways, likely necessary to colonise hostile environments by *S. aureus* and *S. epidermidis* and reflect the possibility of a more specialised capacity of *S. capitis*. Research is ongoing as to whether a common staphylococcal ancestor contained all 16 TCSs with certain species have suffered genome reduction events or whether TCS were acquired over time (Rapun-Araiz et al., 2020). The BraSR TCS that responds to antimicrobial peptides is mostly

restricted to the 'Epidermidis' branch of the staphylococcal species and has high similarity to GraSR suggestive of gene duplication (Coates-Brown et al., 2018).

The accessory genome of many bacteria contains mobile genetic elements, therefore this study aimed to characterise the repertoire of plasmids and phage in the *S. capitis* genome. The representation of phage in the *Staphylococcus* genus is predominantly made up of those infecting *S. aureus*, although the number of CoNS phage characterised is increasing as their public health relevance increases their study and sequencing (Deghorain et al., 2012). The isolation and characterisation of under-sampled CoNS infecting viruses may reveal the presence of new genes and relationships that shape the evolution of staphylococcal phage genomes, especially as CoNS undergo more significant recombination events (Oliveira et al., 2019).

All 88 genomes of this study carried identifiable mobile genetic elements including those identified by PHASTER. Prophage *Siphoviridae*, StB20 that was isolated from *S. capitis* (Deghorain et al., 2012), *S. aureus* isolated phage Staus-ST3982-2 and phage Ipla7 isolated from *S. epidermidis* were identified by PHASTER (Figure 3.15 & Table 3.6) (Deghorain et al., 2012, Kwan et al., 2005, Gutiérrez et al., 2012, Diene et al., 2017, Arndt et al., 2016). Further, more extensive characterisation of the phage found in these *S. capitis* genomes may lead to the identification of novel CoNS phage as the methods used here are restricted to previously characterised phage. In addition to this, the characterisation of phage from a larger set of commensal and clinical *S. capitis* isolates may aid in implicating phage in the transfer of important virulence factors and more about the interplay between genetic alterations that may occur

between *S. aureus* and CoNS that share the same niche and environmental conditions (Deghorain et al., 2012).

The extensive recombination among the genomes of this study suggests *S. capitis* may have arisen following extensive recombination events, similar to other CoNS (Becker et al., 2014). Additionally, these same recombination events may describe the differences between *S. capitis* subspecies. More specifically HGT could be responsible for the gene set encoding the larger virulence repertoire of *S. capitis* subsp. *ureolyticus* including the presence of subspecies specific genes that encode bacteriocins, PSMs and AMR genes. However, this study found no clear plasmids associated with *S. capitis* when analysed with the PlasmidFinder database (Carattoli et al., 2014), although several samples showed similarity to the plasmid pAYP1020 present in strain AYP1020 (Cameron et al., 2015). To further understand the plasmid content of *S. capitis* more robust methods need to be carried out given that certain databases are limited for their content. Additional long read sequencing may provide more detail as to the MGE content of *S. capitis* isolates included in this study, as some are difficult to resolve with short read sequencing and are often misaligned. Preliminary analysis of subspecies representative PacBio sequence data indicates the presence of plasmid MGEs, where *S. capitis* ssp. *ureolyticus* plasmid genomes contain a higher number of virulence-related genes such as *blaZ*, *blaR* and *arsAB* encoding resistance to β-latams and arsenate. Further plasmid annotation and analysis is required to confirm this trend throughout the dataset.

This study used comparative approaches to study the genomes of *S. capitis* isolated from healthy and dandruff scalp skin. The phylogenetic reconstruction

of scalp isolated *S. capitis* plus multiple contextual reference genomes revealed the distinct clustering of the two subspecies of *S. capitis*. Investigation of the described subspecies classifications highlighted that further analysis is required for robust markers of subspecies classification. Genotypic classification of *S. capitis* ssp. *capitis* and ssp. *ureolyticus* does reveal a difference in potential virulence genes and hints at traits that discriminate these closely related subspecies. No determinants were clearly associated with *S. capitis* strains isolated from dandruff scalps, with genotypes of isolates present throughout the whole species. This study confirms that *S. capitis* contains the genomic armour necessary to colonise, survive and compete on human skin described in the current literature and adds some important descriptions and genes for traits that might progress our understanding of *S. capitis*. Additional isolates from both healthy and clinical settings and geographic locations with accompanying phenotypic data will allow more genotypic and phenotypic resolution.

CHAPTER 4

The role of a polysaccharide deacetylase (Pda) in staphylococcal virulence and survival

4.1 Introduction

Staphylococcal species are among the most ubiquitous bacteria found to colonise the human body. Commensal staphylococci can both protect and cause disease of the host. Dysbiosis of resident skin bacteria, and in particular staphylococcal species, is associated with undesirable skin conditions, therefore bacterial survival and competition determinants are important with respect to staphylococcal survival on the skin.

Bacteria of the *Staphylococcus* genus are versatile commensals, and certain species are opportunist pathogens implicated in a broad spectrum of diseases: skin and ocular infections, osteomyelitis and endocarditis (Szweda et al., 2012). Pathogenicity results from the production of an array of virulence factors that contribute to successful host colonisation, immune evasion, acquisition of nutrients within the host and resistance to antimicrobials (Karinou et al., 2019). Most staphylococcal virulence factors are secreted extracellular proteins, and many are closely associated with the bacterial cell envelope, thereby making it an important target for antimicrobial compounds (Karinou et al., 2019, Monteiro et al., 2019).

The bacterial cell envelope is a complex, multi-layered structure that is essential for survival and pathogenesis, functioning to communicate with and protect from the external environment. *S. aureus* has a typical Gram-positive cell envelope, which consists of a cytoplasmic membrane surrounded by a thick peptidoglycan

layer (Rajagopal, 2015). Peptidoglycan is a dynamic macromolecule structure primarily composed of glycan chains made of alternating *N*-acetylglucosamine and *N*-acetylmuramic acid residues connected by peptide bridges (Vollmer et al., 2008). The peptidoglycan sacculus creates a mesh-like structure enveloping the cell, providing rigidity and flexibility to protect from the external environment, whilst withstanding the high internal osmotic pressure (Monteiro et al., 2019, Vollmer et al., 2008) as well as allowing growth and division through constant polymerisation and hydrolysis (Karinou et al., 2019, Vollmer et al., 2008).

Due to the importance of peptidoglycan in bacterial cell survival, it is naturally the target of many antimicrobial compounds. The decoration and structure of the peptidoglycan structure allows bacteria to protectively respond to these stresses. Crucial to cell survival are penicillin-binding proteins (PBPs) which catalyse the two reactions at the final stages of peptidoglycan synthesis: transglycosylation and transpeptidation that are required to synthesise the glycan strand and crosslink them via peptides (Reed et al., 2015).

S. aureus encodes four different PBPs, of these the PBP1 transpeptidase and PBP2 transpeptidase and glycosyltransferase activities are the minimal requirements for bacterial cell survival (Reed et al., 2015). The two non-essential glycosyltransferases that are involved in peptidoglycan biosynthesis are SgtA and StgB (Reed et al., 2011). The methicillin-resistant *S. aureus* specific PBP2 alternative protein, PBP2A, is a transpeptidase that provides β-lactam resistance (Pinho et al., 2001).

The staphylococcal peptidoglycan layer also incorporates carbohydrate-based polymers into its structure, including most prominently teichoic acids. The two

forms present are further characterised into: lipoteichoic acid (LTA), which is a macroamphiphile with its glycolipid anchored in the membrane and its poly(glycerophosphate) chain extending into the cell wall; and wall teichoic acid (WTA), which is covalently liked to the peptidoglycan (Neuhaus and Baddiley, 2003). Teichoic acids form an important part of the bacterial cell wall, contributing to the physical and chemical properties and to the binding of cations (Neuhaus and Baddiley, 2003, Karinou et al., 2019).

The decoration of the bacterial cell surface and peptidoglycan structure allows bacteria to protectively respond to a range of environmental conditions, including antimicrobial compounds. For a bacterium to successfully colonise and infect its host, it must be capable of evading the innate immune system. Defensins and antimicrobial peptides produced largely by neutrophils of the innate immune system but also by skin epithelial cells protect the human mucosal epithelia and skin against microbial infections. Staphylococci are one of the few bacterial genera that colonise mammals that are lysozyme resistant, which promotes persistence and colonisation of the skin and mucosal areas (Bera et al., 2007, Peschel et al., 2001).

To evade several types of antimicrobial molecule, *S. aureus* possesses structural and biochemical strategies that contribute to its virulence and ability to colonise its host effectively. This includes the modification of WTA with D-alanine by the products of the *dlt* operon and controlled by the phosphorylation of the GraRS two-component system (TCS), which responds to the presence of cationic antimicrobial peptides. The response that provides protection against major antimicrobial defence molecules such as defensins and cathelicidins is to increase

in cell surface charge (Kraus et al., 2008, Neuhaus and Baddiley, 2003, Yang et al., 2012). Two mechanisms are used to alter charge: lysinylation of one of three major phospholipids of phosphatidylglycerol by MprF reduces the net negative charge of the bacterial cell wall (Bayer et al., 2015, Peschel et al., 2001, Oku et al., 2004); Dlt proteins catalyse the alanylation of teichoic acid to achieve the same effect (Peschel et al., 1999, Koprivnjak et al., 2006) and both *dlt* and *mprF* genes are controlled by GraRS and several other TCS (Yang et al., 2012, Kraus et al., 2008).

S. aureus can also modify acetylation of peptidoglycan residues that have important roles in bacterial cell wall function. Enzyme *O*-acetyltransferase (OatA) catalyses *O*-acetyl group addition at the C6-position of *N*-acetylmuramic acid residues, while TagO competitively catalyses the phosphoester-linkage of wall teichoic acid (WTA) to the same carbon of *N*-acetylmuramic, both of which inhibit the action of lysozyme (Bera et al., 2005, Bera et al., 2007) (Figure 4.1).

Human lysozyme (*N*-acetylmuramic hydrolase) is an ubiquitous secretory enzyme found in many body fluids, such as tears, sweat and saliva as part of the innate immune response (Hukić et al., 2017). The protein is hydrophobic with cationic properties and functions to enzymatically hydrolyse the β-1,4 glycosidic bonds between the alternating *N*-acetylglucosamine and *N*-acetylmuramic acid residues of peptidoglycan, resulting in bacterial autolysis (Humann and Lenz, 2009) (Figure 4.1).

Both *O*-acetylation and *O*-linked teichoic acid modifications of *N*-acetylmuramic acid enables *S. aureus* to counteract the bactericidal mechanisms of lysozyme. Among other interactions, lysozyme binds to the C-6 hydroxyl groups of the *N*-

acetylmuramic acid residues of peptidoglycan, forming a binding cleft. The presence of *O*-acetyl groups on this same carbon weakens binding affinity due to steric hindrance (Brott and Clarke, 2019). Similarly, *O*-acetylation of the phosphoester-linked WTA on the same C-6 position of *N*-acetylmuramic acid also acts to physically block the cleavage site between *N*-acetylglucosamine and *N*-acetylmuramic acid due to steric hindrance (Bera et al., 2007). Peptidoglycan *O*-acetylation has therefore been recognised as contributory virulence factor for infection of staphylococci, and other pathogens such as *Neisseria gonorrhoeae* (Moynihan and Clarke, 2010, Bera et al., 2007, Brott and Clarke, 2019) (Figure 4.1).

Since the acetylation of residues within the peptidoglycan layer plays an important role in structural and biochemical interactions of the bacterial cell wall, its role in controlling the physiology of the cells is recognised. *N*-acetylglucosamine deacetylase (Pgd) is an enzyme expressed by the Gram-positive pathogenic bacteria *Listeria monocytogenes* (Boneca et al., 2007), *Enterococcus faecalis* (Benachour et al., 2012*)* and *Streptococcus pneumoniae* (Vollmer and Tomasz, 2000, Vollmer and Tomasz, 2002). Pgd deacetylates the amine group attached to C3 of the glucosamine residue of peptidoglycan, providing lysozyme resistance, with mutants showing increased virulence.

Previous study of Pda in the Horsburgh group suggested that this protein might catalyse addition of acetyl groups to the glucosamine residues of the peptidoglycan backbone, similar to Pgd. The function of Pda was also linked with OatA and TagO, which actively remodel the Gram-positive peptidoglycan structure through the addition of side chains (Alorabi, 2016, Bera et al., 2005,

Bera et al., 2007). To gain further insights into its role experiments were designed to provide determine whether the phenotypic differences found in the *pda* mutant were consistent with a lack of deacetylation, and to determine genetic regulation of *pda.* Analysis was planned to identify how Pda might contribute to *S. aureus* evasion of host defences and potential colonisation advantages when competing with resident commensals of the skin microbiota.

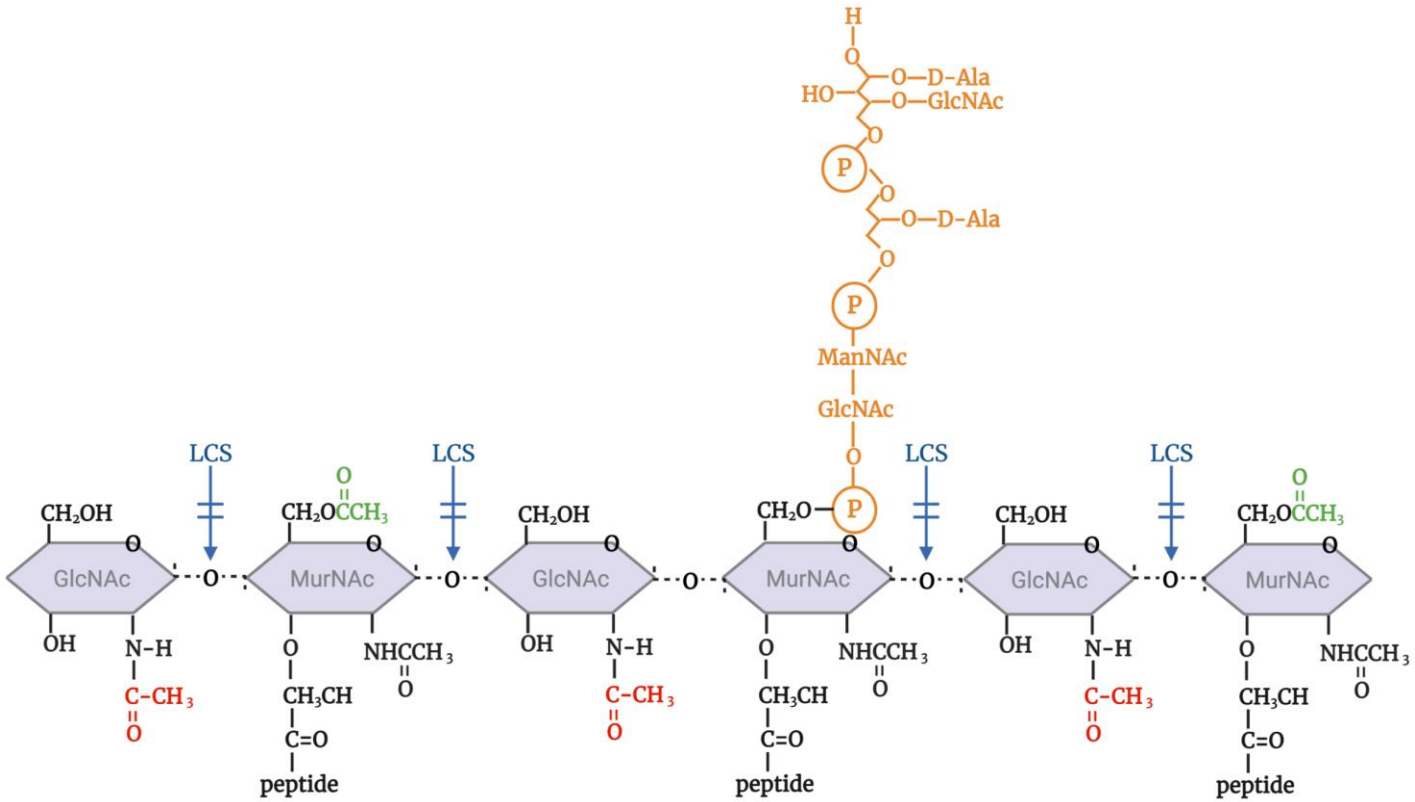


Figure 4.1 *S. aureus* peptidoglycan structure and modifications. Proposed lysozyme cleavage site (LCS, marked in blue) between muramic acid and glucosamine of *S. aureus* peptidoglycan. The muramic acid residues (MurNAc) within *S. aureus* peptidoglycan are modified by the addition of both wall teichoic acid (WTA), marked in orange and acetyl groups ($COCH_3$) to its C6 hydroxyl group (*O*-acetylation). *O*-linked modifications are described by Bera, et al (2005) implicates both OatA and TagO and are marked in green and orange. In contrast to this, Alorabi (2016) proposed that Pda causes *N*-acetylation through the addition of acetyl groups (COCH3) to the amine group attached to C3 of the glucosamine residues (GlcNAc) (marked in red). The decoration of staphylococcal peptidoglycan affects its interaction of lysozyme. Figure adapted from Bera, *et al* (2005) and created in BioRender.com (2020).

4.1.1 Aims of chapter

This chapter aims to study the role of *S. aureus* Pda as a competition determinant with respect to resistance of *S. aureus* and its interactions with other species. This work builds from study of a *S. aureus* polysaccharide deacetylase in the Horsburgh group (Alorabi, 2016). Specifically, two aspects were to be examined: determining the regulation of the *pda* gene, encoding a polysaccharide deacetylase; and how regulation of cell wall charge mediated by Pda, in relation to other described mechanisms, alters the interactions of *S. aureus* with other niche inhabitants.

4.2 Methods and materials

4.2.1 Bacterial strains and culture conditions

For the study of Pda, Ø11 transduction was used to mobilise an allelic replacement *pda* mutation to different *S. aureus* host strains, including strains Newman, SF8300 and SA113. To further understand the possible link between Pda, OatA and TagO, double knockout mutants combining *pda* with either *oatA* or *tagO* mutations were created to compare observed phenotypes with single mutants. The genotypes of mutant strains after transduction were confirmed using PCR and antibiotic resistance testing. The full details of strains are listed in Table 4.1. The *pda* mutant strain was created in a previous study (Alorabi, 2016).

Bacterial strains used in this study were cultured on BHI agar plates or in BHI broth (Lab M) at 37 ˚C, with shaking, overnight for 16-18 h, unless otherwise stated, and supplemented as appropriate with tetracycline (5 μg mL^{-1}) (Sigma Aldrich), chloramphenicol (25 μg mL^{-1}) (Sigma Aldrich) or kanamycin (5 μg mL^{-1}) (Sigma Aldrich).

Table 4.1 Bacterial strains used in this study. Antibiotics (Sigma) are used as follows: kanamycin (Km), 5 μg ml^{-1}; chloramphenicol (Cm), 25 μg ml^{-1} and tetracycline (Tet) 5 μg ml^{-1} with *S. aureus* strains.

ID	Species/Strain Name	Antibiotic resistance	Description
1143	*S. aureus* Newman WT	N/A	Wild-type lab strain (Duthie and Lorenz, 1952).
3005	*S. aureus* Newman *pda*	Tet	Strain with *pda* mutation (Alorabi, 2016).
3021	*S. aureus* Newman *pda/Co*	Cm	*pda* mutant complemented with a plasmid containing the *pda* gene (Alorabi, 2016).
1187	*S. aureus* Newman *oatA*	Km	Strain with *oatA* mutation (Alorabi, 2016).
1190	*S. aureus* Newman *oatA pda*	Tet	Strain with both *oatA* and *pda* mutations (Alorabi, 2016).
1142	*S. aureus* SA113 *WT*	N/A	Wild-type, *S. aureus* strain SA113 (Bera et al., 2005)
1184	*S. aureus* SA113 *pda*	Tet	Strain with *pda* mutation (Alorabi, 2016).
1144	*S. aureus* SA113 *oatA*	Km	Strain with *oatA* mutation (Bera et al., 2005).
1186	*S. aureus* SA113 *oatA pda*	Tet	Strain with both *oatA* and *pda* mutations (Bera et al., 2005).
1145	*S. aureus* SA113 *oatA tagO*	Km	Strain with both *oatA* and *tagO* mutations (Bera et al., 2005).
1138	*S. aureus* SF8300 *WT*	N/A	Clinical isolate (USA300) (Montgomery et al., 2009)
3020	*S. aureus* SF8300 *pda*	Tet	Strain with *pda* mutation (Alorabi, 2016).
3030	*S. aureus* SA113 *pda oatA bshB2*	Tet Km	Strain with *pda, oatA* and *bshB2* mutations (This study)

3031	*S. aureus* SA113 *pda oatA sdgA*	Tet Km Erm	Strain with *pda, oatA* and *sdgA* mutations (This study)
3032	*S. aureus* SA113 *pda oatA sdgB*	Tet Km Erm	Strain with *pda, oatA* and *sdgB* mutations (This study)
3033	*S. aureus* SA113 *pda oatA sdrC*	Tet Km Erm	Strain with *pda, oatA* and *sdrC* mutations (This study)
3034	*S. aureus* SA113 *pda oatA sdrD*	Tet Km Erm	Strain with *pda, oatA* and *sdrD* mutations (This study)
3035	*S. aureus* SA113 *pda oatA sdrE*	Tet Km Erm	Strain with *pda, oatA* and *sdrE* mutations (This study)
3036	*S. aureus* SA113 *pda oatA clfA*	Tet Km Erm	Strain with *pda, oatA* and *clfA* mutations (This study)
3037	*S. aureus* SA113 *pda oatA clfB*	Tet Km Erm	Strain with *pda, oatA* and *clfB* mutations (This study)
3038	*S. aureus* SA113 *pda oatA cap1A*	Tet Km Erm	Strain with *pda, oatA* and *cap1A* mutations (This study)
3039	*S. aureus* SA113 *pda oatA icaA*	Tet Km Erm	Strain with *pda, oatA* and *icaA* mutations (This study)
J014	*S. epidermidis* 14	N/A	Forearm isolate (Kelly, 2013)
1223	*S. epidermidis* 155	N/A	Nasal isolate (Libberton et al., 2014)
1082	*S. epidermidis* Tü3298	N/A	Laboratory strain, known for its production of epidermin and absence of the *icaABCD* operon (Allgaier et al., 1985)

108	*S. capitis* 108/ DSM 20327	N/A	Human scalp isolate type strain DSM 20327 (Kloos and Schleifer, 1975)
3041	*S. capitis* DSM 6717	N/A	Human scalp isolate type strain DSM 6717 (Kloos and Schleifer, 1975)

4.2.2 Confirmation of bacterial strain genotypes by PCR

4.2.2.1 DNA extraction for PCR

DNA was extracted from strains in Table 4.1 to confirm the presence of mutations. Bacteria were cultured in 10 mL BHI broth (Lab M), for 18 h, shaking at 37 ˚C. Subsequently, 1 mL of each overnight culture was then centrifuged for 1 minute at 5000 rpm and resuspended in 180 µl lysis buffer (20 mM Tris- HCl pH8, 2 mM EDTA, 1.2% Triton X-100); the cells were extracted to obtain high-quality genomic DNA of each clone using the QIAGEN DNeasy Blood & Tissue Kit.

4.2.2.2 Strain confirmation by PCR

Purified DNA was used in PCR with specific primer sets to confirm the presence of genes and corresponding resistance cassettes compared with isogenic wild-type parent strains. Primers were used in 25 µL reactions (Table 4.2). The negative control for this experiment contained 10 µL sterile distilled water without bacterial DNA and all PCR amplicon sizes were compared with those amplified from wild-type strain DNA to determine fragment sizes. Genes were amplified under the following conditions: 95 °C initial denaturation for 5 min; 30 cycles of 95 °C for 30 s, 60 °C or 55 °C for 30 s and 72 °C for 2.5 min followed by 72 °C final extension for 10 min. The PCR products were then visualised using gel electrophoresis: 10 µL of each sample were separated on 1% (w/v) agarose gels at 90 mV for 40 min with a 1 kb ladder as size marker (Bioline). The gels were visualised under a UV transilluminator at a wavelength of 302 nm and images were recorded using GeneSnap software (Syngene).

Table 4.2 Oligonucleotide primers employed in the current study. Primer sequences are shown that were used for the amplification of genes of interest to verify presence of appropriate mutations and antibiotic cassettes in bacterial strains.

Primer Name	Oligonucleotide sequence (5' to 3')	Gene target
OatA F	TAGAGGATAGGCTGCAT	*oatA*
OatA R	CTGCACCGTTACTATGGGA	*oatA*
TagO F	GTCTGAATCGACTCCTT	*tagO*
TagO R	ACACGTTTATGGCAGTGC	*tagO*
Pda F	TTAGCGCAACAAGCAAGAGGA	*pda*
Pda (tet cassette Not I) R	ATAACTGCGGCCGCGGCGGATTTTATGACCGATGAAG	*pda*

4.2.3 Investigating Pda activity

4.2.3.1 Peptidoglycan structural analysis by mass spectrometry

Enzyme digestion combined with high performance liquid chromatography (HPLC) separation of muropeptides and mass spectrometry (MS) of peptidoglycan from both wild type and *pda* mutant of *S. aureus* was carried out to determine possible structural effects due to the absence of deacetylase. These experiments were performed by David Sychantha of the Professor Anthony J. Clarke research group, at the University of Guelph, Canada. Peptidoglycan of both wild-type *S. aureus* SF8300 and a *pda* mutant was extracted from strains containing mutations verified in this study.

4.2.3.1.1 Peptidoglycan isolation

Wild-type *S. aureus* SF8300 and the *pda* mutant were cultured in 0.5 L BHI broth (Lab M) at 37 ˚C with aeration (200 rpm), in triplicate. The cells were harvested at mid-exponential phase (OD_{600} = 0.8) with centrifugation (5000 x g, 20 min, 4 ˚C) and the cell pellets were frozen and stored at -20 ˚C. For the extraction of peptidoglycan (PG) sacculi, the cell pellets were resuspended in 125 mL cold phosphate buffered saline (PBS; pH 7). The cells were then added dropwise to an equal volume of boiling sodium dodecyl sulphate (SDS; 8% w/v) in PBS. After 1 hour, the sacculi were cooled to room temperature and collected by centrifugation (45 000 x g, 30 min, 25 ˚C). To remove SDS, the sacculi were washed by centrifugation (45 000 x g, 30 min, 25 ˚C) using 5 volumes (250 mL each) of warm PBS. The sacculi were next disrupted by sonication and enzymatically treated with DNAse and α-amylase (100 µg/mL each) at 37 ˚C for

2 h. This was followed by pronase treatment (100 µg mL $^{-1}$) at 37 ˚C for overnight. The PG was then boiled a second time in SDS (2% w/v) to inactivate pronase for 1 h. To remove covalently bound teichoic acids, the sacculi were treated with 10% tricholoracetic acid (TCA) overnight at room temperature. The pure PG was washed with water and lyophilized.

4.2.3.1.2 Peptidoglycan analysis by mass spectrometry

Muropeptides were prepared by digestion of isolated PG with mutanolysin and lysostaphin (100 µg mL^{-1} each) at 37 ˚C overnight. The enzymes were inactivated with heat (90 ˚C, 30 min) and the sample was centrifuged (20 000 x g, 20 min, RT) to remove insoluble material. The clarified sample was reduced with sodium borohydride (10 mg mL^{-1}) in sodium borate buffer (250 mM, pH 8.0) for 20 mins. After reduction, excess sodium borohydride was destroyed with 20% o-phosphoric acid and the sample was adjusted to pH 4. Analysis of the muropeptides was carried out using liquid chromatography-mass spectrometry (LC-MS) using a Dionex UHPLC and Bruker Amazon SL ion trap systems. The system was fitted with a Thermo hypersil BDS column (5 µm particle size, 250 x 2.1 mm) and a solvent system of (A) water and 0.1% formic acid (v/v) and (B) acetonitrile and 0.1 % (v/v) formic acid was used. Muropeptides were separated using a linear gradient of 0 – 20% B over 120 min. The flow rate was maintained at 0.5 mL/min. Ions were detected in positive mode and the mass-to-charge ratio was scanned across the m/z range of 300-2000 m/z.

4.2.3.2 Creation of cell wall mutant strains by transduction

To identify the surface component Pda might deacetylate, potential targets were considered with a particular focus on described glycosylated surface proteins (Weidenmaier and Lee, 2017). Surface proteins of interest included ClfA, ClfB, SdrCDE and IcaA, as well as the capsular protein Cap1A. Mutations in the cognate genes from a *S. aureus* USA300 mariner tranposon library (Bose et al., 2013) were mobilised by transduction into *S. aureus pda oatA* double mutants. The rationale for the approach was to observe the effect of Pda on its potential target surface protein. Because the *pda* mutant is more resistant to lysozyme compared with *oatA* and *pda oatA* mutants, it was hypothesised that the addition of a Pda target surface protein mutation to the *pda oatA* double mutant would prevent the action of the Pda target surface protein from increasing lysozyme resistance.

Triple mutants of *S. aureus* were created by transduction. Cell lysates of *clfA, clfB, sdrCDE* and *icaA* mutants were made by growing them in BHI for 18 h at 37°C with shaking (250 rpm), 1 mL of which was used to inoculate 25 mL of BHI broth and grown for 2 h, at 37°C with shaking (250 rpm) until culture is OD_{600} 0.5-0.8. 5 mL of cells were then harvested by centrifugation at 4000 rpm for 5 min at 4°C and resuspended in 5 mL BHI broth, 5 mL phage buffer (1 mM MgSO4, 4 mM CaCl2, 50 mM Tris-HCl pH 7.8, NaCl 5.9 g L^{-1} & gelatine 1 g L^{-1}) and 30 µL phage stock (Ø11). This was then incubated at 30 °C for 2 h with gentle mixing (150 rpm) until lysis. Lysates were filter sterilised (0.4 µm).

Cell lysates containing Ø11 were used to mobilise the mutations by generalised transduction with selection in host *S. aureus pda oatA*. The double mutant was

grown in 20 mL BHI broth for 18 h at 37°C with shaking (250 rpm). The culture was split equally into three and cells were harvested by centrifugation at 4000 rpm, for 5 min at 4°C and resuspended in 500 µL BHI broth and 10 mL $CaCl_2$. 500 µL phage lysate, 100 µL phage lysate or 500 µL BHI was added and incubated for 25 min at 30 °C stationary and then 15 min, at 37 °C with agitation. Cells were then incubated with 1 mL 0.02 M ice cold sodium citrate for 5 min, before being washed with a further 1 mL 0.02 M sodium citrate, via centrifugation (4000 rpm, 5 min, 4°C), being careful to remove all supernatant containing phages. Cells were resuspended in 1 mL sodium citrate (0.02 M) and incubated on ice for 2 h, before adding 1 mL BHI & sodium citrate (0.05% w/v) and incubating at 37°C with mixing for 2 h. Recovered cells were then diluted to 10^{-2} and 100 µL of each were plated onto BHI agar containing relevant antibiotics (tet, km & erm) and sodium citrate (0.05% w/v). Selected mutants resistant to erythromycin were used to screen for lysozyme phenotype testing.

4.2.3.2.1 Lysozyme phenotype testing of cell wall mutants

Lysozyme agar diffusion assays were carried out to examine differential survival. Bacterial cultures of strains were grown overnight in 10 mL BHI broth and at 37 °C with shaking. Overnight bacterial cultures (100 µL) were diluted in 4 mL BHI top agar that was kept at 55 °C before addition of bacteria. This bacterial dilution was used based upon previous experiments diluting 10^{0}- 10^{-6} dilutions in top agar and assessing the best bacterial lawns produced. The bacterial dilution was immediately poured over a BHI agar plate and left to set before placing a 3 mm Whatman filter paper disk soaked in lysozyme (50, 100, 150 or 200 mg mL^{-1}) in

the centre of the plate. Plates were dried briefly at room temperature and incubated overnight at 37 °C.

4.2.4 Identifying regulators of Pda activity

4.2.4.1 Experimental evolution and selection

Experimental evolution was carried out using a microtitre plate method (150 µL culture volume). A bacterial culture grown in BHI broth overnight and diluted to an OD_{600}= 0.08-0.1 was used to inoculate wells containing BHI broth (75 µL) and standard two-fold increments of lysozyme. The initial stock solution concentration of lysozyme used for each strain was 100 mg mL^{-1} (Figure 4.2). A range of concentrations were set up across the plate as shown in Table 4.4. All plates were incubated static at 37°C. The first plates were inoculated with the different bacterial strains for the first passage and bacteria from the well with the highest concentration of lysozyme that supported growth comparable with the positive control after 24 h was passaged forward to the next plate. All wells of subsequent passages were inoculated with a 1:1000 dilution of the bacteria passaged forward. This passage was repeated 11 times, in triplicate. Strains were stocked in 20% glycerol at -80°C after each passage.

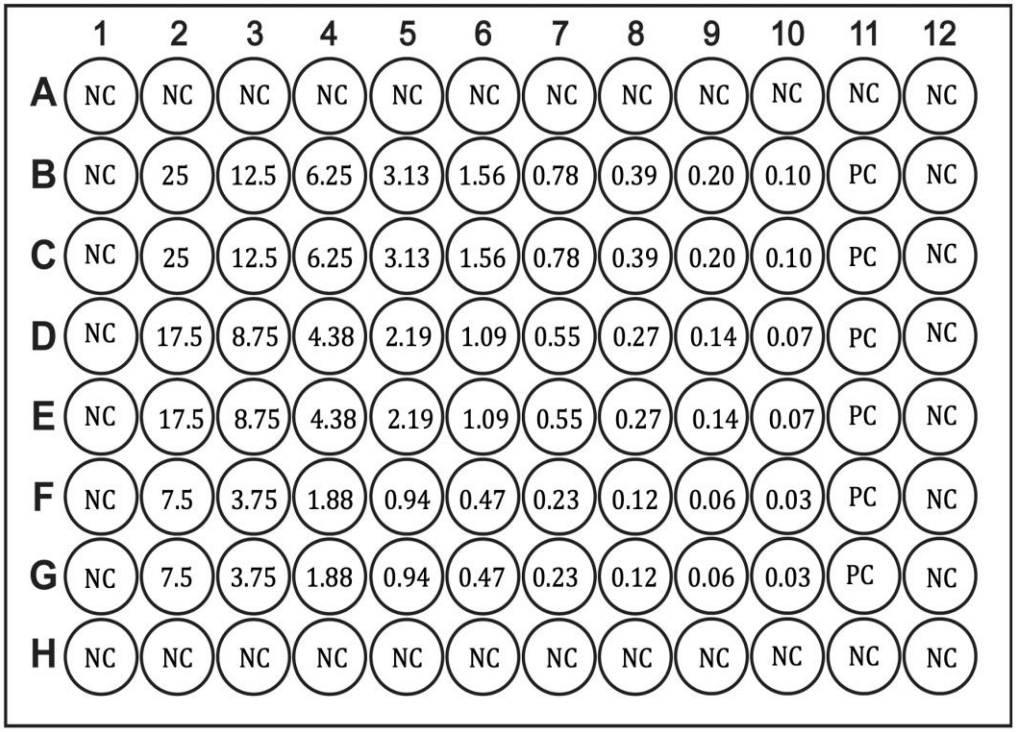


Figure 4.2 Concentrations of lysozyme used in experimental evolution.
Schematic of 96-well microtitre plate with concentrations of lysozyme used. All concentrations are given in mg mL $^{-1}$. PC refers to positive control where wells only contained bacteria and no lysozyme. NC refers to a negative control where wells only contained BHI broth.

4.2.4.2 Lysozyme MIC determination assay

Minimal inhibitory concentration determinations were performed in 96-well microtitre plates (150 µL culture volume). A bacterial culture of *S. aureus* mutant clones from T0 time point, all chosen time points and their respective controls were grown in BHI overnight, with shaking, at 37 °C was diluted to an OD_{600} = 0.08-0.1. This dilution was then used to inoculate wells containing BHI broth (75 µL) and standard twofold increments of lysozyme (100 mg mL^{-1}) to produce a concentration series (Sigma Aldrich). Plates were incubated statically for 18-20 h at 37°C. All experiments were performed in triplicate. The MIC was defined as the lowest concentration at which growth was prevented. All experiments were done in triplicate.

4.2.4.3 Lysozyme agar diffusion assay

Lysozyme agar diffusion assays were carried out to experimentally select for resistance. This experiment was carried out as detailed in section 4.2.3.2.2. If colonies were found within the zone of clearing for any of the different mutants, these were then picked and used to inoculate 10 mL BHI broth. Overnight cultures were then stocked with 20% (v/v) glycerol and used to repeat the agar diffusion assay. All experiments were done in triplicate.

4.2.4.4 DNA extraction for sequencing

DNA was extracted from clones selected from each experimental and control repeat. Clones for each selected and T0 time point plus controls from each experiment were selected from plates and cultured without lysozyme in 10 mL of BHI overnight, shaking at 37 °C. Subsequently, 1 mL of each overnight culture was then centrifuged for 1 min at 5000 rpm and resuspended in 180 µl lysis

buffer (20 mM Tris- HCl pH8, 2 mM EDTA, 1.2% Triton X-100); the cells were extracted to obtain the high-quality genomic DNA of each clone using the QIAGEN DNeasy Blood & Tissue Kit. The purified DNA concentration was measured using Thermo Scientific Nanodrop, Qubit followed by integrity visualisation through gel electrophoresis (1% (w/v) agarose gels at 90 mV for 40 min with a 1 kb ladder). The extracted DNA from clones selected from each experiment and control points was required to be in a final volume of 55 µl and a final concentration of between 1000 – 2000 ng per sample to prepare Illumina TRUSEQ PCR-free DNA libraries with an insert size of 250 bp. All TRUSEQ libraries were sequenced on the Illumina MiSeq platform by the CGR, University of Liverpool.

4.2.4.5 Analysis of lysozyme selection SNPs

Sequence reads were processed by the Centre for Genomic Research (CGR), University of Liverpool. The CGR employs a standard read-filtering pipeline on all sequenced datasets which comprised of: a) the removal of Illumina adaptor sequences using Cutadapt v1.2.1 (Martin, 2011) and b) the trimming of low-quality bases, which removed read segments which do not have a minimum phred quality score of 20, using Sickle v.1.2 (Joshi and Fass, 2011). Following this initial processing by the CGR, read quality was assessed using FastQC v0.11.7 (Andrews, 2010) and MultiQC v1.0 (Ewels et al., 2016). Kraken v2.0 was run to ensure reads were free from contamination using the RefSeq bacterial database (Wood et al., 2019).

If filtered reads were assessed to be good quality and uncontaminated, reads were aligned to their reference genome using the Burrows-Wheeler Alignment

Tool (BWA) v0.7.5.a mem packages (Li and Durbin, 2009). Reference genomes were either downloaded from NCBI databases or assembled and annotated using Unicycler v0.4.7 (Wick et al., 2017) and prokka v1.14.6 (Seemann, 2014). Additional processing of alignment files and creation of BCF files (binary variant call format) for SNP calling was carried out using Samtools v.0.1.18 (Li et al., 2009). Variant SNPs were then called, annotated and effects predicted using SnpEff v4.0 (Cingolani et al., 2012). A bespoke script (unique_SNPs.pl) was then used to filter the SNP data found in evolved, parent and control datasets. Non-synonymous SNPs found in over 80% of reads were deemed important to this study.

4.2.5 Pda as a skin colonisation competition determinant

This study aims to study the role of Pda, a putative surface modification enzyme and competition determinant. Deferred growth inhibition assays were used to assess the competition effect of Pda in *S. aureus* in relation to other staphylococci found on the skin, including *S. capitis* and *S. epidermidis*.

4.2.5.1 Deferred growth inhibition assay

A 25 µl (approximately 10^8 cells) volume of an overnight bacterial culture (inhibitor strains) was spotted onto the middle of an agar plate containing 15 mL of BHI agar (Lab M) prior to incubation for 18 h at 37 °C. Competitor strains were prepared via ten-fold diluted overnight culture approximately 4 x 10^6 CFU ml^{-1}, and a suspension of OD_{600} 0.3 ± 0.05. Competitor strain culture was then sprayed through vaporiser bottles, from a distance of 15 cm above the agar plate, depositing approximately 250 µl of diluted culture over the entire agar surface. Agar plates were then re-incubated for a further 18 h at 37 °C .The assessment of

growth inhibition was based on the diameter of the inhibition zone (ZOI) and a qualitative scoring system based on the degree of growth prevention, from 1 corresponding no detectable inhibition to 4 corresponding to complete inhibition (completely clear ZOI) (Moran et al., 2016). All experiments were carried out in triplicate.

4.3 Results

4.3.1 Confirmation of bacterial strain genotypes by PCR

To study Pda roles, allelic replacement *pda* mutants in different *S. aureus* host strains were created using Ø11 transduction. In addition to this, double mutants combining *pda* with either *oatA* or *tagO* mutations were created to compare observed phenotypes with single mutants to identify links between Pda, OatA and TagO. PCR amplification of *oatA, pda* and *tagO* genes in mutant strains used in this study was performed, and the amplicons were compared with those amplified from wild-type strains to observe fragment size differences.

To study the effects of Pda, an allelic replacement *pda* mutation was mobilised by transduction of *S. aureus* host strains Newman, SF8300 and SA113. Presence of the mutation was confirmed by PCR with primers Pda F and Pda (tet cassette Not I) R (Table 4.2). This primer set amplified a ~1,200 bp region spanning the start of the *pda* gene into the tetracycline cassette which was inserted into the gene (Alorabi, 2016). Figure 4.3 highlights the presence of PCR products of the correct size amplified from strains Newman *pda*, SF8300 *pda* and SA113 *pda* and an absence of PCR products from strains Newman, SF8300 and SA113 wild type, confirming the presence of the *pda* mutation in these three *S. aureus* strains.

Figure 4.3 Gel electrophoresis image of pda PCR amplification products. (L) is the 1 kb molecular size marker (Bioline); (1) negative control of SDW; (2) negative amplification control *S. aureus* wild-type Newman; (3) negative amplification control *S. aureus* wild-type SF8300; (4) negative amplification control *S. aureus* wild-type SA113; (5) start of the *pda* gene and tetracycline resistance cassette in Newman *pda*; (6) start of the *pda* gene and tetracycline resistance cassette in SF8300 *pda* ; (7) start of the *pda* gene and tetracycline resistance cassette in SA113 *pda*.

To further understand the possible link between Pda, OatA and TagO, single mutants, and double mutants combining *pda* and either *oatA* or *tagO* marked mutations, were created. To confirm the presence of these genes and their relevant antibiotic cassettes, PCR amplification of *oatA* and *tagO* was carried out with DNA from mutants and resulting products were compared with those from wild-type strains. All strains in this study (Table 4.1) created by Alorabi *et al*, were confirmed to have the appropriate mutations present due to difference in PCR product size as a result of the presence of antibiotic resistance cassettes within genes of interest (Figure 4.4)(2016).

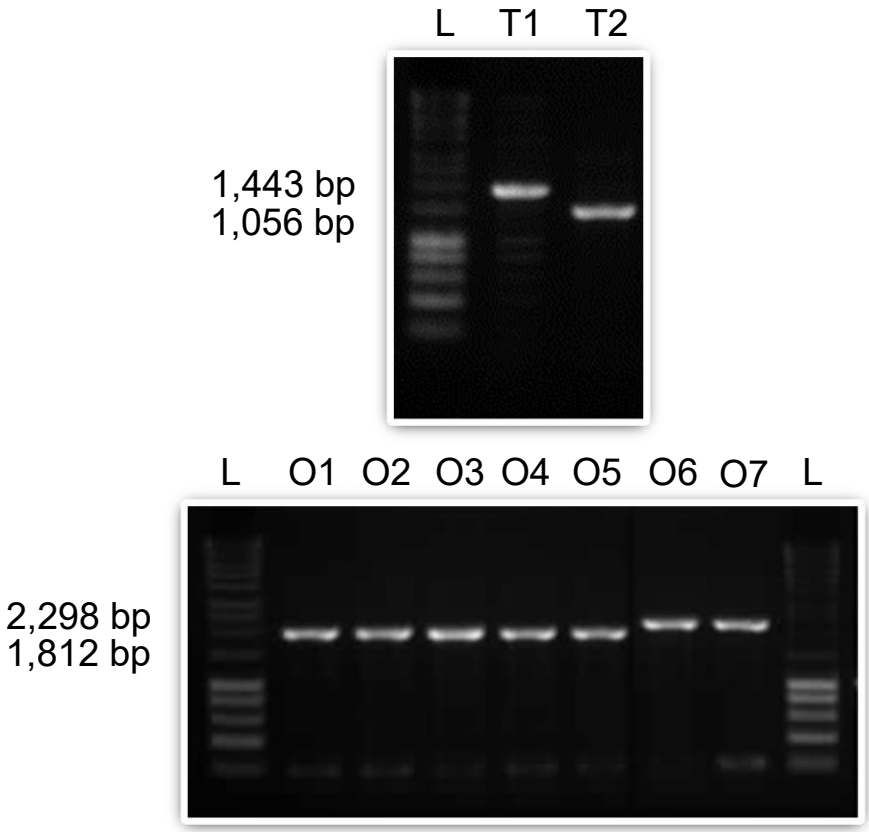


Figure 4.4 Gel electrophoresis images of products of PCR amplification for oatA and tagO gene fragments. (L) is the 1 kb molecular size marker (Bioline); (T1) negative amplification control *S. aureus* wild-type SA113; (T2) *tagO* gene containing erythromycin resistance cassette in SA113 *oatA tagO*; (O1) *oatA* gene containing kanamycin resistance cassette in *Newman oatA*; (O2) *oatA* gene containing kanamycin resistance cassette in *Newman oatA pda*; (O3) *oatA* gene containing kanamycin resistance cassette in SA113 *oatA*; (O4) *oatA* gene containing kanamycin resistance cassette in SA113 *oatA pda*; (O5) *oatA* gene containing kanamycin resistance cassette in SA113 *oatA tagO*; (O6) negative amplification control *S. aureus* wild-type Newman; (O7) negative amplification control *S. aureus* wild-type SA113.

4.3.2 Peptidoglycan structural analysis by mass spectrometry

As a means to indirectly investigate the function of Pda, the chemical structure of peptidoglycan was compared between wild type and *pda* mutant of *S. aureus* strain SF8300. The expectation being that absence of peptidoglycan deacetylase function would produce a clear signature after enzyme digestion followed by high performance liquid chromatography (HPLC) separation of muropeptides and mass spectrometry (MS) of peptidoglycan

The MS spectra of the digested peptidoglycan shows the retention time of muropeptide fragments, with each peak representing fragment sizes of *N*-acetylglucosamine and *N*-acetylmuramic acid residues with different repeat length and decoration of amino acids. Analysis of the MS data comparing spectra from SF8300 wild type with the *pda* mutant revealed that there were no identified differences in muropeptide species as judged by equivalent retention times. From this analysis, it can be expected that peptidoglycan acetylation levels are not determined by Pda deacetylation of glucosamine (Figure 4.5).

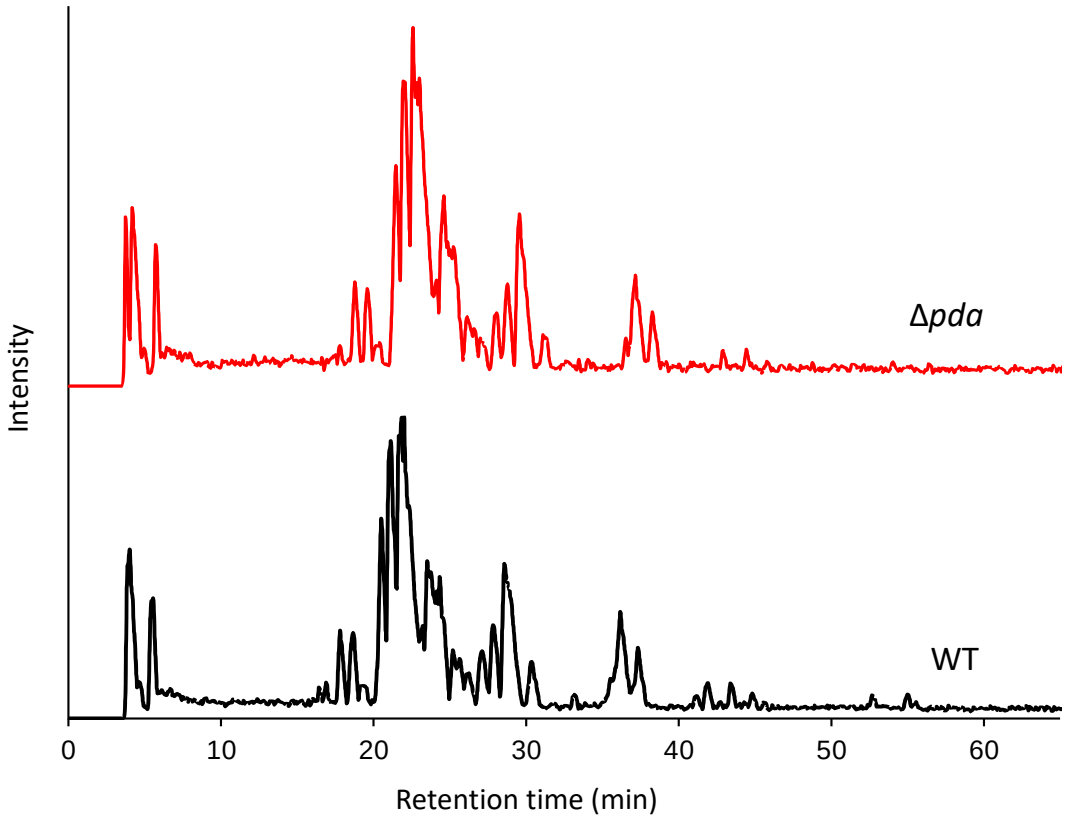


Figure 4.5 LC-MS comparison of peptidoglycan structure of *S. aureus.* LC-MS separation of digested peptidoglycan muropeptides in wild type (black) and Δ*pda* (red) extracted from *S. aureus* strain SF8300.

4.3.3 Lysozyme phenotype testing of cell wall associated mutants

With the aim of identifying potential targets of Pda activity given the absence of discernible modification of peptidoglycan, multiple secreted, glycosylated proteins of *S. aureus* were selected for study from literature descriptions. MSCRAMMs of interest included ClfA, ClfB, SdrCDE, Cap1A, BshB2 and IcaA. Mutations in these genes were mobilised into the *S. aureus* SA113 double mutant Δ*oatA pda* by transduction using Ø11. The lysozyme phenotypes of these mutants were then determined by disk diffusion assay, comparing the results of the triple cell wall mutants with the lysozyme phenotypes of SA113 *oatA*, SA113 *pda* and SA113 *oatA pda* (Figure 4.6).

Lysozyme phenotypes were tested via disk diffusion assay, using 3 mm Whatman filter paper disk soaked in lysozyme at different concentrations (50, 100, 150 & 200 mg mL $^{-1}$). At all concentrations of lysozyme, the SA113 *pda* mutant was more resistant to lysozyme compared with SA113 *oatA* and SA113 *oatA pda*. A Kruksall-Wallis test conducted to examine the differences in lysozyme resistance of each mutant at each concentration provided evidence of differences between all triple cell wall mutants, SA113 *oatA,* SA113 *pda* and SA113 *oatA pda* (p <0.05) at lysozyme concentrations 100 (H= 28.91, p <0.05) , 150 (H= 29.66, p <0.05) , and 200 mg mL $^{-1}$ (H= 26.01, p <0.05) but not at 50 mg mL $^{-1}$ (H= 19.35, p= 0.0805) (Figure 4.6).

A Dunn's post hoc test, looking at pairwise comparisons between mutants did not find significant differences between any of the triple mutants and SA113 *oatA pda* (*padj* > 0.05). Only SA113 *oatA pda SdrD* showed a consistent trend as the mutant

most sensitive to lysozyme across all concentrations, indicating this gene could potentially be affected by Pda despite a lack of statistical support (Figure 4.6).

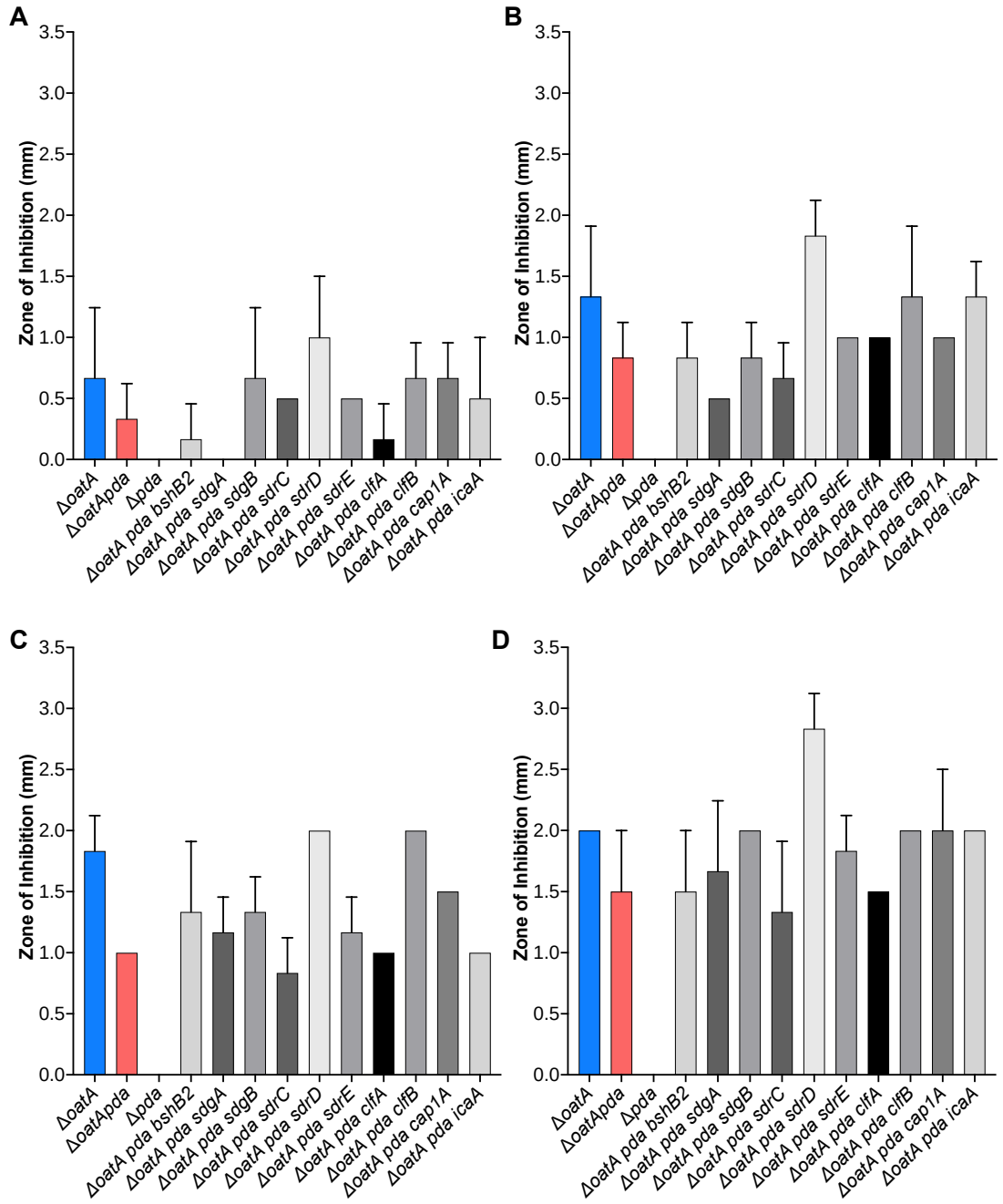


Figure 4.6 Growth inhibition of cell wall mutants from antimicrobial activity of lysozyme. The zone of inhibition (ZOI) for cell wall associated triple mutants combined with *oatA* and *pda* in a disk diffusion assay with (A) 50 mg ml^{-1}; (B) 100 mg ml^{-1}; (C) 150 mg ml^{-1} and (D) 200 mg ml^{-1} concentrations of lysozyme soaked into the disk.

4.3.4 Identifying potential regulators of Pda activity

4.3.4.1 Experimental evolution and selection

Selecting for increased lysozyme resistance was used to potentially reveal contributing partners in *pda* gene regulation. Two separate methods of experimental evolution were undertaken to select for increased lysozyme resistance, microtitre plate passage with differing concentrations of enzyme (Figure 4.7 & Figure 4.8) and a lysozyme diffusion assay (Figure 4.10). Two *S. aureus* strain backgrounds were used: Newman and SA113 to select for altered resistance, without and with gene mutations in *pda, oatA* and *tagO*.

The inactivation of *pda, oatA* and *tagO* in both *S. aureus* Newman and SA113 altered both their initial response to lysozyme and the selection outcome. With both SA113 and Newman strains, the continuous exposure to lysozyme led to an increase in the threshold concentrations that mutants could grow at (Figure 4.7 & Figure 4.8).

Quite different profiles were observed during selection with respect to maximal and final concentrations of lysozyme at which strains cultured. The mutant SA113 *oatA* evolved to be the most resistant to lysozyme, growing at a maximal concentration of 25 mg ml^{-1} by day 11 (Figure 4.7). Somewhat similarly, SA113 *pda, Newman oatA* and *Newman oatA pda* cultures showed growth with a maximal concentration of 25 mg ml^{-1} at intervals during the 11 passages. In comparison, *Newman oatA* and *Newman pda* showed the least resistance over the time course passages (Figure 4.7 & Figure 4.8).

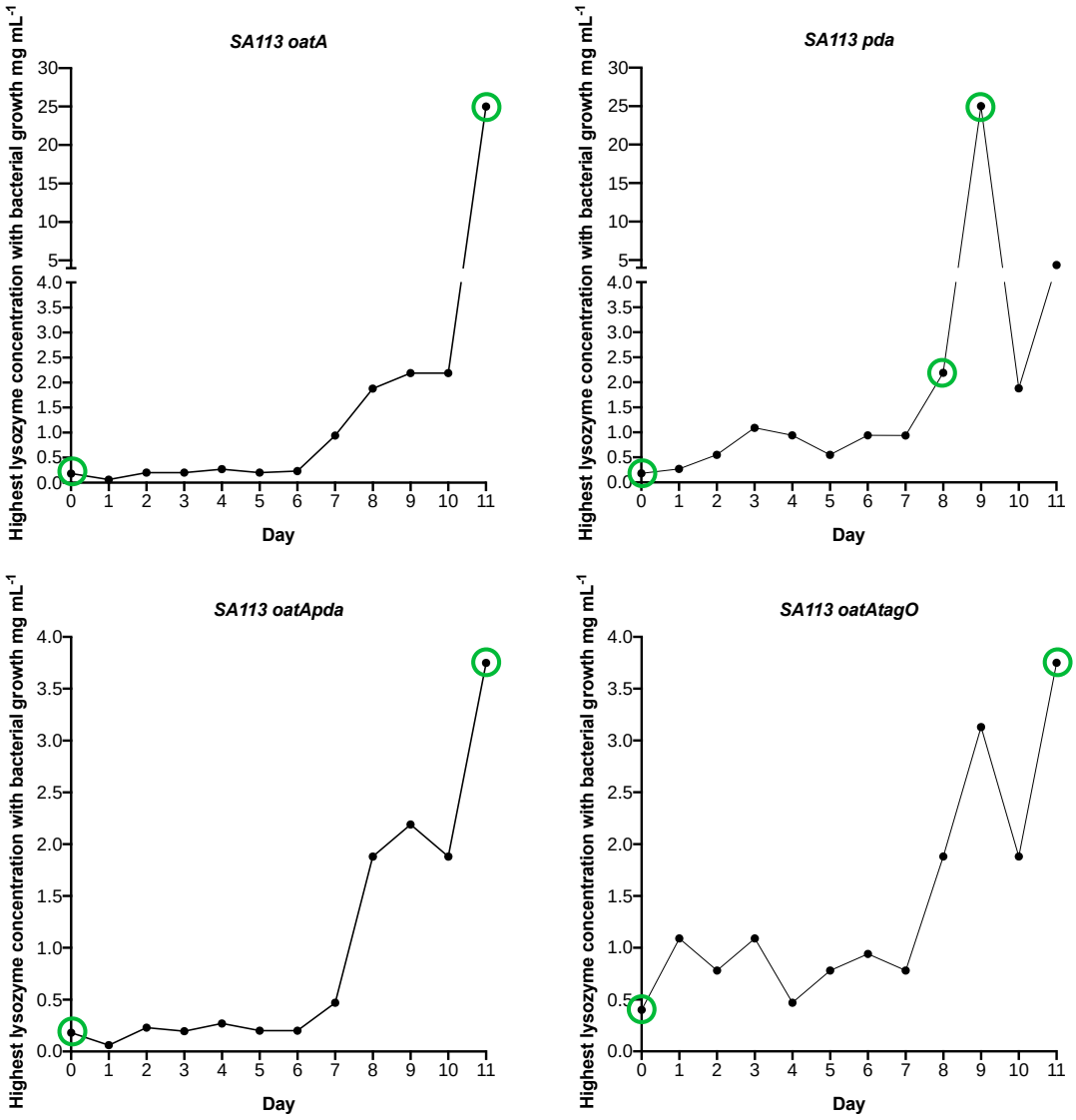
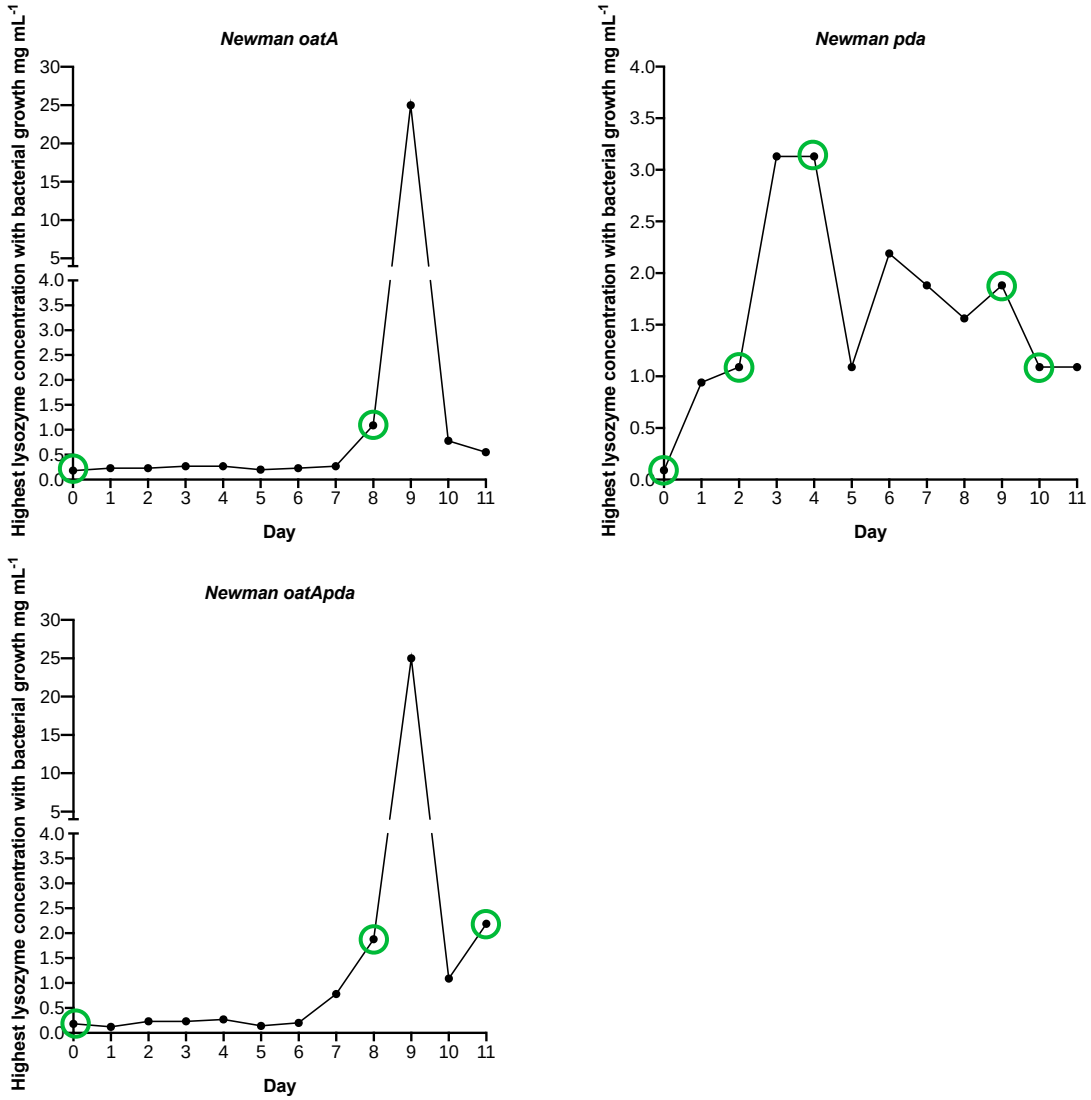


Figure 4.7 Lysozyme resistance selection of *S. aureus* SA113. Experimental evolution using the microtitre plate method for *S. aureus* SA113 mutant's *oatA*, *pda*, *oatA pda* and *oatA tagO*. The highest concentration of lysozyme at which growth occurred, comparable with the positive control after 24 h for each repeat, at each time point. Cells from these wells were passaged forward to the next plate. Time points highlighted in green are those chosen for whole genome sequencing. These data are based on three biological replicates.



<u>Figure 4.8 Lysozyme resistance selection of *S. aureus* Newman.</u> Experimental evolution was carried out using a microtitre plate method for *S. aureus* Newman mutant's *oatA*, *pda* and *oatA pda*. The highest concentration of lysozyme at which growth occurred, comparable with the positive control after 24 hours for each repeat, at each time point. Cells from these wells were passaged forward to the next plate. Time points highlighted in green are those chose for whole genome sequencing. These data are based on three biological replicates.

MIC assays were used to confirm whether fixed genetic changes had been selected during the passaging of the *pda, oatA* and *tagO* mutants of *S. aureus* Newman and SA113 (Figure 4.9). The MIC values were compared of start point and the selected time point isolate, which sustained growth at the highest lysozyme concentration throughout the evolutionary time period (Figure 4.7, Figure 4.8 & Figure 4.9). The MIC values determined for these individual clones selected for MIC assay were higher compared with the values determined for the sampled culture population. Of note, the MIC values for all evolved strains presented a higher MIC compared with day 0 based on triplicate biological replicates, with SA113 *oatA* MIC being increased over 200-fold, indicating that passage with lysozyme selected for genetic change (Figure 4.9).

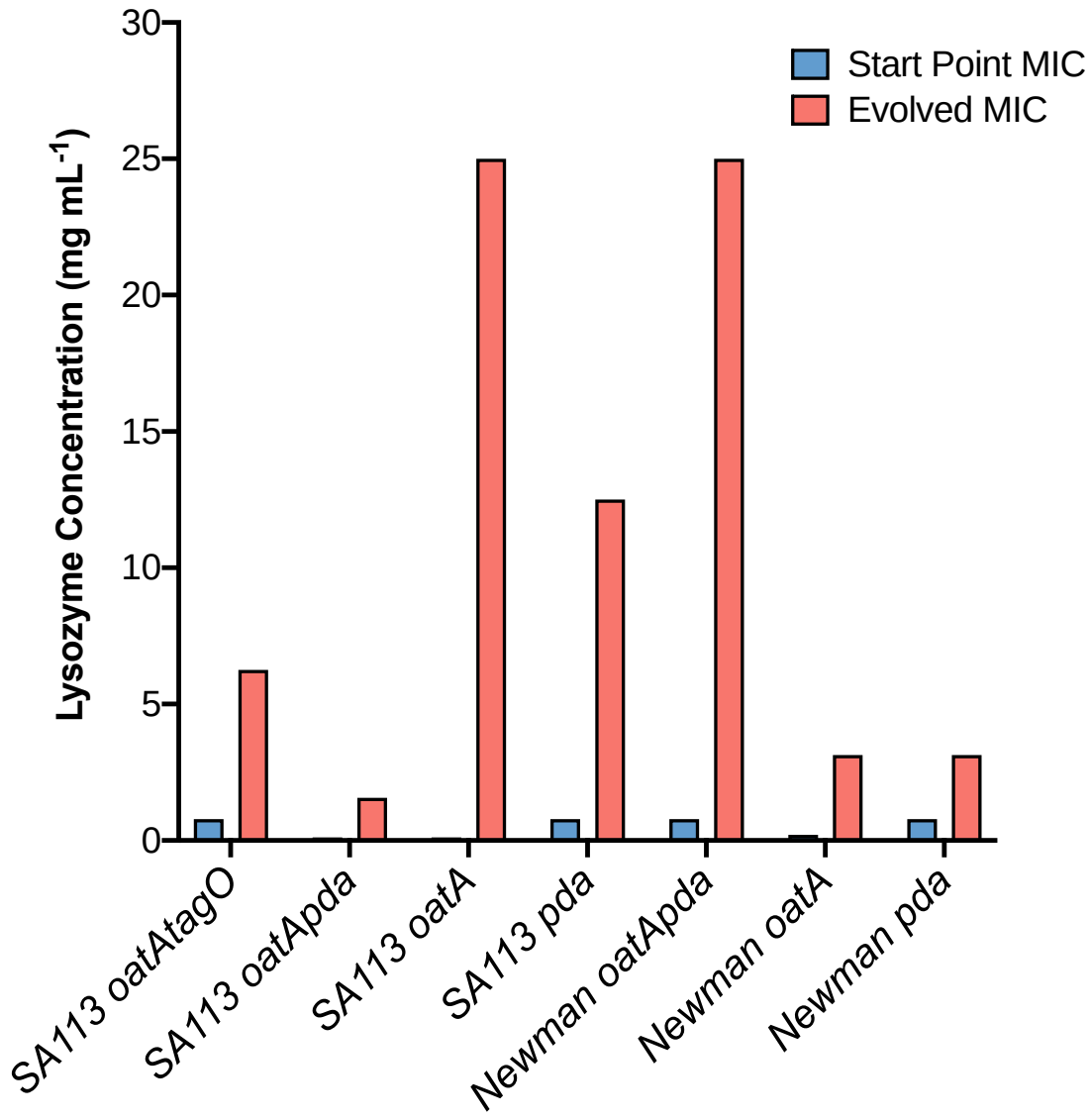


Figure 4.9 MIC of start point mutant and evolved *S. aureus* strains.
Experimental evolution was carried out using a microtitre plate method. The lysozyme MIC start point isolates (blue) and that of evolved isolated (red) are indicated. These data are based on three biological replicates.

Lysozyme resistance was also selected using a lysozyme diffusion assay. This was achieved because growth of SA113 *oatAtagO* was observed within the inhibition zone for the strain, indicating colonies with increased resistance. When these colonies were picked and repeat tested in the assay, they were completely resistant to lysozyme, where the Whatman filter paper discs contained 100 mg ml^{-1} lysozyme (Figure 4.10).

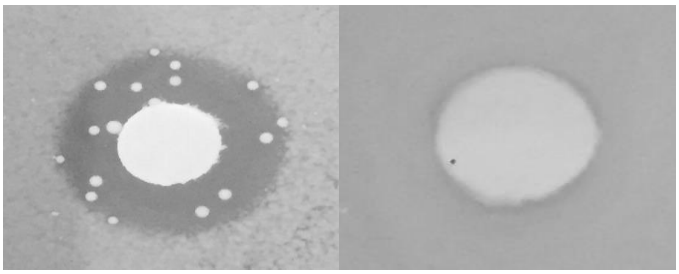


Figure 4.10 Lysozyme selection with agar diffusion assay. Growth inhibition caused by lysozyme was tested on BHI plates with lysozyme Whatman paper discs ((A) 20 mg ml^{-1} & (B) 100 mg ml^{-1}). (A) SA113 *oatA tagO*; (B) Evolved SA113 *oatA tagO* from a single colony within the inhibition zone of A was tested on a second agar plate.

4.3.4.2 Lysozyme selection for SNPs

Further analysis of lysozyme resistance in the *S. aureus pda*, *oatA* and *tagO* mutants from experimental evolution was carried out with two aims. Firstly, to unravel the hierarchical regulation of *pda,* which is not understood, where the described surface charge modifications due to OatA and TagO cell wall decoration are part of a known resistance network. Second, to investigate whether Pda interfaces with the complex mechanisms linking both virulence factors and resistance in *S. aureus* that are associated with the bacterial cell wall.

4.3.4.2.1 Sequencing alignment statistics

Sequencing of isolates with increased resistance to lysozyme was performed to determine SNPs or INDELs that account for changes in resistance and provide indications of potential resistance mechanisms linked to cell surface modifications Pda, OatA and TagO.

Sequence reads were aligned to reference genomes (T0 or WT strains) using the Burrows-Wheeler aligner (BWA) and checked for contamination, by mapping to the RefSeq bacterial database using Kraken v2.0. Analysis of the alignment data (Table 4.3) revealed that for most isolates more than 97% of reads aligned to the reference genomes, whilst others had fewer than 40% aligned. Kraken analysis of isolates with less than 40% of their reads aligned to reference genomes, were also found to be contaminated with reads from other organisms such as *Staphylococcus pasteuri, Bacillaceae* and viruses. Contaminated samples were disregarded in further analysis (Table 4.3).

Table 4.3 Mapping statistics for sequenced *S. aureus* isolates. The number of reads per isolate and the percentage of reads mapped to reference genomes.

Strain	Isolate & Time point	Total number of reads	Total reads mapped to reference genome (%)
SA113	SA113 *oatAtagO* T11	863470	37.48
	SA113 *oatAtagO* T0	1006352	99.90
	SA113 *oatA* T0	1340020	99.92
	SA113 *oatA* T10	671780	99.89
	SA113 *oatApda* T0	981986	99.88
	SA113 *oatApda* T11	1113746	38.10
	SA113 *pda* T0	893536	99.89
	SA113 *pda* T8	706236	36.10
	SA113 *pda* T9	1022124	99.89
	SA113 *oatAtagO* T11 PC	854528	99.92
	SA113 *oatA* T10 PC	816758	99.88
	SA113 *oatApda* T11 PC	511704	96.71
	SA113 *pda* T8 PC	540902	98.37
	SA113 *pda* T9 PC	636768	98.22
	SA113 *oatAtagO* disc diffusion	623892	99.90
Newman	Newman *pda* T0	1334442	99.79
	NEWMAN *pda* T2	445928	99.75
	NEWMAN *pda* T4	775956	99.78
	NEWMAN *oatA* T0	617432	99.82
	NEWMAN *oatA* T8	695752	37.16
	NEWMAN *oatApda* T0	497016	99.33
	NEWMAN *oatApda* T11	697496	36.70
	NEWMAN *oatApda* T8	710044	36.50
	NEWMAN *pda* T10	965756	99.78
	NEWMAN *pda* T9	793714	99.81
	NEWMAN *pda* T9 PC	866418	99.79

	NEWMAN *pda* T10 PC	723252	99.75
	NEWMAN *pda* T2 PC	1135310	99.79
	NEWMAN *pda* T4 PC	779682	99.83
	NEWMAN *oatA* T8 PC	780540	99.84
	NEWMAN *oatApda* T11 PC	789876	98.54
	NEWMAN *oatApda* T8 PC	1000542	98.56

4.3.4.2.2 SNPs and INDELs

This study hypothesised that SNPs resulting from lysozyme selection would account for increased fitness and sequence changes would correspond to changes in functions associated with cell wall surface modifications and synthesis, with anticipated links to the function of OatA, TagO and Pda.

S. aureus strains SA113 and Newman were experimentally evolved by selecting for lysozyme resistance. Genomic variants from evolved isolate sequence reads were detected by SnpEff and filtered using a bespoke script to exclude any identical variants found in wild-type or start point isolates compared with evolved strains. Non-synonymous SNPs were considered as homozygous if > 80 % of reads had alternative bases to the reference genome and were therefore interrogated as allelic variants causing the selected phenotypes of increased lysozyme survival (Table 4.4 & Table 4.5). Sequence variants present in <80 % of reads were disregarded during analysis.

After the removal of sequence reads from isolates that aligned poorly to reference genomes (SA113 *oatA tagO* T11, SA113 *oatA pda* T11, SA113 *pda* T8, Newman *oatA* T8, Newman *oatA pda* T11 & Newman *oatA pda* T8) (Table 4.3) and genomic variants present in <80 % of reads, this left only strain Newman *pda* and SA113 *pda* for analysis to provide insight into the regulation of Pda (Table 4.4 & Table 4.5).

The *S. aureus* genetic polymorphisms found in evolved SA113 *pda* and Newman *pda* isolates were comprised collectively of 11 SNPs and 3 INDELs (Table 4.4 & Table 4.5). There were variants present in genes which potentially associate Pda to biofilm formation (NWMN_0013 & aur_00731, icaR), MSCRAMM genes linked

to virulence and immune evasion (NWMN_1940, sdrH), extracellular proteases (NWMN_1847, SspB) and the bacterial cell wall (NWMN_2395, GtaB & NWMN_0096, CapB). Genomic variations were also observed in other hypothetical proteins that have no known functions described and phage associated proteins (Table 4.4 & Table 4.5).

There were no common SNPs definitively linking any one particular gene to the action of Pda, when comparing variants in both Newman and SA113. However, SNPs in NWMN_0013 and NWMN_2395 were found consistently in all time points in *S. aureus* Newman *pda* and so could be important in lysozyme resistance.

In SA113 *pda* the majority of SNPs were found in hypothetical proteins (aur_02458, aur_02682, aur_02702 & aur_01065), although some were found in phage proteins (aur_00560 & aur_02702) and genes linked to biofilm formation (aur_00731, icaR) and bleomycin resistance (aur_02295). Similarly, variants found in Newman *pda* were also associated with phage proteins (NWMN_1807) and biofilm formation (NWMN_0013) but also included SNPs in additional genes linked to virulence and immune evasion (NWMN_1940, sdrH), extracellular proteases (NWMN_1847, SspB) and the bacterial cell wall (NWMN_2395, GtaB & NWMN_0096, CapB).

Table 4.4 Non-synonymous, homozygous SNPs and INDELs from *S. aureus* Newman *pda* lysozyme evolved isolates. SNPs and INDELs from isolates of *S. aureus* Newman *pda* selected for increased lysozyme MIC were determined and all mutations confirmed and classed as homozygous if > 80 % of reads had alternative bases to the reference genome.

Isolate	Gene	Gene Description	SNP position	Base Change	SNP effect
Newman *pda* T2	NWMN_0013	Hypothetical protein (sequence similarity to cyclic-di-AMP phosphodiesterase GdpP)	19166	C -> A	Stop gained
	NWMN_2395	GtaB- UTP:alpha-glucose 1-phosphate uridyltransferase	2638821	T -> G	Upstream gene variant
			2638823	T -> A	
			2638826	A -> G	
Newman *pda* T4	NWMN_0013	Hypothetical protein (sequence similarity to cyclic-di-AMP phosphodiesterase GdpP)	19166	C -> A	Stop gained
	NWMN_2395	GtaB- UTP:alpha-glucose 1-phosphate uridyltransferase	2638821	T -> G	Upstream gene variant
			2638823	T -> A	
			2638826	A -> G	
	NWMN_1807	Hypothetical protein (sequence similarity to ORF018 of Bacteriophage 77 [Bacteriophage phiNM1])	2018856	AGATCGATCGA -> AGATCGA	Upstream gene variant
			2020041	AGG -> AG	
Newman *pda* T9	NWMN_0013	Hypothetical protein (sequence similarity to cyclic-di-AMP phosphodiesterase GdpP)	19166	C -> A	Stop gained

	NWMN_0096	capB- capsular polysaccharide synthesis enzyme	120739	C -> T	Missense variant
	NWMN_1847	sspB- staphopain thiol proteinase	2059181	T -> A	Stop gained
	NWMN_1940	sdrH- cell wall associated protein, containing serine-aspartate dipeptide repeats	2148330	G -> A	Missense variant
	NWMN_2395	GtaB- UTP:alpha-glucose 1-phosphate uridyltransferase	2638821	T -> G	Upstream gene variant
			2638823	T -> A	
			2638826	A -> G	
Newman *pda* T10	NWMN_0013	Hypothetical protein (sequence similarity to cyclic-di-AMP phosphodiesterase GdpP)	19166	C -> A	Stop gained
	NWMN_0096	capB- capsular polysaccharide synthesis enzyme	120739	C -> T	Missense variant
	NWMN_1847	sspB- staphopain thiol proteinase	2059181	T -> A	Stop gained
	NWMN_1940	sdrH- cell wall associated protein, containing serine-aspartate dipeptide repeats	2148330	G -> A	Missense variant
	NWMN_2395	GtaB- UTP:alpha-glucose 1-phosphate uridyltransferase	2638821	T -> G	
			2638823	T -> A	

			2638826	A -> G	Upstream gene variant
	NWMN_1807	Hypothetical protein (sequence similarity to ORF018 of Bacteriophage 77 [Bacteriophage phiNM1])	2018856	AGATCGATCGA -> AGATCGA	Upstream gene variant
			2020041	AGG -> AG	

Table 4.5 Non-synonymous, homozygous SNPs and INDELs from *S. aureus* SA113 *pda* lysozyme evolved isolates. SNPs and INDELs from isolates of *S. aureus* SA113 *pda* selected for increased lysozyme MIC were determined and all mutations confirmed and classed as homozygous if >80 % of reads had alternative bases to the reference genome.

Isolate	Gene	Gene Description	SNP position	Base Change	SNP effect
SA113 *pda* T9	aur_00731	*icaR*- negative regulator of intercellular adhesin genes *icaABCD.*	206490	G -> C	upstream gene variant
			206511	T -> A	
	aur_02295	Glyoxalase/bleomycin resistance family protein	66029	G -> C	missense variant
	aur_02210	*trmB*- tRNA (guanine-N(7)-)-methyltransferase	68825	GTAA -> G	upstream gene variant
	aur_00560	Phage tail fibre protein	624077	C -> T	missense variant
			624124	G -> A	
	aur_02458	Hypothetical protein (similarity to stage 0 sporulation family protein)	39206	C -> T	upstream gene variant
			39304	C -> T	
	aur_02682	Hypothetical protein	220	A -> G	missense variant
	aur_02702	Hypothetical phage protein	633	T -> C	missense variant
			662	G -> T	
	aur_01065	Hypothetical protein	5390	G -> A	missense variant
			5302	T -> A	

4.3.5 Pda as a skin colonisation competition determinant

Bacterial cell surface modifications promote staphylococcal survival in the host environment, enhancing colonisation and virulence. To study whether Pda of *S. aureus* contributes as a competition determinant, with respect, promoting survival from antimicrobials of other species, competition assays were carried out with staphylococci isolated from skin, including *S. capitis* and *S. epidermidis*.

S. aureus Newman *WT, pda* and the complemented mutant *pda/Co* were used as competitor strains in deferred inhibition assays. These strains were sprayed over a range of cultured inhibitor-producing *S. capitis* and *S. epidermidis* isolates from skin (Table 4.1). The assays revealed that of 12 *S. capitis* and 6 *S. epidermidis* screened for inhibitory effects against *S. aureus* Newman, only 2 *S. capitis* and 3 *S. epidermidis* caused a reduction in growth (Figure 4.11).

Of these inhibitory strains tested*, S. capitis* 47 was isolated from the human dandruff scalp in this study (Chapter 3) whilst *S. capitis* 108 was isolated from a healthy human scalp (Kloos and Schleifer, 1975); both are *S. capitis* ssp. *capitis.* Although they are inhibitory upon *S. aureus* Newman and the isogenic mutants, based on a Kruksall-Wallis test and Dunn's post hoc test, there was not a significant differential survival when comparing strains Newman wild-type with mutants *pda* and *pda/co* (*S. capitis* 47, H= 0.302, p = 0.936; *S. capitis* 108, H = 0.6032, $p > 0.99$) (Figure 4.11 & Figure 4.12).

S. epidermidis strain 14 was isolated from the forearm (Kelly, 2013), 155 from the nasal cavity (Libberton et al., 2014) and Tü3298 is a commonly used lab strain (Allgaier et al., 1985). All three of these *S. epidermidis* strains inhibited *S. aureus* Newman *WT, pda* and *pda/co* with some significant differences. In response to *S.*

epidermidis 155 secreted antimicrobials, both Newman *WT* and Newman *pda/co* were more sensitive, with greater zones of inhibition compared with Newman *pda*, although only Newman *pda/co* was significantly different ($p < 0.05$). The same phenotype was observed with respect to *S. epidermidis* Tü3298, whilst *S. epidermidis 14* caused Newman *WT* to be the most sensitive ($p < 0.05$), then Newman *pda* and Newman *pda/co* presenting as completely resistant. This was unexpected as the complemented strain should reveal a similar phenotype to the WT strain as the *pda* mutation has been resolved.

A Kruksall-Wallis test was conducted to examine the differences in the response of *S. aureus* Newman wild-types and mutant's *pda* and *pda/co* to inhibitory strains. The test supports differences between the three *S. aureus* strains (155, H=6.889, $p = 0.0143$; 14, H= 7.086, $p = 0.0143$; Tü3298, H= 7.724, $p = 0.0036$). A further Dunn's post hoc test, looking at pairwise comparisons between mutants found significant differences between Newman *pda* and Newman *pda/co* in competition with *S. epidermidis* 155 and Tü3298 ($padj < 0.05$). However, pairwise comparisons found significant differences between Newman WT and Newman *pda/co* in response to *S. epidermidis* 14 ($padj < 0.05$) (Figure 4.11).

These results could be explained by the hypothesised change in acetylation pattern of the *S. aureus* bacterial cell wall as a result of the inactivation of Pda might be the cause of differential growth. The hypothesised high levels of acetylation caused by the inactivation of Pda producing a resistant phenotype compared with wild type.

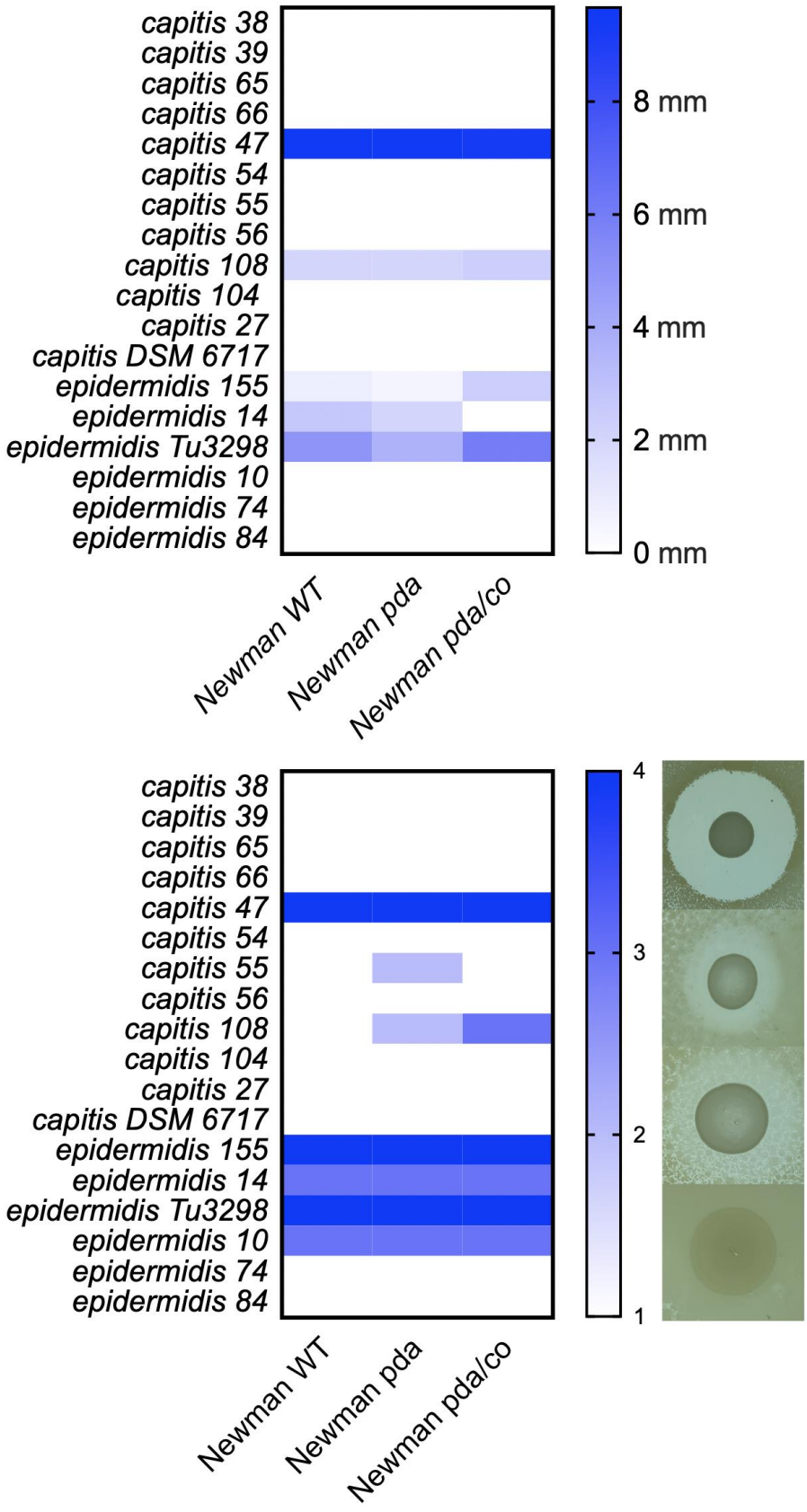


Figure 4.11 Growth inhibition of coagulase-negative staphylococci (CoNS) against *S. aureus*. The growth inhibition area of *S. aureus* is overlaid with colours indicating the relative size of the growth inhibition zone (above) and the assigned score of clarity as a measure of growth prevention (below). These data are the mean values of 3 biological replicates.

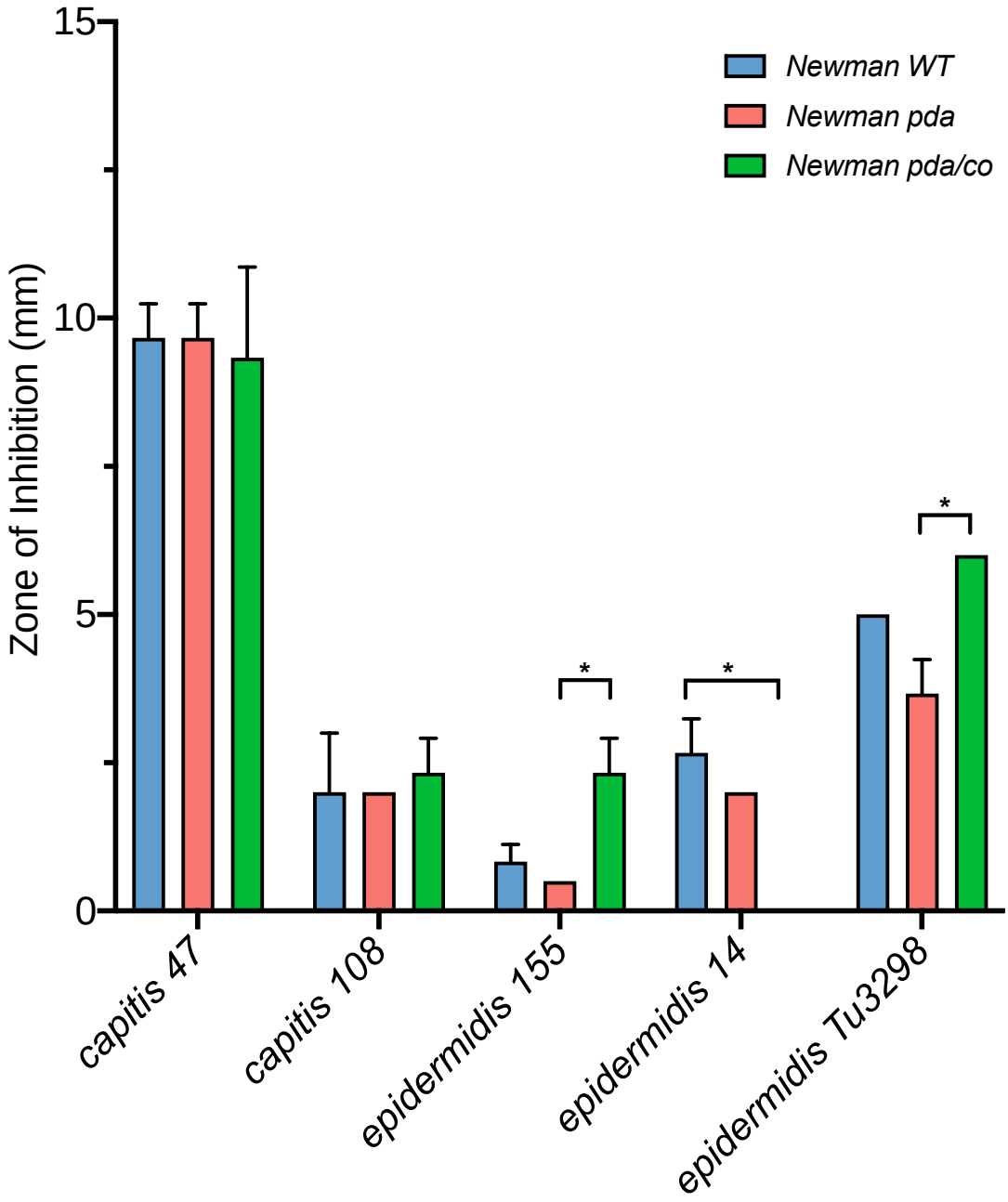


Figure 4.12 Growth inhibition of *S. aureus* Newman wild-type and mutant strains by *S. capitis* and *S. epidermidis* skin isolates. Inhibition zone diameter (mm) of *S. aureus* Newman WT (blue), *pda* (red) and *pda/co*(green) in response to competition from antimicrobials produced by *S. capitis* scalp isolates 47 and 108 and *S. epidermidis* skin isolates 155, 14 and Tü32988. Data are the mean ± standard deviation calculated from 3 biological replicates. * show significance from Kruksall-Wallis test and post hoc Dunn's test (* $p < 0.05$).

4.4 Discussion

S. aureus often forms part of the commensal skin and nasal microbiome in a proportion of the human population, although with a breach of human host defences it can exert its capabilities as a major human pathogen (Lentz et al., 2018). *S. aureus* causes a very diverse range of infections across many body sites, where the effects of bacterial virulence factors have great importance (Hodille et al., 2017).

Cell surface modifications of *S. aureus* are well described and implicated in increased environmental resistance and virulence (Herbert et al., 2007). Multiple classes of antibiotics target peptidoglycan synthesis and structure, causing the destruction of their target pathogen (Chambers and DeLeo, 2009). However, bacterial pathogens have evolved resistance strategies that combat antimicrobial action leading to emerging antibiotic resistance (Lentz et al., 2018).

This study aimed to provide an overview of a putative surface modification enzyme with respect its roles in resistance of *S. aureus* and its interactions with other species. Experiments were designed to investigate the regulation of the *pda* gene and the activity of the encoded polysaccharide deacetylase. This enzyme was hypothesised to alter the decoration of peptidoglycan, mediating cell surface charge changes that, in relation to other described mechanisms, would alter the cellular response to other niche inhabitants and antimicrobial compounds.

A previous study in the research group proposed based on homology and a range of experimental evidence that Pda was a deacetylase enzyme, and might catalyse the *N*-linked deacetylation of the amine group attached to C3 of the glucosamine residues of the peptidoglycan backbone (Alorabi, 2016). To investigate this

hypothesis, a highly targeted experiment was designed to determine possible structural effects of deacetylase absence with collaborators at the University of Guelph. For this, enzyme digestion was combined with high performance liquid chromatography (HPLC) separation of muropeptides and mass spectrometry (MS) of peptidoglycan from both WT and *pda* mutant of *S. aureus* strain SF8300. LCMS spectra highlighted the absence of differences between the WT and mutant, meaning that the glucosamine residues of *S. aureus* are not differentially acetylated in the *pda* mutant compared with WT. Thus, *N*- linked peptidoglycan acetylation levels are not determined by Pda (Figure 4.4).

Although the hypothesised target for Pda deacetylation remains unknown, data comparing *S. aureus* mutants across multiple phenotypic tests, including biofilm assays, MIC assays with different antimicrobials and hydrophobicity assay, indicated that Pda is involved in the alteration of cell surface charge (Alorabi, 2016).

To identify the surface component Pda deacetylates, potential targets were considered with a particular focus on glycosylated surface proteins. Initially, both wall teichoic acid (WTA) and capsule proteins were ruled out as a potential target. *S. aureus* strain SF8300 does not synthesise capsule (Rausch et al., 2019) yet the phenotypes of *pda* inactivation were experimentally demonstrated in this strain background and strains producing a capsule (Alorabi, 2016). LCMS spectra for purified wall teichoic acid from *S. aureus pda* mutant and WT strains were also produced in collaboration with the University of Guelph and this again did not show any differences in acetylation pattern (data not shown; Sychantha, personal communication).

Microbial surface components recognising adhesive matrix molecules (MSCRAMMs) of *S. aureus* were selected for study of Pda activity as they are known to have glycosylation modifications that could be differentially acetylated for their functions linked to adhesion, evasion of the host immune system and biofilm formation (Foster et al., 2014). MSCRAMMs selected for this study included ClfA, ClfB, SdrCDE, BshB2, plus Cap1A and IcaA. Triple mutants were created with by combining mariner mutations of these genes with both *pda* and *oatA* mutations in strain SA113 to enable differences in lysozyme sensitivity to be visualised in assays. The rationale was that adding a deacetylase target mutation to the *pda oatA* double mutant would remove the acetylated target protein proposed to be increasing lysozyme resistance.

The lysozyme phenotype of all triple mutants was compared, and none of the strains had significantly different lysozyme resistance compared with SA113 *oatA pda* ($p > 0.05$). One strain, SA113 *oatA pda sdrD* was consistently more sensitive to lysozyme across all concentrations, indicating that SdrD could influence lysozyme resistance and was a potential target of Pda activity.

SdrD MSCRAMM contributes to *S. aureus* surface adhesion along with SdrC, SdrE and clumping factor proteins ClfA and ClfB. A similar structural organisation is shared among these proteins, consisting of an N-terminal secretory signal peptide, serine-aspartate repeats and a C-terminal containing an LPXTG cell wall anchoring motif (Foster et al., 2014, Ajayi et al., 2018).

These MSCRAMM proteins contribute to *S. aureus* survival, virulence and immune evasion as they interact with host molecules such as fibrinogen, desmoglein 1 and β-neurexin (Askarian et al., 2016, Barbu et al., 2010, Keane et al., 2007). The

protein structure of SdrD is mostly conserved, although some genetic variation has been observed, suggesting adaptation of *S. aureus* to its environment, whilst maintaining its original function (Ajayi et al., 2018). This study may have observed a role for SdrD in survival from lysozyme through the added mutations of *pda* and *oatA* affecting the cell surface charge and overall response to lysozyme. To understand this further, mutation of SdrD along with *pda* and *oatA* and their lysozyme resistance phenotypes should be investigated in different *S. aureus* strain backgrounds as genetic variability in the SdrD gene is dependent on host strain (Ajayi et al., 2018).

In addition to the approach described to unravel Pda activity, the lysozyme phenotypes of a combination of *pda, oatA* and *tagO* mutants was also investigated to study the regulatory system controlling *pda* expression and how this might be modulated in response to environmental factors. The evolution of increased lysozyme resistance by a microtitre plate passage method (Figure 4.7 & Figure 4.8) and lysozyme diffusion assay (Figure 4.9) led to an increase in MIC for the *pda, oatA* and *tagO* mutants used in this study over the 11 passages with increasing lysozyme concentration. High-throughput sequencing and variant analysis of the resulting lysozyme resistant mutants, although this only ultimately included Newman and SA113 *pda,* revealed SNPs and INDELs that may account for changes in resistance.

No common sequence variants were found in both Newman and SA113 *pda* evolved mutants and so no one gene was implicated alone in modulating lysozyme resistance. However, strain differences are not uncommon with *S. aureus* studies related to cell surface function and properties. Therefore, all genes

containing variants identified in this study should be studied further to understand possible strain dependent responses to lysozyme in the absence of *pda.*

This study highlighted genes which alter lysozyme resistance in the absence of *pda* and these could be functionally linked to the *pda* gene and under the same regulatory factor. Notably, SNPs were found in genes implicated in biofilm formation. Nonsense stop-gaining SNPs were identified in NWMN_0013, a hypothetical protein containing a GGDEF domain. The GGDEF domain has been linked to biofilm formation in *S. aureus* and the *icaABCD* operon which is also implicated in biofilm formation and regulated by icaR, a gene found to have upstream modifications to its action in this study (Table 4.4 & Table 4.5) (Cue et al., 2009, Fischer et al., 2014). NWMN_0013 has homology to the gene *gdpP* in *S. aureus* NCTC 8325. This gene produces the enzyme cyclic-di-AMP phosphodiesterase and has roles in controlling cell size, coping with membrane and cell wall stress, and tolerance to inhibitors of peptidoglycan synthesis (Griffiths and O'Neill, 2012, Bowman et al., 2016). Interestingly variants in this gene appeared in all Newman *pda* time points sequenced and a reduction in cell size has also been linked to the absence of *pda*. This gene could therefore be functionally linked to the *pda* gene.

In addition to this, SNPs were also found in MSCRAMM genes which promote adherence of staphylococci, evasion of the host immune system, induce biofilm formation and promote staphylococcal infections (Foster et al., 2014). A missense variant causing an amino acid change from proline to serine in the *sdrH* gene was found in strain Newman *pda*, which has moderately affected protein function. The

Sdr proteins in *S. aureus* are encoded by *sdrCDE*, with additional genes including *sdrF*, *sdrG* and *srdH* being reported in other staphylococci (Xue et al., 2011, McCrea et al., 2000). These genes were also linked to Pda in experiments of this study which looked to identify potential target glycosylated proteins (Figure 4.6).

Other notable variants were found in cell wall associated genes. Upstream modification variants were found to affect *gtaB* (NWMN_2395) in Newman *pda*. This gene is essential for the synthesis of the glycolipid diglucosyl-diacylglycerol, which is synthesised in the bacterial cytoplasm and is a constituent part of lipoteichoic acid (LTA). LTA is an abundant polymer found anchored to the bacterial cell wall and contains D-alanine ester residues due the action of the *dlt* operon, potentially linking Pda to cell surface decoration and charge modification (Gründling and Schneewind, 2007). Altered expression of *gtaB* could have been selected to modulate the LTA expression in the absence of Pda activity and this could be studied further.

SNPs were also identified in the *sspB* gene that encodes an extracellular cysteine protease activated by a proteolytic cascade involving other *S. aureus* proteases – Aur and ScpA. These extracellular proteases were linked to the attenuation of virulence and biofilm formation (Mootz et al., 2013, Shaw et al., 2004). The role of the sspB SNP is unclear but as an extracellular protease it could affect multiple targets and enzymes extracellularly. A further SNP was identified in the gene encoding CapB, which forms a tyrosine kinase complex with CapA controlling peptidoglycan biosynthesis which is functionalised with glycopolymers and capsular polysaccharide (Rausch et al., 2019a), which again links functions of the staphylococcal cell surface with functional category proposed for Pda (Figure 4.5

& Figure 4.6). The capsule was originally ruled out as a target for Pda modification since SF8300 is deficient in capsule proteins, however, there may be multiple targets for Pda, depending on its substrate specificity. SF8300 was not included in the experimental evolution experiment due to its high innate resistance to lysozyme.

Lysozyme resistance in *S. aureus* has currently been attributed to *O*-linked acetylation of peptidoglycan and teichoic acid (Bera et al., 2005, Bera et al., 2007). This study hypothesised that the function of Pda is likely linked to that of OatA and TagO as Pda effects *S. aureus* lysozyme resistance. Unfortunately, mutant strains of OatA and TagO included in lysozyme selection experiments could not be used in this study due to poor alignment to reference sequences so genes potentially linking Pda, OatA and TagO were not found. Therefore, to further understand the connection between these three genes, this experiment could be repeated. Mutants could be analysed using qPCR to reveal whether they have altered expression of *pda* or another related pathway, in the presence of lysozyme. In addition to this the contribution of a regulatory pathway or linked genes should be analysed in strains where *pda* is active and inactivated, to understand whether interactions only occur when *pda* is inactivated.

Although the regulation of Pda is still unknown, there are some candidates that could be tested, including the autolysis regulator MgrA. MgrA is described to regulate multiple genes associated with cell wall remodelling and surface protein expression in *S. aureus* (Crosby et al., 2016). This gene controls the expression of eight surface proteins (Ebh, SraP, Spa, FnbP, SasG, SasC, FmtB and SdrD) affecting clumping, virulence and biofilm formation. That MgrA is involved in the

regulation of SdrD could link Pda in this study. MgrA coordinates with the two-component system ArlRS that controls a regulatory cascade in *S. aureus* that is crucial for pathogenesis (Crosby et al., 2016). ArlRS is a global staphylococcal virulence regulator, important in controlling biofilm formation via Ebh and resisting manganese starvation during host infection (Radin et al., 2016, Walker et al., 2013).

Further potential candidates in Pda regulation include WalKR and VraRS two-component systems. WalKR is a central regulator of bacterial cell wall metabolism, thought to be involved in the maintenance of cell viability, increasing resistance to lysostaphin-induced lysis, peptidoglycan modification and biosynthesis and biofilm formation (Delauné et al., 2012, Monk et al., 2019, Dubrac et al., 2007). VraRS is linked to cell wall synthesis and the locus is transcribed in response to conditions of stress, including presence of antibiotics, oxidative stress and shifts in temperature (Qureshi et al., 2014, Gardete et al., 2006, Galbusera et al., 2011).

There is likely to be some level of coordinated regulation of Pda with not only OatA and TagO that alter surface charge but multiple cell wall determinants, with its regulation being part of a multi-targeted cascade in response to environmental pressures, due to the mutant *pda* strain presenting with phenotypes linked to environmental response, including biofilm formation, differential MIC to a number of antimicrobials and cell wall thickness (Alorabi, 2016).

Further studies are needed to understand the regulation of *pda,* its target and how key regulators interact with each other and target promoters that control the expression *pda* and thus the decoration of peptidoglycan. The methods used

in this study are explorative, they are not the most direct approach to study the function and regulation of Pda.

Experimental evolution is a useful technique in confirming evolutionary predictions and understanding the regulation of imposed selective environments. However, there are limitations to these experiments due to small population sizes, limited timescales, simplified nature of laboratory experiments and the potential to misinterpret the selected forces and other processes at work (Kawecki et al., 2012). A repeat of the experiment outlined in this study in more *S. aureus* strain backgrounds and without contamination in evolution experiments is needed.

Lysozyme phenotype testing should also be repeated using triple mutants including Δ*oatA pda* as well as other genes of interest found in this study and future studies, such as NWMN_0013 (Table 4.4). This would be a more productive way to additionally confirm the effects of Pda on its potential target and how Pda regulation and function is linked to lysozyme resistance and so OatA and TagO.

Additional more direct approaches along with those outlined in this study would be necessary to determine the regulation and function of Pda. A number of strategies were used to reveal genes effecting the expression of the Pda promoter, including transposon screening to identify regulators using a *pda-lacZ* fusion as a transcriptional reporter. Further study is required to complete these experiments, but the work proposed that the integration of the *pda* promoter vector into the *S. aureus* genome upstream of a promoterless β-galactosidase gene (*lacZ*) could be used to screen for possible genetic regulators of *pda.* The β-galactosidase activity in the colonies on X-gal agar makes them sufficiently blue-

coloured for visual screening. In addition to this, the introduction of pRN3208 a Cd^{2+}-resistant, thermosensitive transposon (Tn*551*) delivery vector into the *pda-lacZ* fusion strain, gene mutants that affect the *lacZ* expression can be identified by colour change. Cycles of raised temperature (42°C) are used to generate the library of transposants in the reporter plasmid strain to enable screening for colour variants.

Further experiments were in progress but due to the impact of COVID-19 on lab work these were not completed. One study included metabolomic analysis of a *pda* mutant compared with wild- type, that aimed to confirm differences in expression of metabolism linked proteins in *pda* mutants. The *pda* allelic replacement mutant was found to differentially express multiple metabolism and cell wall proteins compared with the wild-type, when analysed with quantitative label free proteomics (Alorabi, 2016). The dataset revealed that a total of thirty-three proteins were differentially expressed in mutant *pda*, including proteins in pyruvate and acetoin metabolism, which provide biosynthetic intermediates for key cellular processes (Alorabi, 2016, Chaffin et al., 2012).

Another aim of this study was to investigate whether Pda offers any competitive advantage to *S. aureus* with respect to colonisation in the presence of other commensal staphylococci. Bacterial cell surface modifications enhance colonisation, and it was therefore hypothesised that these could also affect colonisation in competition with other staphylococci.

Deferred growth inhibition assays with *S. capitis* and *S. epidermidis* isolated from the skin revealed that the *pda* mutant in *S. aureus* Newman had increased survival from antimicrobial inhibition by *S. epidermidis* strains 155, 15 and Tü3298. This

outcome suggests that Pda acts like other bacterial cell wall modifications that are known to enhance staphylococcal species competition with other commensal bacteria (Foster et al., 2014). The experiments, however, should be considered only a cursory perspective on this area of research and to further understand the contribution of Pda competition, more inhibitor-secreting strains should be screened, including other resident staphylococci such as *Staphylococcus hominis.*

This study overall identifies potential leads for further study. We reveal that Pda, a gene previously associated with deacetylase function contributes to competition and resistance to antimicrobial compounds such as lysozyme. This agrees with studies carried out by Alorabi *et al,* which linked Pda to the action of other known virulence, biofilm and colonisation determinants (2016). However, this study cannot definitively link this gene to any specific regulator or target, therefore further investigation into this and consequently into its role in competition and antimicrobial resistance is required.

CHAPTER 5

General Discussion

5.1 Overall summary

The skin is a largely dry and nutrient-poor environment where staphylococcal species are among the most successful bacterial genera found to colonise the human body. Commensal staphylococci can both protect and cause disease of the host. Dysbiosis of resident skin bacteria and in particular staphylococcal species is associated with undesirable skin conditions like atopic dermatitis, therefore study of bacterial survival and competition determinants are important with respect to staphylococcal survival on the skin.

The focus of this research project was to firstly utilise next generation sequencing to explore and contribute to the current knowledge of the scalp microbiota of healthy and dandruff scalps. This current work is one of the few studies to investigate the dandruff-linked scalp microbiome using shotgun metagenomic technologies. Whole genome sequencing and comparative approaches were used to study *S. capitis* strains isolated from these healthy scalp skin controls and dandruff scalps to investigate the hypothesis that *S. capitis* undergoes selection upon the scalp skin of dandruff sufferers. A thorough study of *S. capitis* genomes isolated from the scalp has so far not been made, and instead isolates from clinical sources have been the major source, meaning this study adds balance to previous investigations. The study also probed the subspecies genotypic classification of *S. capitis* using genomic approaches.

Insight into the genetic determinants of staphylococcal skin survival and competition is essential to understanding bacterial interactions. Therefore, the

current work also sought to study the Pda putative secreted carbohydrate deacetylase enzyme and competition determinant with respect to resistance of *S. aureus* and its interactions with other species.

5.2 Metagenomic profiling of healthy and dandruff scalps

Chapter 2 of this study provides a comprehensive overview of the scalp microbiome and its association with healthy and dandruff scalps, specifically comparing the taxonomic and functional profiles of metagenomic samples isolated from a small human cohort based in the United Kingdom. The data revealed that the taxonomic composition of the scalp microbiome of both healthy and dandruff scalp sites comprises key taxa of the scalp microbiome and mostly correlates with previously published studies.

Malassezia species were identified as the major components of the mycobiome of the scalp, irrespective of scalp condition. The ratio of *M. restricta* to *M. globosa* was higher in dandruff scalps compared to healthy scalps. The *Malassezia* species *M. sympodialis* and *M. globosa* were found in very low proportions in all samples. These findings are consistent with other published studies, although the opposite was observed in several scalp microbiomes (Park et al., 2012, Clavaud et al., 2013, Soares et al., 2015, Xu et al., 2016, Saxena et al., 2018, Gemmer et al., 2002).

Malassezia has long been implicated in the causation of dandruff due to reductions in *Malassezia* numbers and in scalp skin flaking being observed following the use of shampoo containing antifungal compounds, such as zinc pyrithione, ketoconazole, piroctone olamine and climbazole, and the subsequent return of skin flaking once treatment stopped (Turner et al., 2013, Piérard-Franchimont et al., 2002, Schmidt-Rose et al., 2011). The effect of *Malassezia*

populations being linked to skin disorders, due to their high prevalence on both healthy and dandruff-affected skin (Soares et al., 2016).

Such studies have led many to hypothesise that, not only can altered levels of *Malassezia* spp. be responsible for the development of dandruff, but likely a combination of host predisposition and localised microbial dysbiosis contributes too (Soares et al., 2016, Grimshaw et al., 2019). Differences in microbiota species diversity of non-lesional and lesional sites in this study support the suggestion that dandruff is systemic, and not restricted to the skin site exhibiting symptoms. Rather, an instability of the skin microenvironment due to immune response alterations and disruption of the skin barrier are linked with host predisposition (Soares et al., 2016). Understanding the interplay between all the microbial constituents of the scalp microbiome, specifically how each species competes for colonisation, how community stability is naturally maintained and how members exploit and affect the changing physiological conditions will likely aid understanding of the aetiology of dandruff.

Although many scalp studies focus mainly on the mycobiome in relation to dandruff presentation, the contribution of the bacterial microbiome of the scalp should not be underestimated. Here, *C. acnes* was revealed as the major bacterial coloniser found in all scalp states but most abundantly on healthy scalps. Other bacterial colonisers include the *Staphylococcus* species, *S. epidermidis, S. capitis* and *S. aureus* also formed part of the scalp microbiome. *S. epidermidis* were more prevalent on healthy scalps, whereas *S. aureus* and *S. capitis* were most abundant on dandruff scalps. This difference alone could be a contributing factor when this genus is implicated with several other inflammatory skin conditions. The results

of this study match those recently described in others, although there was variation in the ratio of *Cutibacterium* species to *S. epidermidis*. Previous reports proposed that *Cutibacterium* enhancement and suppression of *S. epidermidis* could have potential to lessen dandruff (Xu et al., 2016, Grimshaw et al., 2019, Park et al., 2017, Saxena et al., 2018, Wang et al., 2015).

Published data has suggested that as well as the balance of *S. epidermidis* and *C. acnes*, increased levels of *S. capitis* and *S. aureus* may also be implicated in the cause of dandruff lesions on the scalp (Grimshaw et al., 2019, Xu et al., 2016, Park et al., 2017). Resident skin bacteria express genes that provide them with competitive advantages and increase adhesion and survival on the skin (Grice et al., 2008). It is therefore important to understand the role these bacteria and genes contribute to skin disorders, by firstly exploring the genetic and functional potential of resident bacteria and understanding the effects of current dandruff treatments such as antifungal shampoos. Studies suggest some over the counter antifungal shampoos are also effective against *S. aureus* and *S. epidermidis*, due to the broad-spectrum of antimicrobial activity of some known active ingredients, such as zinc pyrithione (Leong et al., 2019, Schwartz, 2016). Metagenomic studies comparing the scalp taxonomic composition and functional potential of dandruff and healthy scalps, as well as the effects of antifungal shampoos, will shed light on species frequencies in response to the decrease in *Malassezia* spp. associated with a reduction in dandruff symptoms.

The variability in microbiome data between studies is likely attributed to their multiple differences, including cohort size, population origin, sequencing technology and bioinformatic analyses workflow (Saxena et al., 2018, Xu et al.,

2016). Sample size in particular affects the performance of certain methods of statistical analysis (Jonsson et al., 2016). While the human microbiome correlates closely with the health state of its host and is considered stable, host genetic background is an important factor in maintaining the composition of microbial communities. Evidence does exist for genetic effects of human microbiome composition, for example, the gut microbiome of monozygotic twins is significantly more similar than those of dizygotic twins (Kolde et al., 2018, Hansen et al., 2011, Goodrich et al., 2016). These genetic factors are still largely uncharacterised and often mean the use of control groups like this study can have an impact on comparisons (Quince et al., 2017, Kolde et al., 2018).

Investigation into the taxonomic composition of healthy and diseased microbiomes is important to understanding the key organisms associated with cutaneous disorders, as well as identifying novel species. The relevance of the taxonomic composition compared to how taxa affect the structure of the microbiome is often more ambiguous. For example, many studies have proved that despite being an integral part of the resident skin microbiome, *Malassezia* species have a clear connection to skin diseases, such as dandruff and seborrheic dermatitis (Flowers and Grice, 2020). Studies uncovering how *Malassezia* promotes IL-17 and IL-23 dependent inflammation under barrier impairment conditions, as well as secreting proteases such as MgSAP1 inhibiting colonisation of pathogenic *S. aureus* will advance our knowledge about how species can promote health and disease within a microbiome (Sparber et al., 2019, Li et al., 2018).

With the aim to further understand the functional and genetic factors that underpin dandruff, this study adds to others to provide insights into the functional potential of the scalp microbiome. This study found enriched KEGG functional orthologs that could contribute to bacterial and fungal survival on the scalp, and that might associate with the progression to the dandruff state. A key finding was the enrichment of KEGG orthologs linked to bacterial cell wall, and regulatory processes, including biofilm formation, nitrogen metabolism and quorum sensing, determined on dandruff and dandruff healthy scalps. These functional orthologs are linked to survival and formation of microbiomes on cutaneous sites. The enrichment of genes in these functional categories being linked to scalps where dandruff is present is consistent with persistence, and these genes are often linked to pathogenesis (Brandwein et al., 2016, Grice and Segre, 2011, Rausch et al., 2019a, Foster, 2019, Maya-Martinez et al., 2019, Saxena et al., 2018).

The results from this study expand our understanding of the dandruff scalp microbiome and provide new perspectives on the pathogenesis of dandruff. There are clear weaknesses in the study design that were beyond my control. To gain more focused interrogation of the differences highlighted here between the flora found on dandruff and healthy scalp, a larger, longitudinal, more diverse study with additional long-read sequencing would be needed. The extensive metagenomic and interrogative approaches used here would greatly assist further insights.

5.3 Comparative genomics of *Staphylococcus capitis* isolated from healthy and dandruff scalps

CoNS *S. capitis* is frequently found on the human scalp and the forehead and is associated with dandruff presenting scalps, as well as bloodstream infections and neonatal infections in intensive care units (Gras-Le Guen et al., 2007, Maggs and Pennington, 1989, Grimshaw et al., 2019, Carter et al., 2018, Cameron et al., 2015). After its first description, *S. capitis* was divided into two subspecies, *capitis* and *ureolyticus* as members of the 'Epidermidis cluster group', which comprises *S. epidermidis*, *S. saccharolyticus* and *S. caprae* (Lamers et al., 2012, Bannerman and Kloos, 1991). *S. capitis* is mostly studied with reference to the well-described and clinically important *S. epidermidis*. In the absence of an expansive study of *S. capitis* genomes from the scalp, the current work aimed to explore whole genome sequencing of the species to expand knowledge of its population genetic structure, genotypic definition of subspecies. Alongside this analysis, the study sought to identify competitive traits linked to *S. capitis* isolated from dandruff afflicted scalps.

Phylogenetic analysis of *S. capitis* genomes in this study, including both scalp and contextual isolates, revealed distinct clustering based upon their subspecies. The study therefore aimed to dissect the phylogenetic and genomic differences between the two subspecies to explore their clustering. The initial subdivision of *S. capitis* to both ssp. *ureolyticus* and ssp. *capitis* was based upon original descriptions of *S. capitis* ssp. *ureolyticus* urease activity, ability to produce acid from maltose, fatty acid profile, larger colony size and DNA sequence differentiation. This multifactorial discrimination was not explored in full here but biochemical analysis of *S. capitis* subspecies, including culture collection type

strains did not support published differences between subspecies. Further analysis is needed to determine whether weaknesses of the API test system explain this lack of consistent enzyme activity discrimination. Certainly, cursory gene content analysis does not support these phenotypic trait differences.

Further analysis to characterise the genetic relatedness between the two subspecies using ANI revealed no significant differences between genomes of each subspecies. Focusing upon the small number of discriminating gene clusters, the current study found *S. capitis* spp. *ureolyticus* genomes were enriched with antimicrobial resistance gene functions, such as β-lactam resistance genes and β-class phenol soluble modulins; whereas *S. capitis* spp. *capitis* encoded gene clusters linked to survival on the skin, such as an arginine catabolism cluster. Specific analysis of virulence gene profiles of the isolates included in this study determined that *S. capitis* has a species has a similar repertoire of virulence genes to several other CoNS species, with respect to AMR, PSMs and secreted proteases (Moawad et al., 2019, Sun et al., 2020). Investigation of the subspecies classifications highlighted here demonstrate that further analysis is required for robust markers of subspecies classification within the core genome.

It can be hypothesised that HGT events have led to the acquisition of virulence genes within *S. capitis* genomes that appear to be highly recombinogenic. The presence of more virulent strains of *S. capitis* ssp. *ureolyticus*, in greater frequencies in clinical settings combined with less virulent strains isolated from other sources, such as the scalp would indicate a potential contextual basis for the HGT events. Similar to other CoNS, *S. capitis* cannot be considered as a highly clonal species and seemingly evolves through recombination events, in contrast

to *S. aureus* that evolves largely through point mutation (Becker et al., 2014, Carter et al., 2018). Pangenome analysis indicated that the *S. capitis* genome is open, which could arise from a capacity to acquire exogenous DNA whilst living in extensive bacterial communities.

The openness of the *S. capitis* pangenome and the highly recombinogenic genome suggests that its accessory genome has the ability to offer a large repertoire of genes that confer specific advantages under particular environmental conditions to support their colonisation, or most especially virulence in a clinical setting. Many studies utilising the higher resolution afforded by WGS have enabled important differences to be uncovered in *S. capitis* genomes, such as their multidrug resistance profiles across different geographic regions (Cameron et al., 2015, Carter et al., 2018, Cui et al., 2013).

Most published studies focused on *S. capitis* strains isolated from NICU and other clinical settings and presents the emergence of drug resistance in response to the use of antimicrobial and antiseptic therapy to treat infections caused by CoNS isolates (Chong et al., 2016, Mello et al., 2008, Carter et al., 2018). Antimicrobial resistance to vancomycin and fusidic acid has been reported among *S. capitis*, similar to *S. aureus,* suggesting the occurrence of inter-species genetic exchange (Carter et al., 2018, Chong et al., 2016). Although the majority of samples in this study are not clinical isolates, potential virulence linked genes are found throughout genomes, highlighting *S. capitis* versatility and potential for adaptation to cause significant disease. Further understanding the differences between multidrug-resistant *S. capitis* isolates may also uncover more defining features of *S. capitis* subspecies to support the hypothesis that *S. capitis* ssp.

ureolyticus is generally more virulent, where studies have presented AMR in *S. capitis* but not in the context of the subspecies, instead BAPS groups.

In addition to antimicrobial resistance, biofilm formation was suggested to be an important virulence characteristic of *S. capitis* in both a clinical and commensal setting (Cui et al., 2013, Otto, 2010). In keeping with this suggestion, the current work reveals the presence of many biofilm-related genes in the *S. capitis* genome studied, including *icaRADBC* operon, *ebh* and *atlE*. These genes facilitate staphylococcal primary attachment by binding to extracellular matrix molecules, and intercellular aggregation (Otto, 2009). The presence of these genes may confer a selective advantage in both a clinical setting and on the scalp. Additionally, the presence of 14 out of the known 16 TCS found in *S. aureus* in *S. capitis* of this study highlights wide repertoire of signalling systems to adapt to diverse microenvironments (Rapun-Araiz et al., 2020). Improved phenotypic analysis of biofilm formation and multidrug resistance in *S. capitis* isolates, included those in this study will add insight to the factors contributing to pathogenesis and subspecies definition.

This study aimed to explore *S. capitis* genotypes with respect to the causation of dandruff. No genotype was associated with dandruff presentation. *S. capitis* genomes were found to contain genes necessary to colonise, survive and compete on human skin. An improved understanding of the genetic and functional potential of *S. capitis* adds to the current knowledge of how it competes and survives in relation to other scalp microbes, which could lead to further understanding of how the frequencies of certain species within the microbiome

may cause disease. The current work found neither *S. capitis* subspecies associated with dandruff scalps compared to healthy.

Phylogenomic analysis carried out in this study provides a valuable insight into *S. capitis,* isolated from both clinical and commensal environments. Additional isolates from both healthy, dandruff and clinical settings at different geographic locations with accompanying phenotypic data will allow more genotypic and phenotypic resolution to further this study.

5.4 The role of a polysaccharide deacetylase (Pda) in staphylococcal virulence and survival

Previous study of Pda in the Horsburgh group suggested that this protein might catalyse the *N*-linked addition of acetyl groups to the glucosamine residues of the peptidoglycan backbone, similar to Pgd. The function of Pda was also linked with OatA and TagO, which actively remodel the Gram-positive peptidoglycan structure through the addition of side chains (Alorabi, 2016, Bera et al., 2005, Bera et al., 2007). To gain further insights into its role, experiments were designed to determine whether the phenotypic differences found in the *pda* mutant were consistent with a lack of deacetylation and to determine genetic regulation of *pda.*

Enzyme digestion, high performance liquid chromatography (HPLC) separation of muropeptides and mass spectrometry (MS) of peptidoglycan and capsule from both wild type and *pda* mutant of *S. aureus* was carried out to determine possible structural effects of deacetylase absence. The analysis revealed the putative acetylation target of Pda was neither peptidoglycan, teichoic acid nor the capsule. At the outset it was expected, similar to the study carried out by Bera, *et al*.

investigating the role of *oatA,* that additional acetyl groups would be observed in wild type isolates compared to the *pda* mutant, as phenotypic differences were consistent with changes in acetylation pattern of the cell wall (2005).

While the hypothesised target of Pda was clearly shown not to be deacetylation of peptidoglycan, experimental data indicated that Pda was involved in the alteration of cell surface charge (Alorabi, 2016). Given the range of potential strategies to pursue its function, investigation of potential cell wall-associated targets became a focus and the role of Pda in lysozyme resistance was used in an attempt to implicate surface proteins that might be functional targets.

Although a specific regulator or target of Pda was not revealed, leads for further investigation were identified. This study points to SdrD having a potential contribution to lysozyme survival from combined mutations of *sdrD* with *pda* and *oatA* affecting the cell surface charge and overall response to lysozyme. To understand this further, mutation of SdrD along with *pda* and *oatA* and their lysozyme resistance phenotypes should be investigated in different *S. aureus* strain backgrounds as there is genetic variability between different host strains, as revealed by the phenotypic analysis of Alorabi, *et al.* (2016).

This study highlighted genes potentially linked to the regulation of *pda* in the context of other well-described surface charge modifications as well as complex mechanisms attributed to both virulence and resistance in *S. aureus*. SNPs from lysozyme selection experiments were found in genes implicated in biofilm formation, lipoteichoic acid synthesis and extracellular protease production including *sdrH, gtaB* and *sspB* (Xue et al., 2011, Gründling and Schneewind, 2007, Shaw et al., 2004, Mootz et al., 2013).

The outcomes suggest that Pda may act together with described bacterial cell wall modification activities that are known to enhance staphylococcal species competition with other commensal bacteria. Deferred growth inhibition assays with *S. capitis* and *S. epidermidis* isolated from the skin revealed that a *S. aureus* Newman *pda* mutant had increased survival from antimicrobial inhibition by *S. epidermidis* strains 155, 15 and Tü3298 that secrete different antimicrobials. However, these experiments were only a cursory study and to fully understand the contribution of Pda on competition, inhibitor-secreting strains should be screened, including other prominent resident species, such as *S. hominis.*

Staphylococcal spp. and most especially *S. aureus* are important opportunistic pathogens, factors that aid host immune evasion, such as resistance to lysozyme effect adhesion and survival on the skin, such as biofilm formation are essential to staphylococci. Understanding the genetic factors that underpin important phenotypes previously linked to Pda are important to broaden understanding of not only *S. aureus* virulence and colonisation but also potential functions in other bacterial species. The cell wall acetylation mediated by OatA increases *S. aureus* potential for survival, pathogenicity and reinfection by rendering them resistant to killing by lysozyme within macrophages, inhibiting the production of cytokine IL-1β and decreasing the efficiency of helper T-cell priming (Brott and Clarke, 2019, Sanchez et al., 2017, Bera et al., 2005). A further mechanism to manipulate surface acetylation would be of great interest.

The methods described in this study were purposely explorative to develop leads to pursue the target of Pda and were only part of a broader template and not the most direct approach to study the function and regulation of Pda. More direct

approaches were halted. Future investigations such as metabolomic analysis and transposon screening will broaden understanding of the contribution that Pda makes to the cell and reveal the deacetylase target of Pda.

5.5 Concluding remarks

Species of *Staphylococcus* rank among the most abundant bacteria that colonise human skin and are frequent opportunists representing major nosocomial pathogens with a substantial impact on human life and health. A deeper understanding of the mechanisms that enable staphylococci to colonise, compete and persist is crucial for informing on pathogenesis, epidemiology and therapeutics. The work described in this thesis has utilised significant advances in DNA sequencing technologies to more comprehensively interrogate metagenomes and *S. capitis* genomes from the human scalp.

APPENDIX I

Supplementary Tables

Table S 1 A description, usage instructions and input files required for all bespoke scripts used in this thesis.

Thesis section	Author	Name	Description	Usage
3.2.9	Dr Rebecca Ji Bengtsson	extract_protein_seq.py	This script will extract coding protein or gene sequences from a given sequence file, based upon a tab-delimited text file containing a list of CDS identifiers	*for sample in `cat list_of_proteinIDs.txt`; do python extract_protein_seq.py $sample sequence_file.fa/a* **Input files:** • A tab-delimited (.txt) file containing a list of sequence identifiers of interest • Sequence file (amino acid or nucleotide) **Output files:** • Sequence file containing a header of each CDS input followed by the extracted sequence
4.2.4.5 **4.3.4.2.2**	Dr Josephine Moran	unique_SNPs.pl	This script filters SNP data found in evolved, parent and control datasets.	*perl unique_SNPs.pl control_snpEff test_snpEff output_name.csv Snp_eff_chromosome_name min_read_depth(default = 10)*

				Input files: • Files created from variant calling pipeline described in section 4.2.4.5 • VFC file created by snpEFF (Cingolani et al., 2012) with all annotated SNPs found in control (T0 or positive control) sample genomes • VCF file created by snpEFF (Cingolani et al., 2012) with all annotated SNPs found in test sample genomes **Output files:** • CSV file of unique SNPs found in test sample genomes. CSV file contains 17 columns: o Gene name o Gene ID o Reference base o Alternate base o SNP effect o Impact o Codon change o Position in gene o Position in protein o Read depth o % SNP o Reference forward reads o Reference reverse reads

				○ Alternative forward reads ○ Alternative reverse reads • CSV file of SNPs found in both test and control VCF files, containing the same 17 columns as above.

Table S 2 Taxonomic composition of the scalp microbiome for each sample. The relative abundance and number of reads assigned of the top 25 most abundant taxa are shown for each scalp sample. Taxonomic profiles were generated using Kraken v2.0 and relative abundance estimations were generated using Bracken.

Sample	Organism name	Number of assigned reads	Relative abundance (%)
HH_1	*Corynebacterium segmentosum*	2342	28.55
HH_1	*Ralstonia pickettii*	1304	15.89
HH_1	*Malassezia restricta*	1252	15.26
HH_1	*Cutibacterium acnes*	1015	12.37
HH_1	*Corynebacterium kroppenstedtii*	941	11.47
HH_1	*Ralstonia mannitolilytica*	119	1.45
HH_1	*Ralstonia insidiosa*	79	0.96
HH_1	*Paenibacillus larvae*	64	0.78
HH_1	*Staphylococcus epidermidis*	50	0.61
HH_1	*Ustilago maydis*	35	0.43
HH_1	*Ralstonia solanacearum*	26	0.32
HH_1	*Corynebacterium glutamicum*	22	0.27
HH_1	*Corynebacterium striatum*	18	0.22
HH_1	*Lawsonella clevelandensis*	18	0.22
HH_1	*Eremothecium gossypii*	17	0.21
HH_1	*Corynebacterium diphtheriae*	14	0.17
HH_1	*Streptococcus mitis*	13	0.16
HH_1	*Cutibacterium granulosum*	12	0.15
HH_1	*Thielavia terrestris*	11	0.13
HH_1	*Klebsiella pneumoniae*	10	0.12
HH_1	*Veillonella parvula*	9	0.11

HH_1	*Rothia dentocariosa*	8	0.10
HH_1	*Staphylococcus aureus*	8	0.10
HH_1	*Paraburkholderia phymatum*	8	0.10
HH_1	*Corynebacterium simulans*	7	0.09
HH_1	other	802	9.78
HH_2	*Enterococcus hirae*	2831	21.82
HH_2	*Malassezia restricta*	2327	17.93
HH_2	*Cutibacterium granulosum*	1831	14.11
HH_2	*Staphylococcus capitis*	1645	12.68
HH_2	*Enterococcus faecium*	1194	9.20
HH_2	*Staphylococcus epidermidis*	1047	8.07
HH_2	*Cutibacterium acnes*	378	2.91
HH_2	*Ralstonia pickettii*	194	1.50
HH_2	*Paracoccus sp. Arc7-R13*	120	0.92
HH_2	*Cutibacterium avidum*	63	0.49
HH_2	*Veillonella parvula*	56	0.43
HH_2	*Ralstonia mannitolilytica*	43	0.33
HH_2	*Staphylococcus aureus*	41	0.32
HH_2	*Acinetobacter lwoffii*	38	0.29
HH_2	*Haemophilus parainfluenzae*	38	0.29
HH_2	*Propionibacterium sp. oral taxon 193*	28	0.22
HH_2	*Blastococcus saxobsidens*	28	0.22
HH_2	*Geodermatophilus obscurus*	27	0.21
HH_2	*Staphylococcus hominis*	23	0.18
HH_2	*Acinetobacter baumannii*	21	0.16

HH_2	*Actinomyces oris*	20	0.15
HH_2	*Paracoccus marcusii*	18	0.14
HH_2	*Enterococcus cecorum*	18	0.14
HH_2	*Modestobacter marinus*	17	0.13
HH_2	*Ustilago maydis*	16	0.12
HH_2	other	914	7.04
HH_3	*Staphylococcus epidermidis*	1183	33.26
HH_3	*Ralstonia pickettii*	877	24.66
HH_3	*Malassezia restricta*	373	10.49
HH_3	*Staphylococcus capitis*	271	7.62
HH_3	*Cutibacterium acnes*	228	6.41
HH_3	*Ralstonia mannitolilytica*	121	3.40
HH_3	*Cutibacterium granulosum*	74	2.08
HH_3	*Ralstonia insidiosa*	70	1.97
HH_3	*Haemophilus parainfluenzae*	41	1.15
HH_3	*Veillonella parvula*	25	0.70
HH_3	*Ralstonia solanacearum*	23	0.65
HH_3	*Actinomyces oris*	20	0.56
HH_3	*Lawsonella clevelandensis*	15	0.42
HH_3	*Rothia dentocariosa*	14	0.39
HH_3	*Staphylococcus aureus*	8	0.22
HH_3	*Corynebacterium matruchotii*	6	0.17
HH_3	*Staphylococcus pasteuri*	6	0.17
HH_3	*Staphylococcus haemolyticus*	6	0.17
HH_3	*Cupriavidus pinatubonensis*	5	0.14

HH_3	*Acidovorax sp. KKS102*	5	0.14
HH_3	*Propionibacterium sp. oral taxon 193*	4	0.11
HH_3	*Lautropia mirabilis*	4	0.11
HH_3	*Aggregatibacter aphrophilus*	4	0.11
HH_3	*Novibacillus thermophilus*	4	0.11
HH_3	*Tannerella sp. oral taxon HOT-286*	3	0.08
HH_3	other	167	4.69
HH_4	*Malassezia restricta*	2478	26.05
HH_4	*Staphylococcus capitis*	1829	19.23
HH_4	*Cutibacterium acnes*	1441	15.15
HH_4	*Ralstonia pickettii*	494	5.19
HH_4	*Corynebacterium kroppenstedtii*	405	4.26
HH_4	*Lawsonella clevelandensis*	272	2.86
HH_4	*Staphylococcus epidermidis*	146	1.53
HH_4	*Ustilago maydis*	112	1.18
HH_4	*Cutibacterium granulosum*	104	1.09
HH_4	*Ralstonia mannitolilytica*	80	0.84
HH_4	*Streptococcus sanguinis*	57	0.60
HH_4	*Propionibacterium sp. oral taxon 193*	53	0.56
HH_4	*Pseudomonas resinovorans*	50	0.53
HH_4	*Veillonella parvula*	50	0.53
HH_4	*Haemophilus parainfluenzae*	49	0.52
HH_4	*Ralstonia insidiosa*	35	0.37
HH_4	*Fusarium venenatum*	31	0.33
HH_4	*Cutibacterium avidum*	28	0.29

HH_4	*Actinomyces oris*	26	0.27
HH_4	*Thielavia terrestris*	22	0.23
HH_4	*Corynebacterium glutamicum*	21	0.22
HH_4	*Thermothelomyces thermophilus*	20	0.21
HH_4	*Colletotrichum higginsianum*	19	0.20
HH_4	*Mycobacteroides chelonae*	19	0.20
HH_4	*Staphylococcus warneri*	19	0.20
HH_4	*other*	1653	17.38
HH_5	*Malassezia restricta*	2337	46.20
HH_5	*Staphylococcus epidermidis*	1689	33.39
HH_5	*Cutibacterium acnes*	214	4.23
HH_5	*Staphylococcus capitis*	154	3.04
HH_5	*Haemophilus parainfluenzae*	122	2.41
HH_5	*Actinomyces oris*	35	0.69
HH_5	*Veillonella parvula*	30	0.59
HH_5	*Ralstonia pickettii*	21	0.42
HH_5	*Lawsonella clevelandensis*	20	0.40
HH_5	*Staphylococcus aureus*	18	0.36
HH_5	*Lautropia mirabilis*	14	0.28
HH_5	*Ustilago maydis*	13	0.26
HH_5	*Rothia dentocariosa*	12	0.24
HH_5	*Staphylococcus hominis*	12	0.24
HH_5	*Eremothecium gossypii*	8	0.16
HH_5	*Aggregatibacter aphrophilus*	7	0.14
HH_5	*Staphylococcus haemolyticus*	7	0.14

HH_5	*Cutibacterium granulosum*	6	0.12
HH_5	*Streptococcus sanguinis*	5	0.10
HH_5	*Streptococcus oralis*	5	0.10
HH_5	*Klebsiella pneumoniae*	5	0.10
HH_5	*Pyricularia oryzae*	5	0.10
HH_5	*Ralstonia solanacearum*	4	0.08
HH_5	*Zymoseptoria tritici*	4	0.08
HH_5	*Staphylococcus caprae*	4	0.08
HH_5	other	307	6.07
HH_18	*Malassezia restricta*	3520	42.40
HH_18	*Cutibacterium acnes*	3066	36.94
HH_18	*Staphylococcus capitis*	233	2.81
HH_18	*Ralstonia pickettii*	73	0.88
HH_18	*Ustilago maydis*	68	0.82
HH_18	*Pseudomonas resinovorans*	39	0.47
HH_18	*Klebsiella pneumoniae*	38	0.46
HH_18	*Thielavia terrestris*	32	0.39
HH_18	*Thermothelomyces thermophilus*	30	0.36
HH_18	*Colletotrichum higginsianum*	24	0.29
HH_18	*Staphylococcus aureus*	18	0.22
HH_18	*Eremothecium gossypii*	18	0.22
HH_18	*Pasteurella multocida*	12	0.14
HH_18	*Yarrowia lipolytica*	12	0.14
HH_18	*Cercospora beticola*	11	0.13
HH_18	*Neurospora crassa*	11	0.13

HH_18	*Ralstonia solanacearum*	11	0.13
HH_18	*Pyricularia oryzae*	11	0.13
HH_18	*Pochonia chlamydosporia*	10	0.12
HH_18	*Zymoseptoria tritici*	9	0.11
HH_18	*Ralstonia mannitolilytica*	8	0.10
HH_18	*Enterobacter cloacae*	7	0.08
HH_18	*Fusarium pseudograminearum*	7	0.08
HH_18	*Stenotrophomonas maltophilia*	7	0.08
HH_18	*Fusarium venenatum*	7	0.08
HH_18	other	1019	12.28
HH_19	*Cutibacterium acnes*	3155	37.05
HH_19	*Malassezia restricta*	2884	33.87
HH_19	*Cutibacterium granulosum*	820	9.63
HH_19	*Acinetobacter lwoffii*	144	1.69
HH_19	*Acinetobacter johnsonii*	89	1.05
HH_19	*Ralstonia pickettii*	89	1.05
HH_19	*Lawsonella clevelandensis*	77	0.90
HH_19	*Ustilago maydis*	57	0.67
HH_19	*Acinetobacter sp. FDAARGOS_560*	51	0.60
HH_19	*Cutibacterium avidum*	25	0.29
HH_19	*Klebsiella pneumoniae*	18	0.21
HH_19	*Thermothelomyces thermophilus*	16	0.19
HH_19	*Acinetobacter baumannii*	16	0.19
HH_19	*Ralstonia mannitolilytica*	15	0.18
HH_19	*Fusarium venenatum*	13	0.15

HH_19	*Pasteurella multocida*	12	0.14
HH_19	*Moraxella osloensis*	12	0.14
HH_19	*Streptococcus mitis*	10	0.12
HH_19	*Thielavia terrestris*	10	0.12
HH_19	*Staphylococcus aureus*	9	0.11
HH_19	*Colletotrichum higginsianum*	9	0.11
HH_19	*Pyricularia oryzae*	9	0.11
HH_19	*Zymoseptoria tritici*	8	0.09
HH_19	*Acinetobacter sp. M131*	7	0.08
HH_19	*Micrococcus luteus*	7	0.08
HH_19	other	953	11.19
HH_20	*Cutibacterium acnes*	2713	52.64
HH_20	*Ralstonia pickettii*	749	14.53
HH_20	*Staphylococcus capitis*	607	11.78
HH_20	*Malassezia restricta*	159	3.08
HH_20	*Lactobacillus crispatus*	145	2.81
HH_20	*Staphylococcus epidermidis*	120	2.33
HH_20	*Ralstonia mannitolilytica*	96	1.86
HH_20	*Propionibacterium sp. oral taxon 193*	59	1.14
HH_20	*Cutibacterium avidum*	57	1.11
HH_20	*Ralstonia insidiosa*	40	0.78
HH_20	*Streptococcus mitis*	24	0.47
HH_20	*Lawsonella clevelandensis*	16	0.31
HH_20	*Cutibacterium granulosum*	14	0.27
HH_20	*Ralstonia solanacearum*	12	0.23

HH_20	*Dermacoccus nishinomiyaensis*	8	0.16
HH_20	*Bacillus cereus*	7	0.14
HH_20	*Staphylococcus aureus*	7	0.14
HH_20	*Aeribacillus pallidus*	6	0.12
HH_20	*Paenibacillus larvae*	6	0.12
HH_20	*Lactococcus lactis*	6	0.12
HH_20	*Saccharomyces cerevisiae*	6	0.12
HH_20	*Novibacillus thermophilus*	4	0.08
HH_20	*Mycobacterium sp. JLS*	4	0.08
HH_20	*Staphylococcus pasteuri*	4	0.08
HH_20	*Staphylococcus hominis*	4	0.08
HH_20	other	281	5.45
HH_21	*Malassezia restricta*	7650	79.09
HH_21	*Staphylococcus capitis*	1079	11.15
HH_21	*Salmonella enterica*	94	0.97
HH_21	*Ustilago maydis*	52	0.54
HH_21	*Cutibacterium acnes*	37	0.38
HH_21	*Staphylococcus epidermidis*	24	0.25
HH_21	*Lawsonella clevelandensis*	18	0.19
HH_21	*Comamonas testosteroni*	17	0.18
HH_21	*Thielavia terrestris*	13	0.13
HH_21	*Eremothecium gossypii*	11	0.11
HH_21	*Ralstonia pickettii*	11	0.11
HH_21	*Staphylococcus aureus*	8	0.08
HH_21	*Zymoseptoria tritici*	8	0.08

HH_21	*Fusarium venenatum*	8	0.08
HH_21	*Colletotrichum higginsianum*	7	0.07
HH_21	*Cercospora beticola*	7	0.07
HH_21	*Mycobacterium marinum*	7	0.07
HH_21	*Sorangium cellulosum*	7	0.07
HH_21	*Streptomyces sp. M56*	6	0.06
HH_21	*Pochonia chlamydosporia*	6	0.06
HH_21	*Moorea producens*	6	0.06
HH_21	*Cupriavidus taiwanensis*	6	0.06
HH_21	*Pyricularia oryzae*	6	0.06
HH_21	*Burkholderia cepacia*	5	0.05
HH_21	*Ogataea parapolymorpha*	5	0.05
HH_21	other	575	5.94
HH_22	*Staphylococcus epidermidis*	2011	25.90
HH_22	*Staphylococcus capitis*	1891	24.36
HH_22	*Malassezia restricta*	1599	20.60
HH_22	*Lawsonella clevelandensis*	429	5.53
HH_22	*Cutibacterium acnes*	304	3.92
HH_22	*Ralstonia pickettii*	228	2.94
HH_22	*Ustilago maydis*	44	0.57
HH_22	*Corynebacterium glutamicum*	23	0.30
HH_22	*Ralstonia mannitolilytica*	20	0.26
HH_22	*Neurospora crassa*	18	0.23
HH_22	*Staphylococcus hominis*	18	0.23
HH_22	*Eremothecium gossypii*	16	0.21

HH_22	*Fusarium venenatum*	15	0.19
HH_22	*Comamonas testosteroni*	15	0.19
HH_22	*Staphylococcus aureus*	12	0.15
HH_22	*Ralstonia insidiosa*	11	0.14
HH_22	*Ralstonia solanacearum*	10	0.13
HH_22	*Streptococcus gordonii*	9	0.12
HH_22	*Colletotrichum higginsianum*	9	0.12
HH_22	*Streptococcus sanguinis*	9	0.12
HH_22	*Zymoseptoria tritici*	9	0.12
HH_22	*Klebsiella pneumoniae*	9	0.12
HH_22	*Streptococcus oralis*	8	0.10
HH_22	*Rhodococcus erythropolis*	8	0.10
HH_22	*Pochonia chlamydosporia*	8	0.10
HH_22	other	1031	13.28
DD_6	*Cutibacterium acnes*	1952	64.40
DD_6	*Malassezia restricta*	665	21.94
DD_6	*Staphylococcus aureus*	72	2.38
DD_6	*Klebsiella pneumoniae*	71	2.34
DD_6	*Pasteurella multocida*	53	1.75
DD_6	*Ralstonia solanacearum*	26	0.86
DD_6	*Enterobacter cloacae*	18	0.59
DD_6	*Botrytis cinerea*	13	0.43
DD_6	*Burkholderia pseudomallei*	8	0.26
DD_6	*Propionibacterium sp. oral taxon 193*	7	0.23
DD_6	*Candida dubliniensis*	7	0.23

DD_6	*BeAn 58058 virus*	5	0.16
DD_6	*Naumovozyma castellii*	4	0.13
DD_6	*Human endogenous retrovirus K*	4	0.13
DD_6	*Neisseria gonorrhoeae*	4	0.13
DD_6	*Candida albicans*	3	0.10
DD_6	*Corynebacterium matruchotii*	3	0.10
DD_6	*Bacillus cereus*	3	0.10
DD_6	*Thermothelomyces thermophilus*	3	0.10
DD_6	*Paracoccus mutanolyticus*	3	0.10
DD_6	*Candida orthopsilosis*	3	0.10
DD_6	*Neisseria subflava*	3	0.10
DD_6	*Actinomyces oris*	3	0.10
DD_6	*Aspergillus oryzae*	2	0.07
DD_6	*Hydrogenophaga sp. NH-16*	2	0.07
DD_6	other	94	3.10
DD_8	*Malassezia restricta*	1155	65.22
DD_8	*Cutibacterium acnes*	114	6.44
DD_8	*Staphylococcus aureus*	80	4.52
DD_8	*Klebsiella pneumoniae*	63	3.56
DD_8	*Pasteurella multocida*	57	3.22
DD_8	*Staphylococcus epidermidis*	36	2.03
DD_8	*Staphylococcus capitis*	26	1.47
DD_8	*Propionibacterium sp. oral taxon 193*	25	1.41
DD_8	*Ralstonia solanacearum*	22	1.24
DD_8	*Enterobacter cloacae*	16	0.90

DD_8	*Botrytis cinerea*	9	0.51
DD_8	*Candida dubliniensis*	8	0.45
DD_8	*BeAn 58058 virus*	8	0.45
DD_8	*Bacillus cereus*	8	0.45
DD_8	*Burkholderia pseudomallei*	8	0.45
DD_8	*Human endogenous retrovirus K*	7	0.40
DD_8	*Paracoccus mutanolyticus*	5	0.28
DD_8	*Cutibacterium avidum*	5	0.28
DD_8	*Kluyveromyces marxianus*	5	0.28
DD_8	*Human polyomavirus 7*	4	0.23
DD_8	*Legionella pneumophila*	4	0.23
DD_8	*Neisseria gonorrhoeae*	4	0.23
DD_8	*Clostridioides difficile*	3	0.17
DD_8	*Vibrio harveyi*	2	0.11
DD_8	*Candida albicans*	2	0.11
DD_8	*Auritidibacter sp. NML130574*	2	0.11
DD_8	other	93	5.36
DD_10	*Malassezia restricta*	6589	68.92
DD_10	*Cutibacterium acnes*	1316	13.77
DD_10	*Staphylococcus capitis*	1221	12.77
DD_10	*Staphylococcus aureus*	81	0.85
DD_10	*Staphylococcus epidermidis*	56	0.59
DD_10	*Klebsiella pneumoniae*	46	0.48
DD_10	*Pasteurella multocida*	38	0.40
DD_10	*Ralstonia solanacearum*	25	0.26

DD_10	*Enterobacter cloacae*	18	0.19
DD_10	*Candida dubliniensis*	9	0.09
DD_10	*Staphylococcus hominis*	8	0.08
DD_10	*Burkholderia pseudomallei*	8	0.08
DD_10	*Staphylococcus haemolyticus*	8	0.08
DD_10	*Botrytis cinerea*	7	0.07
DD_10	*Staphylococcus saprophyticus*	5	0.05
DD_10	*Staphylococcus pasteuri*	5	0.05
DD_10	*Naumovozyma castellii*	4	0.04
DD_10	*Pseudomonas putida*	4	0.04
DD_10	*Paracoccus mutanolyticus*	3	0.03
DD_10	*Human endogenous retrovirus K*	3	0.03
DD_10	*BeAn 58058 virus*	3	0.03
DD_10	*Legionella pneumophila*	3	0.03
DD_10	*Fusobacterium ulcerans*	3	0.03
DD_10	*Kluyveromyces marxianus*	3	0.03
DD_10	*Synechocystis sp. PCC 6714*	2	0.02
DD_10	*Vibrio harveyi*	2	0.02
DD_10	other	88	0.96
DD_12	*Malassezia restricta*	3057	79.18
DD_12	*Staphylococcus capitis*	385	9.97
DD_12	*Cutibacterium acnes*	198	5.13
DD_12	*Ralstonia pickettii*	56	1.45
DD_12	*Klebsiella pneumoniae*	13	0.34
DD_12	*Staphylococcus aureus*	13	0.34

DD_12	*Pasteurella multocida*	10	0.26
DD_12	*Ralstonia mannitolilytica*	10	0.26
DD_12	*Cutibacterium avidum*	7	0.18
DD_12	*Ralstonia solanacearum*	7	0.18
DD_12	*Staphylococcus epidermidis*	5	0.13
DD_12	*Propionibacterium sp. oral taxon 193*	4	0.10
DD_12	*Aspergillus oryzae*	4	0.10
DD_12	*Ralstonia insidiosa*	4	0.10
DD_12	*Novibacillus thermophilus*	4	0.10
DD_12	*Paracoccus mutanolyticus*	3	0.08
DD_12	*Fusarium venenatum*	3	0.08
DD_12	*Ustilago maydis*	2	0.05
DD_12	*Candida dubliniensis*	2	0.05
DD_12	*Human endogenous retrovirus K*	2	0.05
DD_12	*Candidatus Saccharimonas aalborgensis*	2	0.05
DD_12	*Burkholderia sp. HB1*	2	0.05
DD_12	*Enterobacter cloacae*	2	0.05
DD_12	*Burkholderia pseudomallei*	2	0.05
DD_12	*Neisseria gonorrhoeae*	2	0.05
DD_12	*Neurospora crassa*	2	0.05
DD_12	other	60	1.61
DD_14	*Staphylococcus aureus*	64	14.88
DD_14	*Klebsiella pneumoniae*	60	13.95
DD_14	*Pasteurella multocida*	57	13.26

DD_14	*Malassezia restricta*	43	10.00
DD_14	*Ralstonia solanacearum*	29	6.74
DD_14	*Burkholderia pseudomallei*	27	6.28
DD_14	*Enterobacter cloacae*	20	4.65
DD_14	*Candida dubliniensis*	13	3.02
DD_14	*Human endogenous retrovirus K*	7	1.63
DD_14	*BeAn 58058 virus*	7	1.63
DD_14	*Botrytis cinerea*	6	1.40
DD_14	*Paracoccus mutanolyticus*	5	1.16
DD_14	*Neisseria gonorrhoeae*	4	0.93
DD_14	*Streptomyces sp. TLI_053*	3	0.70
DD_14	*Legionella israelensis*	3	0.70
DD_14	*Ralstonia pickettii*	3	0.70
DD_14	*Neurospora crassa*	3	0.70
DD_14	*Kluyveromyces marxianus*	3	0.70
DD_14	*Colletotrichum higginsianum*	2	0.47
DD_14	*Synechococcus sp. CC9902*	2	0.47
DD_14	*Aquiflexum balticum*	2	0.47
DD_14	*Bacillus cereus*	2	0.47
DD_14	*Fusarium venenatum*	2	0.47
DD_14	*Aspergillus oryzae*	2	0.47
DD_14	*Bradyrhizobium erythrophlei*	2	0.47
DD_14	*Maribacter sp. HTCC2170*	2	0.47
DD_14	other	57	13.72
DD_16	*Klebsiella pneumoniae*	594	68.59

DD_16	*Ralstonia pickettii*	38	4.39
DD_16	*Staphylococcus aureus*	24	2.77
DD_16	*Bacillus cereus*	21	2.42
DD_16	*Streptococcus oralis*	20	2.31
DD_16	*Pasteurella multocida*	16	1.85
DD_16	*Ralstonia solanacearum*	15	1.73
DD_16	*Streptococcus salivarius*	12	1.39
DD_16	*Enterobacter cloacae*	9	1.04
DD_16	*Neisseria subflava*	9	1.04
DD_16	*Ralstonia mannitolilytica*	8	0.92
DD_16	*Ralstonia insidiosa*	6	0.69
DD_16	*Streptococcus mitis*	5	0.58
DD_16	*Malassezia restricta*	5	0.58
DD_16	*Streptococcus gordonii*	4	0.46
DD_16	*Streptococcus sp. oral taxon 064*	4	0.46
DD_16	*Neisseria gonorrhoeae*	4	0.46
DD_16	*Haemophilus parainfluenzae*	4	0.46
DD_16	*Corynebacterium matruchotii*	3	0.35
DD_16	*Klebsiella aerogenes*	3	0.35
DD_16	*Neisseria elongata*	2	0.23
DD_16	*Neisseria sicca*	2	0.23
DD_16	*Veillonella parvula*	2	0.23
DD_16	*Veillonella dispar*	2	0.23
DD_16	*Streptococcus equinus*	2	0.23
DD_16	*Escherichia coli*	2	0.23

DD_16	other	50	6.00
DH_7	*Staphylococcus capitis*	2100	53.86
DH_7	*Malassezia restricta*	447	11.46
DH_7	*Neisseria subflava*	189	4.85
DH_7	*Corynebacterium matruchotii*	185	4.74
DH_7	*Cutibacterium acnes*	111	2.85
DH_7	*Staphylococcus epidermidis*	85	2.18
DH_7	*Actinomyces oris*	59	1.51
DH_7	*Staphylococcus aureus*	38	0.97
DH_7	*Neisseria elongata*	33	0.85
DH_7	*Veillonella parvula*	30	0.77
DH_7	*Haemophilus parainfluenzae*	30	0.77
DH_7	*Neisseria meningitidis*	29	0.74
DH_7	*Klebsiella pneumoniae*	23	0.59
DH_7	*Neisseria flavescens*	20	0.51
DH_7	*Streptococcus sanguinis*	19	0.49
DH_7	*Propionibacterium sp. oral taxon 193*	16	0.41
DH_7	*Rothia dentocariosa*	16	0.41
DH_7	*Ralstonia pickettii*	14	0.36
DH_7	*Pasteurella multocida*	14	0.36
DH_7	*Actinomyces viscosus*	13	0.33
DH_7	*Lautropia mirabilis*	13	0.33
DH_7	*Streptococcus oralis*	12	0.31
DH_7	*Ralstonia solanacearum*	12	0.31
DH_7	*Neisseria sicca*	11	0.28

DH_7	*Ustilago maydis*	10	0.26
DH_7	*Neisseria mucosa*	10	0.26
DH_7	other	360	9.49
DH_9	*Malassezia restricta*	1003	36.39
DH_9	*Lawsonella clevelandensis*	415	15.06
DH_9	*Cutibacterium acnes*	408	14.80
DH_9	*Staphylococcus capitis*	185	6.71
DH_9	*Staphylococcus epidermidis*	155	5.62
DH_9	*Staphylococcus aureus*	63	2.29
DH_9	*Corynebacterium kroppenstedtii*	33	1.20
DH_9	*Klebsiella pneumoniae*	33	1.20
DH_9	*Pasteurella multocida*	27	0.98
DH_9	*Corynebacterium glutamicum*	20	0.73
DH_9	*Propionibacterium sp. oral taxon 193*	15	0.54
DH_9	*Ralstonia solanacearum*	12	0.44
DH_9	*Veillonella parvula*	10	0.36
DH_9	*Enterobacter cloacae*	9	0.33
DH_9	*Rhodococcus erythropolis*	6	0.22
DH_9	*Mycobacterium tuberculosis*	6	0.22
DH_9	*Corynebacterium diphtheriae*	6	0.22
DH_9	*Corynebacterium aquilae*	6	0.22
DH_9	*Cutibacterium avidum*	6	0.22
DH_9	*Botrytis cinerea*	5	0.18
DH_9	*Paracoccus mutanolyticus*	5	0.18
DH_9	*Rothia dentocariosa*	5	0.18

DH_9	*Staphylococcus warneri*	5	0.18
DH_9	*Corynebacterium cystitidis*	4	0.15
DH_9	*Corynebacterium jeikeium*	4	0.15
DH_9	*Mycobacterium sp. EPa45*	4	0.15
DH_9	other	306	11.25
DH_11	*Malassezia restricta*	6014	65.67
DH_11	*Staphylococcus capitis*	1855	20.26
DH_11	*Ralstonia pickettii*	215	2.35
DH_11	*Staphylococcus epidermidis*	59	0.64
DH_11	*Ustilago maydis*	51	0.56
DH_11	*Cutibacterium acnes*	45	0.49
DH_11	*Staphylococcus aureus*	27	0.29
DH_11	*Ralstonia mannitolilytica*	21	0.23
DH_11	*Neurospora crassa*	16	0.17
DH_11	*Eremothecium gossypii*	14	0.15
DH_11	*Pochonia chlamydosporia*	13	0.14
DH_11	*Klebsiella pneumoniae*	12	0.13
DH_11	*Colletotrichum higginsianum*	11	0.12
DH_11	*Yarrowia lipolytica*	10	0.11
DH_11	*Cryptococcus neoformans*	9	0.10
DH_11	*Aspergillus fumigatus*	9	0.10
DH_11	*Cercospora beticola*	9	0.10
DH_11	*Naumovozyma dairenensis*	8	0.09
DH_11	*Zygosaccharomyces rouxii*	8	0.09
DH_11	*Fusarium verticillioides*	8	0.09

DH_11	*Aspergillus oryzae*	8	0.09
DH_11	*Fusarium fujikuroi*	8	0.09
DH_11	*Ralstonia insidiosa*	7	0.08
DH_11	*Pasteurella multocida*	7	0.08
DH_11	*Staphylococcus caprae*	7	0.08
DH_11	*Zymoseptoria tritici*	7	0.08
DH_11	other	700	7.72
DH_13	*Cutibacterium acnes*	2930	37.67
DH_13	*Malassezia restricta*	2566	32.99
DH_13	*Staphylococcus capitis*	1217	15.64
DH_13	*Lawsonella clevelandensis*	332	4.27
DH_13	*Ralstonia pickettii*	67	0.86
DH_13	*Propionibacterium sp. oral taxon 193*	44	0.57
DH_13	*Staphylococcus epidermidis*	30	0.39
DH_13	*Staphylococcus aureus*	22	0.28
DH_13	*Corynebacterium glutamicum*	18	0.23
DH_13	*Klebsiella pneumoniae*	16	0.21
DH_13	*Ralstonia mannitolilytica*	14	0.18
DH_13	*Ustilago maydis*	14	0.18
DH_13	*Pasteurella multocida*	12	0.15
DH_13	*Cutibacterium avidum*	12	0.15
DH_13	*Staphylococcus caprae*	7	0.09
DH_13	*Colletotrichum higginsianum*	7	0.09
DH_13	*Ralstonia solanacearum*	6	0.08
DH_13	*Corynebacterium diphtheriae*	5	0.06

DH_13	*Thielavia terrestris*	5	0.06
DH_13	*Corynebacterium ulcerans*	4	0.05
DH_13	*Enterobacter cloacae*	4	0.05
DH_13	*Rhodococcus erythropolis*	4	0.05
DH_13	*Pochonia chlamydosporia*	4	0.05
DH_13	*Streptomyces hygroscopicus*	4	0.05
DH_13	*Novibacillus thermophilus*	4	0.05
DH_13	*Pseudomonas fluorescens*	3	0.04
DH_13	other	428	5.54
DH_15	*Malassezia restricta*	83	32.81
DH_15	*Pasteurella multocida*	30	11.86
DH_15	*Klebsiella pneumoniae*	26	10.28
DH_15	*Staphylococcus aureus*	26	10.28
DH_15	*Ralstonia solanacearum*	14	5.53
DH_15	*Enterobacter cloacae*	8	3.16
DH_15	*Botrytis cinerea*	7	2.77
DH_15	*Neisseria gonorrhoeae*	4	1.58
DH_15	*Kocuria rhizophila*	3	1.19
DH_15	*Legionella israelensis*	2	0.79
DH_15	*Kluyveromyces marxianus*	2	0.79
DH_15	*Aspergillus oryzae*	2	0.79
DH_15	*BeAn 58058 virus*	2	0.79
DH_15	*Burkholderia pseudomallei*	2	0.79
DH_15	*Thermothelomyces thermophilus*	2	0.79
DH_15	*Fabibacter pacificus*	1	0.40

DH_15	*Blattabacterium cuenoti*	1	0.40
DH_15	*Neisseria meningitidis*	1	0.40
DH_15	*Microcoleus sp. PCC 7113*	1	0.40
DH_15	*Fusarium oxysporum*	1	0.40
DH_15	*Halomonas sp. JS92-SW72*	1	0.40
DH_15	*Paenibacillus sp. JDR-2*	1	0.40
DH_15	*Candida orthopsilosis*	1	0.40
DH_15	*Naumovozyma castellii*	1	0.40
DH_15	*Aquiflexum balticum*	1	0.40
DH_15	*Ustilago maydis*	1	0.40
DH_15	other	29	11.86
DH_17	*Bacillus cereus*	3325	14.09
DH_17	*Streptococcus salivarius*	2533	10.74
DH_17	*Staphylococcus aureus*	2527	10.71
DH_17	*Streptococcus oralis*	2139	9.07
DH_17	*Corynebacterium matruchotii*	1660	7.04
DH_17	*Actinomyces oris*	1249	5.29
DH_17	*Rothia dentocariosa*	1117	4.73
DH_17	*Bacillus thuringiensis*	896	3.80
DH_17	*Neisseria subflava*	818	3.47
DH_17	*Moraxella osloensis*	779	3.30
DH_17	*Bacillus anthracis*	743	3.15
DH_17	*Neisseria elongata*	438	1.86
DH_17	*Lactobacillus fermentum*	353	1.50
DH_17	*Veillonella parvula*	313	1.33

DH_17	*Neisseria sicca*	270	1.14
DH_17	*Neisseria mucosa*	243	1.03
DH_17	*Actinomyces viscosus*	219	0.93
DH_17	*Lautropia mirabilis*	215	0.91
DH_17	*Streptococcus mitis*	208	0.88
DH_17	*Klebsiella pneumoniae*	176	0.75
DH_17	*Neisseria meningitidis*	174	0.74
DH_17	*Veillonella dispar*	164	0.70
DH_17	*Streptococcus sp. FDAARGOS_192*	147	0.62
DH_17	*Streptococcus sanguinis*	137	0.58
DH_17	*Bacillus sp. (in: Bacteria)*	128	0.54
DH_17	*Rothia mucilaginosa*	123	0.52
DH_17	other	2501	11.12

Table S 3 Taxonomic composition of the scalp microbiome for each scalp state. The relative abundance and number of reads assigned of the top 25 most abundant taxa are shown for each scalp state. Taxonomic profiles were generated using Kraken v2.0 and relative abundance estimations were generated using Bracken.

Sample	Organism name	Number of assigned reads	Relative abundance (%)
Dandruff	*Malassezia restricta*	8951	74.97
Dandruff	*Klebsiella pneumoniae*	782	6.55
Dandruff	*Ralstonia pickettii*	441	3.69
Dandruff	*Cutibacterium acnes*	301	2.52
Dandruff	*Botrytis cinerea*	158	1.32
Dandruff	*Staphylococcus aureus*	152	1.27
Dandruff	*Pasteurella multocida*	126	1.06
Dandruff	*Enterobacter cloacae*	94	0.79
Dandruff	*Ralstonia mannitolilytica*	54	0.45
Dandruff	*Staphylococcus capitis*	52	0.44
Dandruff	*Staphylococcus epidermidis*	52	0.44
Dandruff	*Burkholderia pseudomallei*	49	0.41
Dandruff	*Ralstonia solanacearum*	48	0.40
Dandruff	*Candida dubliniensis*	27	0.23
Dandruff	*Ralstonia insidiosa*	26	0.22
Dandruff	*BeAn 58058 virus*	23	0.19
Dandruff	*Neisseria subflava*	22	0.18
Dandruff	*Propionibacterium sp. oral taxon 193*	21	0.18
Dandruff	*Streptococcus oralis*	20	0.17
Dandruff	*Bacillus cereus*	17	0.14
Dandruff	*Lactococcus lactis*	17	0.14

Dandruff	*Cutibacterium avidum*	13	0.11
Dandruff	*Klebsiella aerogenes*	12	0.10
Dandruff	*Corynebacterium matruchotii*	12	0.10
Dandruff	*Human endogenous retrovirus K*	12	0.10
Dandruff	other	457	3.83
Dandruff_Healthy	*Malassezia restricta*	11263	23.48
Dandruff_Healthy	*Cutibacterium acnes*	3341	6.97
Dandruff_Healthy	*Bacillus cereus*	2985	6.22
Dandruff_Healthy	*Corynebacterium matruchotii*	2793	5.82
Dandruff_Healthy	*Staphylococcus aureus*	2741	5.71
Dandruff_Healthy	*Streptococcus salivarius*	2501	5.21
Dandruff_Healthy	*Streptococcus oralis*	2199	4.58
Dandruff_Healthy	*Actinomyces oris*	1876	3.91
Dandruff_Healthy	*Ralstonia pickettii*	1620	3.38
Dandruff_Healthy	*Rothia dentocariosa*	1587	3.31
Dandruff_Healthy	*Neisseria subflava*	1430	2.98
Dandruff_Healthy	*Bacillus thuringiensis*	1225	2.55
Dandruff_Healthy	*Veillonella parvula*	852	1.78
Dandruff_Healthy	*Moraxella osloensis*	818	1.71
Dandruff_Healthy	*Lautropia mirabilis*	651	1.36
Dandruff_Healthy	*Neisseria elongata*	576	1.20
Dandruff_Healthy	*Bacillus anthracis*	549	1.14
Dandruff_Healthy	*Haemophilus parainfluenzae*	403	0.84
Dandruff_Healthy	*Neisseria sicca*	402	0.84
Dandruff_Healthy	*Lactobacillus fermentum*	351	0.73

Dandruff_Healthy	*Neisseria mucosa*	349	0.73
Dandruff_Healthy	*Actinomyces viscosus*	346	0.72
Dandruff_Healthy	*Klebsiella pneumoniae*	291	0.61
Dandruff_Healthy	*Streptococcus sanguinis*	288	0.60
Dandruff_Healthy	*Streptococcus mitis*	267	0.56
Dandruff_Healthy	other	6260	13.05
Healthy	*Cutibacterium acnes*	5368	15.22
Healthy	*Malassezia restricta*	4164	11.80
Healthy	*Ralstonia pickettii*	3438	9.74
Healthy	*Enterococcus hirae*	2854	8.09
Healthy	*Corynebacterium segmentosum*	2334	6.62
Healthy	*Cutibacterium granulosum*	1936	5.49
Healthy	*Haemophilus parainfluenzae*	1529	4.33
Healthy	*Veillonella parvula*	1397	3.96
Healthy	*Corynebacterium kroppenstedtii*	1391	3.94
Healthy	*Staphylococcus epidermidis*	1348	3.82
Healthy	*Enterococcus faecium*	1187	3.36
Healthy	*Actinomyces oris*	1011	2.87
Healthy	*Rothia dentocariosa*	481	1.36
Healthy	*Acinetobacter lwoffii*	376	1.07
Healthy	*Streptococcus sanguinis*	344	0.98
Healthy	*Streptococcus mitis*	336	0.95
Healthy	*Acinetobacter johnsonii*	270	0.77
Healthy	*Lautropia mirabilis*	213	0.60
Healthy	*Streptococcus oralis*	206	0.58

Healthy	*Ralstonia insidiosa*	184	0.52
Healthy	*Propionibacterium sp. oral taxon 193*	174	0.49
Healthy	*Ralstonia mannitolilytica*	167	0.47
Healthy	*Corynebacterium matruchotii*	161	0.46
Healthy	*Staphylococcus capitis*	155	0.44
Healthy	*Cutibacterium avidum*	152	0.43
Healthy	other	4104	11.63

Table S 4 Comparison of metagenomic assembler performance on each scalp sample. Comparison of assembly quality, including N50 and Total number of contigs for each metagenomic sample using three different metagenomic assemblers: IDBA-UD, MEGAHIT and metaSPAdes.

Assembler	Sample	Number of Contigs	N50
IDBA_UD	HH_1	8760	17215
	HH_2	14111	9065
	HH_3	4098	114240
	HH_4	17138	6748
	HH_5	7019	14958
	HH_18	13092	3409
	HH_19	15391	3351
	HH_20	7123	49734
	HH_21	9366	34576
	HH_22	10357	11609
	DH_7	7735	85317
	DH_9	6519	42797
	DH_11	14724	1752
	DH_13	8682	8675
	DH_15	1890	735
	DH_17	22680	2375
	DD_6	7231	1370
	DD_8	5459	93310
	DD_10	14254	1436
	DD_12	4867	5755
	DD_14	3535	759
	DD_16	2138	664
MEGAHIT	HH_1	9472	13806
	HH_2	15519	6857
	HH_3	4187	102042
	HH_4	17738	4714
	HH_5	7158	11450
	HH_18	14486	2793
	HH_19	14511	3446
	HH_20	7425	45244
	HH_21	10429	30740
	HH_22	11679	9004
	DH_7	7054	66988
	DH_9	4585	50643
	DH_11	15451	1609
	DH_13	8659	7483

	DH_15	1140	615
	DH_17	24955	1959
	DD_6	4823	1418
	DD_8	3718	101330
	DD_10	11346	1533
	DD_12	4767	5239
	DD_14	2158	647
	DD_16	1540	615
metaSPAdes	HH_1	31553	16868
	HH_2	51429	7532
	HH_3	16160	98326
	HH_4	43205	5559
	HH_5	28189	11464
	HH_18	29196	3594
	HH_19	39937	3657
	HH_20	23383	78229
	HH_21	19155	43365
	HH_22	28159	11296
	DH_7	42682	91286
	DH_9	43292	54814
	DH_11	32567	1574
	DH_13	21446	10805
	DH_15	29472	715
	DH_17	78566	2131
	DD_6	53168	1344
	DD_8	46825	127813
	DD_10	64771	1433
	DD_12	25308	5846
	DD_14	44278	747
	DD_16	29812	672

Table S 5 Comparison of metagenomic assembler performance on clustered scalp samples. Comparison of assembly quality, including N50 and Total number of contigs of metagenomic samples assembled via scalp sample state using three different metagenomic assemblers: IDBA-UD, MEGAHIT and metaSPAdes.

Assembler	Sample	Number of Contigs	N50
IDBA_UD	Healthy	54043	19465
	Dandruff Healthy	62818	3363
	Dandruff	27054	6355
MEGAHIT	Healthy	44121	16380
	Dandruff Healthy	59622	3758
	Dandruff	20608	6167
metaSPAdes	Healthy	115199	29653
	Dandruff Healthy	138389	4052
	Dandruff	103637	14291

Table S 6 Assembly statistics for *S. capitis* genomes included in this study.
Assembly statistics including number of contigs, total length of assembly, length of the longest contig, N50 and GC content for each *S. capitis* genome included in this study assembled with Unicycler and evaluated with QUAST.

Sample	Number of Contigs	Total Length (bp)	Longest Contig (bp)	N50	GC (%)
AYP1020	2	2503265	2443604	2443604	32.93
C87	14	2473608	978693	884876	32.79
Sample 104	70	2352230	158673	85581	33.08
Sample 105	71	2405657	271368	140477	32.97
Sample 106	70	2450644	263145	92822	33.01
Sample 107	83	2466630	397901	147089	32.78
Sample 108	81	2432997	316452	99980	32.82
Sample 12	47	2423168	974125	677991	32.75
Sample 13	45	2423012	974136	677816	32.75
Sample 15	63	2469398	871219	553046	32.72
Sample 17	44	2396194	931421	847807	32.75
Sample 18	77	2461735	693351	603983	32.7
Sample 19	94	2452999	919070	691172	32.72
Sample 1	60	2435799	931897	461869	32.77
Sample 20	60	2370230	930015	695482	32.82
Sample 21	61	2366226	930456	849981	32.78
Sample 22	57	2377086	929882	865793	32.8
Sample 24	65	2365900	930456	849981	32.78
Sample 25	57	2362657	930450	849975	32.78
Sample 26	57	2357323	930457	460919	32.79
Sample 27	67	2415655	930456	849983	32.75
Sample 28	71	2458319	976578	690689	32.73
Sample 29	52	2458353	975791	850158	32.73
Sample 2	59	2407988	931673	849363	32.72
Sample 30	65	2456533	976579	850285	32.73
Sample 32	74	2457797	976579	850282	32.73
Sample 33	67	2459275	975789	850192	32.73
Sample 34	61	2458048	975765	850272	32.73
Sample 35	73	2457100	976624	850290	32.73
Sample 36	55	2409237	932036	865908	32.74
Sample 37	41	2417692	932036	865908	32.72
Sample 38	49	2416191	932035	865908	32.72
Sample 39	48	2411266	932035	865908	32.73
Sample 3	60	2428579	931896	456064	32.78
Sample 40	53	2457378	932061	865908	32.71
Sample 41	47	2410438	932024	865898	32.73

Sample 42	58	2409236	932036	866178	32.73
Sample 43	48	2409381	932034	865908	32.73
Sample 45	52	2416812	931810	683473	32.68
Sample 47	53	2399989	937183	483584	32.77
Sample 49	72	2405759	931658	683453	32.71
Sample 4	62	2400021	931807	849365	32.74
Sample 50	64	2414604	931917	682642	32.68
Sample 52	45	2415311	931810	683472	32.69
Sample 54	38	2378796	932120	850519	32.77
Sample 55	42	2386840	932187	850519	32.76
Sample 56	45	2388770	932122	850519	32.75
Sample 57	32	2377443	932118	850520	32.78
Sample 58	52	2450152	966805	697785	32.77
Sample 60	33	2379694	932121	850522	32.77
Sample 62	88	2472679	1291080	1291080	32.82
Sample 63	56	2379458	930745	849609	32.74
Sample 65	32	2418196	1245143	1245143	32.8
Sample 66	32	2416528	858631	545288	32.86
Sample 75	46	2349453	930303	849604	32.79
Sample 78	53	2351923	930295	849601	32.78
Sample 79	38	2382817	932221	850455	32.75
Sample 7	70	2381216	929650	849547	32.72
Sample 80	41	2382897	932022	850453	32.76
Sample 81	40	2389037	932021	850453	32.74
Sample 82	40	2383564	932221	850455	32.75
Sample 83	45	2437284	932093	850490	32.77
Sample 85	39	2440108	932092	850488	32.77
Sample 86	37	2439841	932092	850488	32.77
Sample 87	50	2465109	838632	564748	32.88
Sample 88	99	2496011	727687	550107	32.65
Sample 9	75	2492262	556316	395718	32.73
CR02	320	2343249	49691	11854	33.05
CR03	1	2508352	2508352	2508352	32.81
CR05	39	2538782	388289	133499	32.84
CR09	34	2487158	422891	132198	32.82
DMG1800560	53	2502509	336063	104072	32.87
DMG1800566	75	2522349	297344	89331	32.72
DMG1800569	39	2449952	253211	125113	32.89
DMG1800572	81	2472107	150822	57756	32.95
DMG1800574	74	2398139	223322	93955	32.84
DMG1800575	51	2341991	191535	99070	32.72
DMG1800577	47	2385983	272295	135591	32.7

DMG1800578	78	2368048	96976	55485	32.79
DMG1800582	99	2476795	130316	47437	32.88
DMG1800663	63	2486600	191275	59932	32.92
DSM6717	62	2453603	833269	564644	32.87
FDAARGOS_378	2	2485152	2419686	2419686	32.99
LNZR-1	90	2595865	319806	66677	32.67
NCTC11045	70	2418673	782305	680723	32.65
QNI	30	2430101	696440	646907	32.76
SK14	32	2435835	438579	178042	32.77
TW2795	2	2486776	2479312	2479312	33.05

Table S 7 Differential gene clusters found in *S. capitis* subspecies. Gene clusters found in more than half of *S. capitis* ssp. *capitis* or ssp. *ureolyticus* total genomes included in this study. Gene clusters, presence and absence, and functional descriptions were obtained from Panaroo pangenome analysis of assembled genomes.

Gene cluster	Function description	Number of *S. capitis* ssp. *ureolyticus* genomes (total 26)	Number of *S. capitis* ssp. *capitis* genomes (total 58)
group_1288	Hypothetical protein/ putative perfringolysin O regulator protein	21	9
opp1B	Peptide ABC transporter, permease protein 1B	21	16
opp1D	Peptide ABC transporter, permease protein 1D	21	16
group_1627	Staphylopine biosynthesis dehydrogenase	21	16
opp1A	Peptide ABC transporter, permease protein 1A	21	16
opp1C	Peptide ABC transporter, permease protein 1C	21	16
opp1F	Peptide ABC transporter, permease protein 1F	21	16
group_675	MFS transporter	21	16
group_2777	phage leukocidin protein	19	0
group_2271	ATP-binding protein	19	0
group_2005	Hypothetical protein	19	0
group_2740	Hypothetical protein	19	1
group_1402	Beta-class phenol-soluble modulin	19	1

group_1708	Penicillin-hydrolyzing class A beta-lactamse BlaZ	19	1
group_2733	Hypothetical protein	18	0
group_2536	Transcriptional activator rinB	18	2
group_266	Penicillinase respressor BlaI	18	1
group_2205	Hypothetical protein	18	3
ydhK	Arsenic resistance	18	6
group_2598	DUF2829 domain-containing protein (unknown function)	18	11
group_2323	Tail protein	18	14
group_2251	Phage major tail protein, TP901-1 family	18	14
group_2249	Hypothetical protein	18	14
group_2245	AtlE / CHAP domain-containing protein	18	14
group_1972	Tail assembly protein	18	14
group_2739	DUF1064 domain-containing protein (unknown function)	17	2
blaR1_2	Beta-lactam sensor-transducer	17	1
copA_2_copB	Copper transportation	17	7
group_2390	Hypothetical protein	17	14
group_2360	Hypothetical protein	17	14
group_2322	BppU family phage baseplate upper protein	17	14

group_2169	DUF2817 domain-containing protein (unknown function)	17	14
group_2746	Hypothetical protein	16	0
group_2342	Hypothetical protein	16	1
group_2129	Hypothetical protein/phage protein	16	1
group_1679	Dipeptide ABC transporter glycylmethionine-binding lipoprotein	16	1
group_1569	ABC transporter permease	16	1
group_1244	ABC transporter permease	16	1
group_1170	Dipeptide ABC transporter ATP-binding protein	16	1
group_1141	Dipeptide ABC transporter ATP-binding protein	16	1
group_1117	ABC transporter permease	16	1
metN	Putative ATPase, ABC type permease superfamily, uptake of both D- and L-methionine	16	1
group_734	Hypothetical protein	16	1
group_202	ABC transporter subsrtate-binding protein	16	1
group_2704	Recombinase family protein	16	1
group_2321	Hypothetical protein/ phage protein	16	14
group_2170	Phage tail family protein	16	14

group_2328	Phage holin	16	16
group_2768	Transposase (IS431mec/IS6)	16	18
group_2392	XkdX family protein	16	19
group_2029	Beta-class phenol-soluble modulin	15	0
group_2359	BppU family phage baseplate upper protein	15	3
group_2793	DUF1643 domain-containing protein (unknown function)	15	7
group_2792	DUF960 domain-containing protein (unknown function)	15	7
group_2791	Hypothetical protein	15	7
ccrB__ccrB_1	Excision and integration of mec	15	7
group_2765	Hypothetical protein	15	7
group_2764	Recombinase family protein	15	7
group_2405	Hypothetical protein	15	13
group_2622	DUF2483 family protein (only member of the superfamily cl11261 This is a family of proteins found in bacteriophage particularly of the SA bacteriophages 11, Mu50B, family, homologous to phi-ETA orf16.)	14	0
group_1691	Restriction endonucelase subunit S	14	0
group_2795	Hypothetical protein	14	1
group_2182	CHAP domain-containing protein	14	3

group_2502	Hypothetical protein	14	6
group_1974	Hypothetical protein/ transcriptional regulator	14	10
group_844	Unknown function	14	8
group_2721	IS6-like element IS257 family transposase	14	16
group_2411	Hypothetical protein	13	0
group_2259	Phage head-tail connector protein	13	0
group_2178	Phage portal protein	13	0
group_1979	Hypothetical protein	13	0
group_1760	Minor capsid protein	13	0
group_1645	Hypothetical protein	13	0
arsR_arsR_4	Arsenic/antimonite resistance regulatory protein	13	0
group_1455	Bile acid: sodium symporter family protein	13	0
group_2698	DUF536 domain-containing protein (unknown function)	13	2
group_2341	DUF4355 domain-containing protein (unknown function)	13	14
group_1577	DUF3139 domain-containing protein (unknown function)	6	55
group_1689	Hypothetical protein	7	55
group_1364	DUF3139 domain-containing protein (unknown function)	7	55
group_999	Putative transposase	7	55

group_251	DUF3139 domain-containing protein (unknown function)	7	55
group_991	Uncharacterised protein	8	55
oppF_opp3F	Oligopeptide ABC transporter, ATP-binding protein	10	55
oppD	Oligopeptide ABC transporter, ATP-binding protein	10	55
oppB	Oligopeptide transport system permease protein	10	55
group_1828	Dipeptide-binding protein DppE precursor	10	55
group_1789	DUF3899 domain-containing protein (Putative Tryptophanyl-tRNA synthetase.)	10	55
oppC	Oligopeptide transport system permease protein	10	55
group_951	Glutamine amidotransferease	10	55
yjdF	Unknown orphan riboswitch candidate	10	55
yflN	Putative metallo-hydrolase	11	55
group_1918	cCarbon-nitrogen family hydrolase	11	55
group_1155	Hypothetical protein	11	55
fabH_2	3-oxoacyl-[acyl-carrier-protein] synthase 3	11	55
group_421	Helix-turn-helix transcriptional regulator	11	55
azoR_2	FMN-dependent NADH-azoreductase	13	55
group_1595	ArsB_1, arsenical pump membrane protein	0	54

group_1489	Hypothetical protein	0	54
group_685	Hypothetical protein	0	54
stbD		0	54
group_581	Hypothetical protein/ NAD(P)H-binding protein	0	54
argR_1	Transcriptional regulator (linked to arc operon)	2	54
group_1132	Hypothetical protein	0	53
group_2387	Hypothetical protein	12	53
group_1294	Hypothetical protein	0	52
group_3046	Restriction endonuclease subunit S	0	51
group_1467	Recombinase family protein	0	51
group_668	Truncated transposase for IS1272	0	51
group_377	Arsenical resistance operon trans-acting repressor ArsD	0	51
group_193	Arsenical pump-driving ATPase	0	51
cadD	Cadmium resistance	10	50
group_1256	Helix-turn-helix transcriptional regulator	10	50
group_2624	Accessory Sec system glycosylation chaperone GtfB	11	50
group_2601	Accessory Sec system glycosylation chaperone GtfA	11	50

sdrC_pls	YSIRK-type signal domain/LPTXG anchor domain surface protein	10	47
group_1147	Chitosanase	0	48
group_533	Cellulase family glycosylhydrolase	0	48
group_982	HlyD family efflux transporter periplasmic adaptor subunit	2	48
mleS	NAD-dependent malic enzyme	11	46
group_759	LysR family transcriptional regulator	11	48
aes_3	Alpha/beta hydrolase, putative acetyl-hydrolase lipR precursor	11	48
cimH	Citrate transporter, 2-hydroxycarboxylate transporter family protein	11	48
group_2937	Recombinase family protein	0	43
group_1559	Permease	0	43
group_1400	Hypothetical protein	0	43
group_1276	AAA family ATPase	0	43
group_1032	Hypothetical protein	0	43

arsR_1	Regulatory protein of ars operon, arsenate resistance	0	43
group_357	Transposase for transposon Tn552. DDE-type integrase/transposase/recombinase	0	43
mleS_3_mleS	NAD-dependent malic enzyme	0	42
cimH_2_citS	2-hydroxycarboxylate transporter family protein	0	42
group_695	Toxin-antitoxin system, antitoxin component, PHD family	0	42
group_649	Txe/YoeB family addiction module toxin	0	42
group_1278	Hypothetical protein	7	42
group_1079	Efflux RND transporter periplasmic adaptor subunit	0	41
group_556	Cellulase family glycosylhydrolase	0	41
group_538	Recombinase family protein	0	39
group_1325	SEL1-like repeat protein	5	41
group_2287	ABC transporter permease subunit	0	38
group_1541	Hypothetical protein	0	38
drrA	Similarity to SaeR and bacillus PhoP and ResD	0	40
group_1275	Cellulase family glycosylhydrolase	0	40
group_1025	Chitosanase	0	40
group_573	Toxin-antitoxin system, toxin component, Txe/YoeB family	0	39

group_555	Hypothetical protein	1	39
group_2494	Beta-class phenol-soluble modulin	3	37
group_1654	Hypothetical protein	4	37
group_522	Hypothetical protein	0	38
group_2702	Recombinase	0	35
group_1676	Hypothetical protein	0	34
group_1140	Hypothetical protein	0	34
group_996	LysR family transcriptional regulator	0	34
group_2701	ABC transporter permease	2	32
group_2476	Alkylmercury lyase	2	32
group_2699	ABC transporter ATP-binding protein	1	30
group_3024	GNAT family N-acetyltransferase	0	30
group_2863	Signal peptidase I	0	29
group_2820	DUF334 domain-containing protein (unknown function)	0	29
group_1059	Hypothetical protein	0	29
sutR	sutR, sutR, HTH-type transcriptional regulator SutR	1	29

yycB_2	Putative transporter YycB	1	29
repN_1_repN	repN_1_repN, repN_1;repN, Replication initiation protein	2	29
group_2819	Relaxase/mobilization nuclease domain protein	3	27
group_2713	Plasmid mobilisation relaxosome protein MobC	3	27
group_2305	Hypothetical protein	0	28
repE	repE, repE, Replication initiation protein	0	28

Table S 8 Summary of the incomplete prophage present in the *S. capitis* isolates. Incomplete phage are those with a completeness score below 70, questionable phage are those with a completeness score of between 60-90.

Isolate	Genome position	Designation	Phage completeness	Possible phage
Sample 2	contig 1: 377554-390663	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1: 618900-630675	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2: 67181-73201	Incomplete	20	*Staphylococcus* phage StB27 (NC_019914)
Sample 4	contig 1: 377554-390663	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1: 619034-630809	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2: 776164-782184	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 7	contig 3: 152758-180360	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 17	contig 1: 377602-390713	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1: 618745-630520	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2: 774617-780635	Incomplete	20	*Staphylococcus* phage StB27 (NC_019914)

Sample 18	contig 2: 377590-390701	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 5: 67162-73180	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
	contig 6: 109451-121226	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
Sample 19	contig 1: 288436-300211	Incomplete	10	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1: 528376-541487	Incomplete	17	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 4: 93732-99750	Incomplete	13	*Staphylococcus* phage StB27 (NC_019914)
Sample 20	contig 1: 377601-390712	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1: 619011-630786	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 4:67174-73192	Incomplete	13	*Staphylococcus* phage StB27 (NC_019914)
	contig 4 :158007-166403	Incomplete	9	*Staphylococcus* phage SPbeta-like (NC_029119)
Sample 21	contig 1 :299659-311434	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :539710-552820	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)

	contig 2 :67180-73200	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 22	contig 1 :299230-311005	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :539171-552282	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 2 :693890-719867	Incomplete	40	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 2 :792603-798621	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 24	contig 1 :377636-390746	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :619022-630797	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :776782-782802	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 25	contig 1 :377633-390743	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :619019-630794	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :776778-782798	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 26	contig 1 :377637-390747	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)

	contig 1 :619023-630798	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 3 :67212-73232	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 27	contig 1 :299660-311435	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :539711-552821	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67181-73201	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 28	contig 5 :34506-51619	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 29	contig 4 :34505-51618	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 30	contig 7 :358-29310	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 32	contig 7 :34506-47560	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 33	contig 7 :358-29310	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 34	contig 7 :351-29303	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 35	contig 7 :34502-47556	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 36	contig 1 :377469-390579	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618565-630340	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67184-73204	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)

Sample 37	contig 1 :301697-313472	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :541458-554568	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67184-73204	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 38	contig 1 :377468-390578	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618564-630339	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67184-73204	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 39	contig 1 :377468-390578	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618564-630339	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67184-73204	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 40	contig 1 :301697-313472	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :541458-554568	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 2 :67184-73204	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)

Sample 41	contig 1 :301691-313466	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :541452-554562	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 2 :792700-798720	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 42	contig 1 :377469-390579	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618565-630340	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67184-73204	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 43	contig 1 :377468-390578	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618564-630339	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :792705-798725	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 45	contig 1 :377605-390715	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618683-630458	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 3 :67183-73201	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)

Sample 47	contig 1 :377554-390664	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618852-630627	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :924323-935726	Incomplete	30	Escherichia phage RCS47 (NC_042128)
	contig 2 :410365-416385	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 49	contig 1 :301344-313119	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :540953-554063	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 3 :103125-109143	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 50	contig 1 :377591-390701	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :618803-630578	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 3 :67171-73189	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 52	contig 1 :377605-390715	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618683-630458	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 3 :67184-73202	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)

Sample 54	contig 4 :82162-97347	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 55	contig 4 :82162-97347	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
	contig 4 :82162-97347	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 56	contig 5 :82162-97347	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 58	contig 4 :82163-97348	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 63	contig 1 :377265-390374	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618145-629920	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67180-73198	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
	contig 3 :2079-18796	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
Sample 65	contig 1 :280803-287518	Incomplete	10	PHAGE_Staphy_StauST398_2_NC_021323(1)
	contig 1 :591039-598028	Incomplete	20	Prochlorococcus phage P-SSM7 (NC_015290)
	contig 1 :830745-839995	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
Sample 75	contig 1 :377335-390444	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)

	contig 1 :618278-630053	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :67180-73198	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 78	contig 1 :300247-312022	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :539856-552965	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 2 :776406-782424	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 79	contig 1 :377629-390738	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618912-630687	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :777264-783282	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 80	contig 1 :377629-390738	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 1 :618777-630552	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 2 :777262-783280	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 81	contig 1 :301471-313246	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)

	contig 1 :541285-554394	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 2 :777262-783280	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 82	contig 1 :301536-313311	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 1 :541485-554594	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 2 :777264-783282	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
Sample 88	contig 3 :237489-260370	Incomplete	30	*Staphylococcus* phage PT1028 (NC_007045)
Sample 107	contig 7 :7453-30334	Incomplete	30	*Staphylococcus* phage PT1028 (NC_007045)
	contig 27 :614-15800	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
CR05	contig NZ_CTEO01000016.1 6353-19473	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
	contig NZ_CTEO01000037.1 1-21964	Incomplete	20	*Staphylococcus* phage StauST398-1 (NC_021326)
	contig 9 :61013-93299	Incomplete	40	*Staphylococcus* phage PT1028 (NC_007045)
	contig 29 :1607-12897	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
DMG1800574	contig 38 :488-13330	Incomplete	30	*Staphylococcus* phage Ipla7 (NC_018284)

	contig 40 :488-11375	Incomplete	10	*Staphylococcus* phage CNPH82 (NC_008722)
DMG1800575	contig 31 :1026-12642	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
DMG1800577	contig 5 :23013-34788	Incomplete	20	*Bacillus* phage vB_BanS_Tsamsa (NC_023007)
	contig 7 :77436-90546	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
	contig 9 :91042-97062	Incomplete	10	*Staphylococcus* phage StB27 (NC_019914)
DMG1800578	contig 5 :63959-74676	Incomplete	20	*Synechococcus* phage S-SSM7 (NC_015287)
	contig 12 :15918-29028	Incomplete	20	*Staphylococcus* phage SPbeta-like (NC_029119)
DSM6717	contig 3 :99-15259	Incomplete	30	*Staphlyococcus* phage CNPH82 (NC_008722)
	contig 7 :76560-85923	Incomplete	30	*Staphlyococcus* phage CNPH82 (NC_008722)
	contig 10 :28-18839	Questionable	80	*Staphylococcus* phage StB12 (NC_020490)
	contig 11 :104-15545	Questionable	80	*Staphylococcus* phage StB12 (NC_020490)
	contig NZ_JGYJ010000019.1:1317-50895	Incomplete	40	*Staphlyococcus* phage SPbeta-like (NC_029119)

	contig NZ_JGYJ01000052.1:42-13215	Incomplete	10	*Staphlyococcus* phage SPbeta-like (NC_029119)
	contig NZ_JGYJ01000056.1:4115-40619	Incomplete	40	*Staphlyococcus* phage SPbeta-like (NC_029119)
NCTC11045	contig 6 :79338-102653	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)
TW2795	NZ_AP014956.1:2388363-2408096	Incomplete	20	*Staphylococcus* phage PT1028 (NC_007045)

REFERENCES

ABUBUCKER, S., SEGATA, N., GOLL, J., SCHUBERT, A. M., IZARD, J., CANTAREL, B. L., RODRIGUEZ-MUELLER, B., ZUCKER, J., THIAGARAJAN, M., HENRISSAT, B., WHITE, O., KELLEY, S. T., METHÉ, B., SCHLOSS, P. D., GEVERS, D., MITREVA, M. & HUTTENHOWER, C. 2012. Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLOS Computational Biology,* 8**,** e1002358.

ACHTMAN, M. 2008. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. *Annual Review of Microbiology,* 62**,** 53-70.

ADALSTEINSSON, J. A., KAUSHIK, S., MUZUMDAR, S., GUTTMAN-YASSKY, E. & UNGAR, J. 2020. An update on the microbiology, immunology and genetics of seborrheic dermatitis. *Experimental Dermatology,* 29**,** 481-489.

AJAYI, C., ÅBERG, E., ASKARIAN, F., SOLLID, J. U. E., JOHANNESSEN, M. & HANSSEN, A.-M. 2018. Genetic variability in the sdrD gene in *Staphylococcus aureus* from healthy nasal carriers. *BMC Microbiology,* 18**,** 34.

AL HENNAWI, H. E. T., MAHDI, E. M. & MEMISH, Z. A. 2020. Native valve *Staphylococcus capitis* infective endocarditis: a mini review. *Infection,* 48**,** 3-5.

ALEKSEYENKO, A. V., PEREZ-PEREZ, G. I., DE SOUZA, A., STROBER, B., GAO, Z., BIHAN, M., LI, K., METHÉ, B. A. & BLASER, M. J. 2013. Community differentiation of the cutaneous microbiota in psoriasis. *Microbiome,* 1**,** 31-31.

ALLGAIER, H., JUNG, G., WERNER, R. G., SCHNEIDER, U. & ZAHNER, H. 1985. ELUCIDATION OF THE STRUCTURE OF EPIDERMIN, A RIBOSOMALLY SYNTHESIZED, TETRACYCLIC HETERODETIC POLYPEPTIDE ANTIBIOTIC. *Angewandte Chemie-International Edition,* 24**,** 1051-1053.

ALLIGNET, J., ENGLAND, P., OLD, I. & EL SOLH, N. 2002. Several regions of the repeat domain of the *Staphylococcus caprae* autolysin, AtlC, are involved in fibronectin binding. *FEMS microbiology letters,* 213**,** 193-197.

ALNEBERG, J., BJARNASON, B. S., DE BRUIJN, I., SCHIRMER, M., QUICK, J., IJAZ, U. Z., LAHTI, L., LOMAN, N. J., ANDERSSON, A. F. & QUINCE, C. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods,* 11**,** 1144-1146.

ALORABI, J. A. 2016. *Role of peptidoglycan deacetylase in Staphylococcus aureus virulence and survival.* PhD PhD, University of Liverpool.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ANDAM, C. P., CHALLAGUNDLA, L., AZARIAN, T., HANAGE, W. P. & ROBINSON, D. A. 2017. 3 - Population Structure of Pathogenic Bacteria. *In:* TIBAYRENC, M. (ed.) *Genetics and Evolution of Infectious Diseases (Second Edition).* London: Elsevier.

ANDREWS, S. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Online]. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed].

ARGEMI, X., HANSMANN, Y., PROLA, K. & PRÉVOST, G. 2019. Coagulase-Negative Staphylococci Pathogenomics. *International journal of molecular sciences,* 20**,** 1215.

ARGEMI, X., MATELSKA, D., GINALSKI, K., RIEGEL, P., HANSMANN, Y., BLOOM, J., PESTEL-CARON, M., DAHYOT, S., LEBEURRE, J. & PRÉVOST, G. 2018. Comparative genomic analysis of *Staphylococcus lugdunensis* shows a closed pan-genome and multiple barriers to horizontal gene transfer. *BMC genomics,* 19**,** 621-621.

ARNDT, D., GRANT, J. R., MARCU, A., SAJED, T., PON, A., LIANG, Y. & WISHART, D. S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res,* 44**,** W16-21.

ARORA, S., UHLEMANN, A.-C., LOWY, F. D. & HOOK, M. 2016. A Novel MSCRAMM Subfamily in Coagulase Negative Staphylococcal Species. *Frontiers in microbiology,* 7**,** 540-540.

ASKARIAN, F., AJAYI, C., HANSSEN, A. M., VAN SORGE, N. M., PETTERSEN, I., DIEP, D. B., SOLLID, J. U. & JOHANNESSEN, M. 2016. The interaction between *Staphylococcus aureus* SdrD and desmoglein 1 is important for adhesion to host cells. *Sci Rep,* 6**,** 22134.

AYLESWORTH, R. & VANCE, J. C. 1982. *Demodex folliculorum* and *Demodex brevis* in cutaneous biopsies. *Journal of the American Academy of Dermatology,* 7**,** 583-589.

BABA, T., KUWAHARA-ARAI, K., UCHIYAMA, I., TAKEUCHI, F., ITO, T. & HIRAMATSU, K. 2009. Complete Genome Sequence of *Macrococcus caseolyticus* Strain JSCS5402, Reflecting the Ancestral Genome of the Human-Pathogenic Staphylococci. *Journal of Bacteriology,* 191**,** 1180.

BANNERMAN, T. L. & KLOOS, W. E. 1991. *Staphylococcus capitis* subsp. *ureolyticus* subsp. nov. from human skin. *Int J Syst Bacteriol,* 41**,** 144-7.

BARBU, E. M., GANESH, V. K., GURUSIDDAPPA, S., MACKENZIE, R. C., FOSTER, T. J., SUDHOF, T. C. & HÖÖK, M. 2010. β-Neurexin Is a Ligand for the *Staphylococcus aureus* MSCRAMM SdrC. *PLOS Pathogens,* 6**,** e1000726.

BAYER, A. S., MISHRA, N. N., CHEN, L., KREISWIRTH, B. N., RUBIO, A. & YANG, S.-J. 2015. Frequency and Distribution of Single-Nucleotide Polymorphisms within mprF in Methicillin-Resistant *Staphylococcus aureus* Clinical Isolates and Their Role in Cross-Resistance to Daptomycin and Host Defense Antimicrobial Peptides. *Antimicrobial agents and chemotherapy,* 59**,** 4930-4937.

BEASLEY, F. C., VINÉS, E. D., GRIGG, J. C., ZHENG, Q., LIU, S., LAJOIE, G. A., MURPHY, M. E. P. & HEINRICHS, D. E. 2009. Characterization of staphyloferrin A biosynthetic and transport mutants in *Staphylococcus aureus. Molecular Microbiology,* 72**,** 947-963.

BECKER, K., HEILMANN, C. & PETERS, G. 2014. Coagulase-Negative Staphylococci. *Clinical Microbiology Reviews,* 27**,** 870.

BENACHOUR, A., LADJOUZI, R., LE JEUNE, A., HEBERT, L., THORPE, S., COURTIN, P., CHAPOT-CHARTIER, M.-P., PRAJSNAR, T. K., FOSTER, S. J. & MESNAGE, S. 2012. The Lysozyme-Induced Peptidoglycan N-Acetylglucosamine Deacetylase PgdA (EF1843) Is Required for *Enterococcus faecalis* Virulence. *Journal of Bacteriology,* 194**,** 6066-6073.

BENTLEY, S. D. & PARKHILL, J. 2015. Genomic perspectives on the evolution and spread of bacterial pathogens. *Proceedings. Biological sciences,* 282**,** 20150488-20150488.

BERA, A., BISWAS, R., HERBERT, S., KULAUZOVIC, E., WEIDENMAIER, C., PESCHEL, A. & GOETZ, F. 2007. Influence of wall teichoic acid on lysozyme resistance in *Staphylococcus aureus. Journal of Bacteriology,* 189**,** 280-283.

BERA, A., HERBERT, S., JAKOB, A., VOLLMER, W. & GOTZ, F. 2005. Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for lysozyme resistance of *Staphylococcus aureus. Molecular Microbiology,* 55**,** 778-787.

BERUBE, B. J. & WARDENBURG, J. B. 2013. *Staphylococcus aureus* α-toxin: nearly a century of intrigue. *Toxins,* 5**,** 1140-1166.

BHARTI, R. & GRIMM, D. G. 2021. Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics,* 22**,** 178-193.

BIEDERMANN, T., PONTIGGIA, L., BÖTTCHER-HABERZETH, S., THARAKAN, S., BRAZIULIS, E., SCHIESTL, C., MEULI, M. & REICHMANN, E. 2010. Human Eccrine Sweat Gland Cells Can Reconstitute a Stratified Epidermis. *Journal of Investigative Dermatology,* 130**,** 1996-2009.

BISWAS, R., VOGGU, L., SIMON, U. K., HENTSCHEL, P., THUMM, G. & GÖTZ, F. 2006. Activity of the major staphylococcal autolysin Atl. *FEMS microbiology letters,* 259**,** 260-268.

BLANPAIN, C. & FUCHS, E. 2009. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nat Rev Mol Cell Biol,* 10**,** 207-17.

BLICHARZ, L., RUDNICKA, L. & SAMOCHOCKI, Z. 2019. *Staphylococcus aureus*: an underestimated factor in the pathogenesis of atopic dermatitis? *Postepy dermatologii i alergologii,* 36**,** 11-17.

BLIN, K., SHAW, S., STEINKE, K., VILLEBRO, R., ZIEMERT, N., LEE, S. Y., MEDEMA, M. H. & WEBER, T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research,* 47**,** W81-W87.

BOEHNCKE, W.-H. & SCHÖN, M. P. 2015. Psoriasis. *The Lancet,* 386**,** 983-994.

BOGUNIEWICZ, M. & LEUNG, D. Y. 2011. Atopic dermatitis: a disease of altered skin barrier and immune dysregulation. *Immunological reviews,* 242**,** 233-246.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England),* 30**,** 2114-2120.

BONECA, I. G., DUSSURGET, O., CABANES, D., NAHORI, M.-A., SOUSA, S., LECUIT, M., PSYLINAKIS, E., BOURIOTIS, V., HUGOT, J.-P., GIOVANNINI, M., COYLE, A., BERTIN, J., NAMANE, A., ROUSSELLE, J.-C., CAYET, N., PREVOST, M.-C., BALLOY, V., CHIGNARD, M., PHILPOTTT, D. J., COSSART, P. & GIRARDIN, S. E. 2007. A critical role for peptidoglycan N-deacetylation in Listeria evasion from the host innate immune system. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 997-1002.

BOOTHE, D. W., TARBOX, J. A. & TARBOX, M. B. 2017. Atopic Dermatitis: Pathophysiology. *In:* FORTSON E., F. S., STROWD L. (ed.) *Management of Atopic Dermatitis.* Advances in Experimental Medicine and Biology: Springer, Cham.

BORREL, V., GANNESEN, A. V., BARREAU, M., GAVIARD, C., DUCLAIROIR-POC, C., HARDOUIN, J., KONTO-GHIORGHI, Y., LEFEUVRE, L. & FEUILLOLEY, M. G.

2019. Adaptation of acneic and non acneic strains of *Cutibacterium acnes* to sebum-like environment. *MicrobiologyOpen,* 8**,** e00841.

BOSE, J. L., FEY, P. D. & BAYLES, K. W. 2013. Genetic tools to enhance the study of gene function and regulation in *Staphylococcus aureus*. *Applied and environmental microbiology,* 79**,** 2218-2224.

BOUCHAMI, O., DE LENCASTRE, H. & MIRAGAIA, M. 2016. Impact of Insertion Sequences and Recombination on the Population Structure of *Staphylococcus haemolyticus*. *PLOS ONE,* 11**,** e0156653.

BOWERS, R. M., KYRPIDES, N. C., STEPANAUSKAS, R., HARMON-SMITH, M., DOUD, D., REDDY, T. B. K., SCHULZ, F., JARETT, J., RIVERS, A. R., ELOE-FADROSH, E. A., TRINGE, S. G., IVANOVA, N. N., COPELAND, A., CLUM, A., BECRAFT, E. D., MALMSTROM, R. R., BIRREN, B., PODAR, M., BORK, P., WEINSTOCK, G. M., GARRITY, G. M., DODSWORTH, J. A., YOOSEPH, S., SUTTON, G., GLÖCKNER, F. O., GILBERT, J. A., NELSON, W. C., HALLAM, S. J., JUNGBLUTH, S. P., ETTEMA, T. J. G., TIGHE, S., KONSTANTINIDIS, K. T., LIU, W.-T., BAKER, B. J., RATTEI, T., EISEN, J. A., HEDLUND, B., MCMAHON, K. D., FIERER, N., KNIGHT, R., FINN, R., COCHRANE, G., KARSCH-MIZRACHI, I., TYSON, G. W., RINKE, C., KYRPIDES, N. C., SCHRIML, L., GARRITY, G. M., HUGENHOLTZ, P., SUTTON, G., YILMAZ, P., MEYER, F., GLÖCKNER, F. O., GILBERT, J. A., KNIGHT, R., FINN, R., COCHRANE, G., KARSCH-MIZRACHI, I., LAPIDUS, A., MEYER, F., YILMAZ, P., PARKS, D. H., MURAT EREN, A., SCHRIML, L., BANFIELD, J. F., HUGENHOLTZ, P., WOYKE, T. & THE GENOME STANDARDS, C. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology,* 35**,** 725-731.

BOWMAN, L., ZEDEN, M. S., SCHUSTER, C. F., KAEVER, V. & GRÜNDLING, A. 2016. New Insights into the Cyclic Di-adenosine Monophosphate (c-di-AMP) Degradation Pathway and the Requirement of the Cyclic Dinucleotide for Acid Stress Resistance in *Staphylococcus aureus. J Biol Chem,* 291**,** 26970-26986.

BOYKIN, L. M., KUBATKO, L. S. & LOWREY, T. K. 2010. Comparison of methods for rooting phylogenetic trees: A case study using Orcuttieae (*Poaceae: Chloridoideae*). *Molecular phylogenetics and evolution,* 54**,** 687-700.

BRADY, A. & SALZBERG, S. L. 2009. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods,* 6**,** 673-676.

BRANDWEIN, M., STEINBERG, D. & MESHNER, S. 2016. Microbial biofilms and the human skin microbiome. *NPJ biofilms and microbiomes,* 2**,** 3-3.

BREITWIESER, F. P., LU, J. & SALZBERG, S. L. 2019. A review of methods and databases for metagenomic classification and assembly. *Briefings in bioinformatics,* 20**,** 1125-1136.

BRINKAC, L., CLARKE, T. H., SINGH, H., GRECO, C., GOMEZ, A., TORRALBA, M. G., FRANK, B. & NELSON, K. E. 2018. Spatial and Environmental Variation of the Human Hair Microbiota. *Scientific Reports,* 8**,** 9017.

BROTT, A. S. & CLARKE, A. J. 2019. Peptidoglycan O-Acetylation as a Virulence Factor: Its Effect on Lysozyme in the Innate Immune System. *Antibiotics,* 8**,** 94.

BRÜGGEMANN, H., HENNE, A., HOSTER, F., LIESEGANG, H., WIEZER, A., STRITTMATTER, A., HUJER, S., DÜRRE, P. & GOTTSCHALK, G. 2004. The

complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. *Science,* 305**,** 671-3.
BUFFOLI, B., RINALDI, F., LABANCA, M., SORBELLINI, E., TRINK, A., GUANZIROLI, E., REZZANI, R. & RODELLA, L. F. 2014. The human hair: from anatomy to physiology. *International Journal of Dermatology,* 53**,** 331-341.
BURKOVSKI, A. 2018. The role of corynomycolic acids in Corynebacterium-host interaction. *Antonie Van Leeuwenhoek,* 111**,** 717-725.
BUSHNELL, B. *BBtools* [Online]. Available: sourceforge.net/projects/bbmap/ [Accessed].
BYRD, A. L., BELKAID, Y. & SEGRE, J. A. 2018. The human skin microbiome. *Nature Reviews Microbiology,* 16**,** 143-155.
BYRD, A. L., DEMING, C., CASSIDY, S. K. B., HARRISON, O. J., NG, W.-I., CONLAN, S., BELKAID, Y., SEGRE, J. A. & KONG, H. H. 2017. Staphylococcus aureus and *Staphylococcus epidermidis* strain diversity underlying pediatric atopic dermatitis. *Science Translational Medicine,* 9**,** eaal4651.
CAMERON, D., JIANG, J.-H., HASSAN, K., ELBOURNE, L., TUCK, K., PAULSEN, I. & PELEG, A. 2015. Insights on virulence from the complete genome of Staphylococcus capitis. *Frontiers in Microbiology,* 6.
CANDI, E., SCHMIDT, R. & MELINO, G. 2005. The cornified envelope: a model of cell death in the skin. *Nature Reviews Molecular Cell Biology,* 6**,** 328-340.
CAPONE, K. A., DOWD, S. E., STAMATAS, G. N. & NIKOLOVSKI, J. 2011. Diversity of the human skin microbiome early in life. *J Invest Dermatol,* 131**,** 2026-32.
CARATTOLI, A., ZANKARI, E., GARCÍA-FERNÁNDEZ, A., VOLDBY LARSEN, M., LUND, O., VILLA, L., MØLLER AARESTRUP, F. & HASMAN, H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother,* 58**,** 3895-903.
CARTER, G. P., USSHER, J. E., DA SILVA, A. G., BAINES, S. L., HEFFERNAN, H., RILEY, T. V., BROADBENT, R., VAN DER LINDEN, A., LEE, J., MONK, I. R., STINEAR, T. P., HOWDEN, B. P. & WILLIAMSON, D. A. 2018. Genomic Analysis of Multiresistant *Staphylococcus capitis* Associated with Neonatal Sepsis. *Antimicrobial agents and chemotherapy,* 62**,** e00898-18.
CHAFFIN, D. O., TAYLOR, D., SKERRETT, S. J. & RUBENS, C. E. 2012. Changes in the *Staphylococcus aureus* transcriptome during early adaptation to the lung. *PloS one,* 7**,** e41329-e41329.
CHAMBERS, H. F. & DELEO, F. R. 2009. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nature Reviews Microbiology,* 7**,** 629-641.
CHERNOMOR, O., VON HAESELER, A. & MINH, B. Q. 2016. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology,* 65**,** 997-1008.
CHEUNG, G. Y. C., JOO, H.-S., CHATTERJEE, S. S. & OTTO, M. 2014. Phenol-soluble modulins – critical determinants of staphylococcal virulence. *FEMS Microbiology Reviews,* 38**,** 698-719.
CHO, S.-H., STRICKLAND, I., BOGUNIEWICZ, M. & LEUNG, D. Y. 2001. Fibronectin and fibrinogen contribute to the enhanced binding of *Staphylococcus aureus* to atopic skin. *Journal of Allergy and Clinical Immunology,* 108**,** 269-274.
CHOI, E.-J., LEE, H. G., BAE, I.-H., KIM, W., PARK, J., LEE, T. R. & CHO, E.-G. 2018. *Propionibacterium acnes*-derived extracellular vesicles promote acne-like

phenotypes in human epidermis. *Journal of Investigative Dermatology,* 138**,** 1371-1379.
CHONG, J., CAYA, C., LÉVESQUE, S. & QUACH, C. 2016. Heteroresistant Vancomycin Intermediate Coagulase Negative Staphylococcus in the NICU: A Systematic Review. *PLOS ONE,* 11**,** e0164136.
CHRISTNER, M., FRANKE, G. C., SCHOMMER, N. N., WENDT, U., WEGERT, K., PEHLE, P., KROLL, G., SCHULZE, C., BUCK, F. & MACK, D. 2010. The giant extracellular matrix-binding protein of *Staphylococcus epidermidis* mediates biofilm accumulation and attachment to fibronectin. *Molecular microbiology,* 75**,** 187-207.
CINGOLANI, P., PLATTS, A., WANG LE, L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin),* 6**,** 80-92.
CLARKE, S. R., MOHAMED, R., BIAN, L., ROUTH, A. F., KOKAI-KUN, J. F., MOND, J. J., TARKOWSKI, A. & FOSTER, S. J. 2007. The *Staphylococcus aureus* surface protein IsdA mediates resistance to innate defenses of human skin. *Cell host & microbe,* 1**,** 199-212.
CLAUDEL, J.-P., AUFFRET, N., LECCIA, M.-T., POLI, F., CORVEC, S. & DRÉNO, B. 2019. *Staphylococcus epidermidis*: A Potential New Player in the Physiopathology of Acne? *Dermatology,* 235**,** 287-294.
CLAVAUD, C., JOURDAIN, R., BAR-HEN, A., TICHIT, M., BOUCHIER, C., POURADIER, F., EL RAWADI, C., GUILLOT, J., MÉNARD-SZCZEBARA, F., BRETON, L., LATGÉ, J.-P. & MOUYNA, I. 2013. Dandruff is associated with disequilibrium in the proportion of the major bacterial and fungal populations colonizing the scalp. *PloS one,* 8**,** e58203-e58203.
CLAY, H. B., PARL, A. K., MITCHELL, S. L., SINGH, L., BELL, L. N. & MURDOCK, D. G. 2016. Altering the Mitochondrial Fatty Acid Synthesis (mtFASII) Pathway Modulates Cellular Metabolic States and Bioactive Lipid Profiles as Revealed by Metabolomic Profiling. *PLOS ONE,* 11**,** e0151171.
COATES, R., MORAN, J. & HORSBURGH, M. J. 2014. Staphylococci: colonizers and pathogens of human skin. *Future Microbiol,* 9**,** 75-91.
COGEN, A. L., YAMASAKI, K., SANCHEZ, K. M., DORSCHNER, R. A., LAI, Y., MACLEOD, D. T., TORPEY, J. W., OTTO, M., NIZET, V. & KIM, J. E. 2010. Selective antimicrobial action is provided by phenol-soluble modulins derived from *Staphylococcus epidermidis*, a normal resident of the skin. *Journal of Investigative Dermatology,* 130**,** 192-200.
COHAN, F. M. 2005. Periodic Selection and Ecological Diversity in Bacteria. *In:* NURMINSKY, D. (ed.) *Selective Sweep.* Boston, MA: Springer US.
COHAN, F. M. & PERRY, E. B. 2007. A Systematics for Discovering the Fundamental Units of Bacterial Diversity. *Current Biology,* 17**,** R373-R386.
COHEN, A. D., DREIHER, J. & BIRKENFELD, S. 2009. Psoriasis associated with ulcerative colitis and Crohn's disease. *J Eur Acad Dermatol Venereol,* 23**,** 561-5.
COLEMAN, M. L. & CHISHOLM, S. W. 2010. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings of the National Academy of Sciences,* 107**,** 18634.

CONE, L. A., SONTZ, E. M., WILSON, J. W. & MITRUKA, S. N. 2005. *Staphylococcus capitis* endocarditis due to a transvenous endocardial pacemaker infection: Case report and review of *Staphylococcus capitis* endocarditis. *International Journal of Infectious Diseases,* 9**,** 335-339.

COSTELLO, E. K., LAUBER, C. L., HAMADY, M., FIERER, N., GORDON, J. I. & KNIGHT, R. 2009. Bacterial community variation in human body habitats across space and time. *Science (New York, N.Y.),* 326**,** 1694-1697.

CRAMTON, S. E., ULRICH, M., GÖTZ, F. & DÖRING, G. 2001. Anaerobic Conditions Induce Expression of Polysaccharide Intercellular Adhesin in *Staphylococcus aureus* & *Staphylococcus epidermidis*. *Infection and Immunity,* 69**,** 4079.

CROSBY, H. A., SCHLIEVERT, P. M., MERRIMAN, J. A., KING, J. M., SALGADO-PABÓN, W. & HORSWILL, A. R. 2016. The *Staphylococcus aureus* Global Regulator MgrA Modulates Clumping and Virulence by Controlling Surface Protein Expression. *PLOS Pathogens,* 12**,** e1005604.

CROSBY, H. A., TIWARI, N., KWIECINSKI, J. M., XU, Z., DYKSTRA, A., JENUL, C., FUENTES, E. J. & HORSWILL, A. R. 2020. The *Staphylococcus aureus* ArlRS two-component system regulates virulence factor expression through MgrA. *Molecular Microbiology,* 113**,** 103-122.

CROUCHER, N. J., PAGE, A. J., CONNOR, T. R., DELANEY, A. J., KEANE, J. A., BENTLEY, S. D., PARKHILL, J. & HARRIS, S. R. 2014. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research,* 43**,** e15-e15.

CUCARELLA, C., SOLANO, C., VALLE, J., AMORENA, B., LASA, Í. & PENADÉS, J. R. 2001. Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. *Journal of bacteriology,* 183**,** 2888-2896.

CUE, D., LEI, M. G., LUONG, T. T., KUECHENMEISTER, L., DUNMAN, P. M., DONNELL, S., ROWE, S., GARA, J. P. & LEE, C. Y. 2009. Rbf Promotes Biofilm Formation by *Staphylococcus aureus* via Repression of icaR, a Negative Regulator of icaADBC. *Journal of Bacteriology,* 191**,** 6363.

CUI, B., SMOOKER, P. M., ROUCH, D. A., DALEY, A. J. & DEIGHTON, M. A. 2013. Differences between two clinical *Staphylococcus capitis* subspecies as revealed by biofilm, antibiotic resistance, and pulsed-field gel electrophoresis profiling. *Journal of clinical microbiology,* 51**,** 9-14.

DAWSON, T. L. 2007. *Malassezia globosa* and *restricta*: Breakthrough Understanding of the Etiology and Treatment of Dandruff and Seborrheic Dermatitis through Whole-Genome Analysis. *Journal of Investigative Dermatology Symposium Proceedings,* 12**,** 15-19.

DEANGELIS, Y. M., GEMMER, C. M., KACZVINSKY, J. R., KENNEALLY, D. C., SCHWARTZ, J. R. & DAWSON, T. L. 2005. Three Etiologic Facets of Dandruff and Seborrheic Dermatitis: Malassezia Fungi, Sebaceous Lipids, and Individual Sensitivity. *Journal of Investigative Dermatology Symposium Proceedings,* 10**,** 295-297.

DEANGELIS, Y. M., SAUNDERS, C. W., JOHNSTONE, K. R., REEDER, N. L., COLEMAN, C. G., KACZVINSKY, J. R., JR., GALE, C., WALTER, R., MEKEL, M., LACEY, M. P., KEOUGH, T. W., FIENO, A., GRANT, R. A., BEGLEY, B., SUN, Y., FUENTES, G., YOUNGQUIST, R. S., XU, J. & DAWSON, T. L., JR. 2007. Isolation and expression of a *Malassezia globosa* lipase gene, LIP1. *J Invest Dermatol,* 127**,** 2138-46.

DEGHORAIN, M., BOBAY, L.-M., SMEESTERS, P. R., BOUSBATA, S., VERMEERSCH, M., PEREZ-MORGA, D., DRÈZE, P.-A., ROCHA, E. P. C., TOUCHON, M. & VAN MELDEREN, L. 2012. Characterization of Novel Phages Isolated in Coagulase-Negative Staphylococci Reveals Evolutionary Relationships with Staphylococcus aureus Phages. *Journal of Bacteriology,* 194**,** 5829.

DELAUNÉ, A., DUBRAC, S., BLANCHET, C., POUPEL, O., MÄDER, U., HIRON, A., LEDUC, A., FITTING, C., NICOLAS, P., CAVAILLON, J.-M., ADIB-CONQUY, M. & MSADEK, T. 2012. The WalKR System Controls Major Staphylococcal Virulence Genes and Is Involved in Triggering the Host Inflammatory Response. *Infection and Immunity,* 80**,** 3438.

DENG, Y., WANG, H., ZHOU, J., MOU, Y., WANG, G. & XIONG, X. 2018. Patients with acne vulgaris have a distinct gut microbiota in comparison with healthy controls. *Acta dermato-venereologica,* 98**,** 783-790.

DIENE, S. M., CORVAGLIA, A. R., FRANÇOIS, P., VAN DER MEE-MARQUET, N. & REGIONAL INFECTION CONTROL GROUP OF THE CENTRE, R. 2017. Prophages and adaptation of *Staphylococcus aureus* ST398 to the human clinic. *BMC genomics,* 18**,** 133-133.

DIENE, S. M., MERHEJ, V., HENRY, M., EL FILALI, A., ROUX, V., ROBERT, C., AZZA, S., GAVORY, F., BARBE, V. & LA SCOLA, B. 2013. The rhizome of the multidrug-resistant *Enterobacter aerogenes* genome reveals how new "killer bugs" are created because of a sympatric lifestyle. *Molecular biology and evolution,* 30**,** 369-383.

DOBSON, A., COTTER, P. D., ROSS, R. P. & HILL, C. 2012. Bacteriocin production: a probiotic trait? *Applied and environmental microbiology,* 78**,** 1-6.

DOMINGUEZ-BELLO, M. G., COSTELLO, E. K., CONTRERAS, M., MAGRIS, M., HIDALGO, G., FIERER, N. & KNIGHT, R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences,* 107**,** 11971.

DOUGHTY, E. L., SERGEANT, M. J., ADETIFA, I., ANTONIO, M. & PALLEN, M. J. 2014. Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ,* 2**,** e585-e585.

DRÉNO, B., ARAVIISKAIA, E., BERARDESCA, E., GONTIJO, G., SANCHEZ VIERA, M., XIANG, L. F., MARTIN, R. & BIEBER, T. 2016. Microbiome in healthy skin, update for dermatologists. *Journal of the European Academy of Dermatology and Venereology,* 30**,** 2038-2047.

DRÉNO, B., DAGNELIE, M. A., KHAMMARI, A. & CORVEC, S. 2020. The Skin Microbiome: A New Actor in Inflammatory Acne. *American Journal of Clinical Dermatology,* 21**,** 18-24.

DUBIN, G., CHMIEL, D., MAK, P., RAKWALSKA, M., RZYCHON, M. & DUBIN, A. 2001. Molecular cloning and biochemical characterisation of proteases from *Staphylococcus epidermidis*. *Biological chemistry,* 382**,** 1575-1582.

DUBRAC, S., BONECA, I. G., POUPEL, O. & MSADEK, T. 2007. New Insights into the WalK/WalR (YycG/YycF) Essential Signal Transduction Pathway Reveal a Major Role in Controlling Cell Wall Metabolism and Biofilm Formation in *Staphylococcus aureus*. *Journal of Bacteriology,* 189**,** 8257.

EARL, J. P., ADAPPA, N. D., KROL, J., BHAT, A. S., BALASHOV, S., EHRLICH, R. L., PALMER, J. N., WORKMAN, A. D., BLASETTI, M. & SEN, B. 2018. Species-

level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. *Microbiome,* 6**,** 1-26.
EDDY, S. R. 2011. Accelerated Profile HMM Searches. *PLOS Computational Biology,* 7**,** e1002195.
ELIAS, P. M. 2007. The skin barrier as an innate immune element. *Seminars in Immunopathology,* 29**,** 3.
ELSTON, C. A. & ELSTON, D. M. 2014. Demodex mites. *Clinics in Dermatology,* 32**,** 739-743.
ENRIGHT, M. C., DAY, N. P. J., DAVIES, C. E., PEACOCK, S. J. & SPRATT, B. G. 2000. Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. *Journal of Clinical Microbiology,* 38**,** 1008.
EREN, A. M., ESEN Ö, C., QUINCE, C., VINEIS, J. H., MORRISON, H. G., SOGIN, M. L. & DELMONT, T. O. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ,* 3**,** e1319.
EWELS, P., MAGNUSSON, M., LUNDIN, S. & KÄLLER, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics,* 32**,** 3047-3048.
EYERICH, K., EYERICH, S. & BIEDERMANN, T. 2015. The multi-modal immune pathogenesis of atopic eczema. *Trends in immunology,* 36**,** 788-801.
FAHLÉN, A., ENGSTRAND, L., BAKER, B. S., POWLES, A. & FRY, L. 2012. Comparison of bacterial microbiota in skin biopsies from normal and psoriatic skin. *Archives of Dermatological Research,* 304**,** 15-22.
FAWLEY, W. N., UNDERWOOD, S., FREEMAN, J., BAINES, S. D., SAXTON, K., STEPHENSON, K., OWENS, R. C., JR. & WILCOX, M. H. 2007. Efficacy of hospital cleaning agents and germicides against epidemic *Clostridium difficile* strains. *Infection Control and Hospital Epidemiology,* 28**,** 920-925.
FEIL, E. J., COOPER, J. E., GRUNDMANN, H., ROBINSON, D. A., ENRIGHT, M. C., BERENDT, T., PEACOCK, S. J., SMITH, J. M., MURPHY, M., SPRATT, B. G., MOORE, C. E. & DAY, N. P. J. 2003. How clonal is *Staphylococcus aureus*? *Journal of bacteriology,* 185**,** 3307-3316.
FEIL, E. J. & ENRIGHT, M. C. 2004. Analyses of clonality and the evolution of bacterial pathogens. *Current Opinion in Microbiology,* 7**,** 308-313.
FELDGARDEN, M., BROVER, V., HAFT, D. H., PRASAD, A. B., SLOTTA, D. J., TOLSTOY, I., TYSON, G. H., ZHAO, S., HSU, C.-H., MCDERMOTT, P. F., TADESSE, D. A., MORALES, C., SIMMONS, M., TILLMAN, G., WASILENKO, J., FOLSTER, J. P. & KLIMKE, W. 2019. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrobial agents and chemotherapy,* 63**,** e00483-19.
FELSENSTEIN, J. 1974. THE EVOLUTIONARY ADVANTAGE OF RECOMBINATION. *Genetics,* 78**,** 737.
FINDLEY, K. & GRICE, E. A. 2014. The skin microbiome: a focus on pathogens and their association with skin disease. *PLoS pathogens,* 10**,** e1004436-e1004436.
FINDLEY, K., OH, J., YANG, J., CONLAN, S., DEMING, C., MEYER, J. A., SCHOENFELD, D., NOMICOS, E., PARK, M., BECKER, J., BENJAMIN, B., BLAKESLEY, R., BOUFFARD, G., BROOKS, S., COLEMAN, H., DEKHTYAR, M., GREGORY, M.,

GUAN, X., GUPTA, J., HAN, J., HARGROVE, A., HO, S.-L., JOHNSON, T., LEGASPI, R., LOVETT, S., MADURO, Q., MASIELLO, C., MASKERI, B., MCDOWELL, J., MONTEMAYOR, C., MULLIKIN, J., PARK, M., RIEBOW, N., SCHANDLER, K., SCHMIDT, B., SISON, C., STANTRIPOP, M., THOMAS, J., THOMAS, P., VEMULAPALLI, M., YOUNG, A., KONG, H. H., SEGRE, J. A. & PROGRAM, N. I. H. I. S. C. C. S. 2013. Topographic diversity of fungal and bacterial communities in human skin. *Nature,* 498**,** 367-370.

FISCHER, A., KAMBARA, K., MEYER, H., STENZ, L., BONETTI, E. J., GIRARD, M., LALK, M., FRANCOIS, P. & SCHRENZEL, J. 2014. GdpS contributes to *Staphylococcus aureus* biofilm formation by regulation of eDNA release. *International Journal of Medical Microbiology,* 304**,** 284-299.

FITZ-GIBBON, S., TOMIDA, S., CHIU, B.-H., NGUYEN, L., DU, C., LIU, M., ELASHOFF, D., ERFE, M. C., LONCARIC, A., KIM, J., MODLIN, R. L., MILLER, J. F., SODERGREN, E., CRAFT, N., WEINSTOCK, G. M. & LI, H. 2013. *Propionibacterium acnes* Strain Populations in the Human Skin Microbiome Associated with Acne. *Journal of Investigative Dermatology,* 133**,** 2152-2160.

FLOWERS, L. & GRICE, E. A. 2020. The Skin Microbiota: Balancing Risk and Reward. *Cell Host & Microbe,* 28**,** 190-200.

FLUCKIGER, U., WOLZ, C. & CHEUNG, A. L. 1998. Characterization of a Homolog of *Staphylococcus epidermidis*. *Infection and Immunity,* 66**,** 2871.

FLURIN, L., GREENWOOD-QUAINTANCE, K. E. & PATEL, R. 2019. Microbiology of polymicrobial prosthetic joint infection. *Diagnostic Microbiology and Infectious Disease,* 94**,** 255-259.

FOSTER, T. J. 2019. The MSCRAMM Family of Cell-Wall-Anchored Surface Proteins of Gram-Positive Cocci. *Trends Microbiol,* 27**,** 927-941.

FOSTER, T. J. 2020. Surface Proteins of *Staphylococcus epidermidis*. *Frontiers in Microbiology,* 11**,** 1829.

FOSTER, T. J., GEOGHEGAN, J. A., GANESH, V. K. & HÖÖK, M. 2014. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. *Nat Rev Microbiol,* 12**,** 49-62.

FOURNIER, P.-E., GOURIET, F., GIMENEZ, G., ROBERT, C. & RAOULT, D. 2013. Deciphering genomic virulence traits of a *Staphylococcus epidermidis* strain causing native-valve endocarditis. *Journal of clinical microbiology,* 51**,** 1617-1621.

FRY, L., BAKER, B. S., POWLES, A. V., FAHLEN, A. & ENGSTRAND, L. 2013. Is chronic plaque psoriasis triggered by microbiota in the skin? *Br J Dermatol,* 169**,** 47-52.

FUCHS, E. & RAGHAVAN, S. 2002. Getting under the skin of epidermal morphogenesis. *Nat Rev Genet,* 3**,** 199-209.

FURUE, M., CHIBA, T., TSUJI, G., ULZII, D., KIDO-NAKAHARA, M., NAKAHARA, T. & KADONO, T. 2017. Atopic dermatitis: immune deviation, barrier dysfunction, IgE autoreactivity and new therapies. *Allergology international,* 66**,** 398-403.

GALBUSERA, E., RENZONI, A., ANDREY, D. O., MONOD, A., BARRAS, C., TORTORA, P., POLISSI, A. & KELLEY, W. L. 2011. Site-specific mutation of *Staphylococcus aureus* VraS reveals a crucial role for the VraR-VraS sensor in the emergence of glycopeptide resistance. *Antimicrobial agents and chemotherapy,* 55**,** 1008-1020.

GALPERIN, M. Y., MAKAROVA, K. S., WOLF, Y. I. & KOONIN, E. V. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res,* 43**,** D261-9.

GAO, Z., TSENG, C.-H., STROBER, B. E., PEI, Z. & BLASER, M. J. 2008. Substantial Alterations of the Cutaneous Bacterial Biota in Psoriatic Lesions. *PLOS ONE,* 3**,** e2719.

GARDETE, S., WU, S. W., GILL, S. & TOMASZ, A. 2006. Role of VraSR in antibiotic resistance and antibiotic-induced stress response in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy,* 50**,** 3424-3434.

GAVAZZONI DIAS, M. F. R., DE ALMEIDA, A. M., CECATO, P. M. R., ADRIANO, A. R. & PICHLER, J. 2014. The Shampoo pH can Affect the Hair: Myth or Reality? *International journal of trichology,* 6**,** 95-99.

GEMMER, C. M., DEANGELIS, Y. M., THEELEN, B., BOEKHOUT, T. & DAWSON, T. L. 2002. Fast, Noninvasive Method for Molecular Detection and Differentiation of Malassezia Yeast Species on Human Skin and Application of the Method to Dandruff Microbiology. *Journal of Clinical Microbiology,* 40**,** 3350.

GENE ONTOLOGY CONSORTIUM 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Res,* 43**,** D1049-56.

GEOGHEGAN, J. A., IRVINE, A. D. & FOSTER, T. J. 2018. *Staphylococcus aureus* and Atopic Dermatitis: A Complex and Evolving Relationship. *Trends in Microbiology,* 26**,** 484-497.

GEORGIADES, K. & RAOULT, D. 2010. Defining pathogenic bacterial species in the genomic era. *Front Microbiol,* 1**,** 151.

GILABERTE, Y., PRIETO-TORRES, L., PASTUSHENKO, I. & JUARRANZ, Á. 2016. Chapter 1 - Anatomy and Function of the Skin. *In:* HAMBLIN, M. R., AVCI, P. & PROW, T. W. (eds.) *Nanoscience in Dermatology.* Boston: Academic Press.

GOODRICH, J. K., DAVENPORT, E. R., BEAUMONT, M., JACKSON, M. A., KNIGHT, R., OBER, C., SPECTOR, T. D., BELL, J. T., CLARK, A. G. & LEY, R. E. 2016. Genetic determinants of the gut microbiome in UK twins. *Cell host & microbe,* 19**,** 731-743.

GORIS, J., KONSTANTINIDIS, K. T., KLAPPENBACH, J. A., COENYE, T., VANDAMME, P. & TIEDJE, J. M. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology,* 57**,** 81-91.

GRANSLO, H. N., KLINGENBERG, C., FREDHEIM, E. G. A., RØNNESTAD, A., MOLLNES, T. E. & FLÆGSTAD, T. 2010. Arginine Catabolic Mobile Element Is Associated With Low Antibiotic Resistance and Low Pathogenicity in *Staphylococcus epidermidis* From Neonates. *Pediatric Research,* 68**,** 237-241.

GRAS-LE GUEN, C., FOURNIER, S., ANDRE-RICHET, B., CAILLON, J., CHAMOUX, C., ESPAZE, E., RICHET, H., ROZE, J. C. & LEPELLETIER, D. 2007. Almond oil implicated in a S*taphylococcus capitis* outbreak in a neonatal intensive care unit. *Journal of Perinatology,* 27**,** 713-717.

GRIBBON, E. M., CUNLIFFE, W. J. & HOLLAND, K. T. 1993. Interaction of *Propionibacterium acnes* with skin lipids in vitro. *Microbiology,* 139**,** 1745-1751.

GRICE, E. A. 2014. The skin microbiome: potential for novel diagnostic and therapeutic approaches to cutaneous disease. *Seminars in cutaneous medicine and surgery,* 33**,** 98-103.
GRICE, E. A. & DAWSON, T. L. 2017. Host–microbe interactions: Malassezia and human skin. *Current Opinion in Microbiology,* 40**,** 81-87.
GRICE, E. A., KONG, H. H., CONLAN, S., DEMING, C. B., DAVIS, J., YOUNG, A. C., BOUFFARD, G. G., BLAKESLEY, R. W., MURRAY, P. R., GREEN, E. D., TURNER, M. L. & SEGRE, J. A. 2009. Topographical and Temporal Diversity of the Human Skin Microbiome. *Science,* 324**,** 1190.
GRICE, E. A., KONG, H. H., RENAUD, G., YOUNG, A. C., PROGRAM, N. C. S., BOUFFARD, G. G., BLAKESLEY, R. W., WOLFSBERG, T. G., TURNER, M. L. & SEGRE, J. A. 2008. A diversity profile of the human skin microbiota. *Genome research,* 18**,** 1043-1050.
GRICE, E. A. & SEGRE, J. A. 2011. The skin microbiome. *Nature Reviews Microbiology,* 9**,** 244-253.
GRIFFITHS, J. M. & O'NEILL, A. J. 2012. Loss of function of the gdpP protein leads to joint β-lactam/glycopeptide tolerance in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy,* 56**,** 579-581.
GRIMALT, R. 2007. A Practical Guide to Scalp Disorders. *Journal of Investigative Dermatology Symposium Proceedings,* 12**,** 10-14.
GRIMSHAW, S. G., SMITH, A. M., ARNOLD, D. S., XU, E., HOPTROFF, M. & MURPHY, B. 2019. The diversity and abundance of fungi and bacteria on the healthy and dandruff affected human scalp. *PLOS ONE,* 14**,** e0225796.
GRÜNDLING, A. & SCHNEEWIND, O. 2007. Genes required for glycolipid synthesis and lipoteichoic acid anchoring in *Staphylococcus aureus*. *J Bacteriol,* 189**,** 2521-30.
GUPTA, A. K., BATRA, R., BLUHM, R., BOEKHOUT, T. & DAWSON, T. L., JR. 2004. Skin diseases associated with *Malassezia* species. *J Am Acad Dermatol,* 51**,** 785-98.
GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-1075.
GUTIÉRREZ, D., MARTÍNEZ, B., RODRÍGUEZ, A. & GARCÍA, P. 2012. Genomic characterization of two *Staphylococcus epidermidis* bacteriophages with anti-biofilm potential. *BMC genomics,* 13**,** 228-228.
HACHEM, J.-P., CRUMRINE, D., FLUHR, J., BROWN, B. E., FEINGOLD, K. R. & ELIAS, P. M. 2003. pH directly regulates epidermal permeability barrier homeostasis, and stratum corneum integrity/cohesion. *Journal of Investigative Dermatology,* 121**,** 345-353.
HACKER, J., BLUM-OEHLER, G., MÜHLDORFER, I. & TSCHÄPE, H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Molecular microbiology,* 23**,** 1089-1097.
HADFIELD, J., CROUCHER, N. J., GOATER, R. J., ABUDAHAB, K., AANENSEN, D. M. & HARRIS, S. R. 2017. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics,* 34**,** 292-293.
HANIFIN, J. M. & ROGGE, J. L. 1977. Staphylococcal infections in patients with atopic dermatitis. *Archives of dermatology,* 113**,** 1383-1386.
HANSEN, E. E., LOZUPONE, C. A., REY, F. E., WU, M., GURUGE, J. L., NARRA, A., GOODFELLOW, J., ZANEVELD, J. R., MCDONALD, D. T. & GOODRICH, J. A. 2011. Pan-genome of the dominant human gut-associated archaeon,

Methanobrevibacter smithii, studied in twins. *Proceedings of the National Academy of Sciences,* 108**,** 4599-4606.

HARDING, C., WATKINSON, A., RAWLINGS, A. & SCOTT, I. 2000. Dry skin, moisturization and corneodesmolysis. *International journal of cosmetic science,* 22**,** 21-52.

HARRIS, T. M., BOWEN, A. C., HOLT, D. C., SAROVICH, D. S., STEVENS, K., CURRIE, B. J., HOWDEN, B. P., CARAPETIS, J. R., GIFFARD, P. M. & TONG, S. Y. C. 2018. Investigation of trimethoprim/sulfamethoxazole resistance in an emerging sequence type 5 methicillin-resistant *Staphylococcus aureus* clone reveals discrepant resistance reporting. *Clinical Microbiology and Infection,* 24**,** 1027-1029.

HARTFORD, O., O'BRIEN, L., SCHOFIELD, K., WELLS, J. & FOSTER, T. J. 2001. The Fbe (SdrG) protein of *Staphylococcus epidermidis* HB promotes bacterial adherence to fibrinogen. *Microbiology,* 147**,** 2545-2552.

HEILMANN, C., HUSSAIN, M., PETERS, G. & GÖTZ, F. 1997. Evidence for autolysin-mediated primary attachment of *Staphylococcus epidermidis* to a polystyrene surface. *Molecular microbiology,* 24**,** 1013-1024.

HEILMANN, C., SCHWEITZER, O., GERKE, C., VANITTANAKOM, N., MACK, D. & GÖTZ, F. 1996. Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*. *Molecular microbiology,* 20**,** 1083-1091.

HEILMANN, C., ZIEBUHR, W. & BECKER, K. 2019. Are coagulase-negative staphylococci virulent? *Clinical Microbiology and Infection,* 25**,** 1071-1080.

HERBERT, S., BERA, A., NERZ, C., KRAUS, D., PESCHEL, A., GOERKE, C., MEEHL, M., CHEUNG, A. & GOETZ, F. 2007. Molecular basis of resistance to muramidase and cationic antimicrobial peptide activity of lysozyme in staphylococci. *Plos Pathogens,* 3**,** 981-994.

HERZ, U., SCHNOY, N., BORELLI, S., WEIGL, L., KÄSBOHRER, U., DASER, A., WAHN, U., KÖTTGEN, E. & RENZ, H. 1998. A human-SCID mouse model for allergic immune responses: bacterial superantigen enhances skin inflammation and suppresses IgE production. *Journal of investigative dermatology,* 110**,** 224-231.

HODILLE, E., ROSE, W., DIEP, B. A., GOUTELLE, S., LINA, G. & DUMITRESCU, O. 2017. The Role of Antibiotics in Modulating Virulence in *Staphylococcus aureus*. *Clinical Microbiology Reviews,* 30**,** 887.

HOLLAND, K. T., GREENMAN, J. & CUNLIFFE, W. J. 1979. Growth of cutaneous propionibacteria on synthetic medium; growth yields and exoenzyme production. *J Appl Bacteriol,* 47**,** 383-94.

HOLLAND, L. M., CONLON, B., APOS & GARA, J. P. 2011. Mutation of tagO reveals an essential role for wall teichoic acids in *Staphylococcus epidermidis* biofilm development. *Microbiology,* 157**,** 408-418.

HUANG, J. T., ABRAMS, M., TLOUGAN, B., RADEMAKER, A. & PALLER, A. S. 2009. Treatment of *Staphylococcus aureus* colonization in atopic dermatitis decreases disease severity. *Pediatrics,* 123**,** e808-e814.

HUERTA-CEPAS, J., FORSLUND, K., COELHO, L. P., SZKLARCZYK, D., JENSEN, L. J., VON MERING, C. & BORK, P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol,* 34**,** 2115-2122.

HUERTA-CEPAS, J., SZKLARCZYK, D., HELLER, D., HERNÁNDEZ-PLAZA, A., FORSLUND, S. K., COOK, H., MENDE, D. R., LETUNIC, I., RATTEI, T., JENSEN, L. J., VON MERING, C. & BORK, P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research,* 47**,** D309-D314.

HUKIĆ, M., SELJMO, D., RAMOVIC, A., IBRIŠIMOVIĆ, M. A., DOGAN, S., HUKIC, J. & BOJIC, E. F. 2017. The Effect of Lysozyme on Reducing Biofilms by *Staphylococcus aureus, Pseudomonas aeruginosa*, and *Gardnerella vaginalis*: An In Vitro Examination. *Microbial Drug Resistance,* 24**,** 353-358.

HUMAN MICROBIOME JUMPSTART REFERENCE STRAINS CONSORTIUM 2010. A catalog of reference genomes from the human microbiome. *Science.*

HUMANN, J. & LENZ, L. L. 2009. Bacterial peptidoglycan-degrading enzymes and their impact on host muropeptide detection. *Journal of innate immunity,* 1**,** 88-97.

HUSON, D. H., AUCH, A. F., QI, J. & SCHUSTER, S. C. 2007. MEGAN analysis of metagenomic data. *Genome research,* 17**,** 377-386.

HUSSAIN, M., STEINBACHER, T., PETERS, G., HEILMANN, C. & BECKER, K. 2015. The adhesive properties of the *Staphylococcus lugdunensis* multifunctional autolysin AtlL and its role in biofilm formation and internalization. *International Journal of Medical Microbiology,* 305**,** 129-139.

HUTTENHOWER, C., GEVERS, D., KNIGHT, R., ABUBUCKER, S., BADGER, J. H., CHINWALLA, A. T., CREASY, H. H., EARL, A. M., FITZGERALD, M. G., FULTON, R. S., GIGLIO, M. G., HALLSWORTH-PEPIN, K., LOBOS, E. A., MADUPU, R., MAGRINI, V., MARTIN, J. C., MITREVA, M., MUZNY, D. M., SODERGREN, E. J., VERSALOVIC, J., WOLLAM, A. M., WORLEY, K. C., WORTMAN, J. R., YOUNG, S. K., ZENG, Q., AAGAARD, K. M., ABOLUDE, O. O., ALLEN-VERCOE, E., ALM, E. J., ALVARADO, L., ANDERSEN, G. L., ANDERSON, S., APPELBAUM, E., ARACHCHI, H. M., ARMITAGE, G., ARZE, C. A., AYVAZ, T., BAKER, C. C., BEGG, L., BELACHEW, T., BHONAGIRI, V., BIHAN, M., BLASER, M. J., BLOOM, T., BONAZZI, V., PAUL BROOKS, J., BUCK, G. A., BUHAY, C. J., BUSAM, D. A., CAMPBELL, J. L., CANON, S. R., CANTAREL, B. L., CHAIN, P. S. G., CHEN, I. M. A., CHEN, L., CHHIBBA, S., CHU, K., CIULLA, D. M., CLEMENTE, J. C., CLIFTON, S. W., CONLAN, S., CRABTREE, J., CUTTING, M. A., DAVIDOVICS, N. J., DAVIS, C. C., DESANTIS, T. Z., DEAL, C., DELEHAUNTY, K. D., DEWHIRST, F. E., DEYCH, E., DING, Y., DOOLING, D. J., DUGAN, S. P., MICHAEL DUNNE, W., SCOTT DURKIN, A., EDGAR, R. C., ERLICH, R. L., FARMER, C. N., FARRELL, R. M., FAUST, K., FELDGARDEN, M., FELIX, V. M., FISHER, S., FODOR, A. A., FORNEY, L. J., FOSTER, L., DI FRANCESCO, V., FRIEDMAN, J., FRIEDRICH, D. C., FRONICK, C. C., FULTON, L. L., GAO, H., GARCIA, N., GIANNOUKOS, G., GIBLIN, C., GIOVANNI, M. Y., GOLDBERG, J. M., GOLL, J., GONZALEZ, A., GRIGGS, A., et al. 2012. Structure, function and diversity of the healthy human microbiome. *Nature,* 486**,** 207-214.

HYATT, D., CHEN, G.-L., LOCASCIO, P. F., LAND, M. L., LARIMER, F. W. & HAUSER, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics,* 11**,** 119-119.

IEBBA, V., TOTINO, V., GAGLIARDI, A., SANTANGELO, F., CACCIOTTI, F., TRANCASSINI, M., MANCINI, C., CICERONE, C., CORAZZIARI, E. &

PANTANELLA, F. 2016. Eubiosis and dysbiosis: the two sides of the microbiota. *New Microbiol,* 39**,** 1-12.

IRVINE, A. D., MCLEAN, W. I. & LEUNG, D. Y. 2011. Filaggrin mutations associated with skin and allergic diseases. *New England Journal of Medicine,* 365**,** 1315-1327.

IWASE, T., UEHARA, Y., SHINJI, H., TAJIMA, A., SEO, H., TAKADA, K., AGATA, T. & MIZUNOE, Y. 2010. *Staphylococcus epidermidis* Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization. *Nature,* 465**,** 346-9.

JAIN, C., RODRIGUEZ-R, L. M., PHILLIPPY, A. M., KONSTANTINIDIS, K. T. & ALURU, S. 2018a. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications,* 9**,** 5114.

JAIN, M., KOREN, S., MIGA, K. H., QUICK, J., RAND, A. C., SASANI, T. A., TYSON, J. R., BEGGS, A. D., DILTHEY, A. T. & FIDDES, I. T. 2018b. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology,* 36**,** 338-345.

JARI OKSANEN, F. G. B., MICHAEL FRIENDLY, ROELAND KINDT, PIERRE LEGENDRE, DAN MCGLINN, PETER R. MINCHIN, R. B. O'HARA, GAVIN L. SIMPSON, PETER SOLYMOS, M. HENRY H. STEVENS, EDUARD SZOECS AND HELENE WAGNER 2019. vegan: Community Ecology Package. R package version 2.5-6. R package version 2.5-6 ed.

JOHNSON, T., KANG, D., BARNARD, E. & LI, H. 2016. Strain-level differences in porphyrin production and regulation in *Propionibacterium acnes* elucidate disease associations. *Msphere,* 1.

JONSSON, V., ÖSTERLUND, T., NERMAN, O. & KRISTIANSSON, E. 2016. Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC genomics,* 17**,** 1-14.

JOO, H.-S. & OTTO, M. 2015. Mechanisms of resistance to antimicrobial peptides in staphylococci. *Biochimica et biophysica acta,* 1848**,** 3055-3061.

JOSHI & FASS. 2011. *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.2)* [Online]. Available: https://github.com/najoshi/sickle. [Accessed].

KALYAANAMOORTHY, S., MINH, B. Q., WONG, T. K. F., VON HAESELER, A. & JERMIIN, L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods,* 14**,** 587-589.

KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**,** 27-30.

KANEHISA, M., GOTO, S., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research,* 42**,** D199-D205.

KANEHISA, M., SATO, Y. & MORISHIMA, K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol,* 428**,** 726-731.

KANG, D., LI, F., KIRTON, E. S., THOMAS, A., EGAN, R. S., AN, H. & WANG, Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ Preprints,* 7**,** e27522v1.

KANG, D., SHI, B., ERFE, M. C., CRAFT, N. & LI, H. 2015a. Vitamin B12 modulates the transcriptome of the skin microbiota in acne pathogenesis. *Science translational medicine,* 7**,** 293ra103-293ra103.

KANG, D. D., FROULA, J., EGAN, R. & WANG, Z. 2015b. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ,* 3**,** e1165.

KARINOU, E., SCHUSTER, C. F., PAZOS, M., VOLLMER, W. & GRÜNDLING, A. 2019. Inactivation of the Monofunctional Peptidoglycan Glycosyltransferase SgtB Allows *Staphylococcus aureus*. To Survive in the Absence of Lipoteichoic Acid. *Journal of Bacteriology,* 201**,** e00574-18.

KAWECKI, T. J., LENSKI, R. E., EBERT, D., HOLLIS, B., OLIVIERI, I. & WHITLOCK, M. C. 2012. Experimental evolution. *Trends in Ecology & Evolution,* 27**,** 547-560.

KEANE, F. M., LOUGHMAN, A., VALTULINA, V., BRENNAN, M., SPEZIALE, P. & FOSTER, T. J. 2007. Fibrinogen and elastin bind to the same region within the A domain of fibronectin binding protein A, an MSCRAMM of *Staphylococcus aureus*. *Molecular Microbiology,* 63**,** 711-723.

KELLY, J. E. 2013. *Genomic and metagenomic analyses of the skin microbiota.* PhD, University of Liverpool.

KENNY, J. G., WARD, D., JOSEFSSON, E., JONSSON, M., HINDS, J., REES, H. H., LINDSAY, J. A., TARKOWSKI, A. & HORSBURGH, M. J. 2009. The *Staphylococcus aureus* response to unsaturated long chain free fatty acids: survival mechanisms and virulence implications. *PLoS One,* 4**,** e4344.

KINENE, T., WAINAINA, J., MAINA, S. & BOYKIN, L. M. 2016. Rooting Trees, Methods for. *Encyclopedia of Evolutionary Biology***,** 489-493.

KLOOS, W. E., BALLARD, D. N., GEORGE, C. G., WEBSTER, J. A., HUBNER, R. J., LUDWIG, W., SCHLEIFER, K. H., FIEDLER, F. & SCHUBERT, K. 1998. Delimiting the genus Staphylococcus through description of Macrococcus caseolyticus gen. nov., comb. nov. and *Macrococcus equipercicus* sp. nov., *Macrococcus bovicus* sp. nov. and *Macrococcus carouselicus* sp. nov. *International Journal of Systematic and Evolutionary Microbiology,* 48**,** 859-877.

KLOOS, W. E. & SCHLEIFER, K. H. 1975. Isolation and Characterization of Staphylococci from Human Skin II. Descriptions of Four New Species: *Staphylococcus warneri*, *Staphylococcus capitis*, *Staphylococcus hominis*, and *Staphylococcus simulans*. *International Journal of Systematic and Evolutionary Microbiology,* 25**,** 62-79.

KMIECIAK, W., SZEWCZYK, E. M. & CISZEWSKI, M. 2016. Searching for Beta-Haemolysin hlb Gene in *Staphylococcus pseudintermedius* with Species-Specific Primers. *Current microbiology,* 73**,** 148-152.

KOHLER, T., WEIDENMAIER, C. & PESCHEL, A. 2009. Wall teichoic acid protects *Staphylococcus aureus* against antimicrobial fatty acids from human skin. *Journal of bacteriology,* 191**,** 4482-4484.

KOLARSICK, P. A., KOLARSICK, M. A. & GOODWIN, C. 2011. Anatomy and physiology of the skin. *Journal of the Dermatology Nurses' Association,* 3**,** 203-213.

KOLDE, R., FRANZOSA, E. A., RAHNAVARD, G., HALL, A. B., VLAMAKIS, H., STEVENS, C., DALY, M. J., XAVIER, R. J. & HUTTENHOWER, C. 2018. Host

genetic variation and its microbiome interactions within the Human Microbiome Project. *Genome Medicine,* 10**,** 6.

KONG, H. H., OH, J., DEMING, C., CONLAN, S., GRICE, E. A., BEATSON, M. A., NOMICOS, E., POLLEY, E. C., KOMAROW, H. D., PROGRAM, N. C. S., MURRAY, P. R., TURNER, M. L. & SEGRE, J. A. 2012. Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome research,* 22**,** 850-859.

KONG, H. H. & SEGRE, J. A. 2012. Skin Microbiome: Looking Back to Move Forward. *Journal of Investigative Dermatology,* 132**,** 933-939.

KOPRIVNJAK, T., MLAKAR, V., SWANSON, L., FOURNIER, B., PESCHEL, A. & WEISS, J. P. 2006. Cation-Induced Transcriptional Regulation of the Operon of Staphylococcus aureus. *Journal of Bacteriology,* 188**,** 3622.

KOŚCIUCZUK, E. M., LISOWSKI, P., JARCZAK, J., STRZAŁKOWSKA, N., JÓŹWIK, A., HORBAŃCZUK, J., KRZYŻEWSKI, J., ZWIERZCHOWSKI, L. & BAGNICKA, E. 2012. Cathelicidins: family of antimicrobial peptides. A review. *Molecular biology reports,* 39**,** 10957-10970.

KOZITSKAYA, S., OLSON, M. E., FEY, P. D., WITTE, W., OHLSEN, K. & ZIEBUHR, W. 2005. Clonal analysis of *Staphylococcus epidermidis* isolates carrying or lacking biofilm-mediating genes by multilocus sequence typing. *Journal of clinical microbiology,* 43**,** 4751-4757.

KRAUS, D., HERBERT, S., KRISTIAN, S. A., KHOSRAVI, A., NIZET, V., GÖTZ, F. & PESCHEL, A. 2008. The GraRS regulatory system controls *Staphylococcus aureus* susceptibility to antimicrobial host defenses. *BMC Microbiology,* 8**,** 85.

KUMAR, R., JANGIR, P. K., DAS, J., TANEJA, B. & SHARMA, R. 2017. Genome Analysis of *Staphylococcus capitis* TE8 Reveals Repertoire of Antimicrobial Peptides and Adaptation Strategies for Growth on Human Skin. *Scientific reports,* 7**,** 10447-10447.

KWAN, T., LIU, J., DUBOW, M., GROS, P. & PELLETIER, J. 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 5174.

LAI, Y., COGEN, A. L., RADEK, K. A., PARK, H. J., MACLEOD, D. T., LEICHTLE, A., RYAN, A. F., DI NARDO, A. & GALLO, R. L. 2010. Activation of TLR2 by a small molecule produced by *Staphylococcus epidermidis* increases antimicrobial defense against bacterial skin infections. *Journal of Investigative Dermatology,* 130**,** 2211-2221.

LAI, Y., VILLARUZ, A. E., LI, M., CHA, D. J., STURDEVANT, D. E. & OTTO, M. 2007. The human anionic antimicrobial peptide dermcidin induces proteolytic defence mechanisms in staphylococci. *Mol Microbiol,* 63**,** 497-506.

LAMERS, R. P., MUTHUKRISHNAN, G., CASTOE, T. A., TAFUR, S., COLE, A. M. & PARKINSON, C. L. 2012. Phylogenetic relationships among *Staphylococcus* species and refinement of cluster groups based on multilocus data. *BMC Evol Biol,* 12**,** 171.

LANE, D. J., PACE, B., OLSEN, G. J., STAHL, D. A., SOGIN, M. L. & PACE, N. R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America,* 82**,** 6955-6959.

LANG, S., LIVESLEY, M., LAMBERT, P., ELLIOTT, J. & ELLIOTT, T. 1999. The genomic diversity of coagulase-negative staphylococci associated with nosocomial infections. *Journal of Hospital Infection,* 43**,** 187-193.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9**,** 357-359.

LATASA, C., ROUX, A., TOLEDO-ARANA, A., GHIGO, J.-M., GAMAZO, C., PENADÉS, J. R. & LASA, I. 2005. BapA, a large secreted protein required for biofilm formation and host colonization of *Salmonella enterica* serovar *Enteritidis*. Molecular Microbiology, 58**,** 1322-1339.

LEE, Y. B., BYUN, E. J. & KIM, H. S. 2019. Potential Role of the Microbiome in Acne: A Comprehensive Review. *Journal of clinical medicine,* 8**,** 987.

LEEMING, J. P., HOLLAND, K. T. & CUNLIFFE, W. J. 1984. The microbial ecology of pilosebaceous units isolated from human skin. *J Gen Microbiol,* 130**,** 803-7.

LEMRISS, H., DUMONT, Y., LEMRISS, S., MARTINS-SIMOES, P., BUTIN, M., LAHLOU, L., RASIGADE, J. P., EL KABBAJ, S., LAURENT, F. & IBRAHIMI, A. 2016. Genome Sequences of Multiresistant *Staphylococcus capitis* Pulsotype NRCS-A and Methicillin-Susceptible *S. capitis* Pulsotype NRCS-C. *Genome Announc,* 4.

LENTZ, C. S., SHELDON, J. R., CRAWFORD, L. A., COOPER, R., GARLAND, M., AMIEVA, M. R., WEERAPANA, E., SKAAR, E. P. & BOGYO, M. 2018. Identification of a *S. aureus* virulence factor by activity-based protein profiling (ABPP). *Nature Chemical Biology,* 14**,** 609-617.

LEONG, C., SCHMID, B., BUTTAFUOCO, A., GLATZ, M. & BOSSHARD, P. P. 2019. In vitro efficacy of antifungal agents alone and in shampoo formulation against dandruff-associated *Malassezia* spp. and *Staphylococcus* spp. *International Journal of Cosmetic Science,* 41**,** 221-227.

LETUNIC, I. & BORK, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res,* 44**,** W242-5.

LEYDEN, J. J., MARPLES, R. R. & KLIGMAN, A. M. 1974. *Staphylococcus aureus* in the lesions of atopic dermatitis. *British Journal of Dermatology,* 90**,** 525-525.

LEYDEN, J. J., MCGINLEY, K. J. & KLIGMAN, A. M. 1976. ROLE OF MICROORGANISMS IN DANDRUFF. *Archives of Dermatology,* 112**,** 333-338.

LI, D., LIU, C.-M., LUO, R., SADAKANE, K. & LAM, T.-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics,* 31**,** 1674-1676.

LI, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics,* 25**,** 1754-1760.

LI, H., GOH, B. N., TEH, W. K., JIANG, Z., GOH, J. P. Z., GOH, A., WU, G., HOON, S. S., RAIDA, M. & CAMATTARI, A. 2018. Skin commensal *Malassezia globosa* secreted protease attenuates *Staphylococcus aureus* biofilm formation. *Journal of Investigative Dermatology,* 138**,** 1137-1145.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, W. & GODZIK, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics,* 22**,** 1658-1659.

LI, X., LEI, M., SONG, Y., GONG, K., LI, L., LIANG, H. & JIANG, X. 2014. Whole genome sequence and comparative genomic analysis of multidrug-resistant *Staphylococcus capitis* subsp. *urealyticus* strain LNZR-1. *Gut pathogens,* 6**,** 45-45.

LIBBERTON, B., COATES, R. E., BROCKHURST, M. A. & HORSBURGH, M. J. 2014. Evidence that intraspecific trait variation among nasal bacteria shapes the distribution of *Staphylococcus aureus*. *Infection and immunity,* 82**,** 3811-3815.

LINA, B., VANDENESCH, F., ETIENNE, J., KREISWIRTH, B. & FLEURETTE, J. 1992. Comparison of coagulase-negative staphylococci by pulsed-field gel electrophoresis. *FEMS Microbiology Letters,* 92**,** 133-138.

LINDGREN, J. K., THOMAS, V. C., OLSON, M. E., CHAUDHARI, S. S., NUXOLL, A. S., SCHAEFFER, C. R., LINDGREN, K. E., JONES, J., ZIMMERMAN, M. C., DUNMAN, P. M., BAYLES, K. W. & FEY, P. D. 2014. Arginine Deiminase in *Staphylococcus epidermidis* Functions To Augment Biofilm Maturation through pH Homeostasis. *Journal of Bacteriology,* 196**,** 2277.

LOU, Q., QI, Y., MA, Y. & QU, D. 2014. Two-component signal transduction system SaeRS positively regulates *Staphylococcus epidermidis* glucose metabolism. *TheScientificWorldJournal,* 2014**,** 908121-908121.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15**,** 550.

LU, J., BREITWIESER, F. P., THIELEN, P. & SALZBERG, S. L. 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science,* 3**,** e104.

MADDISON, W. P., DONOGHUE, M. J. & MADDISON, D. R. 1984. Outgroup analysis and parsimony. *Systematic biology,* 33**,** 83-103.

MAGGS, A. F. & PENNINGTON, T. H. 1989. Temporal study of staphylococcal species on the skin of human subjects in isolation and clonal analysis of *Staphylococcus capitis* by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *J Clin Microbiol,* 27**,** 2627-32.

MAIDEN, M. C., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K. & CAUGANT, D. A. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences,* 95**,** 3140-3145.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011,* 17**,** 3.

MASSEY, R. C., HORSBURGH, M. J., LINA, G., HÖÖK, M. & RECKER, M. 2006. The evolution and maintenance of virulence in *Staphylococcus aureus*: a role for host-to-host transmission? *Nature Reviews Microbiology,* 4**,** 953-958.

MAYA-MARTINEZ, R., ALEXANDER, J. A. N., OTTEN, C. F., AYALA, I., VOLLMER, D., GRAY, J., BOUGAULT, C. M., BURT, A., LAGURI, C., FONVIELLE, M., ARTHUR, M., STRYNADKA, N. C. J., VOLLMER, W. & SIMORRE, J.-P. 2019. Recognition

of Peptidoglycan Fragments by the Transpeptidase PBP4 From Staphylococcus aureus. *Frontiers in Microbiology,* 9.

MCBRIDE, M. E., DUNCAN, W. C. & KNOX, J. 1977. The environment and the microbial ecology of human skin. *Applied and environmental microbiology,* 33**,** 603-608.

MCCREA, K. W., HARTFORD, O., DAVIS, S., EIDHIN, D. N., LINA, G., SPEZIALE, P., FOSTER, T. J. & HÖÖK, M. 2000. The serine-aspartate repeat (Sdr) protein family in *Staphylococcus epidermidis*. *Microbiology (Reading),* 146 (Pt 7)**,** 1535-1546.

MCGINLEY, K. J., LEYDEN, J. J., MARPLES, R. R., PATH, M. R. C. & KLIGMAN, A. M. 1975. Quantitative Microbiology of the Scalp in Non-Dandruff, Dandruff, and Seborrheic Dermatitis. *Journal of Investigative Dermatology,* 64**,** 401-405.

MCMURDIE, P. J. & HOLMES, S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE,* 8**,** e61217.

MCVEAN, G. A. & CHARLESWORTH, B. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics,* 155**,** 929-944.

MEDINI, D., DONATI, C., TETTELIN, H., MASIGNANI, V. & RAPPUOLI, R. 2005. The microbial pan-genome. *Curr Opin Genet Dev,* 15**,** 589-94.

MELLO, D., DALEY, A. J., RAHMAN, M. S., QU, Y., GARLAND, S., PEARCE, C. & DEIGHTON, M. A. 2008. Vancomycin Heteroresistance in Bloodstream Isolates of *Staphylococcus capitis*. *Journal of Clinical Microbiology,* 46**,** 3124.

MENDIBURU, F. D. 2020. agricolae: Statistical Procedures for Agricultural Research. R package version 1.3-3 ed.

MENZEL, P., NG, K. L. & KROGH, A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications,* 7**,** 11257.

MERAY, Y., GENÇALP, D. & GÜRAN, M. 2018. Putting It All Together to Understand the Role of *Malassezia* spp. in Dandruff Etiology. *Mycopathologia,* 183**,** 893-903.

MERRIMAN, J. A., MUELLER, E. A., CAHILL, M. P., BECK, L. A., PALLER, A. S., HANIFIN, J. M., ONG, P. Y., SCHNEIDER, L., BABINEAU, D. C., DAVID, G., LOCKHART, A., ARTIS, K., LEUNG, D. Y. M. & SCHLIEVERT, P. M. 2016. Temporal and Racial Differences Associated with Atopic Dermatitis *Staphylococcus aureus* and Encoded Virulence Factors. *mSphere,* 1**,** e00295-16.

MILLS, K. J., HU, P., HENRY, J., TAMURA, M., TIESMAN, J. P. & XU, J. 2012. Dandruff/seborrhoeic dermatitis is characterized by an inflammatory genomic signature and possible immune dysfunction: transcriptional analysis of the condition and treatment effects of zinc pyrithione. *British Journal of Dermatology,* 166**,** 33-40.

MIRAGAIA, M., THOMAS, J., COUTO, I., ENRIGHT, M. & DE LENCASTRE, H. 2007. Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. *Journal of bacteriology,* 189**,** 2540-2552.

MIURA, Y., ISHIGE, I., SOEJIMA, N., SUZUKI, Y., UCHIDA, K., KAWANA, S. & EISHI, Y. 2010. Quantitative PCR of *Propionibacterium acnes* DNA in samples

aspirated from sebaceous follicles on the normal skin of subjects with or without acne. *Journal of medical and dental sciences,* 57**,** 65-74.

MOAWAD, A. A., HOTZEL, H., AWAD, O., ROESLER, U., HAFEZ, H. M., TOMASO, H., NEUBAUER, H. & EL-ADAWY, H. 2019. Evolution of Antibiotic Resistance of Coagulase-Negative Staphylococci Isolated from Healthy Turkeys in Egypt: First Report of Linezolid Resistance. *Microorganisms,* 7**,** 476.

MONK, I. R., SHAIKH, N., BEGG, S. L., GAJDISS, M., SHARKEY, L. K. R., LEE, J. Y. H., PIDOT, S. J., SEEMANN, T., KUIPER, M., WINNEN, B., HVORUP, R., COLLINS, B. M., BIERBAUM, G., UDAGEDARA, S. R., MOREY, J. R., PULYANI, N., HOWDEN, B. P., MAHER, M. J., MCDEVITT, C. A., KING, G. F. & STINEAR, T. P. 2019. Zinc-binding to the cytoplasmic PAS domain regulates the essential WalK histidine kinase of *Staphylococcus aureus*. *Nature Communications,* 10**,** 3067.

MONTEIRO, J. M., COVAS, G., RAUSCH, D., FILIPE, S. R., SCHNEIDER, T., SAHL, H.-G. & PINHO, M. G. 2019. The pentaglycine bridges of *Staphylococcus aureus* peptidoglycan are essential for cell integrity. *Scientific Reports,* 9**,** 5010.

MOOTZ, J. M., MALONE, C. L., SHAW, L. N. & HORSWILL, A. R. 2013. Staphopains modulate *Staphylococcus aureus* biofilm integrity. *Infection and immunity,* 81**,** 3227-3238.

MORAN, J. C., CRANK, E. L., GHABBAN, H. A. & HORSBURGH, M. J. 2016. Deferred Growth Inhibition Assay to Quantify the Effect of Bacteria-derived Antimicrobials on Competition. *JoVE***,** e54437.

MOYNIHAN, P. J. & CLARKE, A. J. 2010. O-acetylation of peptidoglycan in gram-negative bacteria identification and characterization of peptidoglycan o-acetyltransferase in *Neisseria gonorrhoeae*. *Journal of Biological Chemistry,* 285**,** 13264-13273.

MUELLER, N. T., BAKACS, E., COMBELLICK, J., GRIGORYAN, Z. & DOMINGUEZ-BELLO, M. G. 2015. The infant microbiome development: mom matters. *Trends in Molecular Medicine,* 21**,** 109-117.

MULCAHY, M. E., GEOGHEGAN, J. A., MONK, I. R., O'KEEFFE, K. M., WALSH, E. J., FOSTER, T. J. & MCLOUGHLIN, R. M. 2012. Nasal colonisation by *Staphylococcus aureus* depends upon clumping factor B binding to the squamous epithelial cell envelope protein loricrin. *PLoS Pathog,* 8**,** e1003092.

NAKAMURA, Y., OSCHERWITZ, J., CEASE, K. B., CHAN, S. M., MUÑOZ-PLANILLO, R., HASEGAWA, M., VILLARUZ, A. E., CHEUNG, G. Y., MCGAVIN, M. J. & TRAVERS, J. B. 2013. Staphylococcus δ-toxin induces allergic skin disease by activating mast cells. *Nature,* 503**,** 397-401.

NALMAS, S., BISHBURG, E., MEURILLIO, J., KHOOBIAR, S. & COHEN, M. 2008. *Staphylococcus capitis* prosthetic valve endocarditis: Report of two rare cases and review of literature. *Heart & Lung,* 37**,** 380-384.

NARRA, H. P. & OCHMAN, H. 2006. Of What Use Is Sex to Bacteria? *Current Biology,* 16**,** R705-R710.

NAUSHAD, S., BARKEMA, H., LUBY, C., CONDAS, L., NOBREGA, D., CARSON, D. & DE BUCK, J. 2016. Comprehensive Phylogenetic Analysis of Bovine Non-aureus Staphylococci Species Based on Whole-Genome Sequencing. *Frontiers in Microbiology,* 7.

NEUHAUS, F. C. & BADDILEY, J. 2003. A Continuum of Anionic Charge: Structures and Functions of Alanyl-Teichoic Acids in Gram-Positive Bacteria. *Microbiology and Molecular Biology Reviews,* 67**,** 686.

NURK, S., BANKEVICH, A., ANTIPOV, D., GUREVICH, A., KOROBEYNIKOV, A., LAPIDUS, A., PRJIBELSKY, A., PYSHKIN, A., SIROTKIN, A., SIROTKIN, Y., STEPANAUSKAS, R., MCLEAN, J., LASKEN, R., CLINGENPEEL, S. R., WOYKE, T., TESLER, G., ALEKSEYEV, M. A. & PEVZNER, P. A. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. *In:* DENG, M., JIANG, R., SUN, F. & ZHANG, X., eds. Research in Computational Molecular Biology, 2013// 2013 Berlin, Heidelberg. Springer Berlin Heidelberg, 158-170.

NURK, S., MELESHKO, D., KOROBEYNIKOV, A. & PEVZNER, P. A. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome research,* 27**,** 824-834.

O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBBERTSE, B., SMITH-WHITE, B., AKO-ADJEI, D., ASTASHYN, A., BADRETDIN, A., BAO, Y., BLINKOVA, O., BROVER, V., CHETVERNIN, V., CHOI, J., COX, E., ERMOLAEVA, O., FARRELL, C. M., GOLDFARB, T., GUPTA, T., HAFT, D., HATCHER, E., HLAVINA, W., JOARDAR, V. S., KODALI, V. K., LI, W., MAGLOTT, D., MASTERSON, P., MCGARVEY, K. M., MURPHY, M. R., O'NEILL, K., PUJAR, S., RANGWALA, S. H., RAUSCH, D., RIDDICK, L. D., SCHOCH, C., SHKEDA, A., STORZ, S. S., SUN, H., THIBAUD-NISSEN, F., TOLSTOY, I., TULLY, R. E., VATSAN, A. R., WALLIN, C., WEBB, D., WU, W., LANDRUM, M. J., KIMCHI, A., TATUSOVA, T., DICUCCIO, M., KITTS, P., MURPHY, T. D. & PRUITT, K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res,* 44**,** D733-45.

O'RIORDAN, K. & LEE, J. C. 2004. Staphylococcus aureus capsular polysaccharides. *Clinical microbiology reviews,* 17**,** 218-234.

O'SULLIVAN, J. N., REA, M. C., O'CONNOR, P. M., HILL, C. & ROSS, R. P. 2019. Human skin microbiota is a rich source of bacteriocin-producing staphylococci that kill human pathogens. *FEMS Microbiology Ecology,* 95.

O'NEILL, A. M., NAKATSUJI, T., HAYACHI, A., WILLIAMS, M. R., MILLS, R. H., GONZALEZ, D. J. & GALLO, R. L. 2020. Identification of a Human Skin Commensal Bacterium that Selectively Kills *Cutibacterium acnes*. *Journal of Investigative Dermatology,* 140**,** 1619-1628.e2.

OGAWA, E., SATO, Y., MINAGAWA, A. & OKUYAMA, R. 2018. Pathogenesis of psoriasis and development of treatment. *The Journal of Dermatology,* 45**,** 264-272.

OH, J., BYRD, A. L., DEMING, C., CONLAN, S., BARNABAS, B., BLAKESLEY, R., BOUFFARD, G., BROOKS, S., COLEMAN, H., DEKHTYAR, M., GREGORY, M., GUAN, X., GUPTA, J., HAN, J., HO, S.-L., LEGASPI, R., MADURO, Q., MASIELLO, C., MASKERI, B., MCDOWELL, J., MONTEMAYOR, C., MULLIKIN, J., PARK, M., RIEBOW, N., SCHANDLER, K., SCHMIDT, B., SISON, C., STANTRIPOP, M., THOMAS, J., THOMAS, P., VEMULAPALLI, M., YOUNG, A., KONG, H. H., SEGRE, J. A. & PROGRAM, N. C. S. 2014. Biogeography and individuality shape function in the human skin metagenome. *Nature,* 514**,** 59-64.

OH, J., BYRD, ALLYSON L., PARK, M., KONG, HEIDI H. & SEGRE, JULIA A. 2016. Temporal Stability of the Human Skin Microbiome. *Cell,* 165**,** 854-866.

OKONECHNIKOV, K., CONESA, A. & GARCÍA-ALCALDE, F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics (Oxford, England),* 32**,** 292-294.

OKU, Y., KUROKAWA, K., ICHIHASHI, N. & SEKIMIZU, K. 2004. Characterization of the *Staphylococcus aureus* mprF gene, involved in lysinylation of phosphatidylglycerol. *Microbiology,* 150**,** 45-51.

OLIVEIRA, H., SAMPAIO, M., MELO, L. D. R., DIAS, O., POPE, W. H., HATFULL, G. F. & AZEREDO, J. 2019. Staphylococci phages display vast genomic diversity and evolutionary relationships. *BMC Genomics,* 20**,** 357.

OLSON, N. D., TREANGEN, T. J., HILL, C. M., CEPEDA-ESPINOZA, V., GHURYE, J., KOREN, S. & POP, M. 2019. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Briefings in Bioinformatics,* 20**,** 1140-1150.

OMER, H., MCDOWELL, A. & ALEXEYEV, O. A. 2017. Understanding the role of *Propionibacterium acnes* in acne vulgaris: The critical importance of skin sampling methodologies. *Clinics in dermatology,* 35**,** 118-129.

OTTO, M. 2008. Staphylococcal Biofilms. *In:* T., R. (ed.) *Bacterial Biofilms.* Current Topics in Microbiology and Immunology: Springer, Berlin, Heidelberg.

OTTO, M. 2009. *Staphylococcus epidermidis* — the 'accidental' pathogen. *Nature Reviews Microbiology,* 7**,** 555-567.

OTTO, M. 2010. Staphylococcus colonization of the skin and antimicrobial peptides. *Expert review of dermatology,* 5**,** 183-195.

OTTO, M. 2013. Coagulase-negative staphylococci as reservoirs of genes facilitating MRSA infection: Staphylococcal commensal species such as *Staphylococcus epidermidis* are being recognized as important sources of genes promoting MRSA colonization and virulence. *Bioessays,* 35**,** 4-11.

OVERHOLT, W. A., HÖLZER, M., GEESINK, P., DIEZEL, C., MARZ, M. & KÜSEL, K. 2020. Inclusion of Oxford Nanopore long reads improves all microbial and viral metagenome-assembled genomes from a complex aquifer system. *Environmental Microbiology,* 22**,** 4000-4013.

PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T. G., FOOKES, M., FALUSH, D., KEANE, J. A. & PARKHILL, J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 31**,** 3691-3693.

PALMER, C. N. A., IRVINE, A. D., TERRON-KWIATKOWSKI, A., ZHAO, Y., LIAO, H., LEE, S. P., GOUDIE, D. R., SANDILANDS, A., CAMPBELL, L. E., SMITH, F. J. D., O'REGAN, G. M., WATSON, R. M., CECIL, J. E., BALE, S. J., COMPTON, J. G., DIGIOVANNA, J. J., FLECKMAN, P., LEWIS-JONES, S., ARSECULERATNE, G., SERGEANT, A., MUNRO, C. S., EL HOUATE, B., MCELREAVEY, K., HALKJAER, L. B., BISGAARD, H., MUKHOPADHYAY, S. & MCLEAN, W. H. I. 2006. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nature Genetics,* 38**,** 441-446.

PAPPAS, A., JOHNSEN, S., LIU, J.-C. & EISINGER, M. 2009. Sebum analysis of individuals with and without acne. *Dermato-endocrinology,* 1**,** 157-161.

PARK, H. K., HA, M.-H., PARK, S.-G., KIM, M. N., KIM, B. J. & KIM, W. 2012. Characterization of the fungal microbiota (mycobiome) in healthy and dandruff-afflicted human scalps. *PloS one,* 7**,** e32847-e32847.

PARK, J. Y., KIM, J. W., MOON, B. Y., LEE, J., FORTIN, Y. J., AUSTIN, F. W., YANG, S.-J. & SEO, K. S. 2015. Characterization of a novel two-component regulatory system, HptRS, the regulator for the hexose phosphate transport system in *Staphylococcus aureus*. *Infection and immunity,* 83**,** 1620-1628.

PARK, M., CHO, Y.-J., LEE, Y. W. & JUNG, W. H. 2018. Understanding the Mechanism of Action of the Anti-Dandruff Agent Zinc Pyrithione against *Malassezia restricta*. *Scientific reports,* 8**,** 12086-12086.

PARK, T., KIM, H.-J., MYEONG, N. R., LEE, H. G., KWACK, I., LEE, J., KIM, B. J., SUL, W. J. & AN, S. 2017. Collapse of human scalp microbiome network in dandruff and seborrhoeic dermatitis. *Experimental Dermatology,* 26**,** 835-838.

PARKS, D. H., IMELFORT, M., SKENNERTON, C. T., HUGENHOLTZ, P. & TYSON, G. W. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research,* 25**,** 1043-1055.

PARKS, D. H., RINKE, C., CHUVOCHINA, M., CHAUMEIL, P.-A., WOODCROFT, B. J., EVANS, P. N., HUGENHOLTZ, P. & TYSON, G. W. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology,* 2**,** 1533-1542.

PARSONS, J. B., YAO, J., FRANK, M. W., JACKSON, P. & ROCK, C. O. 2012. Membrane disruption by antimicrobial fatty acids releases low-molecular-weight proteins from *Staphylococcus aureus*. *Journal of bacteriology,* 194**,** 5294-5304.

PATRA, V., BYRNE, S. N. & WOLF, P. 2016. The Skin Microbiome: Is It Affected by UV-induced Immune Suppression? *Frontiers in microbiology,* 7**,** 1235-1235.

PATRA, V., WAGNER, K., ARULAMPALAM, V. & WOLF, P. 2019. Skin Microbiome Modulates the Effect of Ultraviolet Radiation on Cellular Response and Immune Function. *iScience,* 15**,** 211-222.

PÉCASTAINGS, S., ROQUES, C., NOCERA, T., PERAUD, C., MENGEAUD, V., KHAMMARI, A. & DRÉNO, B. 2018. Characterisation of Cutibacterium acnes phylotypes in acne and in vivo exploratory evaluation of Myrtacine®. *Journal of the European Academy of Dermatology and Venereology,* 32**,** 15-23.

PEHRSSON, E. C., TSUKAYAMA, P., PATEL, S., MEJÍA-BAUTISTA, M., SOSA-SOTO, G., NAVARRETE, K. M., CALDERON, M., CABRERA, L., HOYOS-ARANGO, W. & BERTOLI, M. T. 2016. Interconnected microbiomes and resistomes in low-income human habitats. *Nature,* 533**,** 212-216.

PENG, Y., LEUNG, H. C., YIU, S. M. & CHIN, F. Y. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics,* 28**,** 1420-8.

PESCHEL, A., JACK, R. W., OTTO, M., COLLINS, L. V., STAUBITZ, P., NICHOLSON, G., KALBACHER, H., NIEUWENHUIZEN, W. F., JUNG, G., TARKOWSKI, A., VAN KESSEL, K. P. M. & VAN STRIJP, J. A. G. 2001. *Staphylococcus aureus* Resistance to Human Defensins and Evasion of Neutrophil Killing via the Novel Virulence Factor Mprf Is Based on Modification of Membrane Lipids with l-Lysine. *Journal of Experimental Medicine,* 193**,** 1067-1076.

PESCHEL, A. & OTTO, M. 2013. Phenol-soluble modulins and staphylococcal infection. *Nature Reviews Microbiology,* 11**,** 667-673.

PESCHEL, A., OTTO, M., JACK, R. W., KALBACHER, H., JUNG, G. & GÖTZ, F. 1999. Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *J Biol Chem,* 274**,** 8405-10.

PEVZNER, P. A., TANG, H. & WATERMAN, M. S. 2001. An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences,* 98**,** 9748.

PIÉRARD-FRANCHIMONT, C., GOFFIN, V., DECROIX, J. & PIÉRARD, G. E. 2002. A multicenter randomized trial of ketoconazole 2% and zinc pyrithione 1% shampoos in severe dandruff and seborrheic dermatitis. *Skin Pharmacology and Physiology,* 15**,** 434-441.

PINHO, M. G., DE LENCASTRE, H. & TOMASZ, A. 2001. An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant staphylococci. *Proceedings of the National Academy of Sciences,* 98**,** 10886.

POLAK-WITKA, K., RUDNICKA, L., BLUME-PEYTAVI, U. & VOGT, A. 2020. The role of the microbiome in scalp hair follicle biology and disease. *Experimental Dermatology,* 29**,** 286-294.

PORAYATH, C., SURESH, M. K., BISWAS, R., NAIR, B. G., MISHRA, N. & PAL, S. 2018. Autolysin mediated adherence of *Staphylococcus aureus* with Fibronectin, Gelatin and Heparin. *International journal of biological macromolecules,* 110**,** 179-184.

PRAKASH, T. & TAYLOR, T. D. 2012. Functional assignment of metagenomic data: challenges and applications. *Briefings in bioinformatics,* 13**,** 711-727.

PRESCOTT, S. L., LARCOMBE, D.-L., LOGAN, A. C., WEST, C., BURKS, W., CARABALLO, L., LEVIN, M., ETTEN, E. V., HORWITZ, P., KOZYRSKYJ, A. & CAMPBELL, D. E. 2017. The skin microbiome: impact of modern environments on skin ecology, barrier integrity, and systemic immune programming. *The World Allergy Organization journal,* 10**,** 29-29.

PROCTOR, L. M., CREASY, H. H., FETTWEIS, J. M., LLOYD-PRICE, J., MAHURKAR, A., ZHOU, W., BUCK, G. A., SNYDER, M. P., STRAUSS, J. F., WEINSTOCK, G. M., WHITE, O., HUTTENHOWER, C. & THE INTEGRATIVE, H. M. P. R. N. C. 2019. The Integrative Human Microbiome Project. *Nature,* 569**,** 641-648.

PROKSCH, E., BRANDNER, J. M. & JENSEN, J. M. 2008. The skin: an indispensable barrier. *Experimental dermatology,* 17**,** 1063-1072.

QIN, N., DING, W., YAO, J., SU, K., WU, L. & LI, L. 2012. Genome sequence of *Staphylococcus capitis* QN1, which causes infective endocarditis. *Journal of bacteriology,* 194**,** 4469-4470.

QIN, Z., OU, Y., YANG, L., ZHU, Y., TOLKER-NIELSEN, T., MOLIN, S. & QU, D. 2007. Role of autolysin-mediated DNA release in biofilm formation of *Staphylococcus epidermidis*. *Microbiology,* 153**,** 2083-2092.

QUINCE, C., WALKER, A. W., SIMPSON, J. T., LOMAN, N. J. & SEGATA, N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology,* 35**,** 833-844.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

QURESHI, N. K., YIN, S. & BOYLE-VAVRA, S. 2014. The Role of the Staphylococcal VraTSR Regulatory System on Vancomycin Resistance and vanA Operon

Expression in Vancomycin-Resistant *Staphylococcus aureus*. *PLOS ONE,* 9**,** e85873.

RADIN, J. N., KELLIHER, J. L., PÁRRAGA SOLÓRZANO, P. K. & KEHL-FIE, T. E. 2016. The Two-Component System ArlRS and Alterations in Metabolism Enable *Staphylococcus aureus* to Resist Calprotectin-Induced Manganese Starvation. *PLOS Pathogens,* 12**,** e1006040.

RAJAGOPAL, M., WALKER, S. 2015. Envelope Structures of Gram-Positive Bacteria. *In:* BAGNOLI F., R. R. (ed.) *Protein and Sugar Export and Assembly in Gram-positive Bacteria.* Current Topics in Microbiology and Immunology: Springer, Cham.

RAPUN-ARAIZ, B., HAAG, A. F., SOLANO, C. & LASA, I. 2020. The impact of two-component sensorial network in staphylococcal speciation. *Current Opinion in Microbiology,* 55**,** 40-47.

RASIGADE, J.-P., RAULIN, O., PICAUD, J.-C., TELLINI, C., BES, M., GRANDO, J., BEN SAÏD, M., CLARIS, O., ETIENNE, J., TIGAUD, S. & LAURENT, F. 2012. Methicillin-Resistant *Staphylococcus capitis* with Reduced Vancomycin Susceptibility Causes Late-Onset Sepsis in Intensive Care Neonates. *PLOS ONE,* 7**,** e31548.

RAUSCH, M., DEISINGER, J. P., ULM, H., MÜLLER, A., LI, W., HARDT, P., WANG, X., LI, X., SYLVESTER, M., ENGESER, M., VOLLMER, W., MÜLLER, C. E., SAHL, H. G., LEE, J. C. & SCHNEIDER, T. 2019a. Coordination of capsule assembly and cell wall biosynthesis in *Staphylococcus aureus*. *Nature Communications,* 10**,** 1404.

RAUSCH, P., RÜHLEMANN, M., HERMES, B. M., DOMS, S., DAGAN, T., DIERKING, K., DOMIN, H., FRAUNE, S., VON FRIELING, J., HENTSCHEL, U., HEINSEN, F.-A., HÖPPNER, M., JAHN, M. T., JASPERS, C., KISSOYAN, K. A. B., LANGFELDT, D., REHMAN, A., REUSCH, T. B. H., ROEDER, T., SCHMITZ, R. A., SCHULENBURG, H., SOLUCH, R., SOMMER, F., STUKENBROCK, E., WEILAND-BRÄUER, N., ROSENSTIEL, P., FRANKE, A., BOSCH, T. & BAINES, J. F. 2019b. Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome,* 7**,** 133.

REDDY, T. B., THOMAS, A. D., STAMATIS, D., BERTSCH, J., ISBANDI, M., JANSSON, J., MALLAJOSYULA, J., PAGANI, I., LOBOS, E. A. & KYRPIDES, N. C. 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res,* 43**,** D1099-106.

REED, P., ATILANO, M. L., ALVES, R., HOICZYK, E., SHER, X., REICHMANN, N. T., PEREIRA, P. M., ROEMER, T., FILIPE, S. R., PEREIRA-LEAL, J. B., LIGOXYGAKIS, P. & PINHO, M. G. 2015. Staphylococcus aureus Survives with a Minimal Peptidoglycan Synthesis Machine but Sacrifices Virulence and Antibiotic Resistance. *PLOS Pathogens,* 11**,** e1004891.

REED, P., VEIGA, H., JORGE, A. M., TERRAK, M. & PINHO, M. G. 2011. Monofunctional Transglycosylases Are Not Essential for *Staphylococcus aureus* Cell Wall Synthesis. *Journal of Bacteriology,* 193**,** 2549.

RENNEBERG, J., RIENECK, K. & GUTSCHIK, E. 1995. Evaluation of Staph ID 32 system and Staph-Zym system for identification of coagulase-negative staphylococci. *Journal of clinical microbiology,* 33**,** 1150-1153.

RIEG, S., SABOROWSKI, V., KERN, W. V., JONAS, D., BRUCKNER-TUDERMAN, L. & HOFMANN, S. C. 2014. Expression of the sweat-derived innate defence antimicrobial peptide dermcidin is not impaired in *Staphylococcus aureus* colonization or recurrent skin infections. *Clin Exp Dermatol,* 39**,** 209-12.

RIGOULAY, C., ENTENZA, J. M., HALPERN, D., WIDMER, E., MOREILLON, P., POQUET, I. & GRUSS, A. 2005. Comparative analysis of the roles of HtrA-like surface proteases in two virulent *Staphylococcus aureus* strains. *Infection and immunity,* 73**,** 563-572.

RINKE, C., SCHWIENTEK, P., SCZYRBA, A., IVANOVA, N. N., ANDERSON, I. J., CHENG, J.-F., DARLING, A., MALFATTI, S., SWAN, B. K. & GIES, E. A. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature,* 499**,** 431-437.

RITTIÉ, L. 2016. Cellular mechanisms of skin repair in humans and other mammals. *Journal of cell communication and signaling,* 10**,** 103-120.

ROGERS, J. S., MOORE, A. E., MELDRUM, H. & HARDING, C. R. 2003. Increased scalp skin lipids in response to antidandruff treatment containing zinc pyrithione. *Archives of Dermatological Research,* 295**,** 127-129.

ROHDE, H., BURDELSKI, C., BARTSCHT, K., HUSSAIN, M., BUCK, F., HORSTKOTTE, M. A., KNOBLOCH, J. K. M., HEILMANN, C., HERRMANN, M. & MACK, D. 2005. Induction of *Staphylococcus epidermidis* biofilm formation via proteolytic processing of the accumulation-associated protein by staphylococcal and host proteases. *Molecular microbiology,* 55**,** 1883-1895.

ROHDE, H., FRANKENBERGER, S., ZÄHRINGER, U. & MACK, D. 2010. Structure, function and contribution of polysaccharide intercellular adhesin (PIA) to *Staphylococcus epidermidis* biofilm formation and pathogenesis of biomaterial-associated infections. *Eur J Cell Biol,* 89**,** 103-11.

ROTH, R. R. & JAMES, W. D. 1988. Microbial ecology of the skin. *Annu Rev Microbiol,* 42**,** 441-64.

ROUSSELLE, P., MONTMASSON, M. & GARNIER, C. 2019. Extracellular matrix contribution to skin wound re-epithelialization. *Matrix Biology,* 75-76**,** 12-26.

SABAT, A., BUDIMIR, A., NASHEV, D., SÁ-LEÃO, R., VAN DIJL, J., LAURENT, F., GRUNDMANN, H., FRIEDRICH, A. & MARKERS, E. S. G. O. E. 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eurosurveillance,* 18**,** 20380.

SAHL, H. G. & BIERBAUM, G. 1998. Lantibiotics: biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. *Annu Rev Microbiol,* 52**,** 41-79.

SALTER, S. J., COX, M. J., TUREK, E. M., CALUS, S. T., COOKSON, W. O., MOFFATT, M. F., TURNER, P., PARKHILL, J., LOMAN, N. J. & WALKER, A. W. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology,* 12**,** 87.

SANCHEZ, M., KOLAR, S. L., MÜLLER, S., REYES, C. N., WOLF, A. J., OGAWA, C., SINGHANIA, R., DE CARVALHO, D. D., ARDITI, M., UNDERHILL, D. M., MARTINS, G. A. & LIU, G. Y. 2017. O-Acetylation of Peptidoglycan Limits Helper T Cell Priming and Permits *Staphylococcus aureus* Reinfection. *Cell Host & Microbe,* 22**,** 543-551.e4.

SANGWAN, N., XIA, F. & GILBERT, J. A. 2016. Recovering complete and draft population genomes from metagenome datasets. *Microbiome,* 4**,** 8.

SAUNDERS, C. W., SCHEYNIUS, A. & HEITMAN, J. 2012. Malassezia Fungi Are Specialized to Live on Skin and Associated with Dandruff, Eczema, and Other Skin Diseases. *PLOS Pathogens,* 8**,** e1002701.

SAXENA, R., MITTAL, P., CLAVAUD, C., DHAKAN, D. B., HEGDE, P., VEERANAGAIAH, M. M., SAHA, S., SOUVERAIN, L., ROY, N., BRETON, L., MISRA, N. & SHARMA, V. K. 2018. Comparison of Healthy and Dandruff Scalp Microbiome Reveals the Role of Commensals in Scalp Health. *Frontiers in cellular and infection microbiology,* 8**,** 346-346.

SCHARSCHMIDT, T. C. & FISCHBACH, M. A. 2013. What Lives On Our Skin: Ecology, Genomics and Therapeutic Opportunities Of the Skin Microbiome. *Drug Discov Today Dis Mech,* 10.

SCHMIDT-ROSE, T., BRAREN, S., FÖLSTER, H., HILLEMANN, T., OLTROGGE, B., PHILIPP, P., WEETS, G. & FEY, S. 2011. Efficacy of a piroctone olamine/climbazol shampoo in comparison with a zinc pyrithione shampoo in subjects with moderate to severe dandruff. *International journal of cosmetic science,* 33**,** 276-282.

SCHMIEDER, R. & EDWARDS, R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England),* 27**,** 863-864.

SCHOENFELDER, S. M. K., LANGE, C., ECKART, M., HENNIG, S., KOZYTSKA, S. & ZIEBUHR, W. 2010. Success through diversity – How *Staphylococcus epidermidis* establishes as a nosocomial pathogen. *International Journal of Medical Microbiology,* 300**,** 380-386.

SCHOULS, L. M., SPALBURG, E. C., VAN LUIT, M., HUIJSDENS, X. W., PLUISTER, G. N., VAN SANTEN-VERHEUVEL, M. G., VAN DER HEIDE, H. G. J., GRUNDMANN, H., HECK, M. E. O. C. & DE NEELING, A. J. 2009. Multiple-Locus Variable Number Tandem Repeat Analysis of *Staphylococcus Aureus*: Comparison with Pulsed-Field Gel Electrophoresis and spa-Typing. *PLOS ONE,* 4**,** e5082.

SCHWARTZ, J. R. 2016. Zinc Pyrithione: A Topical Antimicrobial With Complex Pharmaceutics. *J Drugs Dermatol,* 15**,** 140-4.

SCHWARTZ, J. R., DEANGELIS, Y. M. & DAWSON JR, T. L. 2012. Dandruff and seborrheic dermatitis: a head scratcher. *Practical Modern Hair Science,* 1**,** 389-413.

SCZYRBA, A., HOFMANN, P., BELMANN, P., KOSLICKI, D., JANSSEN, S., DRÖGE, J., GREGOR, I., MAJDA, S., FIEDLER, J., DAHMS, E., BREMGES, A., FRITZ, A., GARRIDO-OTER, R., JØRGENSEN, T. S., SHAPIRO, N., BLOOD, P. D., GUREVICH, A., BAI, Y., TURAEV, D., DEMAERE, M. Z., CHIKHI, R., NAGARAJAN, N., QUINCE, C., MEYER, F., BALVOČIŪTĖ, M., HANSEN, L. H., SØRENSEN, S. J., CHIA, B. K. H., DENIS, B., FROULA, J. L., WANG, Z., EGAN, R., DON KANG, D., COOK, J. J., DELTEL, C., BECKSTETTE, M., LEMAITRE, C., PETERLONGO, P., RIZK, G., LAVENIER, D., WU, Y.-W., SINGER, S. W., JAIN, C., STROUS, M., KLINGENBERG, H., MEINICKE, P., BARTON, M. D., LINGNER, T., LIN, H.-H., LIAO, Y.-C., SILVA, G. G. Z., CUEVAS, D. A., EDWARDS, R. A., SAHA, S., PIRO, V. C., RENARD, B. Y., POP, M., KLENK, H.-P., GÖKER, M., KYRPIDES, N. C., WOYKE, T., VORHOLT, J. A., SCHULZE-LEFERT, P., RUBIN, E. M., DARLING, A. E., RATTEI, T. & MCHARDY, A. C.

2017. Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. *Nature Methods,* 14**,** 1063-1071.
SECOR, P. R., JENNINGS, L. K., JAMES, G. A., KIRKER, K. R., PULCINI, E. D., MCINNERNEY, K., GERLACH, R., LIVINGHOUSE, T., HILMER, J. K., BOTHNER, B., FLECKMAN, P., OLERUD, J. E. & STEWART, P. S. 2012. Phevalin (aureusimine B)Production by Staphylococcus aureus Biofilm and Impacts on Human Keratinocyte Gene Expression. *PLOS ONE,* 7**,** e40973.
SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.
SEEMANN, T. 2015. snippy: fast bacterial variant calling from NGS reads.
SEEMANN, T. 2020. *ABRicate (version 0.9.8): Mass screening of contigs for antimicrobial resistance or virulence genes. I* [Online]. Available: https://github.com/tseemann/abricate [Accessed].
SEGATA, N., IZARD, J., WALDRON, L., GEVERS, D., MIROPOLSKY, L., GARRETT, W. S. & HUTTENHOWER, C. 2011. Metagenomic biomarker discovery and explanation. *Genome biology,* 12**,** R60-R60.
SEGATA, N., WALDRON, L., BALLARINI, A., NARASIMHAN, V., JOUSSON, O. & HUTTENHOWER, C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods,* 9**,** 811-814.
SELANDER, R. K., CAUGANT, D. A., OCHMAN, H., MUSSER, J. M., GILMOUR, M. N. & WHITTAM, T. S. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and environmental microbiology,* 51**,** 873.
SHANNON, C. E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal,* 27**,** 379-423.
SHAPIRO, B. J. 2016. How clonal are bacteria over time? *Current Opinion in Microbiology,* 31**,** 116-123.
SHARIFI, F. & YE, Y. 2017. From Gene Annotation to Function Prediction for Metagenomics. *Methods Mol Biol,* 1611**,** 27-34.
SHARMA-KUINKEL, B. K., MANN, E. E., AHN, J.-S., KUECHENMEISTER, L. J., DUNMAN, P. M. & BAYLES, K. W. 2009. The *Staphylococcus aureus* LytSR Two-Component Regulatory System Affects Biofilm Formation. *Journal of Bacteriology,* 191**,** 4767.
SHARMA-KUINKEL, B. K., RUDE, T. H. & FOWLER, V. G., JR. 2016. Pulse Field Gel Electrophoresis. *Methods in molecular biology (Clifton, N.J.),* 1373**,** 117-130.
SHAW, L., GOLONKA, E., POTEMPA, J. & FOSTER, S. J. 2004. The role and regulation of the extracellular proteases of *Staphylococcus aureus*. *Microbiology,* 150**,** 217-228.
SHIBAGAKI, N., SUDA, W., CLAVAUD, C., BASTIEN, P., TAKAYASU, L., IIOKA, E., KUROKAWA, R., YAMASHITA, N., HATTORI, Y., SHINDO, C., BRETON, L. & HATTORI, M. 2017. Aging-related changes in the diversity of women's skin microbiomes associated with oral bacteria. *Scientific Reports,* 7**,** 10567.
SIMÃO, F. A., WATERHOUSE, R. M., IOANNIDIS, P., KRIVENTSEVA, E. V. & ZDOBNOV, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* 31**,** 3210-3212.
SIMPSON, E. H. 1949. Measurement of Diversity. *Nature,* 163**,** 688-688.

SINGH, A., GOERING, R. V., SIMJEE, S., FOLEY, S. L. & ZERVOS, M. J. 2006. Application of Molecular Techniques to the Study of Hospital Infection. *Clinical Microbiology Reviews,* 19**,** 512.

SLABY, B. M., HACKL, T., HORN, H., BAYER, K. & HENTSCHEL, U. 2017. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *The ISME journal,* 11**,** 2465-2478.

SLATKIN, M. 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics,* 9**,** 477-485.

SOARES, R. C., CAMARGO-PENNA, P. H., DE MORAES, V. C. S., DE VECCHI, R., CLAVAUD, C., BRETON, L., BRAZ, A. S. K. & PAULINO, L. C. 2016. Dysbiotic Bacterial and Fungal Communities Not Restricted to Clinically Affected Skin Sites in Dandruff. *Frontiers in Cellular and Infection Microbiology,* 6.

SOARES, R. C., ZANI, M. B., ARRUDA, A. C. B. B., ARRUDA, L. H. F. D. & PAULINO, L. C. 2015. Malassezia Intra-Specific Diversity and Potentially New Species in the Skin Microbiota from Brazilian Healthy Subjects and Seborrheic Dermatitis Patients. *PLOS ONE,* 10**,** e0117921.

SOLDEN, L., LLOYD, K. & WRIGHTON, K. 2016. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Current Opinion in Microbiology,* 31**,** 217-226.

SPARBER, F., DE GREGORIO, C., STECKHOLZER, S., FERREIRA, F. M., DOLOWSCHIAK, T., RUCHTI, F., KIRCHNER, F. R., MERTENS, S., PRINZ, I. & JOLLER, N. 2019. The skin commensal yeast Malassezia triggers a type 17 response that coordinates anti-fungal immunity and exacerbates skin inflammation. *Cell host & microbe,* 25**,** 389-403. e6.

SPARBER, F. & LEIBUNDGUT-LANDMANN, S. 2017. Host Responses to Malassezia spp. in the Mammalian Skin. *Frontiers in immunology,* 8**,** 1614-1614.

SPITTAELS, K.-J., ONGENA, R., ZOUBOULIS, C. C., CRABBÉ, A. & COENYE, T. 2020. *Cutibacterium acnes* Phylotype I and II Strains Interact Differently With Human Skin Cells. *Frontiers in Cellular and Infection Microbiology,* 10.

SPRATT, B. G. 2004. Exploring the concept of clonality in bacteria. *Genomics, Proteomics, and Clinical Bacteriology.* Springer.

SPRATT, B. G. & MAIDEN, M. C. 1999. Bacterial population genetics, evolution and epidemiology. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences,* 354**,** 701-710.

STENMARK, B., HELLMARK, B. & SÖDERQUIST, B. 2019. Genomic analysis of Staphylococcus capitis isolated from blood cultures in neonates at a neonatal intensive care unit in Sweden. *European Journal of Clinical Microbiology & Infectious Diseases,* 38**,** 2069-2075.

STEWART, E. J. 2012. Growing unculturable bacteria. *Journal of bacteriology,* 194**,** 4151-4160.

SUGAI, M., FUJIWARA, T., OHTA, K., KOMATSUZAWA, H., OHARA, M. & SUGINAKA, H. 1997. *epr*, which encodes glycylglycine endopeptidase resistance, is homologous to *femAB* and affects serine content of peptidoglycan cross bridges in *Staphylococcus capitis* and *Staphylococcus aureus*. *Journal of Bacteriology,* 179**,** 4311.

SUN, F., CHO, H., JEONG, D.-W., LI, C., HE, C. & BAE, T. 2011. Aureusimines in *Staphylococcus aureus* Are Not Involved in Virulence. *PLOS ONE,* 5**,** e15703.

SUN, Z., ZHOU, D., ZHANG, X., LI, Q., LIN, H., LU, W., LIU, H., LU, J., LIN, X., LI, K., XU, T., BAO, Q. & ZHANG, H. 2020. Determining the Genetic Characteristics of Resistance and Virulence of the "Epidermidis Cluster Group" Through Pan-Genome Analysis. *Frontiers in cellular and infection microbiology,* 10**,** 274-274.

SUNAGAWA, S., MENDE, D. R., ZELLER, G., IZQUIERDO-CARRASCO, F., BERGER, S. A., KULTIMA, J. R., COELHO, L. P., ARUMUGAM, M., TAP, J. & NIELSEN, H. B. 2013. Metagenomic species profiling using universal phylogenetic marker genes. *Nature methods,* 10**,** 1196-1199.

SUZUKI, S., KAKUTA, M., ISHIDA, T. & AKIYAMA, Y. 2014. GHOSTX: An Improved Sequence Homology Search Algorithm Using a Query Suffix Array and a Database Suffix Array. *PLOS ONE,* 9**,** e103833.

SYED, A. K., REED, T. J., CLARK, K. L., BOLES, B. R. & KAHLENBERG, J. M. 2015. *Staphlyococcus aureus* phenol-soluble modulins stimulate the release of proinflammatory cytokines from keratinocytes and are required for induction of skin inflammation. *Infection and immunity,* 83**,** 3428-3437.

SZWEDA, P., SCHIELMANN, M., KOTLOWSKI, R., GORCZYCA, G., ZALEWSKA, M. & MILEWSKI, S. 2012. Peptidoglycan hydrolases-potential weapons against Staphylococcus aureus. *Applied Microbiology and Biotechnology,* 96**,** 1157-1174.

TAGLIALEGNA, A., NAVARRO, S., VENTURA, S., GARNETT, J. A., MATTHEWS, S., PENADES, J. R., LASA, I. & VALLE, J. 2016. Staphylococcal Bap Proteins Build Amyloid Scaffold Biofilm Matrices in Response to Environmental Signals. *PLOS Pathogens,* 12**,** e1005711.

TAKEUCHI, F., WATANABE, S., BABA, T., YUZAWA, H., ITO, T., MORIMOTO, Y., KURODA, M., CUI, L., TAKAHASHI, M. & ANKAI, A. 2005. Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. *Journal of bacteriology,* 187**,** 7292-7308.

TANAKA, A., CHO, O., SAITO, C., SAITO, M., TSUBOI, R. & SUGITA, T. 2016. Comprehensive pyrosequencing analysis of the bacterial microbiota of the skin of patients with seborrheic dermatitis. *Microbiology and Immunology,* 60**,** 521-526.

TEAM, R. D. C. 2010. R: A language and environment for statistical computing. R Foundation for Statisitcal Computing

TETT, A., PASOLLI, E., FARINA, S., TRUONG, D. T., ASNICAR, F., ZOLFO, M., BEGHINI, F., ARMANINI, F., JOUSSON, O., DE SANCTIS, V., BERTORELLI, R., GIROLOMONI, G., CRISTOFOLINI, M. & SEGATA, N. 2017. Unexplored diversity and strain-level structure of the skin microbiome associated with psoriasis. *npj Biofilms and Microbiomes,* 3**,** 14.

TEVELL, S., HELLMARK, B., NILSDOTTER-AUGUSTINSSON, Å. & SÖDERQUIST, B. 2017. *Staphylococcus capitis* isolated from prosthetic joint infections. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology,* 36**,** 115-122.

THE UNIPROT CONSORTIUM 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research,* 47**,** D506-D515.

THOMAS, T., GILBERT, J. & MEYER, F. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial informatics and experimentation,* 2**,** 3-3.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res,* 22**,** 4673-80.

THURLOW, LANCE R., JOSHI, GAURI S., CLARK, JUSTIN R., SPONTAK, JEFFREY S., NEELY, CRYSTAL J., MAILE, R. & RICHARDSON, ANTHONY R. 2013. Functional Modularity of the Arginine Catabolic Mobile Element Contributes to the Success of USA300 Methicillin-Resistant *Staphylococcus aureus*. *Cell Host & Microbe,* 13**,** 100-107.

THURLOW, L. R., JOSHI, G. S. & RICHARDSON, A. R. 2012. Virulence strategies of the dominant USA300 lineage of community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA). *FEMS Immunology & Medical Microbiology,* 65**,** 5-22.

TONKIN-HILL, G., LEES, J. A., BENTLEY, S. D., FROST, S. D. W. & CORANDER, J. 2018. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome open research,* 3**,** 93-93.

TONKIN-HILL, G., MACALASDAIR, N., RUIS, C., WEIMANN, A., HORESH, G., LEES, J. A., GLADSTONE, R. A., LO, S., BEAUDOIN, C., FLOTO, R. A., FROST, S. D. W., CORANDER, J., BENTLEY, S. D. & PARKHILL, J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology,* 21**,** 180.

TÖRMÄ, H., LINDBERG, M. & BERNE, B. 2008. Skin barrier disruption by sodium lauryl sulfate-exposure alters the expressions of involucrin, transglutaminase 1, profilaggrin, and kallikreins during the repair phase in human skin in vivo. *Journal of investigative dermatology,* 128**,** 1212-1219.

TRIA, F. D. K., LANDAN, G. & DAGAN, T. 2017. Phylogenetic rooting using minimal ancestor deviation. *Nature Ecology & Evolution,* 1**,** 0193.

TRUONG, D. T., FRANZOSA, E. A., TICKLE, T. L., SCHOLZ, M., WEINGART, G., PASOLLI, E., TETT, A., HUTTENHOWER, C. & SEGATA, N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods,* 12**,** 902-903.

TSAI, Y.-C., CONLAN, S., DEMING, C., SEGRE, J. A., KONG, H. H., KORLACH, J. & OH, J. 2016. Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing. *mBio,* 7**,** e01948-15.

TURNER, G., MATHESON, J., LI, G. Z., FEI, X. Q., ZHU, D. & BAINES, F. 2013. Enhanced efficacy and sensory properties of an anti-dandruff shampoo containing zinc pyrithione and climbazole. *International journal of cosmetic science,* 35**,** 78-83.

TURNER, G. A., HOPTROFF, M. & HARDING, C. R. 2012. Stratum corneum dysfunction in dandruff. *International Journal of Cosmetic Science,* 34**,** 298-306.

TYSON, G. W., CHAPMAN, J., HUGENHOLTZ, P., ALLEN, E. E., RAM, R. J., RICHARDSON, P. M., SOLOVYEV, V. V., RUBIN, E. M., ROKHSAR, D. S. & BANFIELD, J. F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature,* 428**,** 37-43.

UBUKATA, K., KONNO, M. & FUJII, R. 1975. Transduction of drug resistance to tetracycline, chloramphenicol, macrolides, lincomycin and clindamycin

with phages induced from *Streptococcus pyogenes*. *The Journal of antibiotics,* 28**,** 681-688.
VENUS, M., WATERMAN, J. & MCNAB, I. 2010. Basic physiology of the skin. *Surgery (Oxford),* 28**,** 469-472.
VISSER, M. J., KELL, D. B. & PRETORIUS, E. 2019. Bacterial dysbiosis and translocation in psoriasis vulgaris. *Frontiers in cellular and infection microbiology,* 9**,** 7.
VOLLMER, W., BLANOT, D. & DE PEDRO, M. A. 2008. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews,* 32**,** 149-167.
VOLLMER, W. & TOMASZ, A. 2000. The pgdA gene encodes for a peptidoglycanN-acetylglucosamine deacetylase in *Streptococcus pneumoniae*. *Journal of Biological Chemistry,* 275**,** 20496-20501.
VOLLMER, W. & TOMASZ, A. 2002. Peptidoglycan N-acetylglucosamine deacetylase, a putative virulence factor in *Streptococcus pneumoniae*. *Infection and Immunity,* 70**,** 7176-7178.
VOLLMERS, J., WIEGAND, S. & KASTER, A.-K. 2017. Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! *PLOS ONE,* 12**,** e0169662.
VON WINTERSDORFF, C. J. H., PENDERS, J., VAN NIEKERK, J. M., MILLS, N. D., MAJUMDER, S., VAN ALPHEN, L. B., SAVELKOUL, P. H. M. & WOLFFS, P. F. G. 2016. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Frontiers in microbiology,* 7**,** 173-173.
VUONG, C., KOCIANOVA, S., VOYICH, J. M., YAO, Y., FISCHER, E. R., DELEO, F. R. & OTTO, M. 2004. A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *J Biol Chem,* 279**,** 54881-6.
VUONG, C., SAENZ, H. L., GÖTZ, F. & OTTO, M. 2000. Impact of the agr quorum-sensing system on adherence to polystyrene in *Staphylococcus aureus*. *J Infect Dis,* 182**,** 1688-93.
WALKER, A. W., DUNCAN, S. H., LOUIS, P. & FLINT, H. J. 2014. Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends in Microbiology,* 22**,** 267-274.
WALKER, J. N., CROSBY, H. A., SPAULDING, A. R., SALGADO-PABÓN, W., MALONE, C. L., ROSENTHAL, C. B., SCHLIEVERT, P. M., BOYD, J. M. & HORSWILL, A. R. 2013. The *Staphylococcus aureus* ArlRS Two-Component System Is a Novel Regulator of Agglutination and Pathogenesis. *PLOS Pathogens,* 9**,** e1003819.
WANG, C., LI, M., DONG, D., WANG, J., REN, J., OTTO, M. & GAO, Q. 2007a. Role of ClpP in biofilm formation and virulence of *Staphylococcus epidermidis*. *Microbes and Infection,* 9**,** 1376-1383.
WANG, L., CLAVAUD, C., BAR-HEN, A., CUI, M., GAO, J., LIU, Y., LIU, C., SHIBAGAKI, N., GUÉNICHE, A., JOURDAIN, R., LAN, K., ZHANG, C., ALTMEYER, R. & BRETON, L. 2015. Characterization of the major bacterial–fungal populations colonizing dandruff scalps in Shanghai, China, shows microbial disequilibrium. *Experimental Dermatology,* 24**,** 398-400.
WANG, R., BRAUGHTON, K. R., KRETSCHMER, D., BACH, T.-H. L., QUECK, S. Y., LI, M., KENNEDY, A. D., DORWARD, D. W., KLEBANOFF, S. J. & PESCHEL, A. 2007b. Identification of novel cytolytic peptides as key virulence

determinants for community-associated MRSA. *Nature medicine,* 13**,** 1510-1514.

WANG, R., KHAN, B. A., CHEUNG, G. Y., BACH, T.-H. L., JAMESON-LEE, M., KONG, K.-F., QUECK, S. Y. & OTTO, M. 2011. *Staphylococcus epidermidis* surfactant peptides promote biofilm maturation and dissemination of biofilm-associated infection in mice. *The Journal of clinical investigation,* 121**,** 238-248.

WANG, Y., KUO, S., SHU, M., YU, J., HUANG, S., DAI, A., TWO, A., GALLO, R. L. & HUANG, C.-M. 2014. *Staphylococcus epidermidis* in the human skin microbiome mediates fermentation to inhibit the growth of *Propionibacterium acnes*: implications of probiotics in acne vulgaris. *Applied microbiology and biotechnology,* 98**,** 411-424.

WARNER, R. R., SCHWARTZ, J. R., BOISSY, Y. & DAWSON, T. L. 2001. Dandruff has an altered stratum corneum ultrastructure that is improved with zinc pyrithione shampoo. *Journal of the American Academy of Dermatology,* 45**,** 897-903.

WATANABE, S., AIBA, Y., TAN, X.-E., LI, F.-Y., BOONSIRI, T., THITIANANPAKORN, K., CUI, B., SATO'O, Y., KIGA, K., SASAHARA, T. & CUI, L. 2018. Complete genome sequencing of three human clinical isolates of *Staphylococcus caprae* reveals virulence factors similar to those of *S. epidermidis* and *S. capitis. BMC Genomics,* 19**,** 810.

WEIDENMAIER, C. & LEE, J. C. 2017. Structure and Function of Surface Polysaccharides of *Staphylococcus aureus. In:* BAGNOLI, F., RAPPUOLI, R. & GRANDI, G. (eds.) *Staphylococcus aureus: Microbiology, Pathology, Immunology, Therapy and Prophylaxis.* Cham: Springer International Publishing.

WHEELER, W. C. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics,* 6**,** 363-367.

WHITLOCK, S. M., ENOS, C. W., ARMSTRONG, A. W., GOTTLIEB, A., LANGLEY, R. G., LEBWOHL, M., MEROLA, J. F., RYAN, C., SIEGEL, M. P., WEINBERG, J. M., WU, J. J. & VAN VOORHEES, A. S. 2018. Management of psoriasis in patients with inflammatory bowel disease: From the Medical Board of the National Psoriasis Foundation. *J Am Acad Dermatol,* 78**,** 383-394.

WICK, R. R., JUDD, L. M., GORRIE, C. L. & HOLT, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS computational biology,* 13**,** e1005595-e1005595.

WICKHAM, H. 2016. *ggplot2: Elegant Graphics for Data Analysis.*, Springer-Verlag New York.

WIKRAMANAYAKE, T. C., BORDA, L. J., MITEVA, M. & PAUS, R. 2019. Seborrheic dermatitis—Looking beyond Malassezia. *Experimental Dermatology,* 28**,** 991-1001.

WILLIAMS, H. C., DELLAVALLE, R. P. & GARNER, S. 2012. Acne vulgaris. *The Lancet,* 379**,** 361-372.

WIRTH, T., BERGOT, M., RASIGADE, J.-P., PICHON, B., BARBIER, M., MARTINS-SIMOES, P., JACOB, L., PIKE, R., TISSIERES, P., PICAUD, J.-C., KEARNS, A., SUPPLY, P., BUTIN, M., LAURENT, F., ADAMKOVA, V., BARKHAM, T., BECKER, K., BENNETT, D., CLARIS, O., CREECH, C. B., DE LENCASTRE, H., DEIGHTON, M., DENIS, O., FERGUSON, J., HUANG, Y.-C., KLINGENBERG, C., INGEBRETSEN, A., LAFERRIÈRE, C., DOS SANTOS, K. R. N., SCHRENZEL, J.,

SPILIOPOULOU, I., STEFANI, S., TAEKSOO, K., TARKKA, E., FRIEDRICH, A., VANDENBROUCKE-GRAULS, C., USSHER, J., VANDENESCH, F., WESTBLADE, L., LINDSAY, J., VANDENESCH, F., LARSEN, A. R., ZANGER, P., KAHL, B. C., AYMERICH, C. P., THE INTERNATIONAL CONSORTIUM FOR STAPHYLOCOCCUS CAPITIS NEONATAL, S. & THE, E. S. G. O. E. 2020. Niche specialization and spread of *Staphylococcus capitis* involved in neonatal sepsis. *Nature Microbiology,* 5**,** 735-745.

WISPLINGHOFF, H., ROSATO, A. E., ENRIGHT, M. C., NOTO, M., CRAIG, W. & ARCHER, G. L. 2003. Related clones containing SCCmec type IV predominate among clinically significant *Staphylococcus epidermidis* isolates. *Antimicrobial agents and chemotherapy,* 47**,** 3574-3579.

WONGSURAWAT, T., JENJAROENPUN, P., TAYLOR, M. K., LEE, J., TOLARDO, A. L., PARVATHAREDDY, J., KANDEL, S., WADLEY, T. D., KAEWNAPAN, B. & ATHIPANYASILP, N. 2019. Rapid sequencing of multiple RNA viruses in their native form. *Frontiers in microbiology,* 10**,** 260.

WOOD, D. E., LU, J. & LANGMEAD, B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology,* 20**,** 257.

WOOD, D. E. & SALZBERG, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology,* 15**,** R46.

WU, G., ZHAO, H., LI, C., RAJAPAKSE, M. P., WONG, W. C., XU, J., SAUNDERS, C. W., REEDER, N. L., REILMAN, R. A., SCHEYNIUS, A., SUN, S., BILLMYRE, B. R., LI, W., AVERETTE, A. F., MIECZKOWSKI, P., HEITMAN, J., THEELEN, B., SCHRÖDER, M. S., DE SESSIONS, P. F., BUTLER, G., MAURER-STROH, S., BOEKHOUT, T., NAGARAJAN, N. & DAWSON, T. L., JR. 2015. Genus-Wide Comparative Genomics of Malassezia Delineates Its Phylogeny, Physiology, and Niche Adaptation on Human Skin. *PLOS Genetics,* 11**,** e1005614.

XU, J., SAUNDERS, C. W., HU, P., GRANT, R. A., BOEKHOUT, T., KURAMAE, E. E., KRONSTAD, J. W., DEANGELIS, Y. M., REEDER, N. L., JOHNSTONE, K. R., LELAND, M., FIENO, A. M., BEGLEY, W. M., SUN, Y., LACEY, M. P., CHAUDHARY, T., KEOUGH, T., CHU, L., SEARS, R., YUAN, B. & DAWSON, T. L. 2007. Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proceedings of the National Academy of Sciences,* 104**,** 18730.

XU, Z., WANG, Z., YUAN, C., LIU, X., YANG, F., WANG, T., WANG, J., MANABE, K., QIN, O., WANG, X., ZHANG, Y. & ZHANG, M. 2016. Dandruff is associated with the conjoined interactions between host and microorganisms. *Scientific Reports,* 6**,** 24877.

XUE, H., LU, H. & ZHAO, X. 2011. Sequence diversities of serine-aspartate repeat genes among *Staphylococcus aureus* isolates from different hosts presumably by horizontal gene transfer. *PloS one,* 6**,** e20332-e20332.

YAHARA, K., SUZUKI, M., HIRABAYASHI, A., SUDA, W., HATTORI, M., SUZUKI, Y. & OKAZAKI, Y. 2021. Long-read metagenomics using PromethION uncovers oral bacteriophages and their interaction with host bacteria. *Nature Communications,* 12**,** 27.

YAMASAKI, K. & GALLO, R. L. 2008. Antimicrobial peptides in human skin disease. *European journal of dermatology : EJD,* 18**,** 11-21.

YANG, S.-J., BAYER, A. S., MISHRA, N. N., MEEHL, M., LEDALA, N., YEAMAN, M. R., XIONG, Y. Q. & CHEUNG, A. L. 2012. The Staphylococcus aureus two-

component regulatory system, GraRS, senses and confers resistance to selected cationic antimicrobial peptides. *Infection and immunity,* 80**,** 74-81.

YAO, Y., STURDEVANT, D. E. & OTTO, M. 2005. Genomewide analysis of gene expression in *Staphylococcus epidermidis* biofilms: insights into the pathophysiology of *S. epidermidis* biofilms and the role of phenol-soluble modulins in formation of biofilms. *Journal of Infectious Diseases,* 191**,** 289-298.

YUE, Y., HUANG, H., QI, Z., DOU, H.-M., LIU, X.-Y., HAN, T.-F., CHEN, Y., SONG, X.-J., ZHANG, Y.-H. & TU, J. 2020. Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. *BMC Bioinformatics,* 21**,** 334.