TÉCNICO LISBOA

# Addressing the relationship between *Staphylococcus aureus* abundance in the human skin microbiome and Atopic Dermatitis: a metagenomics approach

**Rafaela Zoé Carvalho Tavares**

Thesis to obtain the Master of Science Degree in

**Biological Engineering**

Supervisor: Prof. Dr. Rodrigo da Silva Costa

Supervisor: Prof. Dr. Isabel Maria De Sá Correira Leite de Almeida

## Examination Committee

Chairperson: Prof. Dr. Arsénio do Carmo Sales Mendes Fialho

Supervisor: Prof. Dr. Rodrigo da Silva Costa

Member of the Committe: Dr. Tina Keller Costa

**October 2018**

## Acknowledgements

I would like to thank Prof. Dr. Isabel Sá Correia and Prof. Dr. Rodrigo Costa for the opportunity to study about a topic that I feel pationate about, and also for the advices and instructions during this journey. I would also like to thank Prof. Dr. Carmo Fonseca, Prof. Dr. João Ferreira and Prof. Dr. Paulo Leal Filipe from FMUL for sharing their knowledge with me, in order to complement my critical opinion on Atopic Dermatitis.

During the semester I had the opportunity to participate in several meetings of Biological Sciences Research Group (BSRG) from iBB, and for that I would like to thank all of my collegues and Dr. Tina Keller Costa for sharing with me their knowledge and for the inputs to my thesis.

Last but not least, I would like to thank my family for the support during not only this period, but also during my journey in Instituto Superior Técnico.

## Abstract

The increased interest in the relationship between human microbiome and disease has contributed to the development of lower-cost sequencing techniques and software for the analysis of metagenomes. Several studies revealed a relationship between the severity of the disease Atopic Dermatitis and the bacterium *Staphylococcus aureus*. This thesis makes use of a metagenomics-driven approach to determine the extent to which this interaction affects the composition of the total microbiome of the human skin. To determine whether microbial community structure shifts according to higher *S. aureus* abundance in the skin microbiome, an *in silico* survey was performed on 183 samples representing various degrees of *S. aureus* incidence. Principal Coordinates Analysis (PCoA), revealed, in general, significant dissimilarities for skin samples with different percentage values of *S. aureus* and from different body sites, namely Volar Forearm, Antecubital and Popliteal creases. *Propionibacteirum acnes*, *Corynebacterium* sp., *Staphylococcus* sp. *and Staphylococcus succinus* were revealed to be among the most responsive taxa to *S. aureus* incidence regardless of skin location. This study revealed a thus far unknown, positive correlation between *S. aureus* and *S. succninus* abundances in the skin microbiome. Genome-wide analyses performed *in silico* revealed that, although these species possess several genes in common, there are significant differences in the number and types of genes involved in virulence, disease and defense, suggesting similar physiological aptitude along with divergent virulence strategies adopted by both species to dwell in the human skin microbiome.



**Keywords:** Metagenomics, Atopic Dermatitis, Human Skin Microbiome, *Staphylococcus aureus*, *Staphylococcus succinus*.

## Resumo

O aumento do interesse na relação entre o microbioma humano e doenças contribuiu ao desenvolvimento de técnicas de sequenciamento de baixo custo e softwares para a análise de metagenomas. Vários estudos revelaram uma relação entre a gravidade da doença Dermatite Atópica e a bactéria *Staphylococcus aureus*. Consequentemente, foi proposto, nesta tese, provar essa relação com base na análise metagenômica. Para investigar como a estrutura da comunidade microbiana muda em função da maior abundância de *S. aureus* no microbioma da pele, uma pesquisa *in silico* foi realizada em 183 amostras com vários graus de incidência de *S. aureus*. Com recurso à Análise de Coordenadas Principais (PCoA), revelaram-se semelhanças entre o microbioma das amostras de acordo com o valor percentual de *S. aureus* e o local de origem da amostra: antebraço, vincos antecubital e poplíteos. Em geral, foram reveladas dissimilaridades significativas entre os locais e entre amostras com diferentes percentagens de *S. aureus*. Os taxa microbianos mais responsivos ao aumento da abundância de *S. aureus* foram *Propionibacterium acnes*, *Corynebacterium* sp., *Staphylococcus* sp. e *Staphylococcus succinus*, independentemente da localização. Verificou-se a existência de uma correlação positiva entre as abundâncias de *S. aureus* e *S. succninus* em todas as amostras. Investigações genómicas determinaram que, embora estas espécies possuam vários genes em comum, existem diferenças significativas entre o número de genes envolvidos em virulência, doença e defesa. Isto sugere semelhante aptidão fisiológica, porém mecanismos de virulência distintos, adoptados por estas espécies no processo de colonização da pele.



**Palavras-Chave:** Metagenómica, Dermatite Atópica, Microbioma da Pele Humana, *Staphylococcus aureus*, *Staphylococcus succinus*.

# INDEX

## List of Tables

## List of Figures

## List of Abbreviations

**AD** - Atopic Dermatitis

**AMP** – Antimicrobial Peptide

**Ac** – Antecubital Creases

**Bp** – Base Pair

**c-di-GMP** – Cyclic Diguanylate

**DNA** - Deoxyrironucleic Acid

**FLG** - Filaggrin

**HMP** – Human Microbiome Project

**IPR** – InterPro Accession Number

**L1s** – L1 Retrotransposons

**OTU** – Operational Taxonomic Unit

**PAST** – Paleontological Statistics

**PCoA** – Principal Coordinates Analysis

**PCR** – Polymerase Chain Reaction

**Pc** – Popliteal Creases

**QC** – Quality Control

**RAST** – Rapid Annotation using Subsystem Technology

**rRNA** – Ribosomal Ribonucleic Acid

**RNA** – Ribonucleic Acid

**SCORAD** – Scoring Atopic Dermatitis

**SIMPER** - Similarity Percentage Analysis

**Taxa** – Taxonomy

**TBDRs** – TonB-dependent Receptors

**Vf** – Volar Forearm

**WGS** – Whole Genome Sequencing

# 1. Introduction

## 1.1 Microbiome and Metagenome

### 1.1.1 Microbiome

Microbiome is the assembly of all microorganisms (bacteria, archaea, eurkaryotes, and viruses), their genome (i.e., genes) and the surrounding environmental conditions in any given site (Marchesi & Ravel, 2015).

There is still little understanding on the community dynamics, the interactions between microbes, their hosts and their surrounding environment, and about microbial functions in most microhabitats on Earth, even though steep improvements have been made in recent years with the help of DNA sequencing technologies. The mix of advances in DNA/RNA, proteins, metabolite analytical platforms and computing technologies has changed the way in which microbial community analysis is performed (Alivisatos et al., 2015), usually with highly complex datasets generated which demand the development of hi-end data processing pipelines and the creation of novel methodology and terminology to define different and diverse objects of study, such as “microbiomes”, “metagenomes” and “operational taxonomic units”, to name a few terms that have recently emerged. These new advances in technology permit a faster characterization of microbial samples. However, it is still much more left to comprehend (Martiny et al., 2015).

Over the past years, the microbiome and its role in health and disease has become an essential focus in present research. It has exposed how importantly intertwined we are with our microbial passengers. Microbiomes can impact our health in many ways, influencing not only our metabolism but also mental and skin health. Some diseases may cause alterations to the microbiome such as allergies, asthma, cardiovascular diseases and cancer (Mulcahy-O’Grady & Workentine, 2016).

Studies of bacterial communities within the human body have found these communities to be far more diverse than it was previously thought (Fierer et al., 2008). In fact, this is true for both host-associated communities, including the microbiomes of nearly all plants and animals, and for all sorts of habitats, from low to high latitudes (Teixeira et al., 2010) (Gonzalez et al., 2011) (Hacquard et al., 2015).

### 1.1.2 Metagenome

Metagenome is the gathering of genomes and genes out of members in the microbiota (Marchesi & Ravel, 2015).

The studies of metagenomes have gained importance in the past years due to its usefulness in search of the properties of bacterial communities (Cho & Blaser, 2012). They have brought insight into the microbial communities regarding different environments such as aquatic ecosystems, human

skin and others. Several new tools were developed in order to understand the taxonomic composition of metagenomes (Lindgreen et al., 2016).

## 1.2 The Human Microbiome

Studies of the diversity of the human microbiome begun with Antonie van Leewenhoek during the 1680s, when he discovered that, when comparing his oral and fecal microbiota, there were differences between the samples depending on its microenvironment and health and disease state (K Ursell et al., 2013).

Defining the human microbiome has brought some complications due to earlier confusions on the terminology. In an anthropocentric perspective, microbiota refers to the microbial taxa associated with humans whereas the microbiome has been regarded elsewhere as the agglomeration of these microbes and their genes (K Ursell et al., 2013). Nowadays, consensus exists concerning the common use of the terms “microbiota” and “microbiome” as the sum of all microorganisms living in a particular site or sample, be it of human origin or not, while the pool of genomes present in any given microbiome is usually referred to as “metagenome”.

The human body contains somatic cells and symbiotic species, such as bacteria, fungi, viruses, protists, archaea and other microscopic organisms. This group is altogether described as the human microbiome, the agglomeration of all the microbes that live in the human body (Rivera-Amill, 2016) (Costello et al,. 2012).

Understanding and identifying the dominant and rare microbial community members, and the level of species diversity and richness in human body habitats is a subject of extreme relevance. This type of information will allow a better understanding of microbial diversity and stability across space and time in the human body as wel as their roles in the microbial community (Li et al., 2013) (Kuczynski et al., 2010). Our lifecycle stages, cultural settings, various body habitats and the variations in the human microbiome are still poorly explored (Caporaso et al., 2011). However, nowadays, there are already some insights into the human microbiome, including novel and emerging perspectives on the majority of so-far unculturable microorganisms that inhabit our body. These novel insights have been primarily enabled through the use of recently-developed, next generation sequencing technologies and hi-end bioinformatic pipelines that allow the processing of large amounts of data at ever-increasing speed (see below).

The majority of people share a small percentage of microbial species in the gut, oral and genital communities. A colossal diversity is found between samples taken from the same habitat in the same person. Even though each individual is unique so is its body sites, and the diversity of microbial communities in each site is usually large (Kuczynski et al., 2011). For example, both men and women hands have different numbers of species-level phylotypes (Kuczynski et al., 2010).

There is an interesting dynamic between the human microbiota and the environment. Human microbes flow in every surface that we have contact with everyday. Studies show that human fingertips can pass on microbes to keyboards, and that these communities are one example of what differentiates individuals. Another interesting fact is that the microbes that come from the ingestion of food, may be giving us an individual microbiome with “new genes to digest new foods” (K Ursell et

al.,2013).

Even though it is known that microbes are present all around the human body, there is still little information on what role they play in our health. This has become particularly more interesting to study since microbial dysbioses, the disruption of normal microbial community structure, is linked with several human diseases (Gevers et al., 2012). The diversity but also the abundance of microbial members is essential to human health and disease, because some disorders are linked to alterations in host-associated microbial communities such as malnutrition, inflammatory diseases and obesity (Costello et al., 2012). Improvements in sequencing technologies have allowed carrying out large-scale studies of microbial communities, which will help to answer the question of what constitutes a healthy human microbiome, characterize the variability among different humans and to provide guidelines to the study of microbes linked to specific diseases (Hamady, 2009) (Gevers et al., 2012). Regardless of extensive variations in methodologies, the characterization of the taxonomic but also functional characteristics in the human microbiome can help to understand the "normal" composition of microbial symbionts plus the connection of microbial communities and human diseases (Kuczynski et al., 2011).

Diseases, drug interactions and our own metabolism are affected by the human microbiome (Kuczynski et al., 2011). Therefore, in congruence with this knowledge, there are some studies made on human microbiome that disclose the fact that healthy individuals also have different microbes in different habitats (Consortium, 2013). However, this diversity is still difficult to explain, even though factors such as the environment, diet, genetics and microbial exposure could be involved, the human body could be seen as an ecosystem, and therefore human health is a result of what we do with our ecosystem (Consortium, 2013) (Costello et al., 2012).

Although there is already some knowledge about the human microbiome, many questions still remain to be answered adequately for several human microniches, such as "how many species live in a given body site?" (Methé et al., 2012) (K Ursell et al., 2013).

### 1.2.1 The Human Microbiome Project (HMP)

In order to improve the knowledge about the human microbiome, The Human Microbiome Project (HMP), was designed with the main purpose of understanding the composition of the normal microbiome of healthy individuals (Cho & Blaser, 2012). Understanding the normal skin bacterial variations such as intrapersonal, interpersonal, temporal and topographical can be used to calculate the statistical power to execute disease-related studies (Grice et al., 2009).

HMP has also examined the largest cohort and set of distinct body habitats, microbial communities and their connections with their human hosts. Therefore it tries to focus on reference genomes that will provide frameworks to metagenomic annotation and analysis, but it also creates a criterion of the microbial community and function from a person, which will set criteria for what we call healthy (Consortium, 2013) (Methé et al., 2012).

This project includes 16S rRNA gene sequence data, and also whole-genome shotgun metagenomic data of human individuals. This will allow the understanding of patterns in the microbial diversity across not only the human body but also in between individuals. The data collected by the

project allows testing of the concept of enterotypes in the human microbiome (Koren et al., 2013).

The 16S rRNA gene is common to bacteria and archaea but not to eukaryotes which possess an 18S subunit analogous to the prokaryotic 16S. The sequence analysis of this gene is an important part of understanding the characterization of bacterial diversity. It also has highly conserved regions that act as a molecular clock and a binding site for PCR primers (Grice & Segre, 2011). High-throughput approaches to sequencing the 16S rRNA gene revealed a higher number of bacterial diversity on the skin than if it was used the culture-based method (Kong, 2012). Therefore, cultivation-independent methods relying on the analysis of the skin microbial metagenome circumvent biases inherent to cultivation-dependent methods in the determination of microbial diversity.

Metagenomic approaches are still not capable of differentiating 16S rRNA genes that come from living or dead organisms. Another obstacle is that favouritism exists in these methods when PCR is used prior to 16S rRNA gene sequencing. They also fail to provide information regarding the gene content of flexible, open pan-genomes or plasmids (Grice & Segre, 2011) although several current improvements are being made in this regard. Shotgun metagenomic sequencing facilitates a strain-level examination of microbes within the microbial community, provided that enough phylogenetic resolution is obtained, which can only be possible when multiple genes are involved in the analysis. It helps to understand the various microbial community levels that go from kingdom to species and strain-level diversification (Byrd et al., 2017), besides providing information on the functional diversity of microbiomes. Some studies have shown that, for human skin samples, whole metagenome sequencing with tape-stripping is not only doable in clinical practices but at the same time it is minimally invasive and avoids the disruption of the skin surface with the use of a surgical blade (Chng et al., 2016).

Previous studies on phylogenetic marker genes have focused on taxonomic characteristics of different skin areas and disease stages. Yet these have provided little information on the ecosystem's functionality. On the other hand, metagenomic shotgun sequencing approaches the entire DNA in a sample, which allows characterizing not only the community's functional capacity but also the genomes for which no targeted amplicon strategies exist (Kong & Segre, 2015).

In the future, with the growth of genomic techniques, the investigation of microbial communities across all sorts of samples will become not only more feasible and non-invasive, which will help the monitoring of microbial communities in the human body and the health of vital organs such as the gut and the skin. Improvements in monitoring the microbial community will also lead to the discovery of novel therapies (Chng et al., 2016) (Cani, 2018).

## 1.3 The Human Skin Microbiome

The human skin is considered to be one of the largest organs in the human body and has three main layers: epidermis, dermis and hypodermis. It has many functions, such as: protection, heat regulation, control of evaporation, excretion and absorption. However, its main role is to serve as a physical barrier that protects our body from toxic substances and foreign organisms (Grice & Segre, 2011).

The skin microbiome contains multiple microorganisms - a density up to $10^7$ per $cm^2$ with diverse communities of bacteria, fungi, mites and viruses (Cooper et al.,2015), where the majority are harmless or even beneficial to their host (Grice & Segre, 2011).

The skin surface has a range of microenvironments with distinct features, like temperature, moisture, pH, sebum and topography. Other aspects that define different microenvironments are the density of hair follicles, glands, skin thickness and folds (Kong & Segre, 2015). Because of this heterogeneity across the human skin, a wide range of bacterial communities with different composition and diversity colonize the human body depending on the topography and physico-chemical characteristics of the skin (mostly associated with moist, dry and sebaceous microenvironments) as it can be seen in the Figure 1 (Grice, 2014).

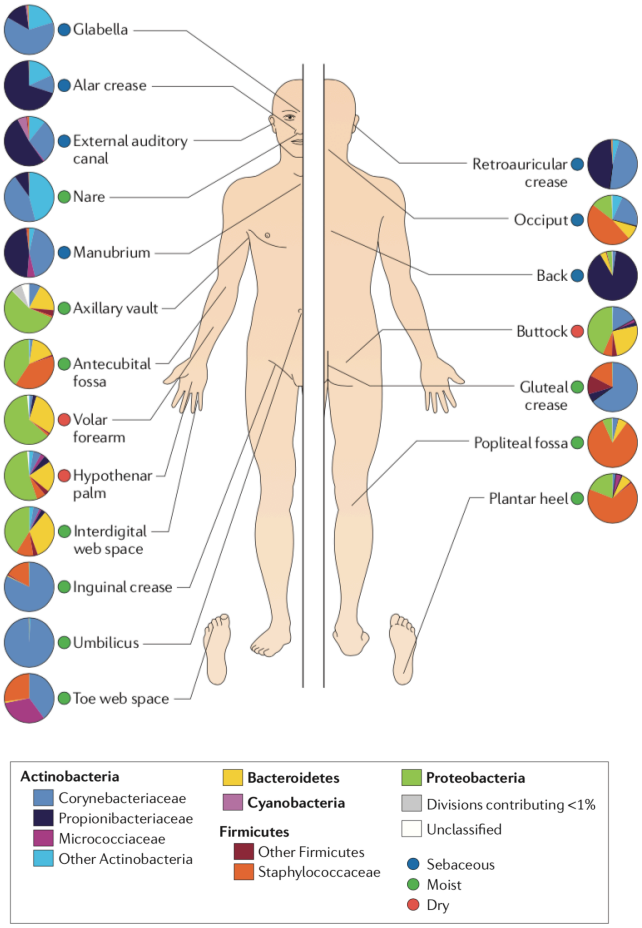



**Figure 1 | Topographycal distribution of bacteria on skin, depending on the microenvironment (sebaceous, moist or dry area).** The skin sites identified are preferred toward infections. For each site, there is a pie chart identifying the relative abundance of the principal microorganisms present. Respective phylum is described in bold (Grice & Segre, 2011).

Nevertheless, since the skin is continuously exposed to the environment, it becomes a struggle to identify what species are transient or residents of the community (Grice & Segre, 2011). There are two main factors that can influence the variation in the skin microbiome: environment and host. Host factors are specific to each individual such as age, gender and skin location. Age is the factor that has the most impact on the microenvironment of the skin. For example, during puberty changes in sebum are very common. Gender differences, physiological and anatomical, such as

hormonal production, sebum and sweat are also thought to influence the composition of microbial communities inhabiting the human skin (Grice & Segre, 2011). Environmental factors, which can impact the colonization of skin by bacteria, are the clothing choice, cosmetics, soaps, hygiene products, moisturizers and antibiotic use. “These products may alter the conditions of the skin barrier however the effects on the skin microbiota remain unclear” (Grice & Segre, 2011). Therefore, it is challenging to define what constitutes a “healthy” bacterial community (Fierer et al., 2008).

The perception that we have of the skin ecosystem is very important to understand the relationship between host and microorganisms. Perturbations disturbing the host-microorganism relationship can be endogenous or exogenous. Endogenous are perturbations that are related to, for example, genetic variations; on the other hand, exogenous perturbations are related to external factors such as hand washing (Figure 2)(Grice & Segre, 2011).

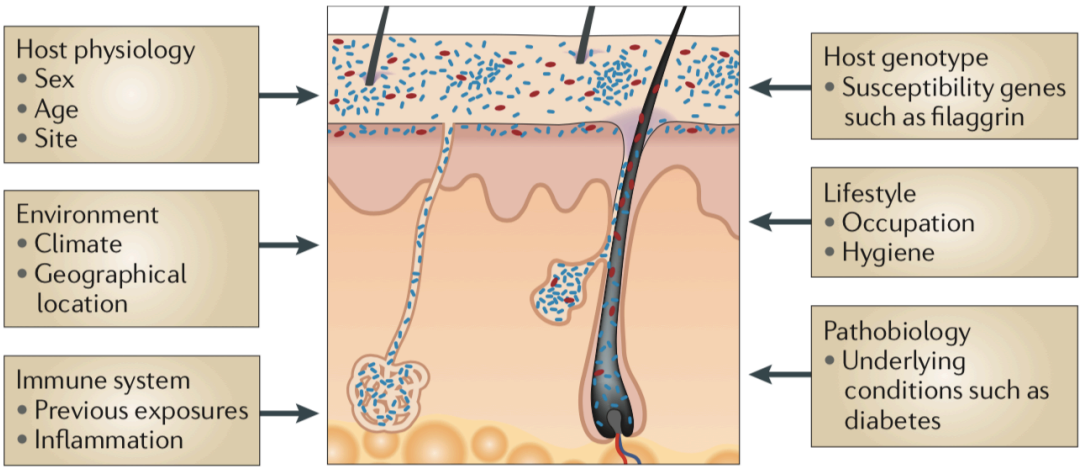



**Figure 2 | Principal exogenous and endogenous factors that contribute to variation of skin microbiome** (Grice & Segre, 2011).

As mentioned above, it has been already acknowledged that there are different bacterial phyla on normal human skin. However, there are a few types that have larger quantities than others, such as: *Actinobacteria*, *Proteobacteria*, *Bacteroidetes* and *Firmicutes* (Thomas et al., 2017). Bacterial taxa frequently found on skin surfaces display an intra- and inter-individual variability: intra-individual variation in skin microbial communities is less pronounced than inter-individual variation (Grice & Segre, 2011).

It is known that the primarily bacterial colonizers of the skin are *Staphylococcus epidermidis* and other coagulase-negative *staphylococci*, which are predominant in sebaceous areas. *Corynebacteria* is mostly found in moist sites, however, *staphylococci* sometimes can also be found in these sites. Furthermore, non-bacterial microorganisms like *Malassezia* spp. (fungal species) also prevail in sebaceous areas. Other non-bacterial species to be considered are the *Demodex* mites, which are microscopic arthropods that tend to feed on sebum and are commonly found during puberty as they prefer to colonize in sebaceous areas of the face, they may also feed on epithelial cells or other organisms that reside in the same space as the *Demodex*.

A significant portion of the skin microbiome can infiltrate the subepidermal area and the deeper dermal stroma where they can cooperate with a variety of host cell types, thus, giving an

opportunity to control the behaviours of cells underneath the surface, some examples of this are: capacity of controlling the inflammation after injuries, increase the expression of host antimicrobial peptides and mature T cell responses (Figure 3)(Gallo, 2016).

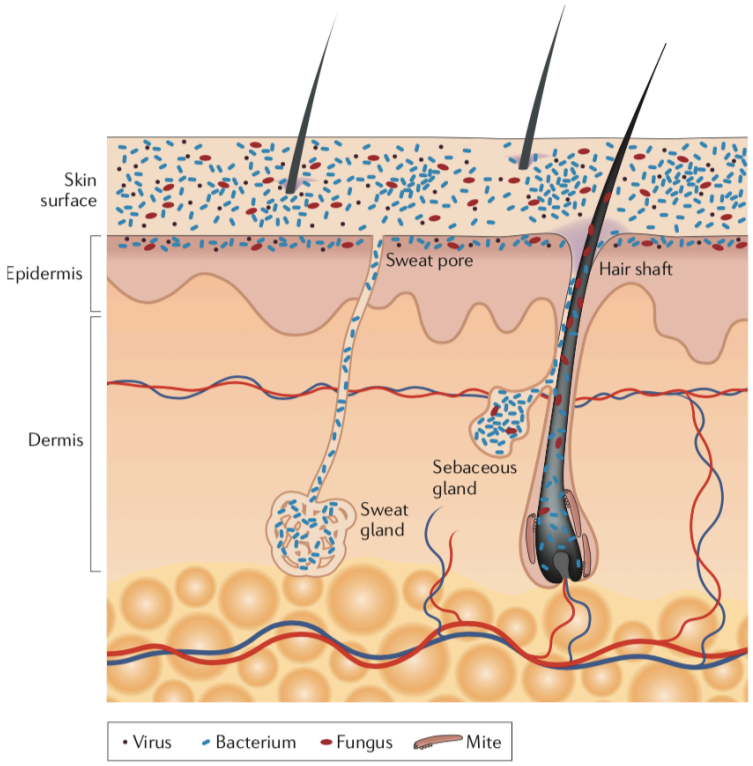



**Figure 3 | Main layers of the skin viewed in cross-section with microorganisms.** Different microorganisms (viruses, bacteria, and fungus) and mites habit not only the skin surface but also epidermis and dermis through hair follicules and glands (Grice & Segre, 2011).

The adaptive immune responses can balance the skin microbiota but this microbiota can also work to educate the immune system. Another very important aspect in immune response is the relationship between the microorganisms. Several studies have demonstrated that many dynamic interactions can benefit the host. The interactions between bacteria and the host are favourable because bacteria can produce antimicrobial peptides (AMP), a defence mechanism, and other molecules that try to resist co-colonisation by pathogens (Kong & Segre, 2015).

Host interactions can be sustained through sensing and signalling mechanisms, metabolic pathways or immunogenic features, which are likely to exhibit site-specificity (Kong & Segre, 2015).

The dysregulation of the skin immune responses is common in many skin disorders but how this affects the microbiotia remains unclear, due to the little understanding of which organisms create a balance, how this relates to the genetic and environmental variation (Grice & Segre, 2011). Skin diseases are linked to many factors like stages of life, topographical location and specific microorganisms (Grice & Segre, 2011) (Kong & Segre, 2015).

As different molecular methods had been developed in order to identify microorganisms, so has our understanding that the skin bacteria can be diverse and variable between human individuals. This understanding was also used to gain insight into the microbial involvement in human skin disorders and to discover antimicrobial and pro-microbial therapeutic approaches (Grice & Segre,

2011). In order to test the therapeutic potential of managing the microbiome in skin disorders, an elucidation of the baseline skin microbiome is needed (Grice et al., 2009).

There are several common skin disorders that demand an underlying microbial contribution like intensifying disease severity or facilitating transitions from opportunistic to pathogenic, as a result of clinical improvement that has been proven to be correlated with antimicrobial treatments. There are many ways in which skin disorders have been associated with specific organisms: when a skin disorder has a correlation to the microbiota, when a skin disorder has an unidentified microbial component or when a skin commensal becomes invasive which origins infections. In recent years, scientists have realized that commensal microorganisms were not simple passengers notwithstanding they have an important role in our physiology including our immune responses and metabolism (Grice & Segre, 2011).

## 1.4 Metagenomics: the Study of Metagenomes

Handelsman and colleagues were the first to coin the term “metagenomics” (Handelsman et al.,1998), however in a DNA recombination (i.e. cloning) context, therefore as “an approach where random fragments of environmental DNA are cloned into a suitable vector for maintenance in a surrogate host for functional screening, looking for gain of function in the surrogate host” (Marchesi & Ravel, 2015). Nowadays, the field of metagenomics has grown large and involves all sorts of cultivation-independent approaches and techniques used to analyse the collective pool of all genomes, the metagenome, in a given sample. Thus, the metagenome is the genetic material that represents the genomes of a microbial community and provides a view into the functional potential of the consortium. It has also become a powerful tool in microbiology research and further clinical applications.

The Metagenomics approach has the ability to identify and characterize bacterial and viral pathogens. Some of its advantages consist in the ability to reduce costs and time, while it increases sensitivity to discover new pathogens that cannot be detected so easily by the culture-based methods. It is a promising approach for the microbial diagnostics field (Mulcahy-O’Grady & Workentine, 2016).
The ability to study the composition and dynamics of the microbial communities has improved with the development of high-throughput sequencing technologies, bringing the capacity to generate metagenomics data enabling taxonomic and functional analysis (Waldor et al., 2015). However, even with all the improvements, high-throughput DNA sequencing generates a high volume of sequencing data generated and so there are still some technical, analytical and computational challenges (Kong & Segre, 2015) (Mitchell et al., 2016).

With passing years, knowledge about sequencing technologies has improved, increasing our capacity to identity and characterize microbial communities colonizing the skin and their mechanisms of interactions, including pathogenic and mutualistic ones (Figure 4). Furthermore, our knowledge of healthy skin microbiota and how it changed skin disorders is growing. In spite of the growth of our knowledge in sequencing technology and healthy cutaneous microbiota, further investigation is still needed to understand how this could help diagnostics, prognostics and therapeutics, as they are promising in order to manage and treat dermatological diseases (Grice, 2014).

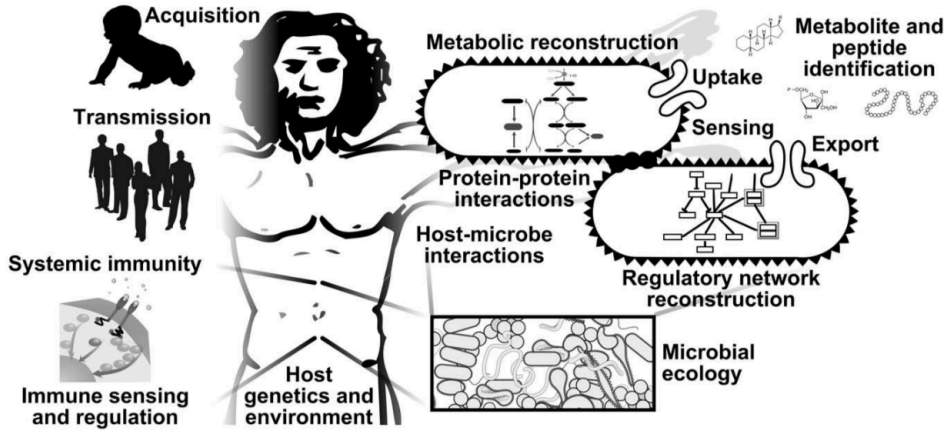



**Figure 4** | **Human Microbiome and bioinformatics efforts.** Bioinformatic tools have been helping the improvement of the study of the human microbiome as well as its microorganism-host and microbe-microbe relationships (Gevers et al., 2012).

Culture-based methods rely on the cultivation and isolation of microorganisms, however, less than 10% are able to survive in standard laboratory conditions. They also exclude microbes that depend on microbe-microbe interactions or synthrophy to thrive. However, it is still a relevant method to provide insights on bacterial, fungal and viral populations on the skin (Grice, 2014), in spite of being in general time-consuming.

With the use of DNA sequencing, the accuracy and the precision of characterization and analysis of microbiomes becomes less biased in comparison with culture-based approaches. Current technologies allow the sequencing of nucleotides at the gigabase scale, resulting in an overview of the structural and functional diversity of a microbial community. One of the major advantages of shotgun metagenomics is the non-targeted nature of the sequencing. Once all the DNA sequences are collected, this holds the potential to answer a series of questions such as the metabolic pathways, antimicrobial resistance genes, presence of pathogens and overall community composition (Mulcahy-O'Grady & Workentine, 2016).

As already mentioned above, a popular approach used to identify bacterial populations in a community is through sequencing the small 16S subunit of the ribosomal RNA gene, as it will allow the identification of the bacteria or archaea present in a sample. This method is also a molecular technique and the basis for current taxonomic classification of microbes based on molecular markers (Grice, 2014). Whereas a 16S rRNA gene-based approach will deliver information on the taxonomic composition of a given microbiome, untargeted metagenomics will explore the distribution and abundance of multiple genes in a given microsetting, providing relevant information on the functional and taxonomic attributes of a microbial community.

### 1.4.1 Metagenomics for Diagnosis of Infectious Disease

There are several benefits that can be obtained from metagenomics data sets in order to improve the clinical applications, for example, fast and easy identification of pathogens. Moreover, this method may be also beneficial to other applications, such as study of antibiotic-resistance genes,

diagnosis and many diseases associated with bacterial, viral and fungal microbiomes. Metagenomics also can support the study of more complex phenotypes that are correlated with disorders and disruptions of the skin microbiome. Even though some disorders do not correlate with single pathogens, they emerge from the relationships of the microbiota present in the samples (Mulcahy-O'Grady & Workentine, 2016).

## 1.5 Atopic Dermatitis

Atopic Dermatitis (AD), also known as atopic eczema, is a chronic skin disorder that has many different causes such as skin barrier dysfunction, decreased immune responses and microbial skin colonization (Kim et al., 2017). It is a cyclical disease, with flare periods that induce cutaneous phenotypes and non-flare periods where the skin has no signs of infections (Figure 5)(Chng et al., 2016). To measure AD severity, it is common to use the SCORing Atopic Dermatitis (SCORAD) index (Figure 5) (Kong, 2012). Some factors such as the environment and human genetics affect the way allergic diseases such as AD express themselves.

Some studies suggest that AD is usually found in areas such as the antecubital and popliteal sites that have similar microorganisms, although different composition of microbial communities.

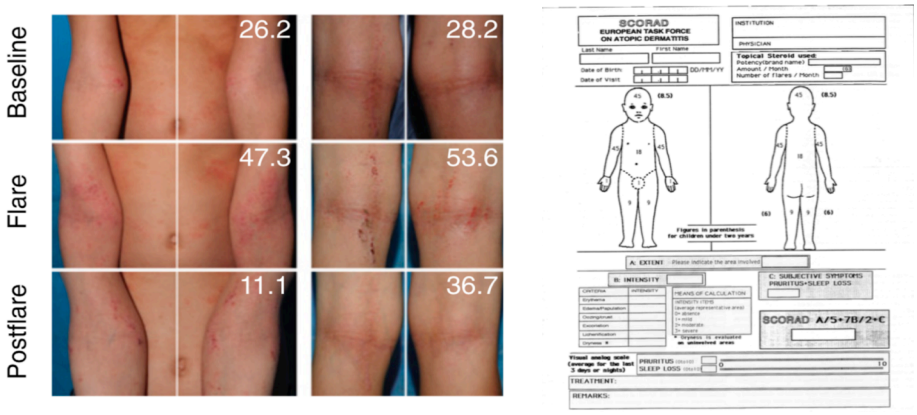



**Figure 5 | AD disease severity and SCORAD.** Representative clinical images of the antecubital (AC, left) and popliteal creases in patients with overall disease severity scores (objective SCORAD) and SCORAD evaluation sheet (Kong, 2012).

Scientists believe that the emergence of AD is linked to a balance of the microbial community. Therefore, a clear understanding of how AD may influence the microbiome and which microorganisms influence severity is needed, in order to decide which is the "best" treatment (Thomas et al., 2017). Currently, there is no established "best" treatment to AD, although studies of the skin microbiome may help in the pursuit of alternative therapies (Grice, 2014).

AD patients have a typical skin microbiome. It has been proven that this microbiota impacts the skin microenvironment allowing pathogenic microbial colonization and stimulating a pro-inflammatory skin phenotype. Bacteria linked with AD have different roles. Therefore, being able to understand those roles and their contribution to disease emergence may be useful in future development of interventions to restore the microbial balance, but also to improve the general health of the skin while avoiding AD flares (Kong et al., 2017).

### 1.5.1 Characteristics of Bacterial Communities Associated with Atopic Dermatitis

To better understand characteristics of microbial communities associated with AD, several scientists measured their diversity using for instance the Shannon diversity index, a measure tool that considers not only the richness (i.e., the total number of bacterial types), but also its evenness, the relative proportion of these bacterial types. It was concluded that there is a strong association between an increase in severity of AD and a lower diversity of the microbial community (Kong, 2012).

In several studies, *Staphylococcus aureus* has been revealed to be associated with AD, because it is usually found in lesional and nonlesional AD skin. The *S. aureus* bacterium is especially abundant during AD flares. However its colonization of the host skin has not been yet clarified as being a cause or effect (Grice, 2014). Neverthless, it is known that there is a correlation between an increase of *S. aureus*, the emergence of AD and a decrease in microbial diversity (Kong, 2012). A decrease in *S. aureus* is, usually, followed by an increase in *Propionibacterium*, *Streptococcus* and *Corynebacterium,* which can be used as an indicator of the recovery from deterioration of AD (Kim et al., 2017). More severe cases of AD flares showed a higher presence of *Staphylococcus aureus* where less severe cases had a *Staphylococcus epidermidis* presence. *S.epidermidis* has been associated to infections due to the fact that it is permanent omnipresent colonizer of the skin (Otto, 2009). It has also been proven that patients with more severe cases have a higher percentage of *S. aureus* strains whereas less severe cases display a more heterogeneous *S. epidermidis* strain community (Gallo, 2016).

*S. epidermidis* engages not only directly but also indirectly with the innate and adaptive immune barrier to pathogens (Gallo, 2016). Researchers have favoured the hypothesis that *S. epidermidis* is not causing the disease but rather keeping a relationship with the host. It is also believed that these bacteria may be beneficial during infection. Some ideas reference the fact that *S.epidermidis* might function as a probiotic that can prevent colonization of bacteria such as *S. aureus*; however, there is still no clear evidence of these ideas (Otto, 2009). Questions still remain on why it is advantageous to *S.epidermidis* not to cause severe infections by maintaining a low level of virulence as it opposes to its “cousin” *S.aureus*. Even though some studies have tried to prove the fact that *S. epidermidis* may out-compete *S. aureus*, there is still no evidence *in vivo* (Otto, 2009).

Another interesting fact is that AD-associated skin microbiome generates excessive ammonia, which explains a high skin pH, which may favour the prevalence of pathogenic bacteria. Some studies suggest that a reduction of ammonia-oxidizing bacteria may increase the risks of AD (Kong et al., 2017). It also has been proven in previous studies that children with a lower diversity in the gut microbiota may develop atopic eczema in later years, suggesting a connection between gut microbiota and the skin (Grice, 2014).

### 1.5.2 Immunne Responses in Atopic Dermatitis

Bacteria, present in human skin, interact with cells below the surface, inducing cells behavior in controlling the immune responses (Gallo, 2016).

Antimicrobial peptides (AMPs) are oligopeptides with a varying number of amino acids, (Bahar & Ren, 2013) also known as immune defense molecules, that come from the production of mast, paneth, epithelials cells, neutrophils and adipocytes. They try to control the growth of microorganisms that make up the skin microbiome. When the balance of the microbiota is altered it can lead to dybiosis, which can consequently lead to a disruption of the immune homeostasis and the increase of disease symptoms. Even though there is some uncertainty of this occurrence, some patients with AD have decreased capacity to produce some AMPs like β-defensis and cathelicidins even in the presence of an inflammation. This deficiency could be related to a resistance in the growth of pathogens. Having high numbers of *S. aureus* in AD could lead to pathophysiology of AD which is related to immune dysfunction, decrease of AMPs, disruption of the skin barrier and an aggravation of allergic reactions (Nakatsuji et al., 2017). *S. epidermidis* can reduce the inflammation after injuries, promoting the expression of AMPs and increasing the development of cutaneous T cells (Nakatsuji et al., 2017). Exotoxins produced by *S.aureus* usually have super antigenic properties that can stimulate T-cells directly (Thomas et al., 2017).

### 1.5.3 Mutations Impact in Atopic Dermatitis

There are several risk factors of AD such as the loss of function derived from a mutation in the the gene encoding filaggrin (FLG), which is an important factor of terminal differentiation and skin barrier function since it regulates pH and hydration of the epidermis (Chng et al., 2016). This type of mutation is associated with the increase of AD severity and *S. aureus* colonization (Gonzalez et al., 2016).

### 1.5.4 Treatments Impacts in the Microbial Community

Nowadays, Atopic Dermatitis can be treated with systematic antibiotics, combination of topical corticosteroids and diluted bleach bath (Kong, 2012).

Some studies claim that using a combination of topical steroids and antibiotic treatments can eliminate *S. aureus* from some patients with AD. Other studies say that with the use of oral antibiotics it is possible to reduce bacterial colonization and improve AD severity. However, it is only temporary as the benefits did no last longer than three months (Baranska-Rybak et al., 2011). The application of antimicrobial coagulase-negative *staphylococci* (CoNS) to patients with AD can reduce the colonization of *S. aureus*. This shows that commensal skin bacteria can protect against pathogens but also how dybiosis in the skin microbiome can progress to a disease (Nakatsuji et al., 2017).

### 1.5.5 Objectives and Research Questions

The main goal of this study was to address the relationship between *Staphylococcus aureus* abundance and shifts in the structure of the human skin microbiome, with implications to our understanding of Atopic Dermatitis (AD), using a metagenomic approach. In order to have an accurate answer to the main question, several other questions had to be made.

It is already known that a high abundance of *S. aureus* in human skin is very common in AD patients. However, is there any relationship between *S. aureus* and the microbial community? In order to answer this question the response of phylum taxonomic abundances in the regions susceptible to AD: Antecubital Creases (Ac), Popliteal Creases (Pc) and Volar Forearm (Vf) samples was analysed, with charts created with the relative frequency data - percentage of the reads per phylum and per sample per site

If there is any relationship between *S. aureus* and the human skin microbiome, "which are the main taxonomic and functional components of the skin microbiome that respond most to high abundance of *S. aureus*?", "Are the effects equal for all sites?", " Does the bacterial community vary between sites and according to the abundance of *S. aureus*?". Through multivariate statistics using Principal Coordinates Analysis (PCoA), the formation of skin sample clusters along with the percentage of *S. aureus* in the samples was examined. Since the ordination diagrams obtained from the PCoAs did not allow for a clear visual distinction between sample groups, One-way PERMANOVA tests were necessary to precisely test for significant differences in skin microbiome structure among the studied groups. For functional analysis at the IPR level, the same multivariate analysis and tests mentioned above were performed.

To identify the bacterial taxa that contribute to the major dissimilarities between the microbiomes according to site of origin or different abundances of *S. aureus*, SIMPER tests were performed at the OTU (Operational Taxonomic Unit) level. For Functional analysis at the IPR level, the same SIMPER tests mentioned above were performed.

# 2. Materials and Methods

## 2.1 Collection and treatment of taxonomic and functional data

### 2.1.1 Collection of the data set from EBI-Metagenomics (MGnify)

The dataset analysed in this thesis was collected from EBI-Metagenomics (MGnify) and can be accessed through the following code: MGYS00000604. This dataset was found through the study of Kong et. al., 2012. It can be also accessed through GenBank (http://www.ncbi.nlm.nih.gov/genbank) as BioProject ID 46333.

When acessing the reference code at EBI-Metagenomics, there are different matrix files summarizing the study. Each downloadable file contains an aggregation of the analysis results from the individual study's runs. To get further information about the analysis and individual runs it is necessary to access the webpage.

For this study the following matrix files were downloaded: phylum level taxonomies, taxonomic assignments at the Operational Taxonomic Unit (OTU, defined at 97% similarity of 16S rRNA genes) leveal and InterPro (IPR) matches for functional analysis using the InterPro protein sequence database. The file names are respectively SRP002480_phylumtaxonomy_abundances_v2.0; SRP002480_taxonomy_abundances_v2.0 and SRP002480_IPR_abundances_v2.0, and can be downloaded from the MGnify platform using the abovementioned accession code.

The downloaded files were very complex and had a lot of information that was not necessary for the aim of this study. All of the files are built upon the same samples using only 16S RNA gene sequencing for taxonomic assignments at phylum (phylum taxonomies table) and Operational Taxonomic Units (OTUs, a proxy for bacterial species determined at 97% 16S rRNA gene similarity) levels, plus a functional table derived from untargeted shotgun sequencing containing InterPro IPR entries functional assignments.

#### 2.1.1.1 EBI-Metagenomics (MGnify)

As it was mentioned in the introduction, the study of metagenomes and its relevance have been increasing. As a result of this, an in order to improve the quality of future studies, the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) created the EBI Metagenomics platform (Hunter et al., 2014), which has been recently updated and renamed as “MGnify”.

EBI-Metagenomics / MGnify is a free website tool and database of metagenomics projects, one of the biggest platforms of analysis and record metagenomic and meta-transcriptomic data. This website tool has the capacity of delivering structural and functional diversity data from all sorts of microbiomes, integrating environmental data (so-called “metadata”), and thus allowing researchers to obtain insights into the roles played by physical-chemical conditions in shaping natural microbial communities, how microbial species may interact with one another within microbiomes and how they interact with their hosts. Scientists can submit different types of data, like 16S rRNA amplicon data,

WGS sequenced metagenomic and meta-transcriptomic reads, and user-submitted sequence assemblies. Disregarding the data source, the platform can standardize the analysis workflow and has the ability to produce taxonomic and functional diversity tables for further studies (Mitchell et al., 2016).

The scheme below introduces the analysis pipeline (v2.0) released in March 2015 (Mitchell et al., 2016).

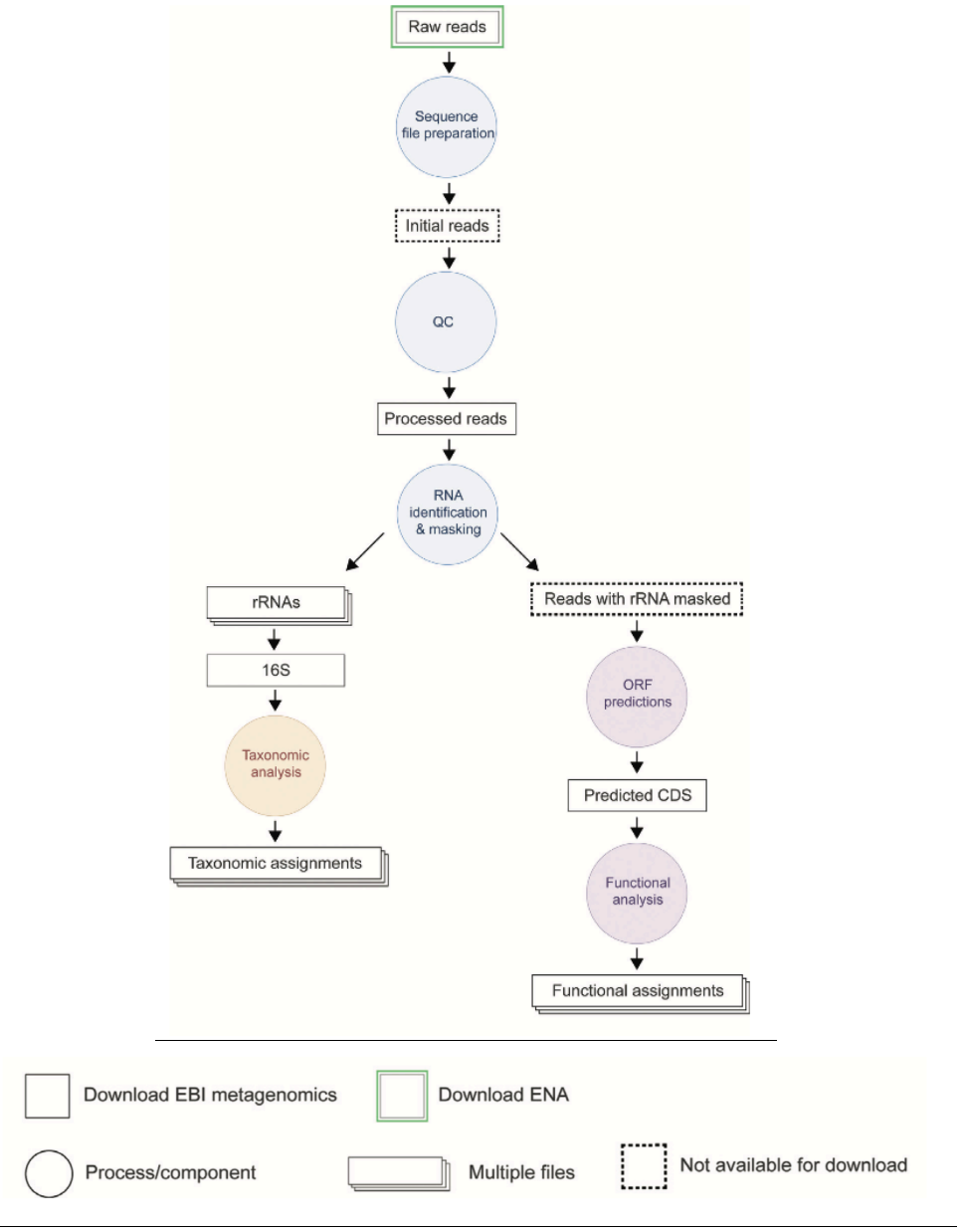



**Figure 6** | **Schematic of the EBI Metagenomics analysis pipeline.** Processes/components are indicated as circles and inputs/outputs are represented by rectangles. The pipeline is divided into two parts: one performing taxonomic classification (based on 16S rRNA) and the other providing functional annotation (based on InterPro databases). A full description of the steps and tools are provided on the EMG. (Mitchell et al., 2016)

### 2.1.2 Treatment of the dataset

To obtain a proper dataset for this study, the sites selected corresponded to the Antecubital (Ac) and Popliteal Creases (Pc), plus the Volar forearm (Vf) and the Nares. The selection of these sites was driven by the fact that Ac and Pc are the areas that are more usual to have AD crises, Vf is important in order to understand if the microbiome proliferates from Ac and the Nares used as a control (Kong, 2012). In spite of being a reference, there was no significant data about the nares to support robust comparisons and conclusions so this area had to be discarded.

Form a statistical point of view, an elimination was made of all the samples where the sum of 16S rRNA gene reads was less than 1000, to eliminate the possibility to include poorly characterized samples in the analysis. This criterion eliminated all the Nare samples present in the raw datasets available at MGNify, and therefore this study contemplates only the Ac, Pc and Vf microbiomes containing different levels of *S. aureus* abundances (see below). Therefore, it was necessary to standardize all the three data sheets containing taxonomic (phylum and OTU levels) and functional (IPR entries) information.

In spite of having information from where the samples were taken, the right and left antecubital area, for example, it was assumed as it being "just the antecubital area". As they are considered to be the same area, they are expected to have the same environment, therefore, the same microbiome.

In the beginning of this study, the excel sheets had 4384 samples characterized either by 16S rRNA gene sequencing procedures or by total metagenome shotgun sequencing. As a consequence of the selection of only those samples characterized by total metagenome sequencing and of the filtering employed to discard samples with less than 1,000 rRNA gene reads from these samples, a total of 183 skin microbiome samples were left, which 60 samples were from Antecubital Creases, 76 from Popliteal Creases and 47 from Volar Forearm. The resulting contingency tables (phylum- and OTU-level taxonomies and IPR assignments) were prepared for downstream statistical analyses as explained below.

#### 2.1.2.1 Relative Frequency and Hellinger transformation

##### 2.1.2.1.1 Relative Frequency

In order to normalize the data, it was necessary to tranform it in percentages. Therefore, as Kenney wrote "If an event occurred r times on the way described as "success", in a series of n independent random trials, all made under the same essential conditions, the ratio r/n is called the relative frequency of success" (Kenney, 1963).

Since the conditions for this study are similar as Kenny described, I starded to pull out all of my data in relative frequency, since my reads were in absolute frequency.

The relative frequency is defined as the ratio between the absolute frequency and the total number of observations.

$$\text{Relative Frequency} = \frac{\text{number of reads per entry}*}{\text{sum of all reads per sample}} \qquad \textbf{(Equation 1)}$$

*at Phylum, OTU and IPR level per sample.

#### 2.1.2.1.2 Hellinger Tranformation

One of the aims of this research was to find a relationship between the microbiome and since there is a large number of Phylum/OTU/IPR, it was important to have my data normalized using the Hellinger Transformation in order to have an improvement of the proportional abundance data, thus correcting for a possible overweight of highly abundant IPR entries, OTUs or phyla when determining the most differentiating taxonomic groups and functions across skin samples with different abundances of *S. aureus*.

The Hellinger transformation is defined as

$$y'_{ij} = \sqrt{\frac{y_{ij}}{y_{i.}}} \qquad \textbf{(Equation 2)}$$

"Where j indexes the species, ij the site/sample, and i is the row sum for the ith sample" (Legendre & Legendre, 2012).

The data, in this stage was already in the form yij/yi, so it was only necessary to do a square root transformation to the data.

## 2.2 Analysis of Phylum Taxonomic Abundance

The analysis of the taxonomic abundance at the phylum level was necessary since it was important to have an overview of the microbiome present in the skin surface, and especially in the critical areas (Ac and Pc) usually affected by Atopic Dermatitis.

Five bar charts were created with the relative frequency data - percentage of the reads per phylum and per sample per site: Ac, Pc and Vf - to have an overview of the main phyla present in the skin microbiome: *Actinobacteria*, *Firmicutes*, *Proteobacteria*, *Bacteriodetes*, Others and Unassigned (the one's that don't have any type of classification).

## 2.3 Taxonomic and Functional Analysis

For multivariate analysis of taxonomic and functional profiles, the PAST software (v3) was used.

### 2.3.1 PAST

Past (PAleontological STatistics) is a free software available for scientific data analysis, that integrates spread-sheed-type data with functions for data manipulation, plotting, univariate and multivariate statistics, ecological analysis, time series and spatial analysis, morphometrics and stratigraphy (Hammer et al., 2001).

In order to improve possible outcomes, the following resources within PAST were used: PCoA with the Bray-Curtis Dissimilarity index, PERMANOVA with Bonferroni corrections to test the validity of clusters formed using PCoA and SIMPER (similarity percentage) analysis to determine which phylum /

OTU / IPR displayed the largest shifts in abundance according to increasing levels of *S. aureus* abundance in the skin.

#### 2.3.1.1 PCoA –Principal Coordinates analysis

Principal Coordinates Analysis (PCoA) is a multivariate statistic, an ordination method used to explore and to visualize similarities or dissimilarities of data; to visualize individual and/or group differences (Hammer et al., 2001)(Gail et al., 2007).

PCoA is typically applied when a reduction and interpretation of large multivariate data sets, with some underlying linear structure, is needed. The aim of PCoA is to calculate a distance matrix and produce a graphical configuration in a low dimensional space (typically two or three dimensions), such that the distances between the points in the configuration reflect the original distances as accurate as possible. The PCoA can be applied either on the variables or on the observations (Gail et al., 2007). The PCoA routine finds the eigenvalues and eigenvectors of a matrix containing the distances between all data points. The eigenvalues, giving a measure of the variance accounted for by the corresponding eigenvectors (coordinates) are given for the first four most important coordinates. The percentages of variance accounted for by these components are also given (Hammer et al., 2001).

##### 2.3.1.1.1 Bray-Curtis Dissimilarity

Bray-Curtis measure for abundance data was used to measure the comparisons between samples in analyses of multivariate data (Somerfield, 2008).
The Bray-Curtis dissimilarity expression is:

$$d_{ij} = \frac{\sum_{k=1}^{p} |y_{ik} - y_{jk}|}{\sum_{k=1}^{p} (y_{ik} + y_{jk})} \qquad \textbf{(Equation 3)}$$

"For the following, let Y = ($y_{ik}$ ) be the (N × p) matrix of i = 1, ..., N observation units (rows) by k = 1, ..., p variables (columns). Each dissimilarity or distance measure will be given as a value $d_{ij}$ between observation units i and j, where j (like i) goes from 1 up to N" (Anderson, 2005).

"The dissimilarity between samples is the sum (over species) of the modulus of the difference between counts and them divided by the sum of the counts in the samples being compared" (Somerfield, 2008).

This is a suitable measure for biological data, which combines the structural information on presence or absence with quantitative counts of the species that are present (Somerfield, 2008).

#### 2.3.1.2 Permanova

Permutational multivariate analysis of variance (PERMANOVA) is a geometric partitioning of variation across multivariate data, defined explicitly in the space of a chosen dissimilarity measure, in response to one or more factors in an analysis of variance design. Statistical inferences are made in a distribution-free setting using permutational algorithms.

"In the simplest situation of a one-way test (i.e. the test of a single factor), with a group and N observation units (replicates) per group, let N be the total number of observation units and let $d_{ij}$ be the distance between observation i and observation j" (Anderson, 2005). Then, the total sum of squares is:

$$SS_T = \frac{1}{N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_{ij}^2 \qquad \textbf{(Equation 4)}$$

To calculate $SS_T$ it is only necessary to add up the squares of all of the distances in the sub-diagonal half of the distance matrix and divide by the total number of observations (N), as illustrated below (Figure 7):

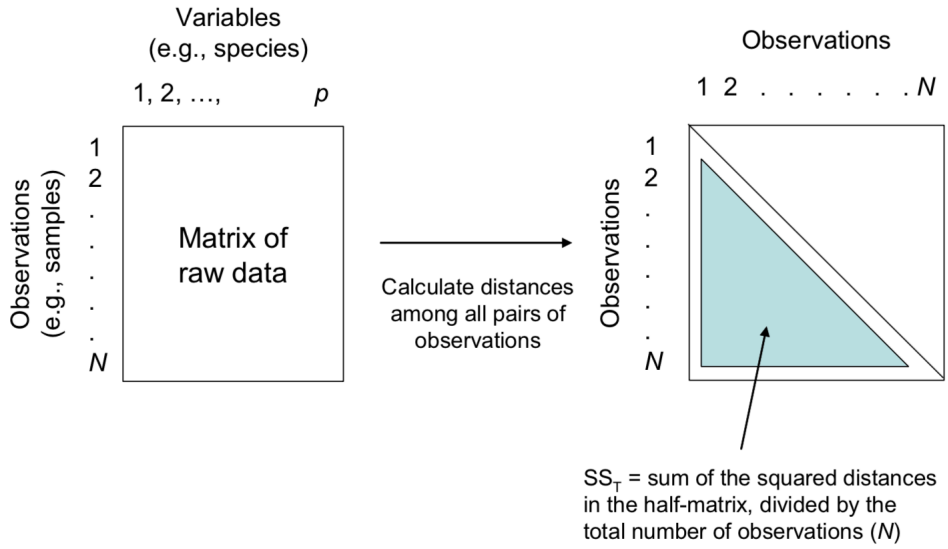



**Figure 7 | Schematic representation of the sum of the squared distances in the half-matrix, divided by the total numer of observations**. (Anderson, 2005)

Thus, schematically, for the case of two groups with equal sample sizes, this is (Figure 8):

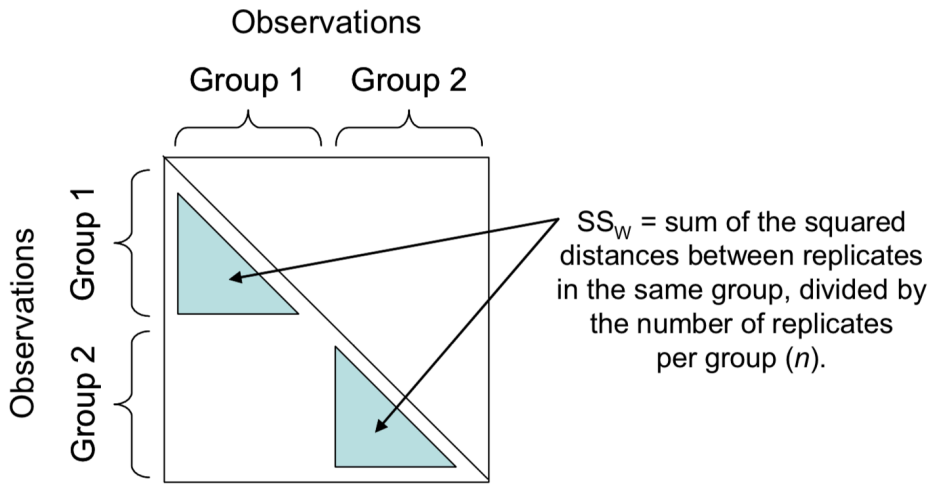



**Figure 8 | Schematic representation of the sum of the squared distances in the half-matrix, divided by the total number of observations for the case of two groups with equal samples sizes.** (Anderson, 2005)

Next, the within-group sum of squares is:

$$SS_W = \frac{1}{n}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_{ij}^2\, \varepsilon_{ij} \qquad \textbf{(Equation 5)}$$

Here the $\varepsilon_{ij}$ takes the value of 1 if observation i and observation j are in the same group, any other way it would take the value of zero (Anderson, 2005).

The among-group sum of squares is the difference: $SS_A = SS_T - SS_W$ .

A pseudo F-ratio associated with the test of this factor is:

$$F = \frac{SS_A/(a-1)}{SS_W/(N-a)} \qquad \textbf{(Equation 6)}$$

"Where (a −1) are the degrees of freedom associated with the factor and (N − a) are the residual degrees of freedom" (Anderson, 2005).

When considering one-way analysis, the permutation procedure is used when the distribution of F is under a null hypothesis, which means that there is no effect of the factor. If the null hypothesis were true, then the F-statistic actually obtained with the real ordering of the data relative to the treatments will be similar to the values obtained under permutation. If, however, there is a significant effect of treatments, then the value of F obtained with the real ordering will appear large relative to the distribution of values obtained under permutation. In that case, the value of F for the data is unlikely to have been obtained if the null hypothesis were true (Anderson, 2005).

The frequency distribution of the values of $F^{\pi}$ is discrete: which means that the number of ways that is possible for data to be re-ordered is finite. The probability associated with the test statistic under a true null hypothesis is calculated as the proportion of the $F^{\pi}$ values that are greater than or equal to the value of F observed for the real data. P-value can be calculated as:

$$P = \frac{(N.of\ F^{\pi} \geq F)+1}{(Total\ no.of F^{\pi})+1} \qquad \textbf{(Equation 7)}$$

When trying to understand multivariate data it is important to know that the observations of the data matrix are randomly permuted. However, this does not mean these values are shuffled just anywhere.

Concerning the one-way test, by enumerating all the possible permutations, we will have the correct P-value that is linked to the null hypothesis (Anderson, 2005).

There are some assumptions to take into account, when discussing one-way analysis. PERMANOVA only assumes the observations units that are convertable under a true null hypothesis. It the observations tied in with each other then by randomly shuffling them will only destroy the inherent structure. Also, in the majority of cases there is the assumption that observation units are independent of each other. On the other hand, it is not expected that individual variables are independent of each other (Anderson, 2005).

**Bonferroni corrected p-values for Pairwise**

Bonferroni test is a method used to compare multiple tests used in statistical analysis, demonstrating where there is any statistical significance of the dependent variable or not. In this study it was used to count act the problem of multiple comparisons by correcting against type 1 error.

This test is based on the idea that if you test N dependent or independent hypotheses, one way of maintaining the error rate is to test each individual hypothesis at a statistical significance level that is deflated by a factor of (1/n). Therefore, for a significance level for the whole family of tests of α, the Bonferroni correction would be to test each of the individual tests at a significance level of α/n (Simes, 1986).

The Bonferroni inequality is usually used when doing several tests of significance to set an upper bound on the overall significance level $\alpha$ (Simes, 1986).

"If $T_1$, . . ., $T_n$ is a set of n statistics with corresponding p-values $P_l$, ... , $P_n$ for testing hypotheses $H_1$,…, $H_n$, the classical Bonferroni multiple test procedure is usually performed by rejecting $H_o$= ($H_1$,…, $H_n$) if any p-value is less than $\alpha$ / n. Furthermore, the specific hypothesis $H_i$ is rejected for each $P_i \leq \alpha$ / n (i = 1, .. ., n)" (Simes, 1986).

The Bonferroni inequality,

$$pr\left(\cup_{i=1}^{n}(P_i \leq \frac{\alpha}{n})\right) \leq \alpha, (0 \leq \alpha \leq 1) \qquad \textbf{(Equation 8)}$$

It ensures that, the probability of rejecting at least one hypothesis when all others are true, is no greater than $\alpha$ (Simes, 1986).

### 2.3.1.3 Simper – Similarity Percentage Analysis

SIMilarity PERcentages decompose the similarities into the contributions from each species, mostly used as a post-hoc test in multivariate abundance. In other words, it finds the average contributions from each species to the calculated Bray-Curtis dissimilarities (in the case of this thesis) among sample groups. It is commonly used to answer the question 'in which taxa is this difference most evident?'.

The strength of between-group is the effect with variance; this means that the SIMPER algorithm will identify taxa with a strong between-group effect and with large within-group variance, regardless of the presence of a between-group effect (Warton et al., 2012).

### 2.3.2 Linear regressions for the results of taxonomic analysis

Following SIMPER analyses, there was an interest to determine the linear correlation between the relative abundances of each of the top ten SIMPER entries and those of *Staphylococcus aureus*. Therefore, several linear regressions were made to understand those relationships.

Linear regression is an empirical model that is commonly used for predictive analysis. It is a very usefull tool to understand what is the relationship between a variable of interest and a set of related, predictor variables (Kenney, 1963).

Furthermore it is very important to compreend the correlation coefficient, as it is what indicates whether the variables have any linear relationship or not.

## 2.4 Collection, analysis and comparison of *S. succinus* and S. *aureus* genomes.

In this study, the RAST annotation pipeline was used to compare four *S. succinus* genomes with one genome representing a multi-resistant *S. aureus* strain and to compare the number of genes present in each subsystem. The strains used for the comparison were *S. aureus* USA300 and *S. succinus* SNUC1280, 14BME20, DSM15096 and DSM14617.

### 2.4.1 RAST

RAST, Rapid Annotations using Subsystems Technology, is an automatic annotation free server, built upon the framework provided by the SEED system (Figure 9) (Overbeek et al., 2014). It returns an analysis of the genes and subsystems of the genome in question, as supported by comparative and other forms of evidence (Glass & Meyer, 2011).

Users can upload raw sequence data in FASTA format and they receive access to an annotated genome in an environment that supports comparison with an integration of hundreds of existing genomes; the sequences will be normalized and processed and summaries automatically generated. The server provides several methods to access the different data types, including phylogenetic and metabolic reconstructions, and the ability to compare the metabolism and annotations of one or more metagenomes and genomes. In addition, the server offers a comprehensive search capability (Aziz et al., 2008) (Glass & Meyer, 2011).

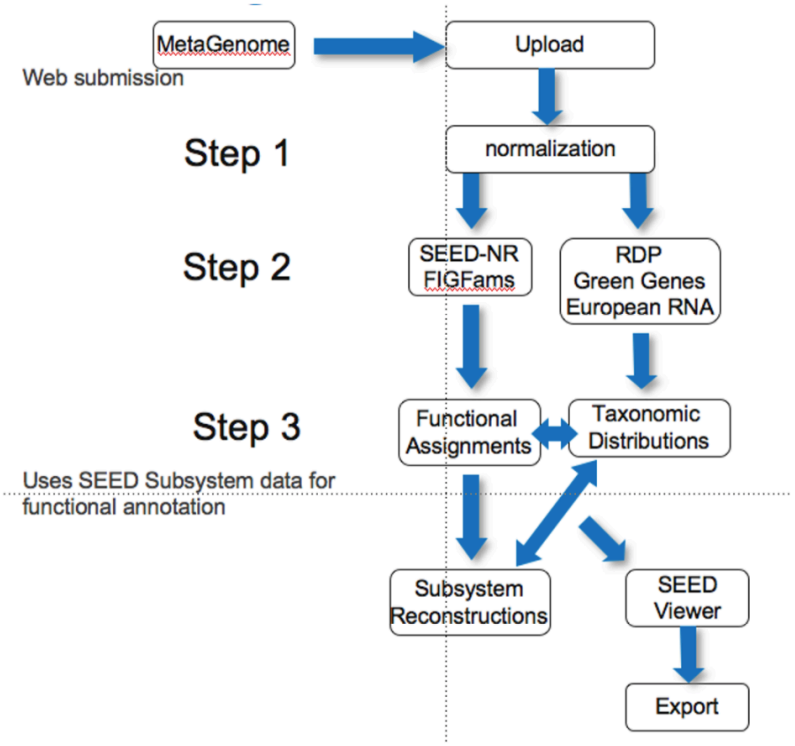



**Figure 9 | Overview of the workflow implemented in the metagenomics RAST pipeline.** During the process three different steps are executed. A full description of the steps is provided on the following article. (Glass & Meyer, 2011)

Various tools have been built into the framework, allowing users to compare their data against other metagenomes or complete genomes taken from the SEED environment. The subsystems heat map and the taxonomic heat map provide comparative metagenomics summaries that cover the differences between samples (Glass & Meyer, 2011).

The subsystem comparison tools identify the number of pegs in each metagenome that are connected to a subsystem via protein level similarity. Based on these connections, each subsystem present in a sample is scored by counting the number of sequences that are similar to a protein in each subsystem. This score is divided by the total number of sequences from the sample that are similar to any protein in a subsystem, to give a fraction of sequences present in each subsystem. This approach allows comparisons between samples that have different numbers of sequences. Since the fractions tend to be small (a few sequences hit each subsystem, but there are now over 600 subsystems in the SEED), the scores can be factored for display purposes. Furthermore, a nonquantitative approach is provided to group the subsystem scores, emphasizing those subsystems that are most different between the samples. Moreover, the display can be limited to specific areas of metabolism, or other subsystem groups, as desired by the user (Glass & Meyer, 2011).

The subsystem spreadsheet is populated with all genomes that have those functional roles and provides links to the relevant protein pages. The subsystem info tab provides an expert annotator's notes on the creation of the subsystem. Although they are not comprehensive, the SEED subsystems are a particularly useful way to quickly determine the proteins that are involved in a related function and to determine known variations in functionality between organisms (Overbeek et al., 2014).

# 3. Results and Discussion

## 3.1 Analysis of Phylum Taxonomic Abundance

In order to confirm if the relative skin phylum taxonomic abundance with the literature was in concordance the data set, a bar chart representing the relative abundance of bacterial phyla across all the skin samples analysed was created. Figure 10 is therefore an overview of the relative abundance of the different phyla observed across all the 183 skin microbiome samples analysed, regardless of higher or lower incidence of *Staphylococcus aureus* in the samples.

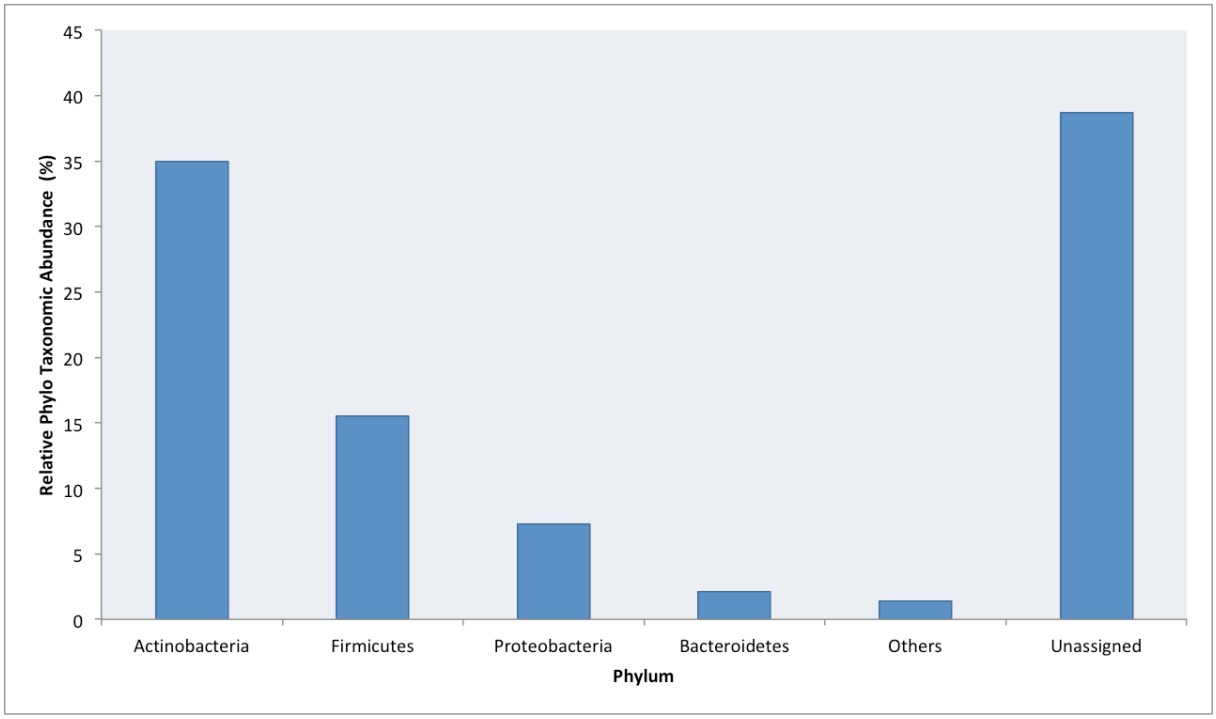



**Figure 10 | Global Analysis of Taxonomic Abundance at the phylum level.** Relative Frequency of the sum of the reads per Phylum in all skin microbiome samples analysed (N = 183).

As it can be observed in Figure 10, the data set is in concordance with the literature, since the most frequent phyla across all skin samples are: *Actinobacteria* and *Firmicutes*, followed by *Proteobacteria* and *Bacteroidetes* (Thomas et al., 2017). As we can see, there are a lot of reads that were not classified - one of the big challenges of metagenomics approach. More investigation is needed in order to discover more bacteria.

Bacterial taxa found on skin surfaces display intra-individual variability, so it is also important to know how the main six phyla are distributed in the samples (Grice & Segre, 2011).

To have a better understanding of how the six groups highlighted in Figure 10 were distributed according with skin location, another bar chart was created with an overview of Phylum taxonomic abundances per Sample (Figure 11).

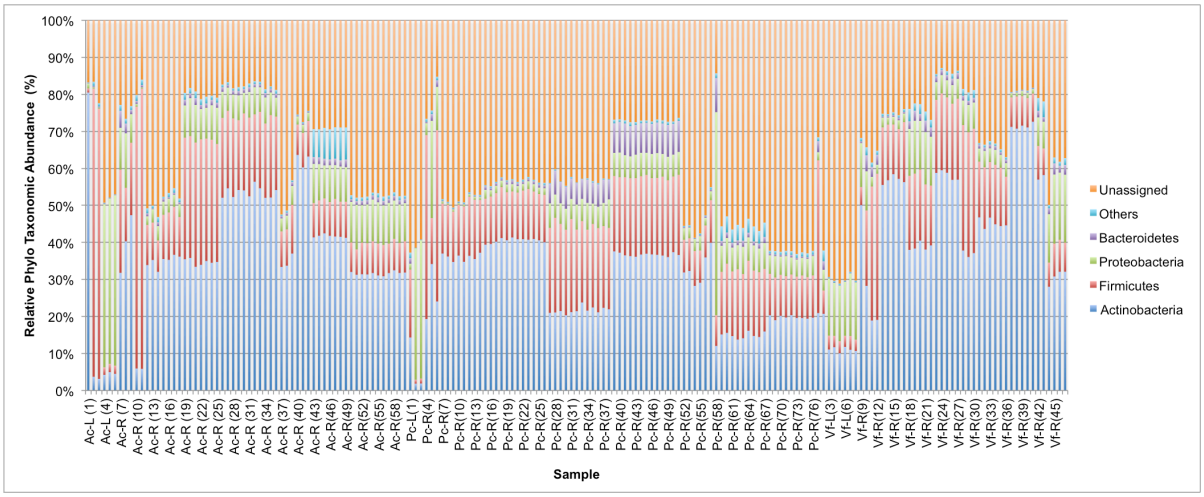



**Figure 11 | Overview of Taxonomic Abundance at Phylum Level per Sample.** Relative Frequency of the Abundance of Phylum per Sample. Ac- Antecubital Crease; Pc – Popliteal Crease, Vf – Volar Forearm. (R) . right side; (L), left side. Chart for all the 183 samples analysed in this study are shown. Sample lables for only about 1/3 of the total amount of samples are shown for the sake of simplicity.

The samples in this dataset are not identified as if they were collected from patients with Atopic Dermatitis or from healthy patients. However, it is known that there are samples collected from patients with Atopic Dermatitis, and also in different disease stages, therefore it can be assumed that this between-group difference in abundance is because of that, because as it is said in the literature, when a patient has a flare, there is a substantial growth of *Staphyococcus aureus*, which belongs to the phylum *Firmicutes*, resulting in the decrease of other species, for example, from *Actinobacteria*, like *Corynebacterium*. And this is actually seen in Figure 11. When there is a higher relative frequency of *Firmicutes*, the relative frequency of *Actinobacteria* decreases (Kong, 2012).

Neverthless, it is known that there is a correlation between a decrease of *S. aureus* and an increase in *Propionibacterium*, *Streptococcus* and *Corynebacterium*, which can be used as an indicator of the recovery from deterioration of AD (Kim et al., 2017).

To better understand/discover more about the relative phylum abundance between the groups (Ac, Pc and Vf) and if there was any impact of the relative frequency of *S. aureus* present on phylum distributions (and later, on OTU and IPR distributions, see below), the samples were divided into five groups according to the percentage of incidence of *S. aureus* (Table 1).

**Table 1| Division of the samples per relative frequency of *S. aureus*.**

| Groups (% *S. aureus*) | No. Samples | Percentage(%) |
|---|---|---|
| **0-1%** | 50 | 27,32 |
| **1-2%** | 43 | 23,50 |
| **2-5%** | 52 | 28,42 |
| **>5-10%** | 29 | 15,85 |
| **>10%** | 9 | 4,92 |
| **Total** | 183 | 100 % |

As it can be observed from Table 1, there are more samples with one to five percent of *S. aureus* and there are only 9 samples (from 183) with a relative frequency of *S. aureus* higher than 10%. Although there is no information in the matrix files, these latter samples are probably from people with flare.

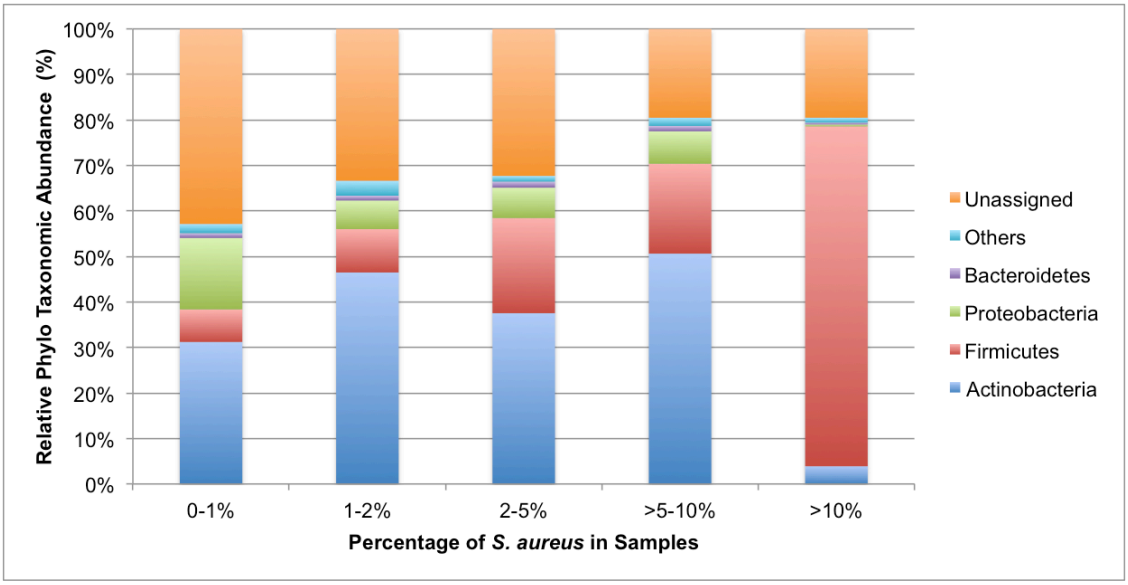



**Figure 12 | Relative Phylum Taxonomic Abundances *versus* Percentage of *S. aureus* in Antecubital Creases Samples.** Chart for all Antecubital Creases samples analysed in this study are shown (N=60).

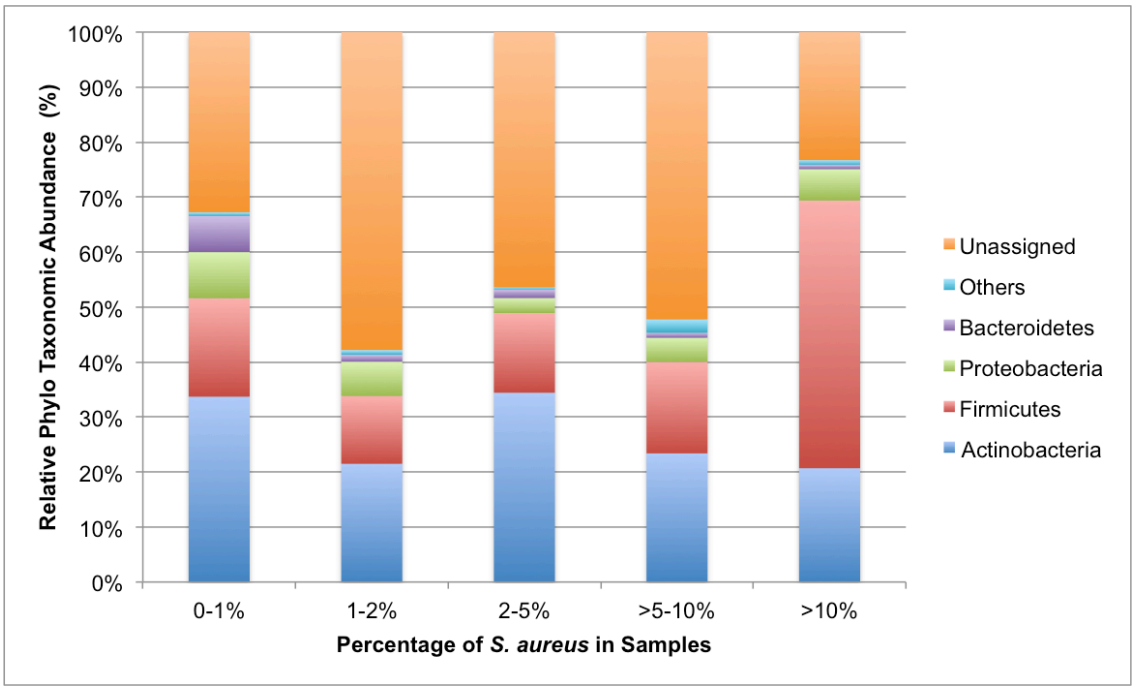



**Figure 13 | Relative Phylum Taxonomic Abundances *versus* Percentage of *S. aureus* in Popliteal Creases Samples.** Chart for all Popliteal Creases samples analysed in this study are show (N=76).

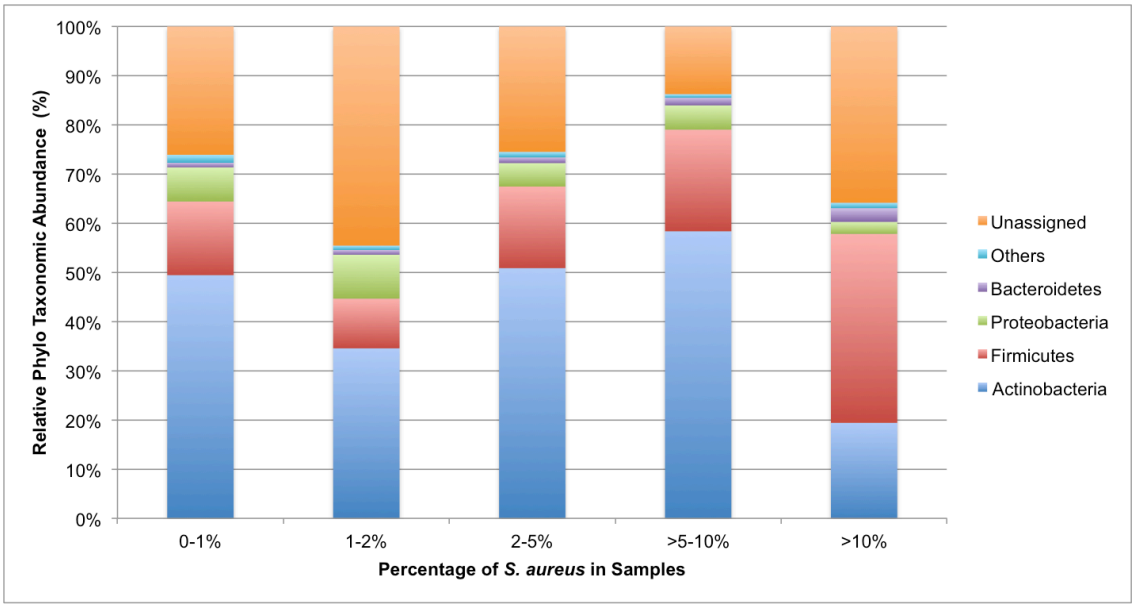



**Figure 14 | Relative Phylum Taxonomic Abundances versus Percentage of *S.aureus* in Volar Forearm Samples.** Chart for all Volar Forearm samples analysed in this study are show (N=47).

Observing the Figures 12, 13 and 14, it can be concluded that when there are more than ten percent of *S. aureus*, there is a decrese in *Actinobacteria*. Higher frequencies of *S. aureus* may promote higher incidence of other *Firmicutes* bacteria given the very high proportions of this phylum when *S. aureus* is present in abundances higher than 10%. To determine whether this is true and to infer which bacterial species in the complex skin microbiome may respond positively or negatively to higher *S. aureus* abundances, a more refined analysis of OTU abundances and distributions across the data was performed.

## 3.2 Taxonomic Analysis at the OTU level

### 3.2.1 Taxonomic Analysis at the OTU level of All the Samples

One of the big challenges in the study of microbial communities in Atopic Dermatitis is to define and characterize what are the relevant features of the microbiome. In this study, the focus is on the microbial community changes between the “favourite” areas of AD and to know if *S. aureus* influences the community.

In order to understand the similarities or dissimilarities of the data, Principal Coordinates Analysis (PCoA) on Bray-Curtis dissimilarity matrices were performed.

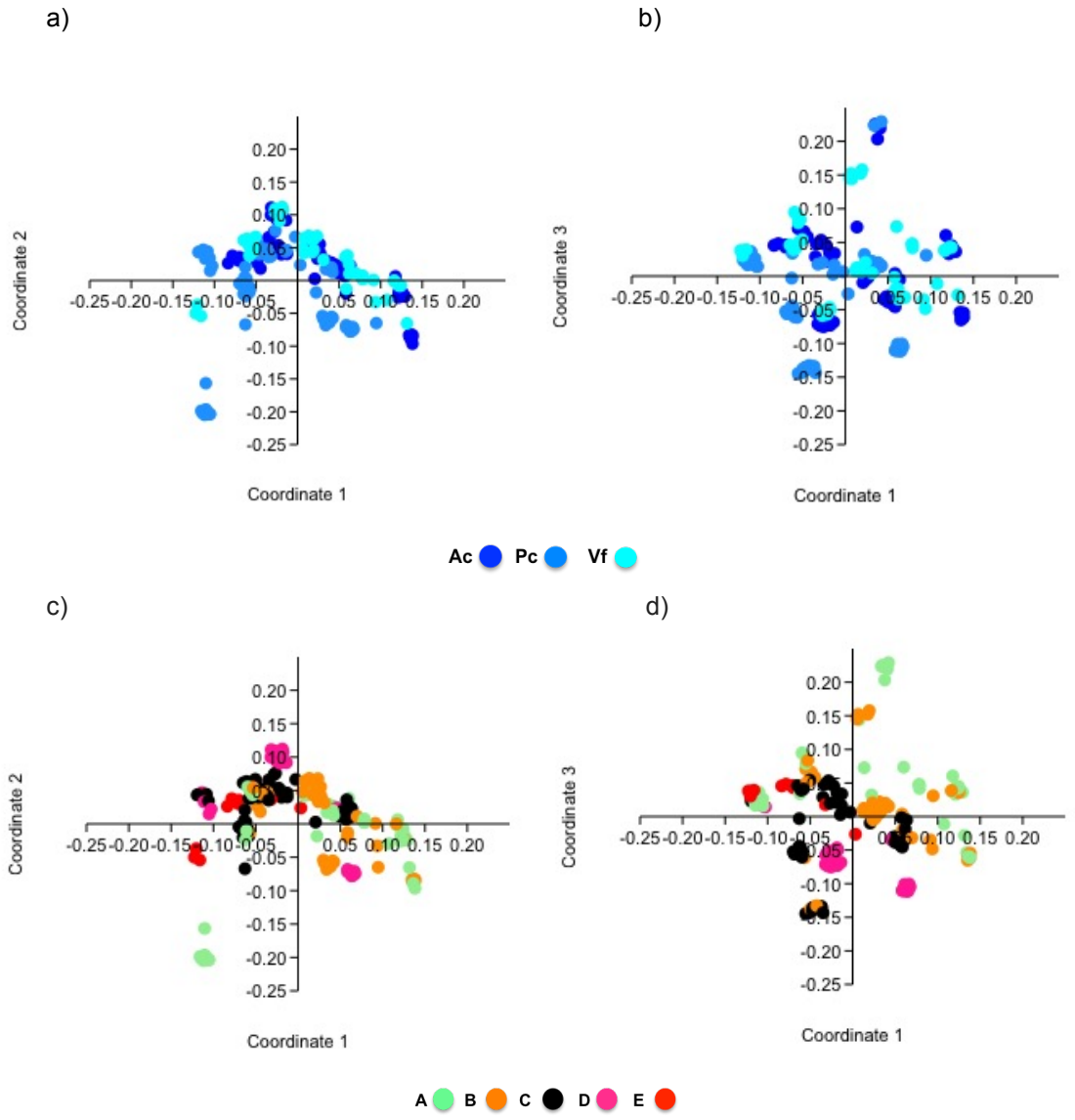



**Figure 15 | Principal Coordinates Analysis (PCoA) of OTU profiles retrieved for All Samples (N = 183).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of taxonomic profiles of microbial communities between the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and Volar Forearm (Vf); Coordinate 1 versus Coordinate 2. (b) PCoA of taxonomic profiles of microbial communities between the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and volar forearm (Vf); Coordinate 1 versus Coordinate 3. (c) PCoA of taxonomic profiles of microbial communities according to the relative percentage of *S. aureus* in the samples: A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%), Coordinate 1 versus Coordinate 2. (d) PCoA of taxonomic profiles of microbial communities according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

By comparing PCoA results from Figures 15 a) and b), we can see that there are a large overlap of samples from different areas, suggesting that, overall, there is no clear partition of microbiome structure across different skin locations and, therefore, simple exploration of the PCoA graphs were inconclusive in this regard. Although, complementing these results with those in Figure 15 c) and d) we can observe that, the emergence of clusters representing samples with similar proportions of *Staphylococcus aureus* in somewhat clearer, although some overlap of samples displaying different *S. aureus* percentages in the ordination diagram still persisted. Althouhg this result

suggests that the microbiota is more similar when samples from the same site share the same average of percentage of *S.aureus*, simple exploration of the ordination diagram was not helpful in allowing for clear conclusions in this regard perhaps because of the low percent of data variation explained by the coordinates 1 to 3 used to built the diagrams.

To determine whether the structure of the skin bacterial communities changed according to their site of origin and the frequency of *S. aureus*, PERMANOVA tests were carried out. Results in Tables 2 and 3 pertain to tests for differences according to sample origin (Ac, Pc and Vf).

**Table 2 | Summary of One-Way PERMANOVA carried out for the whole dataset (N=183).** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 28.33 |
| Within-group sum of squares: | 26.2 |
| F: | 7.329 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the three sites (Ac, Pc and Vf), disregarding differences in *S. aureus* abundances, with a p-value of 0.0001.

**Table 3 | Pairwise of One-Way PERMANOVA carried out for the whole dataset.** Significant values are in bold.

| | Ac | Pc | Vf |
|---|---|---|---|
| Ac | | **0.0003** | **0.0378** |
| Pc | **0.0003** | | **0.0003** |
| Vf | **0.0378** | **0.0003** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the three sites (Ac, Pc and Vf), disregarding differences in *S. aureus* abundances, with a p-value of 0.0001.

From the summary (Table 2) it can be concluded that the results of the test are relevant since the p-value is 0.0001.

In Table 3, all of the values are in bold, since all the values are statistically relevant. In other words, there are significant dissimilarities between all the areas: Ac, Pc and Vf, in spite of the overlap between sample groups observed in Figures 15a and b.

In order to identify bacterial taxa that contribute to the difference between preferred sites of disease manifestation, SIMPER analysis at the OTU level was performed. Mean change in abundance with areas and percentage contribution to differences in the Bray-Curtis dissimilarity metric were tabulated below (Table 4).

**Table 4 | Results from SIMPER analysis of all samples.** SIMPER analysis identifying the percentage contribution of each OTU according to the Bray Curtis dissimilarity metric between Ac, Pc and Vf.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean Ac | Mean Pc | Mean Vf |
|---|---|---|---|---|---|---|
| Root;k__Bacteria;**p__Actinobacteria;**c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;**g__Propionibacterium;s__acnes** | 1.459 | 2.606 | 2.606 | 0.486 | 0.241 | 0.508 |
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;**g__Corynebacterium;s__** | 1.077 | 1.923 | 4.53 | 0.347 | 0.486 | 0.394 |
| Root;k__Bacteria;**p__Proteobacteria**;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;**g__Pseudomonas;s__** | 0.601 | 1.074 | 5.603 | 0.0992 | 0.0705 | 0.119 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__aureus** | 0.5512 | 0.9848 | 6.588 | 0.18 | 0.149 | 0.148 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;**g__Streptococcus;s__** | 0.5197 | 0.9284 | 7.517 | 0.194 | 0.125 | 0.187 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__** | 0.4258 | 0.7608 | 8.277 | 0.175 | 0.166 | 0.151 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__succinus** | 0.3794 | 0.6778 | 8.955 | 0.136 | 0.119 | 0.115 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Anaerococcus;s__** | 0.3438 | 0.6142 | 9.569 | 0.068 | 0.125 | 0.0843 |
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;**g__Propionibacterium;s__granulosum** | 0.3357 | 0.5998 | 10.17 | 0.107 | 0.052 | 0.117 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Peptoniphilus;s__** | 0.3316 | 0.5925 | 10.76 | 0.0414 | 0.0961 | 0.0468 |
| (...) | | | | | | |
| Root;k__Bacteria;**p__Firmicutes;**c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__epidermidis** | 0.2486 | 0.4442 | 15.67 | 0.0897 | 0.0692 | 0.0747 |

Similarity percentage analysis of the OTU diferences between the different sites Ac, Pc and Vf. The first collumn identifies the OTU explained by that row, the second collumn represents the average of dissimilarity, the third collumn shows the % dissimilarity explained by that OTU, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the OTU thus far represented in the table and the last three show mean abundance at ste Ac, mean abundance at site Pc and mean abundance at site Vf.

Table 4 shows the ten most site-differentiating OTUs revealed by SIMPER analysis, along with a row for *Staphylococcus epidermidis*, since it is a species believed to have a relationship with *S. aureus* in the process of skin colonization, althought the precise type of this relationship is currently unknown.

After analysing Table 4 to better understand the relationship between the OTUs and their sites, interesting results were found about the main species possessing diferent abundance in the three different areas. For instance, *P. acnes* has a low mean abundance in Pc and *Pseudomonas* sp. has a higher mean abundance in Vf comparing to the other sites.

Since it is already known that *S. aureus* is related to AD and its abundance differs between skin sites, it is even more interesting to know how *S. aureus* and the other 10 species from SIMPER results correlate.

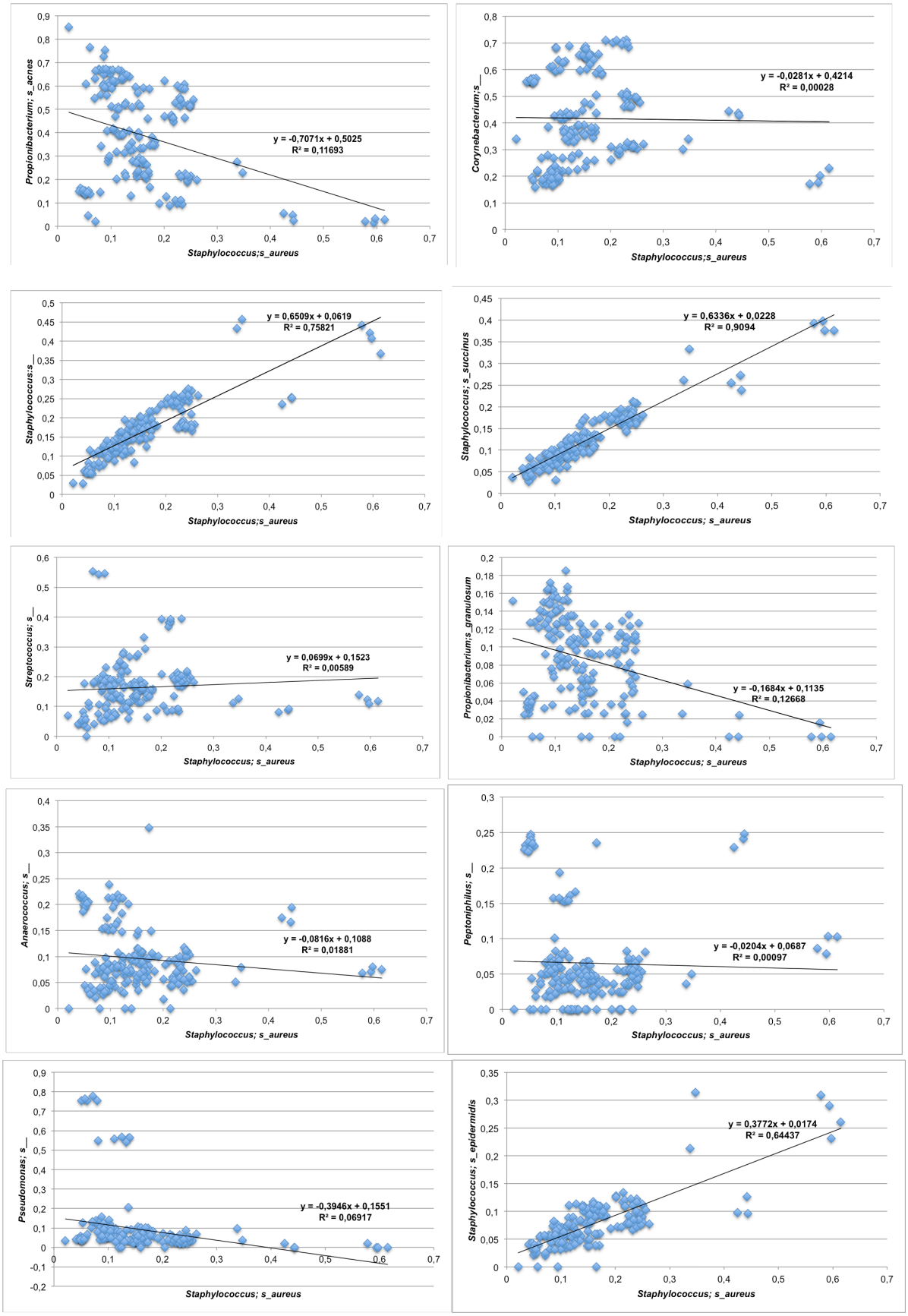



**Figure 16| Linear Regression between the abundance of the ten most site-differentiating OTUs revealed by SIMPER analysis and *S. aureus*.** The data used for this linear regressions were Hellinger transformed.

As it can be observed from all the linear regressions, there are three obvious positive linear correlations between *S. aureus* and *Staphylocuccus* sp., *S. succinus* and *S. epidermidis*.

In concordance with the literature, it is known that *S. aureus* interacts with members of the same genus and that when we have a high abundance of *S. aureus* in a sample, *S. epidermidis* reacts but the answer to why it is until now unknown (Kong, 2012) (Nair et al., 2014).

According to the literature, *S. epidermidis* inhibits the growth of *S. aureus* (Nair et al., 2014). So, maybe the growth of abundance of *S. epidermidis* keeps up with the abundance of *S. aureus* as a possible immune response.

The correlation between *S. aureus* and *S. succinus* was not metioned before in the literature and it is very interesting because it is the best linear correlation according to the regression coefficient ($r^2$=0,909).

The relationship between *S. aureus* and *Propionibacterium acnes* is intriguing, *because P. acnes* is usually related to skin disorders and as we can observe, when there is a high percentage of *S. aureus*, the amount of *P. acnes* is null or its relative abundance is low. In fact, it has already been reported that there is a co-existence of *S. aureus* with other microbes like *P. acnes* in acne lesions. In acne lesions, *S. aureus* invades the human skin as a pathogen provocating tissue damage. Levy et al. showed that the prevalence and resistance patterns of *S. aureus* in individuals with acne are higher compared with those without acne, which indicates that both bacteria are associated with acne but the exact mechanism is still not explicit (Kumar et al., 2016) (Levy et al., 2003).

However, there are still other relationships (e.g. competition) or factors that affect the abundance of *P. acnes*. As we can identify in the graphic related to this relationship, when we are towards low relative abundance of *S. aureus* there is a wide range of abundance of *P. acnes*. So this can indicate that there are other factors affecting the abundance of *P. acnes*. Moreover, our data suggest strongly that *P. acnes* and *S. aureus* possess a negative abundance correlation in skin samples not affected by acne lesions which is highly contrating with the relationship between both species in the presence of acne lesions. The observations made above for *P. acnes* do also hold true for *Propionibacterium granulosum.*

Although *Corynebacterium* spp. is more sensitive to the changes of the microenvironment (Kwaszewska et al., 2014), its abundance decreases only when *S. aureus* abundances are very high, probably due to being less sensitive to microenvironmental changes that take place in the skin as a consequence of AD. The relashionship between *Corynebacterium* and *S. aureus* has been studied, and consensus exists on the competitive nature of this interaction. For instance, a study found that *S. aureus* responds to commensal *Corynebacterium* with a shift to commensalism. This opens up the possibility that commensal *Corynebacterium* spp. are an unexplored source for new antivirulence therapies that limit activation of *S. aureus* infection (Ramsey et al., 2016).

The majority of the interactions between *S. aureus* and other bacterial species, especially non *Staphylococcus* species, are competitive in nature, however this does not mean that these organisms completely inhibit the colonization of *S. aureus*; rather, *S. aureus* employs numerous defense strategies for its survival, counterattacking the competing bacteria and surviving in the same ecological niche (Nair et al., 2014).

From all the other species which have been selected after SIMPER analysis, the correlation between their abundances and that of *S. aureus* was not as pronounced as the ones mentioned above for *Staphylococcus* species or *P. acnes*. This is because either they show high abundances (in some samples only) when *S. aureus* abundance is very low (like *Pseudomonas* or *Peptoniphilus*, for instance, but usually they have low abundance anyway) or they show low abundance only when *S. aureus* is very high (like *Corynebacterium*). Our first SIMPER analysis aimed at detecting OTUs that oscillate considerably between skin sites, but across all the data several of the OTUs fetched this way do not show a clear positive or negative regression with *S. aureus*.

### 3.2.1.1 Taxonomic analysis at the OTU level of All the Samples without *S. aureus*

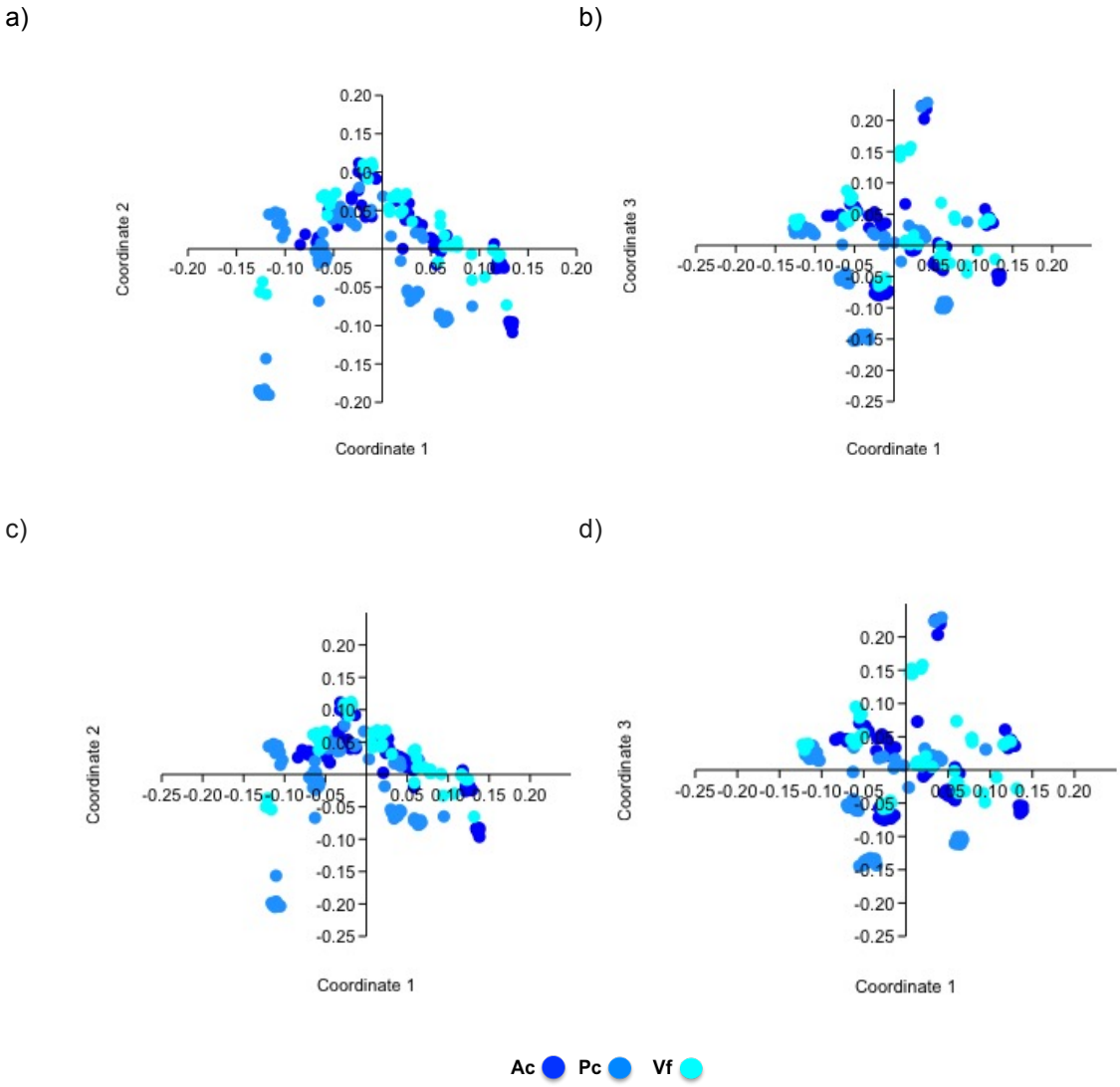



**Figure 17 | Principal Coordinates Analysis (PCoA) at OTU profiles retrived for All Samples without *S. aureus* (N=183)** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of taxonomic profiles of microbial communities without the OTU relatively to *S. aureus* between the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and Volar Forearm (Vf); Coordinate 1 versus Coordinate 2. (b) PCoA of taxonomic profiles of microbial communities without the OTU relatively to *S. aureus* between the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and volar forearm (Vf); Coordinate 1 versus Coordinate 3. (c) PCoA of taxonomic profiles of microbial communities between the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and Volar Forearm (Vf); Coordinate 1 versus Coordinate 2. (d)

PCoA of taxonomic profiles of microbial communities between the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and volar forearm (Vf); Coordinate 1 versus Coordinate 3

Observing the PCoA results from Figure 17, it can be concluded that there are no significant changes in the results shown in Figure 15 when extracting the OTU referent to *S. aureus*. This analysis was made in order to understand whether abundance relationships between the other members of the microbial community were not dwarfed in the previous analysis, especially in cases where *S. aureus* abundances were extremely high. After removing *S. aureus* from the OTU table and rearranging the relative abundances of all other species accordingly, no significant differences in the overall analysis output were observed. This absence of effect after removing *S. aureus* from the data may be due to the fact that, after all, only a few samples in the entire dataset possess extremely high *S. aureus* abundances whereas the majority display low to moderate *S. aureus* levels which may not compromise substantially the abundance relationships of the other microbiome members.

**Table 5 | Summary of One-Way PERMANOVA carried out for the whole dataset without *S. aureus* (N=183).** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 28.74 |
| Within-group sum of squares: | 26.59 |
| F: | 7.285 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the three sites (Ac, Pc and Vf), disregarding differences in *S. aureus* abundances, with a p-value of 0.0001.

**Table 6 | Pairwise of One-Way PERMANOVA carried out for the whole dataset without *S. aureus*.** Significant values are in bold.

| | Ac | Pc | Vf |
|---|---|---|---|
| Ac | | **0.0003** | **0.0441** |
| Pc | **0.0003** | | **0.0003** |
| Vf | **0.0441** | **0.0003** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the three sites (Ac, Pc and Vf), disregarding differences in *S. aureus* abundances, with a p-value of 0.0001.

Overall, the results of PERMANOVA permutation test did not show, as the PCoA analysis, relevant changes, comparing to the results of Table 2 and 3.

Also there were no significant results from SIMPER analysis comparing Table 4 with Table 7.

**Table 7 | Results from SIMPER analysis of all samples without *S. aureus*.** SIMPER analysis identifying the percentage contribution of each OTU according to the Bray Curtis dissimilarity metric between Ac, Pc and Vf.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean Ac | Mean Pc | Mean Vf |
|---|---|---|---|---|---|---|
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;**g__Propionibacterium;s__acnes** | 1.468 | 2.603 | 2.603 | 0.493 | 0.245 | 0.513 |
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;**g__Corynebacterium;s__** | 1.083 | 1.92 | 4.523 | 0.356 | 0.493 | 0.401 |
| Root;k__Bacteria;**p__Proteobacteria**;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;**g__Pseudomonas;s__** | 0.6045 | 1.072 | 5.595 | 0.1 | 0.0712 | 0.12 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;**g__Streptococcus;s__** | 0.5273 | 0.935 | 6.53 | 0.199 | 0.127 | 0.189 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__** | 0.4703 | 0.834 | 7.364 | 0.184 | 0.17 | 0.154 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__succinus** | 0.4207 | 0.7459 | 8.11 | 0.144 | 0.121 | 0.118 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Anaerococcus;s__** | 0.348 | 0.617 | 8.727 | 0.0701 | 0.126 | 0.0863 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Peptoniphilus;s__** | 0.3424 | 0.6071 | 9.334 | 0.0434 | 0.0969 | 0.0489 |
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;**g__Propionibacterium;s__granulosum** | 0.3379 | 0.5991 | 9.933 | 0.108 | 0.0527 | 0.119 |
| Root;k__Bacteria;**p__Cyanobacteria**;c__Chloroplast;o__Streptophyta;f__;**g__;s__** | 0.3134 | 0.5557 | 10.49 | 0.0868 | 0.0505 | 0.0633 |
| (...) | | | | | | |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__epidermidis** | 0.2742 | 0.4862 | 13.03 | 0.0954 | 0.0706 | 0.0762 |

Similarity percentage analysis of the OTU diferences between the different sites Ac, Pc and Vf. The first collumn identifies the OTU explained by that row, the second collumn represents the average of dissimilarity ,the third collumn shows the % dissimilarity explained by that OTU, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the OTU thus far represented in the table and the last three show mean abundance at site Ac, mean abundance at site Pc and mean abundance at site Vf.

## 3.2.2 Taxonomic Analysis at the OTU level of Samples from Antecubital Creases

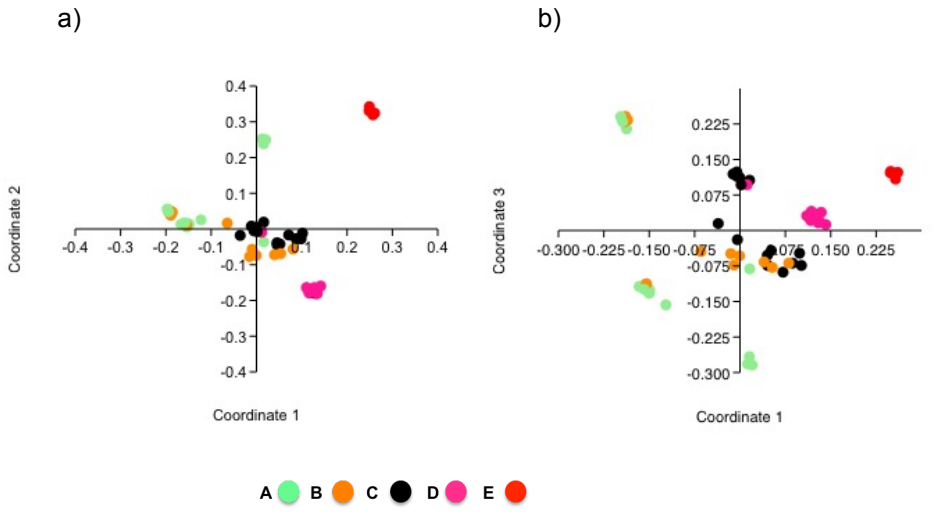



**Figure 18 | Principal Coordinates Analysis (PCoA) at OTU profiles retrived for Antecubital Creases (Ac) Samples (N=60).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of taxonomic profiles of microbial communities of Ac according to the relative percentage of *S. aureus* in the samples; A (0-1%), B (1-2%), C

(2-5%), D (5-10%) and E (More than 10%); Coordinate 1 versus Coordinate 2. (b) PCoA of taxonomic profiles of microbial communities of Ac according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

Observing Figure 18, it is clear that groups D and E are the most obvious and concise clusters altogether, consistently suggesting that high abundance of *S. aureus* impacts considerably the structure of the skin microbiome. Clearly there is an influence of the percentage of *S. aureus*. However, there is no clear partition of microbiome structure between groups A, B and C.

Although in graphic a) it seems that groups A and B have similar feautures, in graphic b), from another perspective, the communities are not so similar. It is very important to be concerned when analysing 2D graphics, because what we are observing may not be what it seems. In the future, a 3D chart will be a necessary tool in order to have a clearer view of the relationship between the groups.

Attending to the results of graphic b), there is a big difference between the samples from the group A at the lower right quadrant and at the upper left quadrant. This reveals that there are other factors, including stochastic variability, shaping the structure of the skin microbiome in situations where the percentage of *S. aureus* is low.

**Table 8 | Summary of One-Way PERMANOVA performed on Antecubital Creases (Ac) Samples (N=60) to test community variation according to increasing abundances of *S. aureus*.** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 8.711 |
| Within-group sum of squares: | 5.627 |
| F: | 7.535 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Ac with a p-value of 0.0001.

**Table 9| Pairwise of One-Way PERMANOVA performed on Ac samples to test for skin community variation according to increasing abundances of *S. aureus*.** Significant values are in bold

| | A | E | C | B | D |
|---|---|---|---|---|---|
| A | | **0.003** | **0.001** | 0.109 | **0.001** |
| E | **0.003** | | **0.003** | **0.009** | **0.008** |
| C | **0.001** | **0.003** | | **0.002** | **0.001** |
| B | 0.109 | **0.009** | **0.002** | | **0.001** |
| D | **0.001** | **0.008** | **0.001** | **0.001** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Ac with a p-value of 0.0001.

In Table 9, almost of the values are in bold. The values in bold are statistically relevant, in other words, there are significant dissimilarities between the samples related to those groups, in spite of the overlap between sample groups observed in Figures 18a and b.

Also, from Table 9 is possible to observe that there is no statistical significance between group A and B. In other words, the microbial community, from the samples with one to two percent of *S. aureus* are not significantly different from samples with zero to one percent. However in Figure 18, it is possible to see that there is a large extent of variability in structure of samples from group B, some of the samples are in fact similar to samples within group A but others are not. These two disparities can be due to other external factors or may be due to the slightest difference of percentage of *S. aureus* comparing to group A.

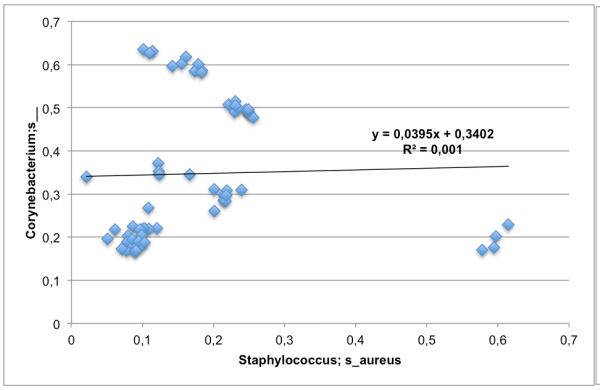
**Table 10** | **Results from SIMPER analysis of Ac samples.** SIMPER analysis identifying the percentage contribution of each OTU according to the Bray Curtis dissimilarity metric between groups A, B, C, D and E.

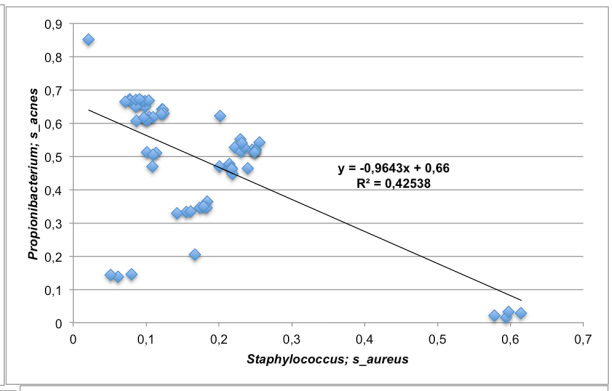
| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean A | Mean B | Mean C | Mean D | Mean E |
|---|---|---|---|---|---|---|---|---|
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;**g__Propionibacterium;s__acnes** | 1.348 | 2.442 | 2.442 | 0.574 | 0.588 | 0.409 | 0.521 | 0.0245 |
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;**g__Corynebacterium;s__** | 1.124 | 2.036 | 4.478 | 0.202 | 0.358 | 0.442 | 0.478 | 0.195 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__aureus** | 0.917 | 1.661 | 6.138 | 0.0803 | 0.112 | 0.19 | 0.242 | 0.596 |
| Root;k__Bacteria;**p__Proteobacteria**;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;**g__Pseudomonas;s__** | 0.6626 | 1.2 | 7.338 | 0.219 | 0.0627 | 0.0651 | 0.0375 | 0.00523 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__** | 0.5995 | 1.086 | 8.424 | 0.103 | 0.135 | 0.212 | 0.191 | 0.409 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__succinus** | 0.5956 | 1.079 | 9.503 | 0.0707 | 0.0817 | 0.159 | 0.173 | 0.386 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;**g__Streptococcus;s__** | 0.4934 | 0.8936 | 10.4 | 0.162 | 0.163 | 0.253 | 0.216 | 0.122 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__epidermidis** | 0.395 | 0.7153 | 11.11 | 0.0475 | 0.0616 | 0.104 | 0.0982 | 0.273 |
| Root;k__Bacteria;**p__Cyanobacteria**;c__Chloroplast;o__Streptophyta;f__;**g__;s__** | 0.3563 | 0.6453 | 11.76 | 0.1 | 0.114 | 0.0814 | 0.0708 | 0 |
| Root;k__Bacteria;**p__Proteobacteria**;c__Gammaproteobacteria;o__Enterobacteriales;**f__Enterobacteriaceae;g__;s__** | 0.3133 | 0.5674 | 12.32 | 0.121 | 0.0628 | 0.0653 | 0.0366 | 0.00393 |

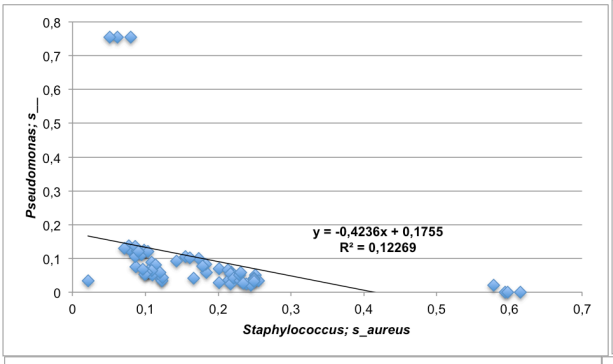
Similarity percentage analysis of the OTU diferences between the different groups of samples according to the relative frequency of *S. aureus*. The first collumn identifies the OTU explained by that row, the second collumn represents the average of dissimilarity ,the third collumn shows the % dissimilarity explained by that OTU, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the OTU thus far represented in the table and the last five show mean abundance of group A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%).

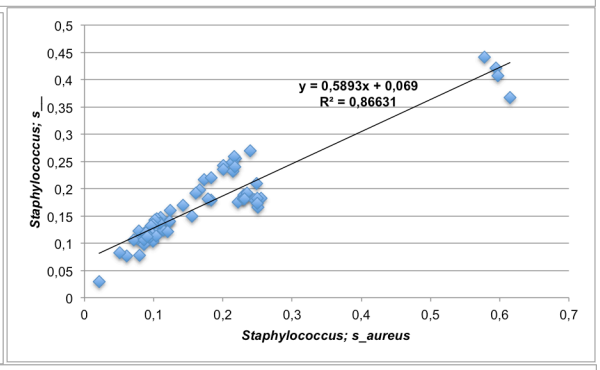
For the same reason as metioned when analysing Table 3, in this table the top ten results from the SIMPER analysis done for the Ac location only are listed, with a row for *Staphylococcus epidermidis*.

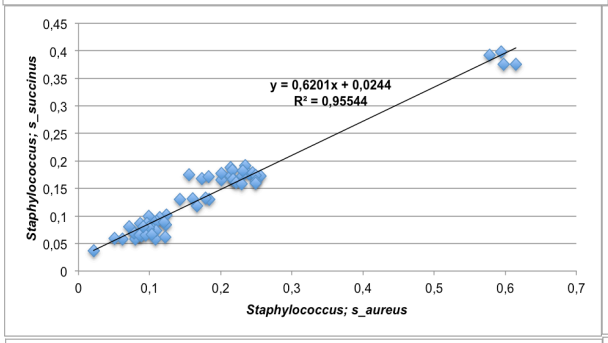
Therefore, it is also demanding to know if there is any interesting correlation between top 10 OTU results from this SIMPER analysis and *S. aureus*.

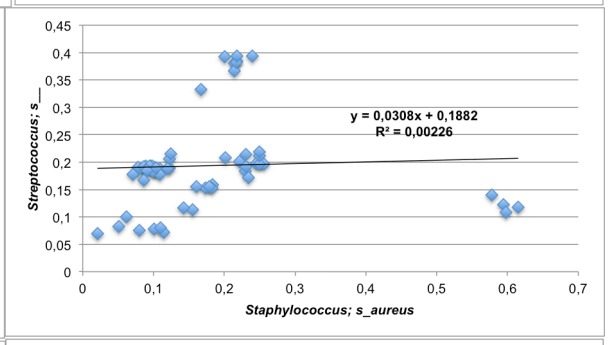

Corynebacterium;s__
y = 0,0395x + 0,3402
R² = 0,001
Staphylococcus; s_aureus

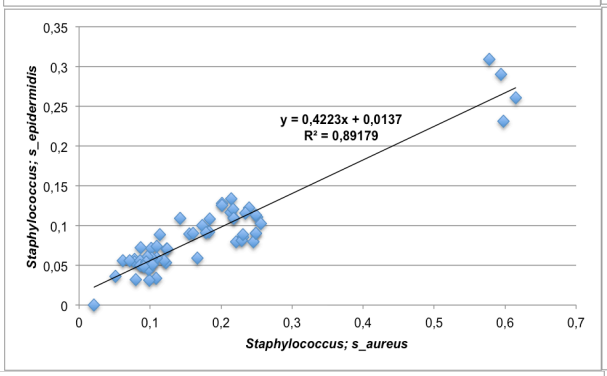

Propionibacterium; s_acnes
y = -0,9643x + 0,66
R² = 0,42538
Staphylococcus; s_aureus

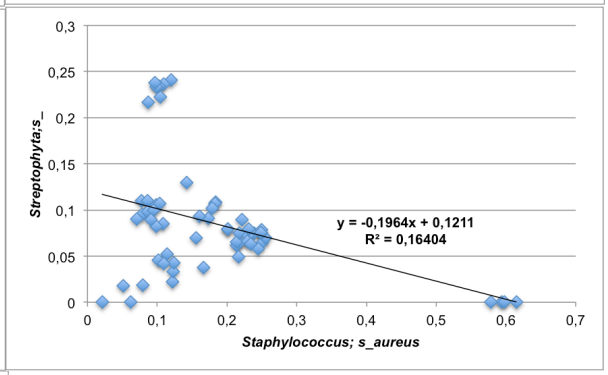

Pseudomonas; s__
y = -0,4236x + 0,1755
R² = 0,12269
Staphylococcus; s_aureus


Staphylococcus; s__
y = 0,5893x + 0,069
R² = 0,86631
Staphylococcus; s_aureus


Staphylococcus; s_succinus
y = 0,6201x + 0,0244
R² = 0,95544
Staphylococcus; s_aureus


Streptococcus; s__
y = 0,0308x + 0,1882
R² = 0,00226
Staphylococcus; s_aureus


Staphylococcus; s_epidermidis
y = 0,4223x + 0,0137
R² = 0,89179
Staphylococcus; s_aureus


Streptophyta;s_
y = -0,1964x + 0,1211
R² = 0,16404
Staphylococcus; s_aureus

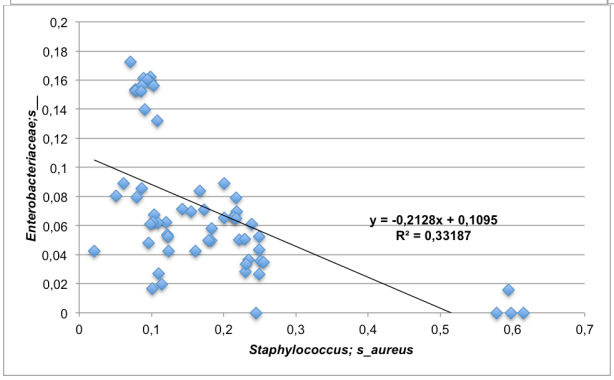

Enterobacteriaceae;s__
y = -0,2128x + 0,1095
R² = 0,33187
Staphylococcus; s_aureus

**Figure 19 | Linear Regression between the abundance of the ten most group-differentiating OTUs revealed by SIMPER analysis of Ac samples and *S. aureus*.** The data used for this linear regressions were Hellinger transformed.

The results of these linear regressions are in agreement with the results obtained in the analysis of the entire dataset. There are three positive correlations between *S. aureus* and *Staphylococcus* sp., *S. succinus* and *S. epidermidis.* And again the better positive correlation is between *S. aureus* and *S. succinus* with a high regression coefficient ($r^2$=0.9554). These conclusions reinforce the relationship between *S. aureus* and these bacteria. The relationship of *P. acnes* with *S. aureus* appears to be correlated negatively as it was observed also in Figure 16, which emphasizes their relationship.

The conclusions of the relationship of *S. aureus* with *Corynebacterium* sp. in Ac are only slightly different from those of the entire dataset. Here we can observe that the behaviour of *Corynebacterium* sp. is generally similar when the abundance of *S. aureus* is low or very high, while Figure 15 revealed a tendency for lower abundances of *Corynebacterium* in samples with *S. aureus* percentages higher than 10%. However, the robustness of these conclusions is affected by the fact that we have less samples with high *S. aureus* abundance in Ac than for the other amounts, precluding us to anticipate which would be the response of the other species in high *S. aureus* percentages with high accuracy. Once again these observations reinforce that *Corynebacterium* "resists" to high abundance of *S. aureus*, which agrees with the literature that says they are competitors (Ramsey et al., 2016).

The relationship between the other species and *S. aureus* all stil not clear, because, as metioned before either they show high abundances (in some samples only) when *S. aureus* abundance is very low or they show low abundance only when *S. aureus* is very high.

### 3.2.3 Taxonomic Analysis at the OTU level of Samples from Popliteal Creases

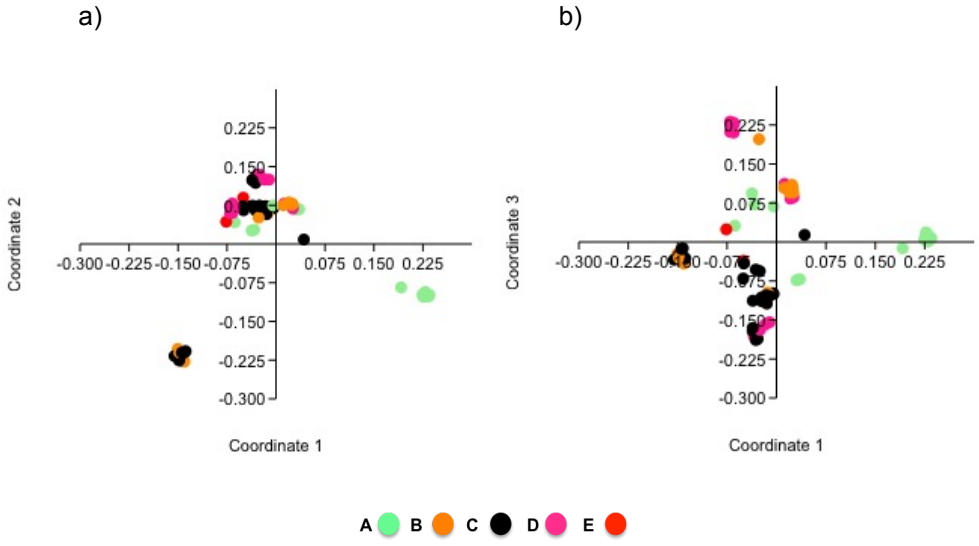



**Figure 20 | Principal Coordinates Analysis (PCoA) at OTU profiles retrived for Popliteal Creases (Pc) Samples (N=76).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of taxonomic profiles of microbial communities of Pc according to the relative percentage of *S. aureus* in the samples; A (0-1%), B (1-2%), C (2-5%), D

(5-10%) and E (More than 10%); Coordinate 1 versus Coordinate 2. (b) PCoA of taxonomic profiles of microbial communities of Pc according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

Taking into account the results obtained via PCoA for the Pc taxonomic data, there is no evidence of formation of clusters, as it was also concluded for Ac PCoA results. The samples are very dispersed so it is impossible to take conclusions about if the microbiome of the samples are or are not similar. For a better understanding, the permutation test PERMANOVA was carried out (Table 11).

**Table 11 | Summary of One-Way PERMANOVA performed on Popliteal Creases (Pc) Samples (N=76) to test for skin community variation according to increasing abundances of *S. aureus*.** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 10.96 |
| Within-group sum of squares: | 8.097 |
| F: | 6.278 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Pc with a p-value of 0.0001.

**Table 12 | Pairwise of One-Way PERMANOVA performed on Pc samples to test for skin community variation according to increasing abundances of *S. aureus*.** Significant values are in bold.

| | C | A | E | D | B |
|---|---|---|---|---|---|
| C | | **0.001** | **0.021** | **0.001** | **0.001** |
| A | **0.001** | | 0.318 | **0.001** | **0.001** |
| E | **0.021** | 0.318 | | 0.567 | 0.47 |
| D | **0.001** | **0.001** | 0.567 | | **0.001** |
| B | **0.001** | **0.001** | 0.47 | **0.001** | |

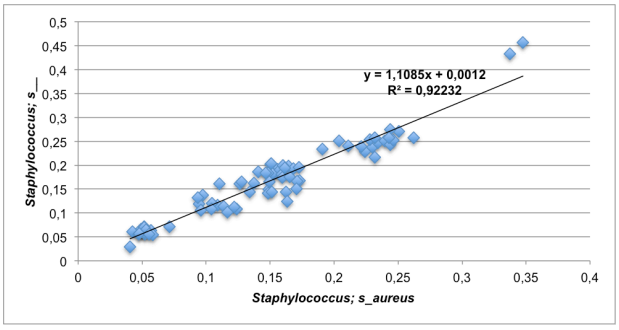
Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Pc with a p-value of 0.0001.

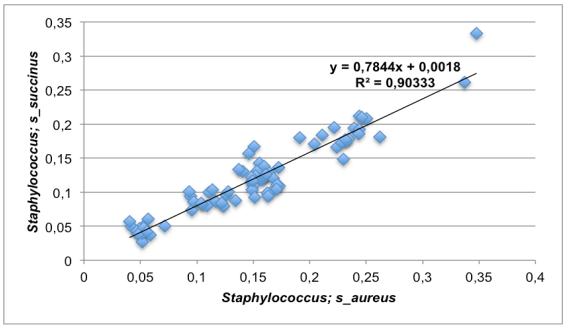
Observing the charts and complementing with the information of PERMANOVA results, it seems that the samples from the individuals with high abundance of *S. aureus*, especially from group D, have a somehow distinct microbial community in comparison with other *S. aureus* abundance groups. However, the trends observed here were not as clear as in the Ac site (Figure 18). Indeed, noticing Figure 20, the dots of group E did not form a cluster clearly separated from the others. The dots that are visible are closer to different groups: A, B and D. This can be grounded by the results of Pairwise PERMANOVA (Table 12), which revealed no significant statistical dissimilarities between group E and A, B and C. Nevertheless, significant differences between the other groups could be observed.

**Table 13 | Results from SIMPER analysis of Pc samples.** SIMPER analysis identifying the percentage contribution of each OTU according to the Bray Curtis dissimilarity metric between groups A, B, C, D and E.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean A | Mean B | Mean C | Mean D | Mean E |
|---|---|---|---|---|---|---|---|---|
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;**g__Corynebacterium;s__** | 0.8881 | 1.603 | 1.603 | 0.508 | 0.415 | 0.539 | 0.434 | 0.32 |
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;**g__Propionibacterium;s__acnes** | 0.6748 | 1.218 | 2.822 | 0.173 | 0.329 | 0.275 | 0.177 | 0.252 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__** | 0.5225 | 0.9434 | 3.765 | 0.0732 | 0.136 | 0.184 | 0.25 | 0.445 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__aureus** | 0.4793 | 0.8654 | 4.63 | 0.061 | 0.121 | 0.165 | 0.239 | 0.342 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Peptoniphilus;s__** | 0.4501 | 0.8126 | 5.443 | 0.18 | 0.111 | 0.0532 | 0.0617 | 0.0433 |
| Root;k__Bacteria;**p__Bacteroidetes**;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;**g__Prevotella;s__** | 0.4042 | 0.7298 | 6.173 | 0.154 | 0.0574 | 0.0634 | 0.0204 | 0.00889 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Anaerococcus;s__** | 0.3857 | 0.6965 | 6.869 | 0.18 | 0.152 | 0.0932 | 0.0971 | 0.0654 |
| Root;k__Bacteria;**p__Proteobacteria**;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;**g__Pseudomonas;s__** | 0.377 | 0.6808 | 7.55 | 0.127 | 0.0863 | 0.036 | 0.0494 | 0.0658 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Finegoldia;s__** | 0.3747 | 0.6765 | 8.227 | 0.131 | 0.155 | 0.0667 | 0.0634 | 0.0473 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__succinus** | 0.3703 | 0.6686 | 8.895 | 0.0541 | 0.0961 | 0.129 | 0.185 | 0.297 |
| (...) | | | | | | | | |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__epidermidis** | 0.243 | 0.4388 | 18.56 | 0.0338 | 0.054 | 0.0777 | 0.0874 | 0.264 |

Similarity percentage analysis of the OTU diferences between the diferent groups of samples according to the relative frequency of *S. aureus*. The first collumn identifies the OTU explained by that row, the second collumn represents the average of dissimilarity ,the third collumn shows the % dissimilarity explained by that OTU, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the OTU thus far represented in the table and the last five show mean abundance of group A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%).



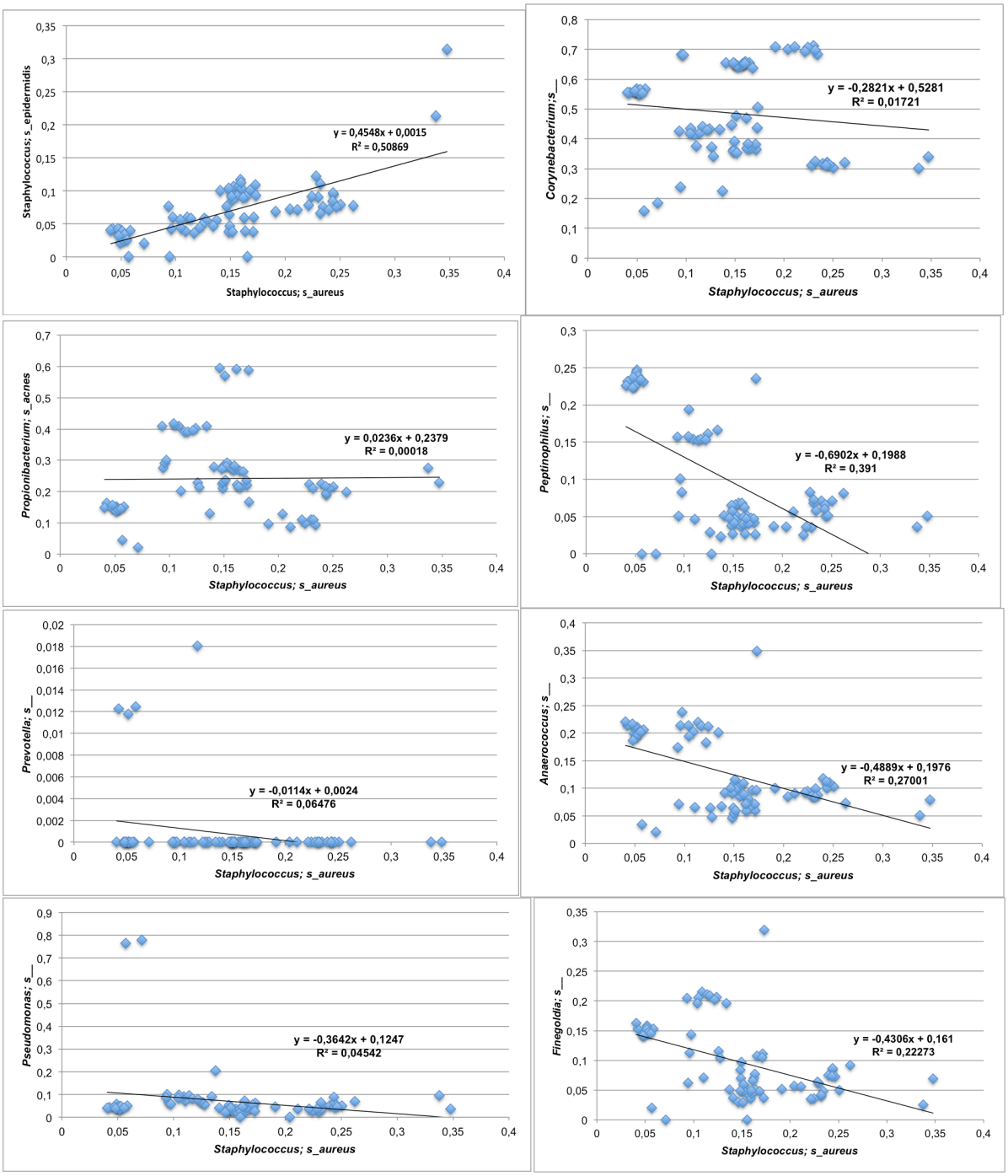



**Figure 21 | Linear Regression between the abundance of the ten most group-differentiating OTUs revealed by SIMPER analysis of Pc samples and *S. aureus*.** The data used for this linear regressions were Hellinger transformed.

The results from these linear regressions are slightly different from those of Ac. There is still a good positive linear correlation between *S. aureus* and *Staphylococcus* sp. and *S. succinus*. Notwithstanding, *S. epidermidis* has still a positive correlation with *S. aureus*, however much less pronounced than in Ac. There is a "big jump" when looking to the samples with high abundance of *S. aureus* and the ones with low and medium abundance. It seems that, when we are towards high percentage of *S. aureus* there is a sudden and sharp increase of *S. epidermidis* abundance. If tooking out the samples from group E, for sure that will result in a smoother positive linear correlation.

About the charts related to the relationship between *S. aureus* and *Corynebacterium* sp. and *P. acnes*, it must be taken into account that there is not enough number of samples from group E. Nevertheless, the results reveal no correlation when we look to the horizontal line that reflects these relationships, contrasting previous observations on the inverse relationship between *S. aureus* and *P. acnes* (all data and Ac).

Concerning *Prevotella* and *Pseudomonas* sp., these OTUs were found to display extremely low abundances already when *S. aureus* abundances were only moderate, suggesting these organisms are highy suppressed by the presence of *S. aureus* on human skin, particularly in Pc sites.

Regarding *Peptinophilus*, *Anaerococcus* and *Finegaldia* the only correlation is that these three species are very abundant when there are low abundance of *S. aureus*. For medium and high percentage of *S. aureus*, the proportion of these three bacteria stabilizes.

### 3.2.4 Taxonomic Analysis at the OTU level of Samples from Volar Forearm

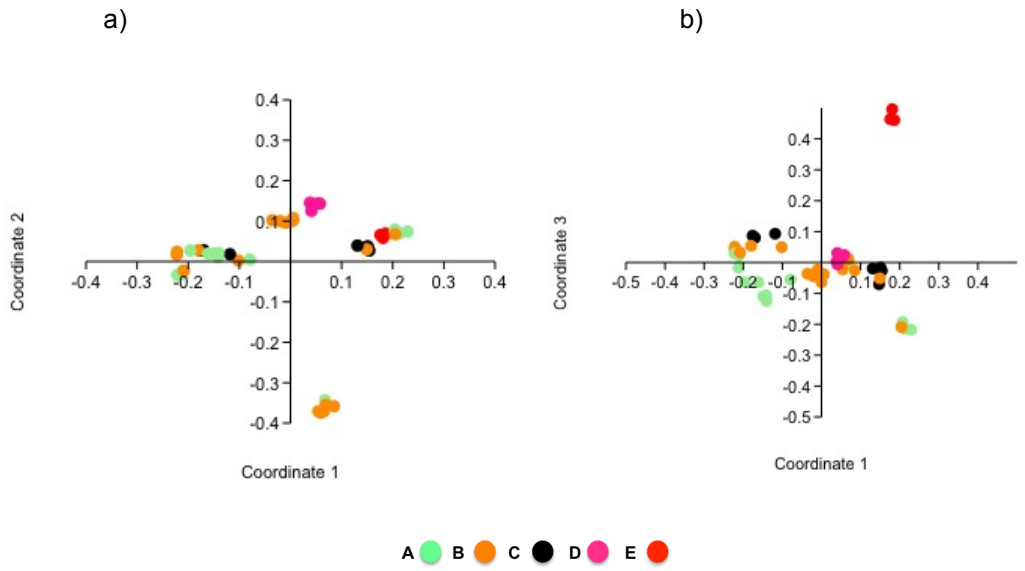



**Figure 22 | Principal Coordinates Analysis (PCoA) at OTU profiles retrived for Volar Forearm (Vf) Samples (N=47).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of taxonomic profiles of microbial communities of Vf according to the relative percentage of *S. aureus* in the samples; A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%); Coordinate 1 versus Coordinate 2. (b) PCoA of taxonomic profiles of microbial communities of Vf according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

This PCoA demonstrates that, overall, there was no clear formation of clusters and it seems that all the samples had a similar microbial community. Exception to this are samples from group E in chart b), whereby it is conclusive that these samples have a microbial community very different from the other samples. For further information about possible differences in microbiome structure, it is necessary to analyse the PERMANOVA results.

**Table 14 | Summary of One-Way PERMANOVA performed on Volar Forearm (Vf) Samples (N=47) to test for skin community variation according to increasing abundances of *S. aureus*.** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 6.526 |
| Within-group sum of squares: | 4.956 |
| F: | 3.326 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Vf with a p-value of 0.0001.

**Table 15 | Pairwise of One-Way PERMANOVA performed on Vf samples to test for skin community variation according to increasing abundances of *S. aureus*.** Significant values are in bold.

| | E | B | A | C | D |
|---|---|---|---|---|---|
| E | | **0.001** | **0.0021** | **0.0082** | **0.0157** |
| B | **0.001** | | **0.0096** | 0.0612 | **0.0021** |
| A | **0.0021** | **0.0096** | | **0.012** | **0.0001** |
| C | **0.0082** | 0.0612 | **0.012** | | **0.0014** |
| D | **0.0157** | **0.0021** | **0.0001** | **0.0014** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Vf with a p-value of 0.0001.

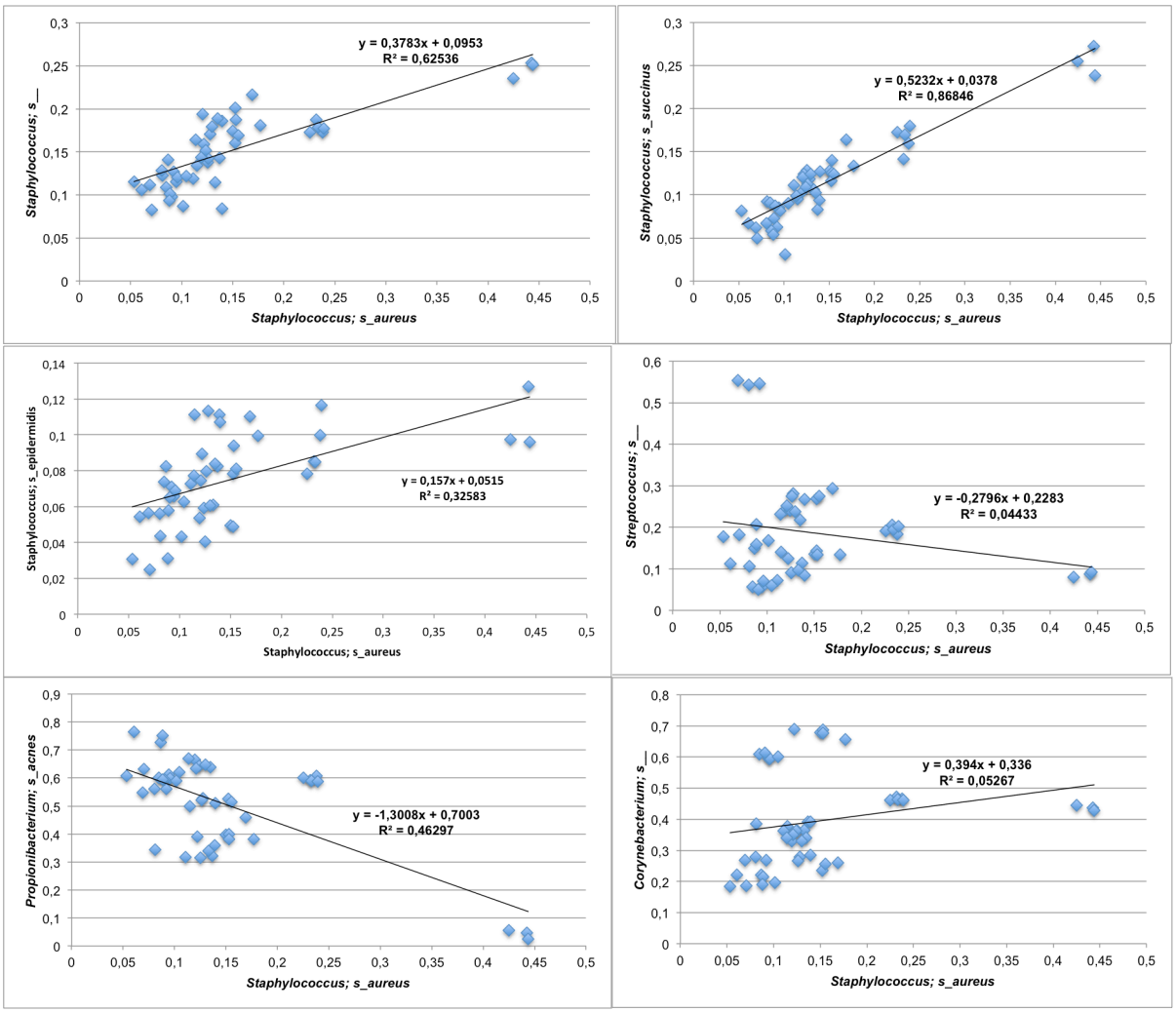
According to Pairwise results, there are no dissimilar significant statistical results between groups B and C, which is in accordance with the results of the PCoA charts. These two groups have short distances in between. All the other groups from Vf site have significant dissimilarities.

**Table 16 | Results from SIMPER analysis of Vf samples.** SIMPER analysis identifying the percentage contribution of each OTU according to the Bray Curtis dissimilarity metric between groups A, B, C, D and E.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean A | Mean B | Mean C | Mean D | Mean E |
|---|---|---|---|---|---|---|---|---|
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;**g__Propionibacterium;s__acnes** | 1.216 | 2.271 | 2.271 | 0.607 | 0.512 | 0.437 | 0.596 | 0.0422 |
| Root;k__Bacteria;**p__Actinobacteria**;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;**g__Corynebacterium;s__** | 1.119 | 2.091 | 4.362 | 0.345 | 0.368 | 0.493 | 0.464 | 0.436 |
| Root;k__Bacteria;**p__Proteobacteria**;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;**g__Pseudomonas;s__** | 0.832 | 1.554 | 5.916 | 0.114 | 0.188 | 0.0634 | 0.0324 | 0.00692 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;**g__Streptococcus;s__** | 0.7283 | 1.36 | 7.276 | 0.212 | 0.178 | 0.198 | 0.196 | 0.0869 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__aureus** | 0.6574 | 1.228 | 8.504 | 0.0812 | 0.124 | 0.159 | 0.234 | 0.437 |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__succinus** | 0.3889 | 0.7264 | 9.231 | 0.0723 | 0.104 | 0.133 | 0.165 | 0.255 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];**g__Anaerococcus;s__** | 0.3574 | 0.6676 | 9.898 | 0.0742 | 0.0749 | 0.101 | 0.0667 | 0.178 |
| Root;k__Bacteria;**p__Firmicutes**;c__Clostridia;o_ | 0.3387 | 0.6327 | 10.53 | 0.0408 | 0.0217 | 0.0362 | 0.0535 | 0.239 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| _Clostridiales;f__[Tissierellaceae];**g__Peptoniphilus;s__** | | | | | | | | |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__** | 0.334 | 0.6239 | 11.15 | 0.112 | 0.146 | 0.184 | 0.178 | 0.247 |
| Root;k__Bacteria;**p__Proteobacteria**;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;**g__;s__** | 0.3333 | 0.6226 | 11.78 | 0.0417 | 0.0792 | 0.0877 | 0.0698 | 0.00692 |
| (...) | | | | | | | | |
| Root;k__Bacteria;**p__Firmicutes**;c__Bacilli;o__Bacillales;f__Staphylococcaceae;**g__Staphylococcus;s__epidermidis** | 0.1758 | 0.3284 | 30.2 | 0.0559 | 0.0769 | 0.0801 | 0.0929 | 0.107 |

Similarity percentage analysis of the OTU diferences between the different groups of samples according to the relative frequency of *S. aureus*. The first collumn identifies the OTU explained by that row, the second collumn represents the average of dissimilarity ,the third collumn shows the % dissimilarity explained by that OTU, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the OTU thus far represented in the table and the last five show mean abundance of group A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%).



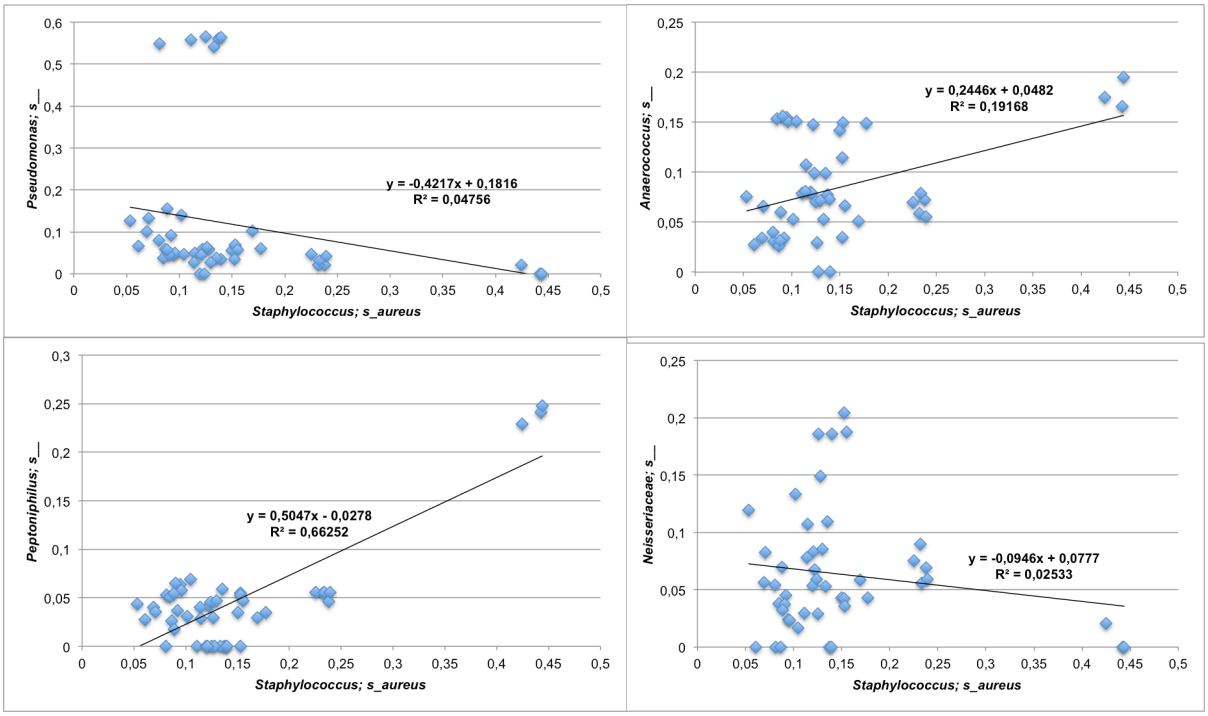



**Figure 23 | Linear Regression between the abundance of the ten most group-differentiating OTUs revealed by SIMPER analysis of Vf samples and *S. aureus*.** The data used for this linear regressions were Hellinger transformed.

As in the other linear regression analyses, positive correlations between *S. aureus* and *Staphylococcus* sp. and *S. succinus* could be identified. It is clear that these two *Staphylococcus* species have a relationship with *S. aureus*. Because they all belong to the same genus, it is expected that these species share a number of physiological similarities tha may underlie equivalent responses to the skin environment and its microbiome, provided that intense competition for the same resources does not compromise their co-existence. However, as it was metioned before for the Pc samples, *S. epidermidis* did not show exactly the same pattern of correlation with *S. aureus* as did the other two *Staphylococcus* species. In fact, in this particular case, we can see a larger extent of variability in the relative abundance of *S. epidermidis* when percentage abundances of *S. aureus* are low or moderate. At the presence of high percentage of *S. aureus*, there are also high percentage values of *S. epidermidis*.

Relatively to *Pseudomonas* and *Streptococcus* sp., it seems that there is no significant influence by the amount of *S. aureus* in the samples. For *Neisseriaceae* sp. it seems also that there is no significant influence when the abundance of *S. aureus* is low to moderate. However, when it is in the presence of more *S. aureus* it disappears. These observations are once more in line with the fact that, usually, when in the presence of high abundance of *S. aureus* the diversity of the total microbial community in the skin is low (Kong, 2012).

In accordance with results from the other sites, in Vf samples, *P. acnes* had a negative linear correlation with *S. aureus*. Regarding *Corynebacterium* sp. no correlation with *S. aureus* was found to exist, because even when it is in the presence of high abundance of *S. aureus* it seems that nothing had changed. For *Anaerococcus*, it was observed that it displays high abundances when the abundance of *S. aureus* is likewise very high. Last but not least, *Peptoniphilus* had a curious positive

linear correlation with *S. aureus*. This last observation was not found in the other areas or even in the global analysis.

## 3.3 Functional Analysis

### 3.3.1 Functional Analysis of All the Samples

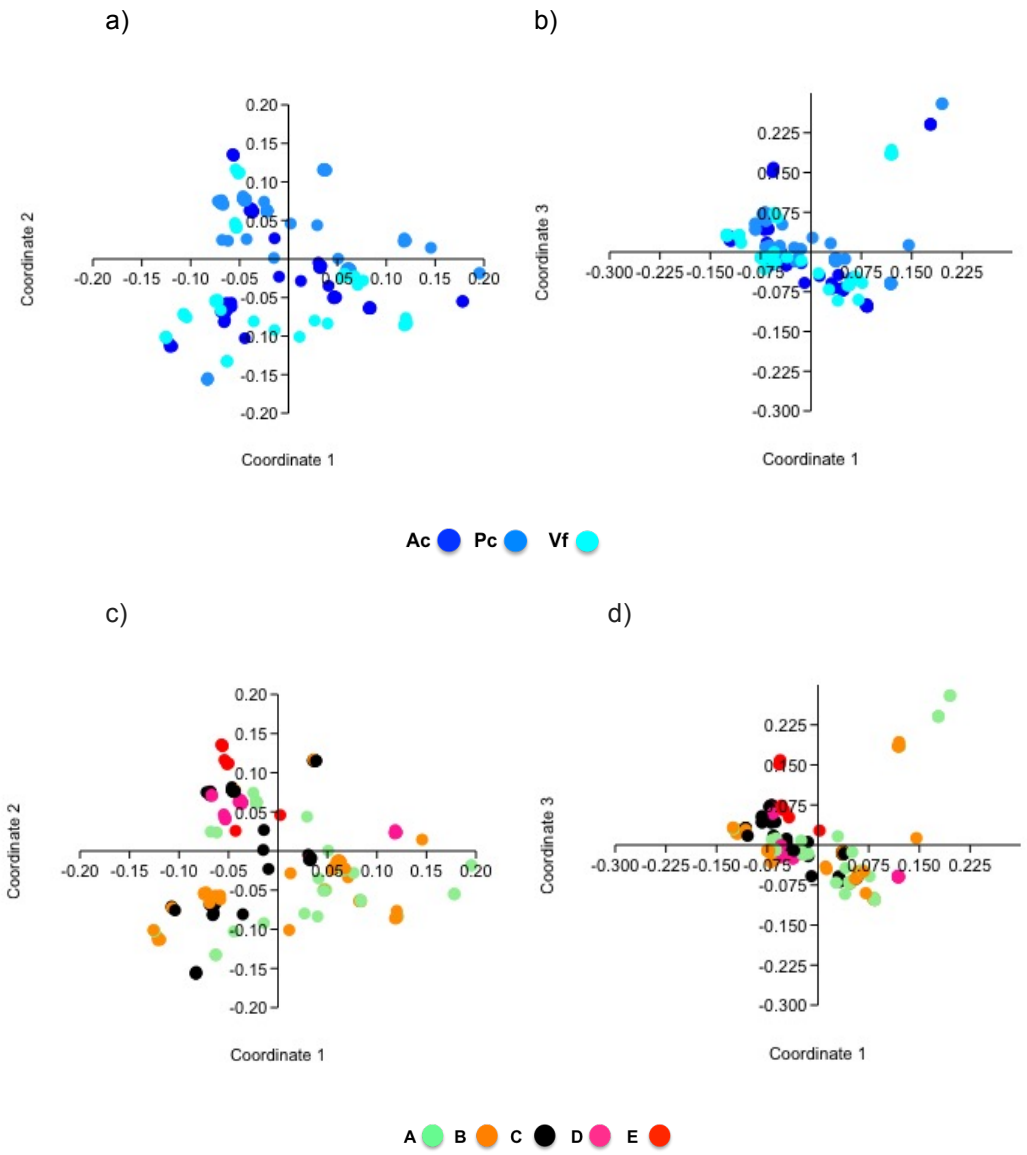



**Figure 24 | Principal Coordinates Analysis (PCoA) of IPR profiles retrieved for All Samples (N=183).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of IPR profies from the microbial community of the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and Volar Forearm (Vf); Coordinate 1 versus Coordinate 2. (b) PCoA of IPR profiles from the microbial community between the three areas: Antecubial Creases (Ac), Popliteal Creases (Pc) and Volar Forearm (Vf); Coordinate 1 versus Coordinate 3. (c) PCoA of IPR profiles from the microbial community according to the relative percentage of *S. aureus* in the samples; A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%); Coordinate 1 versus Coordinate 2. (d) PCoA of IPR profiles from the microbial community according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

No particular functional clusters were formed according to the site of origin of the samples (Figure 24 a) and b)). In graphs c) and d), in general, the the distances between samples of the same group are smaller suggesting that, after all, the clusters are formed, to a higher extent, according to the similarities of the percentage of *S. aureus* in the samples than site of origin. However, still a large overlap between samples of different groups was observed, with exception of group E. Therefore, a permutational test – PERMANOVA – was performed t test for eventual significant differences in skin microbiome functional profiles according to site of origin and % *S. aureus*.

**Table 17 | Summary of One-Way PERMANOVA for skin microbiome functional profiles per site.** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 3.639 |
| Within-group sum of squares: | 3.351 |
| F: | 7.732 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples according to the area (Ac, Pc and Vf), disregarding differences in *S. aureus* abundances, with a p-value of 0.0001.

**Table 18 | Summary of One-Way PERMANOVA for skin microbiome functional profiles according to percentage of *S. aureus* in the samples.**

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 3.639 |
| Within-group sum of squares: | 3.019 |
| F: | 9.125 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples with a p-value of 0.0001

Both One-Way PERMANOVA tests suggest that there were significant differences in functional profiles between groups. Paiwise PERMANOVA tests were then run to determine which groups, for each comparison, were significantly different from the others.

**Table 19| Pairwise of One-Way PERMANOVA for skin microbiome functional profiles per site.** Significant values are in bold.

| | Ac | Pc | Vf |
|---|---|---|---|
| Ac | | **0.0003** | 0.0666 |
| Pc | **0.0003** | | **0.0003** |
| Vf | 0.0666 | **0.0003** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples according to the area (Ac, Pc and Vf), disregarding differences in *S. aureus* abundances, with a p-value of 0.0001.

Through the analysis of Table 19, it is possible to verify that samples of Ac and Vf were not statistically different from the functional point of view. In this case the PERMANOVA test corroborates the pattern of distribution of Ac and Vf samples in the ordination charts (Figure 24), in which a large overlap between samples can be observed.

**Table 20| Pairwise One-Way PERMANOVA for skin microbiome functional profiles according to percentage of *S. aureus* in the samples.** Significant values are in bold.

| | A | E | C | B | D |
|---|---|---|---|---|---|
| A | | **0.001** | **0.001** | 0.906 | **0.001** |
| E | **0.001** | | **0.001** | **0.001** | **0.001** |
| C | **0.001** | **0.001** | | **0.001** | **0.008** |
| B | 0.906 | **0.001** | **0.001** | | **0.001** |
| D | **0.001** | **0.001** | **0.008** | **0.001** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples with a p-value of 0.0001

The same is reflected by analysing Table 20 that there is no statistical significance between the IPR profiles from the microbial communities that are represented as group A and B. All the other groups were found to differ from one another in functional terms in spite of a certain degree of overlap of the samples in the ordination space (Figure 24).

To rank all of the IPRs that contribute the most to the total variation of the dataset, a SIMPER analysis of all samples was carried out. The results are summarized in Table 21.

**Table 21 | Results from SIMPER analysis on functional (IPR) profiles of all samples.** SIMPER analysis identifying the percentage contribution of each IPR according to the Bray Curtis dissimilarity metric between Ac, Pc and Vf.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean Ac | Mean Pc | Mean Vf |
|---|---|---|---|---|---|---|
| **Transposase, L1** | 0.01992 | 0.1011 | 0.1011 | 0.0238 | 0.006 | 0.0103 |
| **EAL domain** | 0.01659 | 0.08415 | 0.1852 | 0.0248 | 0.0251 | 0.0239 |
| **Transposase, mutator type** | 0.01653 | 0.08384 | 0.2691 | 0.0341 | 0.0483 | 0.0353 |
| CAMP factor | 0.01494 | 0.0758 | 0.3449 | 0.0283 | 0.0142 | 0.0292 |
| Domain of unknown function DUF1725 | 0.01461 | 0.07409 | 0.419 | 0.0163 | 0.00575 | 0.00712 |
| WD40/YVTN repeat-like-containing domain | 0.01414 | 0.07173 | 0.4907 | 0.0375 | 0.0311 | 0.041 |
| Amino acid permease/ SLC12A domain | 0.01398 | 0.07094 | 0.5616 | 0.0483 | 0.0332 | 0.0498 |
| **TonB-dependent receptor, beta-barrel** | 0.01393 | 0.07069 | 0.6323 | 0.0277 | 0.028 | 0.0279 |
| RNA polymerase sigma factor 54 interaction domain | 0.01264 | 0.06411 | 0.6964 | 0.02 | 0.0215 | 0.0196 |
| Protein of unknown function DUF576 | 0.01252 | 0.06351 | 0.7599 | 0.0134 | 0.00623 | 0.00998 |

Similarity percentage analysis of IPR shifts in abundance across the different sites Ac, Pc and Vf. The first collumn identifies the IPR explained by that row, the second collumn represents the average of dissimilarity the third collumn shows the % dissimilarity explained by that IPR, the fourth

column is related to the cumulative Bray-Curtis dissimilarity metric for the IPR thus far represented in the table and the last three show mean abundance per sites: Ac, mean abundance at locus Pc and mean abundance at locus Vf.

### 3.3.2 Functional Analysis of Samples from Antecubital Creases

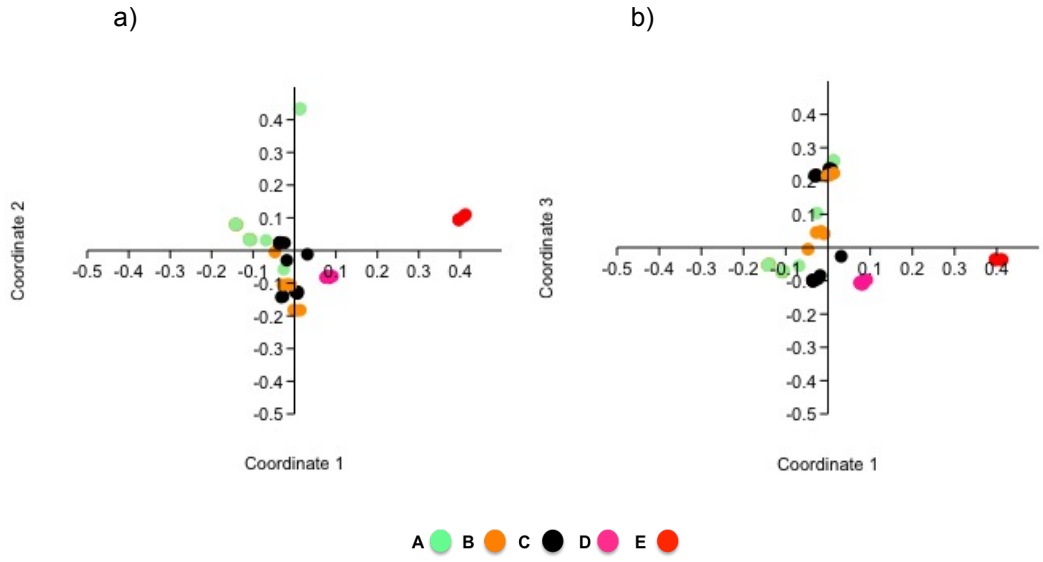



**Figure 25 | Principal Coordinates Analysis (PCoA) of IPR profiles retrieved for Antecubital Creases (Ac) Samples (N=60).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of IPR profiles from the microbial community of Ac according to the relative percentage of *S. aureus* in the samples; A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%); Coordinate 1 versus Coordinate 2. (b) PCoA of IPR profiles from the microbial community of Ac according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

It is easy to identify that groups D and E are more separate than the others, in other words it means that IPRs from group D and E are more similar in between samples than the others that are more scattered. Although dispersed, it is identifiable that de dots from group A, B and C are closer to each other, which may mean that even though the microbial communities have different abundance in *Staphylococcus aureus* that is not reflected in IPRs abundance.

**Table 22 | Summary of One-Way PERMANOVA of skin microbiome functional (IPR) profiles from the Antecubital Creases (Ac) sites.** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 1.197 |
| Within-group sum of squares: | 0.6714 |
| F: | 10.77 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Ac with a p-value of 0.0001

**Table 23 | Pairwise One-Way PERMANOVA of skin microbiome functional (IPR) profiles from the Ac site.** Significant values are in bold.

| | A | E | C | B | D |
|---|---|---|---|---|---|
| A | | **0.002** | **0.001** | **0.013** | **0.001** |
| E | **0.002** | | **0.004** | **0.006** | **0.004** |
| C | **0.001** | **0.004** | | 0.298 | **0.001** |
| B | **0.013** | **0.006** | 0.298 | | **0.001** |
| D | **0.001** | **0.004** | **0.001** | **0.001** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Ac with a p-value of 0.0001

In spite of the previous observation, the only IPR profiles found to show no statistical significance with one another were those from groups B and C. Once again, it is clear the importance of applying the PERMANOVA test to verify patterns of sample distribution in the ordination graphs.

According to Table 23, all of the other comparisons were significantly different.

To rank all of the IPRs that contribute the most to the total variation of the Ac dataset, a SIMPER analysis was performed. The results are summarized in Table 24.

**Table 24 | Results from SIMPER analysis on functional (IPR) profiles of Ac samples.** SIMPER analysis identifying the percentage contribution of each IPR according to the Bray Curtis dissimilarity metric between groups A, B, C, D and E.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean A | Mean B | Mean C | Mean D | Mean E |
|---|---|---|---|---|---|---|---|---|
| **Transposase, L1** | 0.03483 | 0.1746 | 0.1746 | 0.0158 | 0.00514 | 0.0557 | 0.0173 | 0.0042 |
| Domain of unknown function DUF1725 | 0.0241 | 0.1208 | 0.2953 | 0.0118 | 0.00378 | 0.0375 | 0.0111 | 0.00196 |
| Protein of unknown function DUF576 | 0.02261 | 0.1133 | 0.4086 | 0.00212 | 0.00442 | 0.00635 | 0.0228 | 0.09 |
| Endonuclease/exonuclease/phosphatase | 0.0188 | 0.09419 | 0.5028 | 0.0281 | 0.0289 | 0.0554 | 0.0329 | 0.0124 |
| **EAL domain** | 0.01702 | 0.08531 | 0.5881 | 0.0387 | 0.0231 | 0.0232 | 0.0153 | 0.00348 |
| Protein G-related, albumin-binding GA module | 0.0161 | 0.08066 | 0.6688 | 0.00416 | 0.00702 | 0.0103 | 0.0186 | 0.0656 |
| Domain of unknown function DUF1542 | 0.01484 | 0.07435 | 0.7431 | 0.00751 | 0.0117 | 0.0173 | 0.0215 | 0.0618 |
| Leukocidin/porin | 0.01455 | 0.07293 | 0.8161 | 0.000428 | 0.000261 | 0.00297 | 0.0156 | 0.054 |
| **TonB-dependent receptor, beta-barrel** | 0.01439 | 0.0721 | 0.8882 | 0.0407 | 0.0257 | 0.024 | 0.0215 | 0.00959 |
| FIVAR domain | 0.0135 | 0.06766 | 0.9558 | 0.00447 | 0.00744 | 0.0103 | 0.0166 | 0.0555 |

Similarity percentage analysis of IPR shifts in abundance between the different groups of samples according to the relative frequency of *S. aureus*. The first collumn identifies the IPR explained by that row, the second collumn represents the average of dissimilarity the third collumn shows the % dissimilarity explained by that IPR, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the IPR thus far represented in the table and the last five show mean abundance of group A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%).

### 3.3.3 Functional Analysis of Samples from Popliteal Creases

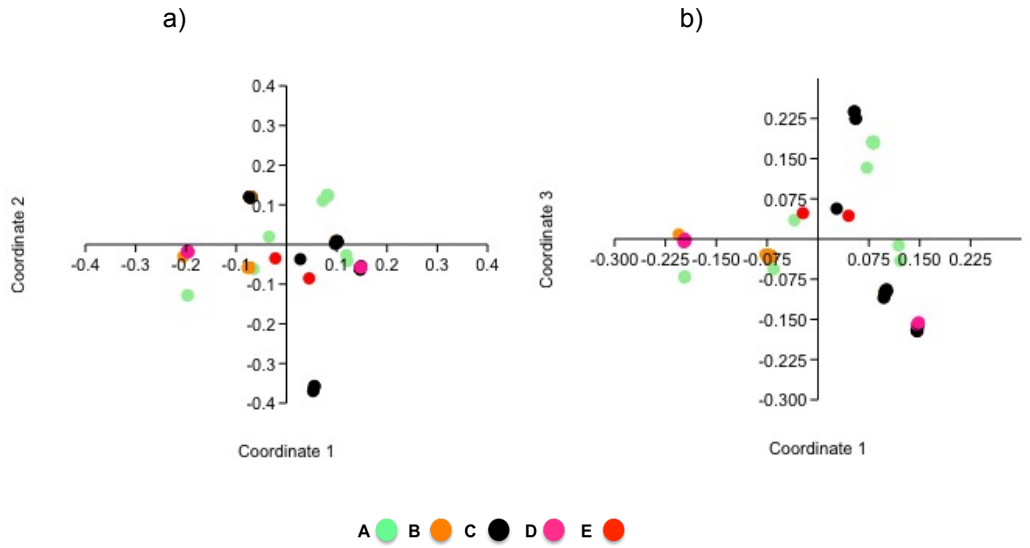



**Figure 26 | Principal Coordinates Analysis (PCoA) of IPR profiles retrieved for Popliteal Creases (Pc) Samples (N=76).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of IPR profiles from the microbial community of Pc according to the relative percentage of *S. aureus* in the samples; A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%); Coordinate 1 versus Coordinate 2. (b) PCoA of IPR profiles from the microbial community of Pc according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

There is a lot of dispersion of the samples which may indicate that the IPRs are dissimilar or so similar between samples that even when comparing samples with the different abundance of *S. aureus* is not possible to identify distinct clusters.

**Table 25 | Summary of One-Way PERMANOVA of skin microbiome functional (IPR) profiles from the Pc site.** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 1.327 |
| Within-group sum of squares: | 1.04 |
| F: | 4.894 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Pc with a p-value of 0.0001

**Table 26 | Pairwise One-Way PERMANOVA of skin microbiome functional (IPR) profiles from the Pc site**. Significant values are in bold.

| | C | A | E | D | B |
|---|---|---|---|---|---|
| C | | **0.001** | 0.152 | **0.004** | **0.007** |
| A | **0.001** | | 0.466 | **0.001** | **0.005** |
| E | 0.152 | 0.466 | | 0.086 | 0.086 |
| D | **0.004** | **0.001** | 0.086 | | 0.09 |
| B | **0.007** | **0.005** | 0.086 | 0.09 | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Pc with a p-value of 0.0001

Observing the results from Tables 25 and 26, it is intriguing that all of the IPRs from the samples of group E, with high abundance of *S. aureus* are statistically similar. These results were in principle not expected, since it was hypothesized that samples with higher abundance of this species that is, until now, the most studied bacterium implicated with AD, would present a disruption in function. Because only two samples belonging to site E were included in the Pc analysis, it is likely that the lack of significance results from low robustness of the permutational test when very few samples (observations) are considered. Another interesting result is that IPRs from samples with 1-2% and 5-10% are more similar to one another than for example with samples with nearest *S. aureus* percentage values.

To rank all of the IPRs that contribute the most to the total variation of the Pc dataset, SIMPER analysis was performed. The results are summarized in Table 27.

**Table 27 | Results from SIMPER analysis on functional (IPR) profiles of Pc samples.** SIMPER analysis identifying the percentage contribution of each IPR according to the Bray Curtis dissimilarity metric between groups A, B, C, D and E.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean A | Mean B | Mean C | Mean D | Mean E |
|---|---|---|---|---|---|---|---|---|
| **Transposase, mutator type** | 0.01755 | 0.09492 | 0.09492 | 0.0516 | 0.0471 | 0.0508 | 0.0412 | 0.034 |
| **EAL domain** | 0.01678 | 0.09077 | 0.1857 | 0.0229 | 0.0321 | 0.0177 | 0.0384 | 0.0189 |
| **TonB-dependent receptor, beta-barrel** | 0.01405 | 0.07597 | 0.2617 | 0.0367 | 0.0326 | 0.0211 | 0.0266 | 0.0258 |
| NAD(P)-binding domain | 0.01309 | 0.0708 | 0.3325 | 0.147 | 0.164 | 0.153 | 0.17 | 0.154 |
| WD40/YVTN repeat-like-containing domain | 0.01281 | 0.0693 | 0.4018 | 0.0272 | 0.0297 | 0.0314 | 0.0374 | 0.0311 |
| Winged helix-turn-helix DNA-binding domain | 0.01271 | 0.06872 | 0.4705 | 0.119 | 0.13 | 0.121 | 0.129 | 0.124 |
| RNA polymerase sigma factor 54 interaction domain | 0.01258 | 0.06802 | 0.5385 | 0.0225 | 0.0252 | 0.0162 | 0.0294 | 0.0137 |
| Aldehyde oxidase/xanthine dehydrogenase, molybdopterin binding | 0.01227 | 0.06635 | 0.6048 | 0.0196 | 0.0274 | 0.0207 | 0.0353 | 0.0146 |
| **GGDEF domain** | 0.01188 | 0.06422 | 0.6691 | 0.0163 | 0.0248 | 0.0149 | 0.0279 | 0.0186 |
| P-loop containing nucleoside triphosphate hydrolase | 0.01123 | 0.06073 | 0.7298 | 0.274 | 0.266 | 0.27 | 0.261 | 0.25 |

Similarity percentage analysis of IPR shifts in abundance between the different groups of samples according to the relative frequency of *S. aureus*. The first collumn identifies the IPR explained by that row, the second collumn represents the average of dissimilarity the third collumn shows the % dissimilarity explained by that IPR, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the IPR thus far represented in the table and the last five show mean abundance of group A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%).

### 3.3.4 Functional Analysis of Samples from Volar Forearm

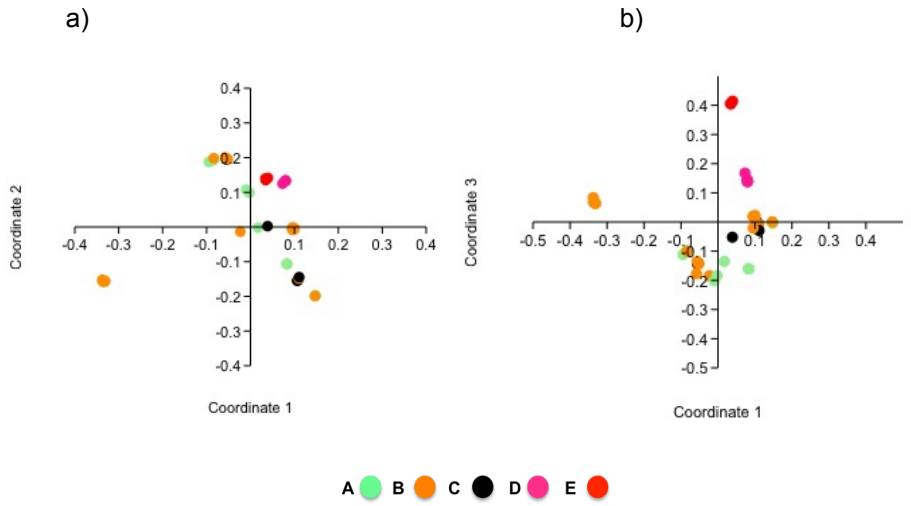



**Figure 27 | Principal Coordinates Analysis (PCoA) of IPR profiles retrieved for Volar Forearm (Vf) Samples (N=47).** PCoA performed on Bray-Curtis dissimilarity index calculated from Hellinger transformed data. (a) PCoA of IPR profiles from the microbial community of Vf according to the relative percentage of *S. aureus* in the samples; A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%); Coordinate 1 versus Coordinate 2. (b) PCoA of IPR profiles from the microbial community of Vf according to the relative percentage of *S. aureus* in the samples; Coordinate 1 versus Coordinate 3.

From the PCoA ordination diagrams, (Figure 27), it is possible to identify that IPR profiles from groups A and B are variable and overlap across the ordination space, that B and C also overlap to some extent and that maybe IPR profiles from groups D and E are also similar.

It can also be seen that, especially in chart b), clusters from the different samples, accounting of their abundance of *S. aureus*, are well identified, especially concerning the differentiation between sites D and E.

**Table 28 | Summary of One-Way PERMANOVA of skin microbiome functional (IPR) profiles from the Volar Forearm (Vf) sites.** Results of the Permutation test.

| | |
|---|---|
| Permutation N: | 9999 |
| Total sum of squares: | 0.827 |
| Within-group sum of squares: | 0.6198 |
| F: | 3.51 |
| p (same): | 0.0001 |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Vf with a p-value of 0.0001

**Table 29 | Pairwise One-Way PERMANOVA of skin microbiome functional (IPR) profiles from the Vf site.** Significant values are in bold.

| | E | B | A | C | D |
|---|---|---|---|---|---|
| E | | **0.006** | **0.019** | 0.072 | 0.188 |
| B | **0.006** | | 1 | 0.842 | 0.062 |
| A | **0.019** | 1 | | 1 | **0.005** |
| C | 0.072 | 0.842 | 1 | | **0.033** |
| D | 0.188 | 0.062 | **0.005** | **0.033** | |

Results of the permutational Analysis of Variance for the distance matrix with 9999 permutations, supporting statistical differences between the samples of Vf with a p-value of 0.0001

With the results from the pairwise PERMANOVA, we can confirm some of the assumptions made when observing the results of PCoA and also get reliable results of other relationships that are not possible to be identified at first sight. For example, the IPR profiles from the samples of group E are not significantly different from the ones of groups C and D. For the relationship between groups C and E if there were more charts with different coordinates or a 3D chart eventually assessments would reveal clearer patterns. Observing the results of the relationship between groups B and A, C, and D, as it was told when analysing the results of PCoA, the IPR profiles overlap considerably, so there were no significant differences in the functional profiles from these sample groups.

To rank all of the IPRs that contribute the most to the total variation of the Vf dataset, a SIMPER analysis was performed. The results are summarized in Table 30.

**Table 30 | Results from SIMPER analysis on functional (IPR) profiles of Vf samples.** SIMPER analysis identifying the percentage contribution of each IPR according to the Bray Curtis dissimilarity metric between groups A, B, C, D and E.

| Taxon | Av. dissim | Contrib. % | Cumulative % | Mean A | Mean B | Mean C | Mean D | Mean E |
|---|---|---|---|---|---|---|---|---|
| **EAL domain** | 0.01821 | 0.09759 | 0.09759 | 0.0236 | 0.0314 | 0.0208 | 0.012 | 0.00717 |
| Protein of unknown function DUF576 | 0.01543 | 0.08267 | 0.1803 | 0.00375 | 0.00469 | 0.00369 | 0.0246 | 0.0611 |
| **TonB-dependent receptor, beta-barrel** | 0.01466 | 0.07852 | 0.2588 | 0.0283 | 0.0329 | 0.0221 | 0.0195 | 0.0234 |
| **Transposase, L1** | 0.01435 | 0.07687 | 0.3357 | 0.00761 | 0.0133 | 0.0161 | 0.00194 | 0.0057 |
| Porin domain | 0.01421 | 0.07612 | 0.4118 | 0.0182 | 0.026 | 0.0122 | 0.0086 | 0.0106 |
| RNA polymerase sigma factor 54 interaction domain | 0.01404 | 0.07525 | 0.487 | 0.019 | 0.026 | 0.0145 | 0.0111 | 0.0108 |
| Transcription regulator HTH, LysR | 0.01397 | 0.07486 | 0.5619 | 0.0275 | 0.0365 | 0.0237 | 0.0238 | 0.0238 |
| Winged helix-turn-helix DNA-binding domain | 0.01287 | 0.06896 | 0.6308 | 0.115 | 0.125 | 0.115 | 0.117 | 0.113 |
| WD40/YVTN repeat-like-containing domain | 0.01258 | 0.06741 | 0.6983 | 0.0476 | 0.0397 | 0.047 | 0.0304 | 0.0217 |
| WD40 repeat | 0.01199 | 0.06423 | 0.7625 | 0.0301 | 0.0241 | 0.0302 | 0.0128 | 0.011 |

Similarity percentage analysis of IPR shifts in abundance between the different groups of samples according to the relative frequency of *S. aureus*. The first collumn identifies the IPR explained by that row, the second collumn represents the average of dissimilarity the third collumn shows the % dissimilarity explained by that IPR, the fourth column is related to the cumulative Bray-Curtis dissimilarity metric for the IPR thus far represented in the table and the last five show mean abundance of group A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (More than 10%).

Observing SIMPER results from tables 20, 23, 26 and 29, there are four IPRs that are common in all, or almost all, SIMPER ranks: Transposase, L1 (IPR004244); Transposase, mutator type (IPR001207); TonB-dependent receptor, beta-barrel (IPR000531); and EAL domain (IPR001633).

Transposase L1 is an enzyme that coworks with Long Interspersed Nuclear Elements (L1s), one of the most abundant retrotransoposons in the human genome (Moran, 1999).

The fact that L1 transposase is one of the results that most contributes to the dissimilarity of the samples may indicate the possibility that some of the samples are from patients who, due to a mutation, developed a propensity for AD. Even though group D and E had lower abundances of L1 transposases and we know that samples from D and E may be from people with AD, some of the other samples are from people with AD but clearly are not in a flare state. For fundamental conclusions, there should be a deeper research into this topic.

In fact, L1 retrotransposons and their ability to promote deleterious insertions lead to their identification as causal agents of disease in 1988. Since then, many scientists found that deleterious L1 insertions occasionally cause genetic disorders (Moran, 1999). However, most of L1s are inactive and uncapable to cause insertional mutagenesis (Ostertag et al., 2015).

“Transposase, mutator type” is an enzyme encoded by autonomous mobile genetic elements such as transposon or insertion sequences (IS). Curiously, there is an insertion sequence, Is256 that has been detected in the genome of several clinical isolates in the clinical methicillin-resistant *S. aureus* (Schreiber et al.,2013).

Since one of the main goals of this study was to detect what are the dissimilarities between microbial communities, according with increasing abundances of *S. aureus*, this may reflect that there are some *S. aureus* more resistant to antibiotics than others. Of course this conclusion is just an assumption and in order to have a more precise answer for the reason why Transposase, mutator type is one of the most differentianting IPRs among groups, a deeper research into metabolic dynamics would be needed.

EAL domain is found in several bacterial signaling proteins and it has been proven that is involved in the degradation of a second messenger, cyclic di-GMP (Krasteva et al., 2012). Cyclic di-GMP is an important second messenger that is involved in the regulation of bacterial life-style transitions relevant for biofilm formation and virulence. The three main domains that are involved with cyclic di-GMP are: EAL, GGDEF and HD-GYP (Pesavento & Hengge, 2009). The GGDEF domain (IPR000160) is also in the top 10 results of SIMPER analysis of Pc and according to the literature, controls the synthesis c-di-GMP (Römling & Amikam, 2006). Usually, the levels of c-di-GMP are associated with the ability of adaptational capacities of the bacteria. A higher level of c-di-GMP is synonym of a good capacity of biofilm formation, however a low level of c-di-GMP promotes biofilm dispersion and bacterial virulence. Frequently, opportunistic pathogens among the Gram-negative bacteria display a high level of c-di-GMP because of their need to adapt to different microenvironments (Krasteva et al., 2012). The SIMPER results reveal lower proportions of the GGDEF domain in the D and E groups, suggesting that the capacity of the skin microbiome to form biofilms under c-di-GMP regulation has been disrupted to some extent, probably due to the much

lower representativeness of Gram-negative bacteria in these samples, a group that is widely known to govern biofilm formation thorugh c-di-GMP.

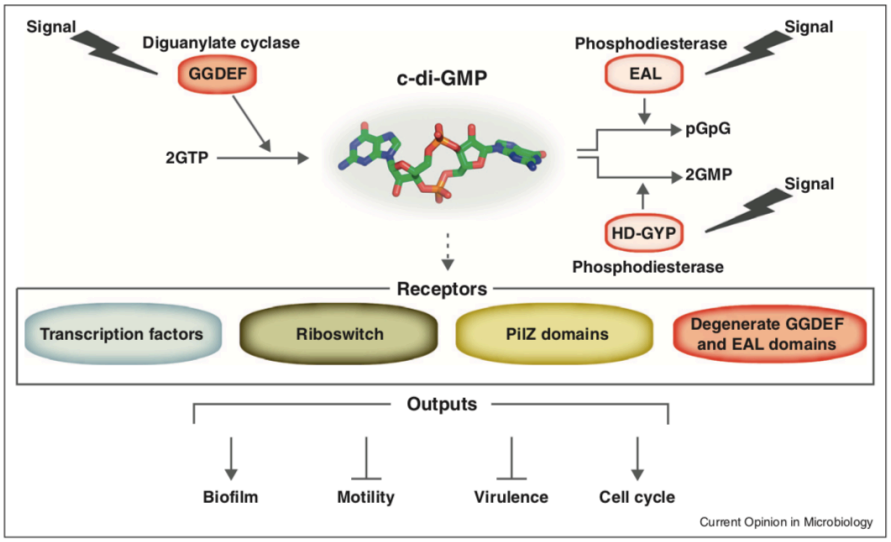



**Figure 28 | Schematization of the roles of Cyclic di-GMP and the relationship with EAL, GGDEF and HD-GYP.** Cyclic di-GMP is an important second messenger that is involved in the regulation of bacterial life-style transitions relevant for biofilm formation and virulence. The three main domains that are involved with cyclic di-GMP biosynthesis are: EAL, GGDEF and HD-GYP. GGDEF and the other two main domains are responsible for, respectively, synthesis and degradation of c-di-GMP in Gram-negative bacteria (Sondermann, Shikuma, & Yildiz, 2012).

However, a recent study on GGDEF and EAL domains and its relationship with c-di-GMP demonstrated that GGDEF protein GdpS, from *Staphylococcus epidermidis*, is able to regulate biofilm formation in *S. epidermidis* even independently of c-di-GMP (Zhu et al.,2017). It is still unclear whether such regulatory mechanism is valid for *S. aureus*, and how this ability can be affected by the microenvironment. The overall decrease in GGDEF and EAL domains in samples with high *S. aureus* percentages indicates that the genetic make-up of the skin microbiome is shifting towards a more virulent state given the overall inverse relationship between biolfim formation and virulence (Figure 28). Relating this information and the results of SIMPER, there is evidence that there will be samples with high or low levels of c-di-GMP, since EAL domain has a high contribution for the dissimilarity between samples. As mentioned above, the Gram-negative vs. Gram-positive abundances within each sample group from A through E are likely to interefere with this outcome.

TonB-dependet receptors (TBDRs), beta barrel are proteins located in the outer membrane of Gram-negative bacteria and have an important role in nutrient transport, such as iron, an essencial micronutrient for bacteria (Mosbahi et al.,2018). They act as pathways in response to outside ligands and import extracellular nutrients into periplasmic space (Wang et al., 2016).

During pathogenesis, the nutrient transport becomes a way of interaction between pathogen and host, which is a critical determining factor in the outcome of infection. As an immune response to the pathogen, hosts limit the availability of iron acquisition (Wang et al., 2016) (Mosbahi et al., 2018). However, to overcome the lack of iron, Gram-negative bacteria produce iron uptake systems that capture iron-containing substrates, such as siderophores. The binding of siderophore to TBDRs

triggers some mechanisms allowing translocation of the iron into the cytoplasm (Mosbahi et al., 2018).

After the analysis of SIMPER IPR ranks it was verified that along with the increase of percentage of *S. aureus* in the samples there was a decrease in abundance of TonB-dependent receptors, beta barrel. It makes sense since with the increase of *S. aureus* abundance, there is a decrease of microbiome diversity, namely of diverse Gram-negative bacteria, such as *P. acnes* and several others, and as a result this mechanism typical of virulent Gram-negative bacteria is less represented in sample groups D and E.

## 3.4 Analysis of *S. succinus* genomes

After linear regression analysis was carried out, based on the results obtained previously using the SIMPER test, a strong and positive correlation between *S. aureus* and *S. succinus* was revealed.

Until now, there is no information about the precise molecular interactions, eventual syntrophic behaviour and niche partitioning underlying the relationship between *S. aureus* and *S. succinus* in the human skin. Therefore, to explore the above mentioned issues making use of currently-available genomic information a comparative analysis of four *S. succinus* genomes with one genome representing a multi-resistant *S. aureus* strain was carried out using the freely-avialable server RAST. The strains used for the comparison were, *S. aureus* USA300 and *S. succinus* SNUC1280, 14BME20, DSM15096 and DSM14617.

### 3.4.1 Comparison of the genomes

The RAST sever provides some general data about the species under analysis. An overview of the general genomic attributes of the five strains inspected is listed below (Table 31).

**Table 31 | Comparison of size, GC content, subsystem coverage, numbers of coding sequences, RNAs and subsystems of strain *S. aureus* USA300 and strains of *S. succinus*: SNUC1280, 14BME20, DSM15096 and DSM14617.** Values available on RAST.

| | USA300 | SNUC1280 | 14BME20 | DSM15096 | DSM14617 |
|---|---|---|---|---|---|
| Size (bp) | 2 917 469 | 2 771 412 | 2 745 675 | 2 871 374 | 2 786 115 |
| GC – content (%) | - | 32.9 | 33.1 | 33 | 32.9 |
| Number of Contigs | 4 | 173 | 1 | 169 | 339 |
| Number of Subsystems | 385 | 406 | 409 | 405 | 406 |
| Subsystem Coverage (%) | 59 | 53 | 53 | 51 | 53 |
| Number of Coding Sequences | 2608 | 2623 | 2586 | 2751 | 2624 |
| Number of RNAs | 105 | 73 | 80 | 72 | 67 |

From this organisms' overview, it is possible to observe that *S. aureus* USA300 had slightly larger genome size (2 917 469 bp), subsystem coverage (59) and number of RNAs (105) than the *S. succinus* strains. Although it has the highest subsystem coverage, the number of subsystems (385) is lower than the average of subsystems in *S. succinus* (~407).

Between *S. succinus* strains, DSM15096 had the largest genome size (2 871 374bp), also the highest number of coding sequences (2751), however not the highest number of RNAs. The number of subsystems did not vary much.

Posteriorly, alignments based on aminoacid sequence homologies inferred from the corresponding genome sequences of *Staphylococcus aureus* (MRSA – Methicillin-resistant Staphylococcus aureus) strain USA 300 and *Staphylococcus succinus* strains 14BME20, DSM14617, DSM15096 and SNUC1280 were performed (Figure 29). The alignments were executed in the RAST server using the "sequence-based comparison" tool with the aim of a general and faster analysis to reveal possible similarities between *S. aureus* and *S. succinus*.

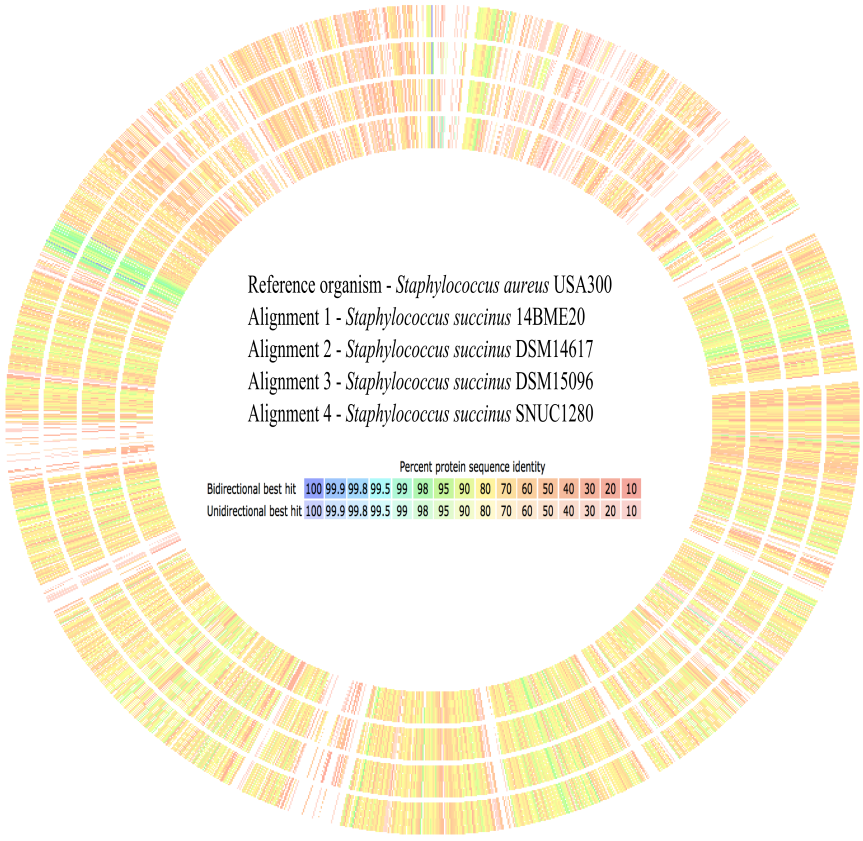



**Figure 29 | Sequence-based, genome-wide alignment between the multi-resistant *Staphylococcus aureus* (MRSA) strain USA 300 and *Staphylococcus succinus* strains 14BME20 (outermost ring), DSM14617, DSM15096 and SNUC1280 (innermost ring).** Alignments were performed on RAST based on aminoacid sequence homologies inferred from the corresponding genome sequences. Several genes are common to *S. aureus* and *S. succinus* strains, but usually they display only moderate levels of aminoacid sequence homology.

Observing the alignment, although there are several genes in common between *S. aureus* and *S. succinus*, they exhibit only moderate levels of aminoacid sequence homology.

### 3.4.2 Comparison of the Subsystem Feature Counts

Since the number of subsystems and subsystem coverage differs between *S. aureus* USA300 and all strains of *S. succinus* SNUC1280, 14BME20, DSM15096 and DSM14617, and in order to understand which are the differences, Table 32 displays the 27 main substystems provided by RAST.

**Table 32 – Comparison of the total number of genes in each of the subsystems of *S. aureus* USA300 and strains of *S. succinus*: SNUC1280, 14BME20, DSM15096 and DSM14617.** Values available on RAST. The values in bold are the ones representing the largest variations between *S. aureus* and *S. succinus* strains.

| Subsystem | USA300 | SNUC1280 | 14BME20 | DSM15096 | DSM14617 |
|---|---|---|---|---|---|
| **Cofactors, Vitamins, Prosthetic groups, Pigments** | 216 | 183 | 189 | 192 | 185 |
| Cell Wall and Capsule | 125 | 95 | 104 | 100 | 96 |
| **Virulence, Disease and Defense** | **97** | **79** | **56** | **62** | **95** |
| Potassium metabolism | 20 | 10 | 5 | 10 | 10 |
| Photosynthesis | 0 | 0 | 0 | 0 | 0 |
| Miscellaneous | 48 | 26 | 34 | 27 | 27 |
| **Phages, Profages, Transposable elements, plasmids** | 26 | 9 | 15 | 11 | 17 |
| **Membrane Transport** | **71** | **28** | **29** | **32** | **25** |
| **Iron acquisition and metabolism** | **53** | **27** | **28** | **26** | **26** |
| RNA metabolism | 148 | 115 | 126 | 114 | 119 |
| Nucleosides and Nucleotides | 107 | 85 | 89 | 89 | 86 |
| Protein Metabolism | 205 | 204 | 206 | 205 | 201 |
| **Cell Division and Cell Cycle** | 38 | 27 | 42 | 29 | 27 |
| **Motility and Chemotaxis** | **1** | **0** | **0** | **0** | **0** |
| Regulation and Cell signalling | 85 | 38 | 41 | 39 | 40 |
| Secondary Metabolism | 7 | 4 | 4 | 4 | 4 |
| DNA Metabolism | 102 | 78 | 70 | 88 | 74 |
| **Fatty acids, Lipids and Isoprenoids** | 99 | 118 | 110 | 106 | 108 |
| Nitrogen Metabolism | 20 | 9 | 10 | 21 | 9 |
| Dormancy and Sporulation | 9 | 9 | 9 | 9 | 9 |
| Respiration | 40 | 32 | 35 | 33 | 33 |
| Stress Response | 85 | 77 | 74 | 75 | 80 |
| Metabolism of Aromathic compounds | 4 | 11 | 11 | 10 | 10 |
| Amino Acids and Derivatives | 327 | 327 | 334 | 325 | 324 |
| Sulfur Metabolism | 15 | 27 | 27 | 27 | 28 |
| Phosphorous Metabolism | 22 | 32 | 31 | 35 | 33 |
| **Carbohydrates** | **272** | **343** | **337** | **350** | **351** |

*S. aureus* and *S. succinus* share several genes across all RAST subsystems except in subsystem Motility and Chemotaxis, although here the difference between the strains is negligible. Although they share the same subsystems as expected, there are some fluctuations between the number of genes. The number of genes varies more in the following subsystems: Virulence, Disease and Defense; Membrane Transport; Iron Acquisition and Metabolism and Carbohydrates.

Regarding the subsystem Virulence, Disease and Defense, there is a difference between *S. aureus* and *S. succinus* and even between *S. succinus* strains. *S. aureus* has 97 genes in this subsystem and *S. succinus* 14BME20 has 56 genes. The strain of *S. succinus* that has more genes is SNUC1280 (79). The other three subsystems reveal that there is a substancial difference between *S. aureus* and *S. succinus* strains, however within *S. succinus* strains there is no significant diference.

### 3.4.2.1 Comparison of the subsystem Virulence, Disease and Defense

Since, in the previous comparison, it was found that there is a substantial difference between *S. aureus* and *S. succinus* and even between *S. succinus* strains in the subsystem of Virulence, Disease and Defense, deeper research was made into this subsystem. The genes from this subsystem are divided in subgroups in Table 33.

**Table 33 | Comparison of genes from the subsystem Virulence, Disease and Defense of *S. aureus* (USA 300) and four *S. succinus* strains (SNUC1280, 14BME20, DSM15096 and DSM14617).** Values available at RAST server.

| Strain | | USA 300 | SNUC1280 | 14BME20 | DSM15096 | DSM14617 |
|---|---|---|---|---|---|---|
| Virulence, Disease and Defense | | **97** | **79** | **56** | **62** | **95** |
| Adhesion | Total | 23 | 1 | 1 | 1 | 1 |
| | **Adhesins in *Staphylococcus*** | **23** | **0** | **0** | **0** | **0** |
| | *Streptococcus* pyogenes recombinatorial zone | 0 | 1 | 1 | 1 | 1 |
| Toxins and superantigens | Total | 0 | 0 | 0 | 0 | 0 |
| Bacteriocins, ribosomally synthesized antibacterial peptides | Total | 6 | 10 | 10 | 10 | 10 |
| | Bacitracin Stress Response | 6 | 4 | 4 | 4 | 4 |
| | **Colicin V and Bacteriocin Production Cluster** | **0** | **6** | **6** | **6** | **6** |
| Resistence to antibiotics and toxic compounds | Total | 59 | 56 | 33 | 39 | 72 |
| | **Methicillin resistance in Staphylococci** | **28** | **22** | **0** | **0** | **23** |
| | Copper homeostasis | 2 | 2 | 2 | 3 | 6 |
| | Bile hydrolysis | 2 | 2 | 2 | 2 | 2 |
| | Cobalt-zinc-cadmium resistance | 6 | 3 | 4 | 4 | 6 |
| | Multidrug Resistance, 2-protein version Found in Gram-positive bacteria | 3 | 3 | 3 | 3 | 3 |
| | Mercuric reductase | 2 | 2 | 2 | 4 | 3 |
| | Mercury resistance operon | 1 | 1 | 1 | 2 | 2 |
| | Aminoglycoside adenylyltransferases | 0 | 0 | 1 | 0 | 0 |
| | Teicoplanin-resistance in Staphylococcus | 4 | 3 | 5 | 3 | 3 |
| | Resistance to fluoroquinolones | 4 | 4 | 4 | 4 | 4 |
| | Arsenic resistance | 3 | 5 | 5 | 9 | 11 |
| | Fosfomycin resistance | 1 | 0 | 2 | 1 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Copper homeostasis: copper tolerance | 0 | 2 | 0 | 2 | 2 |
| | Beta-lactamase | 1 | 1 | 1 | 1 | 1 |
| | Cadmium resistance | 0 | 2 | 0 | 0 | 5 |
| | Resistance to chromium compounds | 0 | 1 | 1 | 1 | 1 |
| | Multidrug Resistance Efflux Pumps | 2 | 3 | 0 | 0 | 0 |

The main difference between *S. aureus* and *S. succinus* strains is in the subgroup of adhesion. *S. aureus* has 23 genes referred to Adhesins in *Staphylococcus* and *S. succinus* strains do not have any gene. Secondly, whereas all *S. succinus* strains are potentially able to produce bacteriocins, this gene cluster was not found in multi-resistant *S. aureus* USA300. Altogether, these results suggest that, while *S. aureus* may be a more competent skin/host colonizer thorugh the action of adhesins, *S. succinus* may possess the ability to outcompete other bacteria in the skin microbiome through the biosynthesis of bacteriocins. Provided that *S. aureus* strains can cope with bacteriocin production by *S. succinus* (what is thus far not clear), both strains could engage in a synergistic interaction where *S. aureus* modifies the physical-chemistry of the skin microbiome favouring both strains to thrive (given their similarities, for instance, in nutrient acquisition) whereas *S. succinus* deters growth of potential competitors through the biosynthesis of inhibitory secondary metabolites. Such hypotheses must be addressed in the future through dedicated studies if we are to illuminate the positive relationships usually observed among *Staphylococcus* species during the emergence of AD.

Further, a subsystem in which several differences between *S. succinus* and *S. aureus* strains were found was “Resistance to antibiotics and toxic compounds”. There are genes of Metchicilin resistance in *Staphylococci* in strains SNUC1280 and DSM14617 but the other strains have no genes.

In summary, although there were many similarities between the genomes of both species, suggesting they are able to explore similar microniches, there were also conspicuous diferences between them regarding the subgroup of genes reffered to adhesion, bacteriocin production (only undertaken by *S. succinus* strains) and resistance to antibiotics and toxic compounds.

# Conclusion

The aim of this study was to address the relationship between *Staphylococcus aureus* abundance and shifts in the structure of the human skin microbiome, with implications to our understanding of Atopic Dermatitis (AD), using a metagenomic approach. To this end, three datasets collected from EBI-Metagenomics (MGnify) database were analysed to deliver do Phylum-level Abundance profiles, OTU-level Taxonomic profiles and Functional (IPR) profiles of the human skin microbiome across gradients of *S. aureus* abundance in the samples, a major indicator of the emergence of AD flares in skin. The datasets had samples collected from Antecubital (Ac), Popliteal (Pc) Creases and Volar forearm from patients with Atopic Dermatitis in several stages.

Phylum Taxonomic Abundance analysis demonstrated that the most abundant phyla were: *Actinobacteria*, *Firmicutes*, followed by *Proteobacteria* and *Bacteroidetes*, reflecting what is referred to in the literature (Kong, 2012). In order to understand the relationship between *S. aureus* and the microbial community, the response of phylum taxonomic abundances according to the percentage of *S. aureus* in the Ac, Pc and Vf samples were evaluated. The samples were divided in five groups according to the percentage of *S. aureus*: A (0-1%), B (1-2%), C (2-5%), D (5-10%) and E (≥10%). This analysis demonstrated that, although the major phyla were the same, there were significant dissimilarities in relative frequency between Ac, Pc and Vf samples and also, that increases in the relative abundance of *S. aureus* (*Firmicutes*) induces a decrease in all other phyla, and therefore, a reduction of microbial diversity.

Taxonomic analysis reveald that the microbiome of samples with the same relative abundance of *S. aureus* were more similar to one another than those with lower *S. aureus* abundances, revealing convergence in structure and reduced complexity in skin microbiomes with high *S. aureus* incidence, as expected. Through multivariate statistics using Principal Coordinates Analysis (PCoA) formation of clusters along with the percentage of *S. aureus* in the samples was examined. Since the ordination diagrams obtained from the PCoAs did not allow for a clear visual distinction between sample groups, a One-way PERMANOVA tests were necessary to precisely test for significant differences in skin microbiome structure among the studied groups. There were statistical significant dissimilarities between all areas, and within areas also except for group A and B in Ac, between group E and A, B and C in Pc and between group B and C in Vf. These results reinforce the fact that the bacterial community varies not only between sites but also according to the abundance of *S. aureus*. To identify the bacterial taxa that contribute to the major dissimilarities between the microbiomes according to site of origin or different abundances of *S. aureus*, SIMPER tests were performed at the OTU (Operational Taxonomic Unit) level. After identyfing the major taxa that contributed to the dissimilarity of the microbiome, it was sought to understand if there would be any correlation with *S. aureus*. It was found that there were positive correlations between *S. aureus* and *Staphylococcus* sp., *S. succinus* and *S. epidermidis*. *S. succinus* was a surprise since it demonstrated the better regression coefficient and it was never metioned before in skin disorders and with *S. aureus*. *Propionibacterium acnes*, *Corynebacterium* spp*., Streptococcus* spp., among others, demonstrated in almost all cases that in the presence of a higher abundance of *S. aureus* there is a decrease to low or null abundance levels of

these species. Some execptions, relatively to *Corynebacterium* spp.*,* were identified in samples from Ac and in Vf samples, *P. acnes* demonstrated a negative linear correlation with *S. aureus*.

For Functional analysis at the IPR level, the same multivariate analysis and tests mentioned above were performed. Scatter plots derived from PCoA were inconclusive, since the samples were very dispersed. Results from One-Way PERMANOVA test were more conclusive and revealed that there was no significantly difference in functional profiles across skin sites or *S. aureus* abundance groups, only between Ac and Vf and group A and B, considering all the samples. Within the Ac site, it was demonstrated that groups B and C were too similar; there were also no significant differences between groups B and D, E and A, B, C and D in Pc samples; and for Vf there were no significant between A and B, C and group B with C and D and finally with group E, C and D. To rank all IPRs that contributed the most for the total variation of the dataset, SIMPER analyses were performed revealing four IPRs that were common in all analyses undertaken for differences among *S. aureus* abundance groups: Transposase, L1; Transposase, mutator type; TonB-dependet receptor, beta-barrel; and EAL domain. All of these IPRs are related to pathogenesis factors that may contribute to AD or responses to pathogenesis at different degrees.

Lastly, after the comparison of *S. succinus* and *S. aureus* genomes it is possible to conclude that although there are several genes in common between these species, these genes exhibit only moderate levels of aminoacid sequence homology. There are significant diferences between the number of genes classified in the subsystem Virulence, Disease and Defense, more precisely genes related to adhesion, bacteriocin production and resistence to antibiotics and toxic compounds.

In conclusion, although it is already being recognized the importance of the study of the relationship between *S. aureus* and other microorganisms, further research is needed particularly to elucidate its relationship with other *Staphylococci* during AD emergence, and whether or not these species act as disease causing agents or their increased frequency is a consequence of a physico-chemically altered skin microniche under AD. The field of metagenomics is improving, and has been demonstrating a huge potential in addressing diseases from the taxonomic and functional standpoints.

# Bibliography


Alivisatos, A. P., Blaser, M. J., Brodie, E. L., Chun, M., Dangl, J. L., Donohue, T. J., … Taha, S. A. (2015). A unified initiative to harness Earth's microbiomes: Transition from description to causality and engineering. *Science*, *350*(6260), 507–508. https://doi.org/10.1126/science.aac8480

Anderson, M. J. (2005). PERMANOVA Permutational multivariate analysis of variance. *Austral Ecology*, 1–24. https://doi.org/10.1139/cjfas-58-3-626

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., … Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, *9*, 1–15. https://doi.org/10.1186/1471-2164-9-75

Bahar, A. A., & Ren, D. (2013). Antimicrobial peptides. *Pharmaceuticals*, *6*(12), 1543–1575. https://doi.org/10.3390/ph6121543

Baranska-Rybak, W., Cirioni, O., Dawgul, M., Sokolowska-Wojdylo, M., Naumiuk, L., Szczerkowska-Dobosz, A., … Kamysz, W. (2011). Activity of Antimicrobial Peptides and Conventional Antibiotics against Superantigen Positive Staphylococcus aureus Isolated from the Patients with Neoplastic and Inflammatory Erythrodermia. *Chemotherapy Research and Practice*, *2011*(February), 1–6. https://doi.org/10.1155/2011/270932

Byrd, A. L., Deming, C., Cassidy, S. K. B., Harrison, O. J., Ng, W.-I., Conlan, S., … Kong, H. H. (2017). *Staphylococcus aureus* and *Staphylococcus epidermidis* strain diversity underlying pediatric atopic dermatitis. *Science Translational Medicine*, *9*(397), eaal4651. https://doi.org/10.1126/scitranslmed.aal4651

Cani, P. D. (2018). Human gut microbiome: Hopes, threats and promises. *Gut*, 1–10. https://doi.org/10.1136/gutjnl-2018-316723

Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., … Knight, R. (2011). Moving pictures of the human microbiome. *Genome Biology*, *12*(5), R50. https://doi.org/10.1186/gb-2011-12-5-r50

Chng, K. R., Tay, A. S. L., Li, C., Ng, A. H. Q., Wang, J., Suri, B. K., … Nagarajan, N. (2016). Whole metagenome profiling reveals skin microbiome-dependent susceptibility to atopic dermatitis flare. *Nature Microbiology*, *1*(9). https://doi.org/10.1038/nmicrobiol.2016.106

Cho, I., & Blaser, M. J. (2012). Applications of Next-Generation Sequencing: The human microbiome: at the interface of health and disease. *Nature Publishing Group*, *13*(4), 260–270. https://doi.org/10.1038/nrg3182

Consortium, T. H. M. P. (2013). Structure, Function and Diversity of the Healthy Human Microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234.Structure

Cooper, A. J., Weyrich, L. S., Dixit, S., & Farrer, A. G. (2015). The skin microbiome: Associations between altered microbial communities and disease. *Australasian Journal of Dermatology*, *56*(4),

268–274. https://doi.org/10.1111/ajd.12253

Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M., & Relman, D. A. (2012). The application of ecological theory towards an understanding of the human microbiome. *Science*, *336*(6086), 1255–1262. https://doi.org/10.1126/science.1224203.The

Fazekas, I., & Liese, F. (1996). Some properties of the Hellinger transform and its application in classification problems. *Computers & Mathematics with Applications*, *31*(8), 107–116. https://doi.org/10.1016/0898-1221(96)00035-1

Fierer, N., Hamady, M., Lauber, C. L., & Knight, R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences*, *105*(46), 17994–17999. https://doi.org/10.1073/pnas.0807920105

Gail, M., Krickeberg, K., Samet, J., Tsiatis, A., & Wong, W. (2007). *Analysing Ecological Data*.

Gallo, R. L. (2016). S. epidermis Influence on Host Immunity: More Than Skin, *17*(2), 143–144. https://doi.org/10.1016/j.chom.2015.01.012.S.

Gevers, D., Pop, M., Schloss, P. D., & Huttenhower, C. (2012). Bioinformatics for the Human Microbiome Project. *PLoS Computational Biology*, *8*(11). https://doi.org/10.1371/journal.pcbi.1002779

Glass, E. M., & Meyer, F. (2011). The Metagenomics RAST Server: A Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, *8*, 325–331. https://doi.org/10.1002/9781118010518.ch37

Gonzalez, A., Clemente, J. C., Shade, A., Metcalf, J. L., Song, S., Prithiviraj, B., … Knight, R. (2011). Our microbial selves: What ecology can teach us. *EMBO Reports*, *12*(8), 755–784. https://doi.org/10.1038/embor.2011.137

Grice, E. A. (2014). The skin microbiome: potential for novel diagnostic and therapeutic approaches to cutaneous disease. *Seminars in Cutaneous Medicine and Surgery*, *33*(2), 98–103. https://doi.org/10.12788/j.sder.0087

Grice, E. A., Kong, H. H., Conlan, S., Deming, C. B., Davis, J., Young, A. C., … Segre, J. A. (2009). Topographical and Temporal Diversity of the Human Skin Microbiome. *Science*, *324*(5931), 1190–1192. https://doi.org/10.1126/science.1171700

Grice, E. A., & Segre, J. A. (2011). The skin microbiome. *Nature Reviews Microbiology*, *9*(4), 244–253. https://doi.org/10.1038/nrmicro2537

Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S., … Schulze-Lefert, P. (2015). Microbiota and host nutrition across plant and animal kingdoms. *Cell Host and Microbe*, *17*(5), 603–616. https://doi.org/10.1016/j.chom.2015.04.009

Hamady, M. (2009). Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Research*, (303), 1141–1152. https://doi.org/10.1101/gr.085464.108.19

Hammer, Ø., Harper, D. A. T. a. T., & Ryan, P. D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, *4(1)*(1), 1–9. https://doi.org/10.1016/j.bcp.2008.05.025

Hammer, Ø., Harper, D., & Ryan, P. (2001). PAST-PAlaeontological STatistics, ver. 1.89. *University of Oslo, Oslo*, (1999), 1–31. Retrieved from http://www.researchgate.net/publication/228393561_PASTPalaeontological_statistics_ver._1.89/file/32bfe5135d45cd6b3b.pdf

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, *5*(10). https://doi.org/10.1016/S1074-5521(98)90108-9

Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., … Sansone, S. A. (2014). *EBI metagenomics - A new resource for the analysis and archiving of metagenomic data*. *Nucleic Acids Research* (Vol. 42). https://doi.org/10.1093/nar/gkt961

K Ursell, Luke; L Metcalf, Jessica; Wegener Parfrey, Laura and Knight, R. (2013). Definig the Human Microbiome. *NIH Manuscripts*, *70*(Suppl 1), 1–12. https://doi.org/10.1111/j.1753-4887.2012.00493.x.Defining

Kenney, J. F. (1963). *Mathematics of Statistics*.

Kim, M.-H., Rho, M., Choi, J.-P., Choi, H.-I., Park, H.-K., Song, W.-J., … Pyun, B. Y. (2017). A Metagenomic Analysis Provides a Culture-Independent Pathogen Detection for Atopic Dermatitis. *Allergy, Asthma & Immunology Research*, *9*(5), 453. https://doi.org/10.4168/aair.2017.9.5.453

Kong, H. H. (2012). Temporal shifts in the skin microbiome associated with disease flare and treatment in children with atopic dermatitis. *Genome Research*, 850–859. https://doi.org/10.1101/gr.131029.111

Kong, H. H., Andersson, B., Clavel, T., Common, J. E., Jackson, S. A., Olson, N. D., … Traidl-Hoffmann, C. (2017). Performing Skin Microbiome Research: A Method to the Madness. *Journal of Investigative Dermatology*, *137*(3), 561–568. https://doi.org/10.1016/j.jid.2016.10.033

Kong, H. H., & Segre, J. a. (2015). *Biogeography and individuality shape function in the human skin metagenome* (Vol. 514). https://doi.org/10.1038/nature13786.Biogeography

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., … Ley, R. E. (2013). A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. *PLoS Computational Biology*, *9*(1). https://doi.org/10.1371/journal.pcbi.1002863

Krasteva, P. V., Giglio, K. M., & Sondermann, H. (2012). Sensing the messenger: The diverse ways that bacteria signal through c-di-GMP. *Protein Science*, *21*(7), 929–948. https://doi.org/10.1002/pro.2093

Kuczynski, J., Costello, E. K., Nemergut, D. R., Zaneveld, J., Lauber, C. L., Knights, D., … Knight, R.

(2010). Direct sequencing of the human microbiome readily reveals community differences. *Genome Biology*, *11*(5). https://doi.org/10.1186/gb-2010-11-5-210

Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., & Knight, R. (2011). Experimental and analytical tools for studying the human microbiome. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3129

Kumar, B., Pathak, R., Mary, P. B., Jha, D., Sardana, K., & Gautam, H. K. (2016). New insights into acne pathogenesis: Exploring the role of acne-associated microbial populations. *Dermatologica Sinica*, *34*(2), 67–73. https://doi.org/10.1016/j.dsi.2015.12.004

Kwaszewska, A., Sobiś-Glinkowska, M., & Szewczyk, E. M. (2014). Cohabitation—relationships of corynebacteria and staphylococci on human skin. *Folia Microbiologica*, *59*(6), 495–502. https://doi.org/10.1007/s12223-014-0326-2

Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, *129*(2), 271–280. https://doi.org/10.1007/s004420100716

Legendre, P., & Legendre, L. (2012). Transformations for community composition data. *Numerical Ecology*, 327–333. https://doi.org/10.1111/j.1461-0248.2010.01552.x

Levy, R. M., Huang, E. Y., Roling, D., Leyden, J. J., & Margolis, D. J. (2003). Effect of antibiotics on the oropharyngeal flora in patients with acne. *Archives of Dermatology*, *139*(4), 467–471. https://doi.org/10.1001/archderm.139.4.467

Li, K., Bihan, M., & Methé, B. A. (2013). Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0063139

Lindgreen, S., Adair, K. L., & Gardner, P. P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports*, *6*, 1–14. https://doi.org/10.1038/srep19233

Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, *3*(1), 31. https://doi.org/10.1186/s40168-015-0094-5

Martiny, J. B. H., Jones, S. E., Lennon, J. T., & Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science*, *350*(6261). https://doi.org/10.1126/science.aac9323

Mercedes E. Gonzalez, et al. (2016). Cutaneous microbiome effects of fluticasone proprionate cream and adjunctive bleach baths in childhood atopic dermatitis.pdf. *J Am Acad Dermatol*.

Methé, B. a, Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., … Ding, Y. (2012). A framework for Human microbiome research. *Nature*, *486*(7402), 215–221. https://doi.org/10.1038/nature11209.A

Mitchell, A., Bucchini, F., Cochrane, G., Denise, H., Ten Hoopen, P., Fraser, M., … Finn, R. D. (2016). EBI metagenomics in 2016 - An expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Research*, *44*(D1), D595–D603. https://doi.org/10.1093/nar/gkv1195

Moran, J. V. (1999). Human L1 retrotransposition: insights and peculiarities learned from a cultured

cell retrotransposition assay. *Genetica*, *107*(1–3), 39–51. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10952196

Mosbahi, K., Wojnowska, M., Albalat, A., & Walker, D. (2018). Bacterial iron acquisition mediated by outer membrane translocation and cleavage of a host protein, 2–7. https://doi.org/10.1073/pnas.1800672115

Mulcahy-O'Grady, H., & Workentine, M. L. (2016). The challenge and potential of metagenomics in the clinic. *Frontiers in Immunology*, *7*(FEB), 1–8. https://doi.org/10.3389/fimmu.2016.00029

Nair, N., Biswas, R., Götz, F., & Biswas, L. (2014). Impact of Staphylococcus aureus on pathogenesis in polymicrobial infections. *Infection and Immunity*, *82*(6), 2162–2169. https://doi.org/10.1128/IAI.00059-14

Nakatsuji, T., Chen, T. H., Narala, S., Chun, K. A., Two, A. M., Yun, T., … Gallo, R. L. (2017). Antimicrobials from human skin commensal bacteria protect against Staphylococcus aureus and are deficient in atopic dermatitis. *Science Translational Medicine*. https://doi.org/10.1126/scitranslmed.aah4680

Ostertag, Eric M., Kazazian Jr, H. H. (2015). Retrotransposons l1 in human disorders.pdf.

Otto, M. (2009). Staphylococcus epidermidis—the'accidental'pathogen. *Nature Reviews Microbiology*, *7*(8), 555–567. https://doi.org/10.1038/nrmicro2182.Staphylococcus

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., … Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1), 206–214. https://doi.org/10.1093/nar/gkt1226

Pesavento, C., & Hengge, R. (2009). Bacterial nucleotide-based second messengers. *Current Opinion in Microbiology*, *12*(2), 170–176. https://doi.org/10.1016/j.mib.2009.01.007

Ramsey, M. M., Freire, M. O., Gabrilska, R. A., Rumbaugh, K. P., & Lemon, K. P. (2016). Staphylococcus aureus Shifts toward commensalism in response to corynebacterium species. *Frontiers in Microbiology*, *7*(AUG), 1–15. https://doi.org/10.3389/fmicb.2016.01230

Rivera-Amill, V. (2016). The Human Microbiome and the Immune System: An Ever Evolving Understanding, *4*(11), 10–14. https://doi.org/10.1016/S2214-109X(16)30265-0.Cost-effectiveness

Römling, U., & Amikam, D. (2006). Cyclic di-GMP as a second messenger. *Current Opinion in Microbiology*, *9*(2), 218–228. https://doi.org/10.1016/j.mib.2006.02.010

Schreiber, F., Szekat, C., Josten, M., Sahl, H. G., & Bierbaum, G. (2013). Antibiotic-induced autoactivation of IS256 in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *57*(12), 6381–6384. https://doi.org/10.1128/AAC.01585-13

Simes, R. J. (1986). Biometrika Trust An Improved Bonferroni Procedure for Multiple Tests of Significance Author(s An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, *73*(3), 751–754. Retrieved from

http://www.jstor.org/stable/2336545%0Ahttp://www.jstor.org/%0Ahttp://www.jstor.org/action/showPublisher?publisherCode=bio.%0Ahttp://www.jstor.org

Somerfield, P. J. (2008). Identification of the Bray-Curtis similarity index: Comment on Yoshioka (2008). *Marine Ecology Progress Series*, *372*, 303–306. https://doi.org/10.3354/meps07841

Sondermann, H., Shikuma, N. J., & Yildiz, F. H. (2012). You've come a long way: C-di-GMP signaling. *Current Opinion in Microbiology*, *15*(2), 140–146. https://doi.org/10.1016/j.mib.2011.12.008

Teixeira, L. C. R. S., Peixoto, R. S., & Rosado, A. S. (2010). Bacterial Diversity in Rhizosphere Soil from Antarctic Vascular Plants of Admiralty Bay in Maritime Antarctica. *Molecular Microbial Ecology of the Rhizosphere*, *2*(8), 1105–1112. https://doi.org/10.1002/9781118297674.ch105

Thomas et al., 2017. (2017). The microbiome and atopic eczema: More than skin deep. *Australasian Journal of Dermatology*, *58*(1), 18–24. https://doi.org/10.1111/ajd.12435

Waldor, M. K., Tyson, G., Borenstein, E., Ochman, H., Moeller, A., Finlay, B. B., … Smith, H. (2015). Where Next for Microbiome Research? *PLoS Biology*, *13*(1), 1–9. https://doi.org/10.1371/journal.pbio.1002050

Wang, R., Xu, H., Du, L., Chou, S. H., Liu, H., Liu, Y., … Qian, G. (2016). A TonB-dependent receptor regulates antifungal HSAF biosynthesis in Lysobacter. *Scientific Reports*, *6*(March), 1–10. https://doi.org/10.1038/srep26881

Warton, D. I., Wright, S. T., & Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, *3*(1), 89–101. https://doi.org/10.1111/j.2041-210X.2011.00127.x

Zhu, T., Zhao, Y., Wu, Y., & Qu, D. (2017). The Staphylococcus epidermidis gdpS regulates biofilm formation independently of its protein-coding function. *Microbial Pathogenesis*, *105*, 264–271. https://doi.org/10.1016/j.micpath.2017.02.045