UNIVERSITÄT ZU LÜBECK

From the Department of Dermatology, Allergology, and Venereology of the

University of Lübeck

Director:

Prof. Dr. med. Detlef Zillikens

Genome-wide mapping of gene-microbiota interactions in susceptibility to epidermolysis bullosa acquisita

Dissertation for Fulfillment

of Requirements for the

Doctoral Degree

of the University of Lübeck

from the Department of Natural Sciences

Submitted by

Girish Srinivas

From Dharmapuri, INDIA

Lübeck 2013

First referee: Prof. Dr. S. Ibrahim

Second referee: Prof. Dr. W. Solbach

Third referee: Prof. Dr. A. Franke

Date of oral examination: 29.10.2013

Approved for printing on 31.10.2013

Dedication

To my beloved parents for their love and support

Acknowledgements

Declaration of Author's Contribution

I present my doctoral research results in this thesis. The study design was conducted by my supervisors Prof. Dr. Saleh Ibrahim, Prof. Dr. John Baines and myself. The interpretation of the results was achieved during numerous discussions. The major part of the practical laboratory and bioinformatics data analysis of this thesis was conducted by me, with some exceptions:

Dr. Steffen Möller guided me through the QTL mapping procedure using HAPPY software. Jun Wang assisted me in linear mixed model analysis. Dr. Sven Künzel performed 454 sequencing. Katja Cloppenborg-Schmidt and Silke Carstensen gave technical support in the lab to perform DNA extraction and PCR of the bacterial 16S rRNA gene from mouse ear samples. Dr. Christoph Ellebrecht maintained the SJL/J mouse strain and collected ear samples from them. Susen Müller and Andreia Marques extensively supported in maintaining advanced intercross lines and also assisted in mouse ear sample collections. Derk Wachsmuth and Werner Wegner gave IT assistance to use the Max Planck Institutes computing clusters. Prof. Dr. Saleh Ibrahim, Prof. Dr. John Baines, Prof. Dr. Detlef Zillikens, and Prof. Dr. Ralf Ludwig gave their valuable suggestions for the discussion section of this study.

Table of Contents

List of Figures

List of Tables

Summary

The skin is in constant contact with the environment and serves a critical barrier function, yet provides a range of niches to inhabiting microbial communities. A multitude of interactions between the skin microbiota, host and environment contribute to community structure and its potential contribution to changes in health status is well known. Susceptibility to chronic inflammatory diseases is determined by the interaction of immunogenetic and environmental risk factors. In particular, resident microbial communities as environmental factors are the subject of intense scrutiny due to numerous observations of differences in community composition or structure are of primary etiological importance or secondary to the altered inflammatory environment remains largely unknown.

Epidermolysis bullosa acquisita (EBA) is a chronic skin blistering disease of autoimmune origin characterized by antibodies to type VII collagen (COL7). This study provides experimental evidence for host gene-microbiota interactions contributing to disease risk in a mouse immunization model of EBA. By using an advanced intercross mouse population, genetic loci contributing to variability in the skin microbiota were simultaneously identified along with susceptibility to EBA and their overlap. QTL mapping of the skin microbiota with susceptibility to EBA demonstrates the involvement of host gene-microbe interactions in disease. Furthermore, treating the abundances of individual bacterial species as covariates with disease lead to the discovery of a novel disease locus. The majority of the identified covariate taxa were characterized by a reduction in abundance being associated with increased disease risk. This provides evidence of a primary role for individual bacterial species abundances in disease susceptibility and underscores their importance in protection from disease. Interestingly, in a parallel study in this thesis, mice that did not develop clinical disease showed a higher diversity in their skin microbial communities before disease induction. This further demonstrates the importance of skin community in predictive of EBA

disease outcome. Thus, further characterization of these putative probiotic species or species assemblages offers promising potential for preventative and therapeutic treatment development.

1. Introduction

Animals throughout their evolutionary process have been constantly in direct contact with diverse microbial communities that help them thrive in different environmental conditions on earth. This microbial community is made up of bacteria, viruses, fungi, and archaea, which directly or indirectly interact with the host. It is now known that animals host thousands of different bacterial species and are comprised of up to 90% of bacterial cells [1]. It also estimated that there are 10 viral particles for every bacterial cell [2]. The presence of these microbes has a great impact on the host in many ways. These microbes provide a vast reservoir of metabolic capabilities that can complement host metabolism well-being [3], [4] as well as developmental and nutritional processes [5]. However, the roles of other microbial communities like viruses, fungi, and archaea that inhabit animals are less studied.

Bacteria colonizes throughout the portions of the human body that are exposed to the environment including the skin [6], gastrointestinal (GI) tracts [7], and lungs [8]. The human colon alone is colonized by 70% of all the microbes in the human body [9], [10]. This is one of the main reasons why most of the microbial studies focus on the gut. The number of bacterial species colonizing the human gut varies widely between different studies, but at the phyla level it is mainly dominated by 2 major phyla: *Firmicutes* and *Bacteroidetes*. There are other phyla such as *Actinobacteria*, *Proteobacteria*, *Cyanobacteria*, and *Fusobacteria* present in less abundance [11]. The studies have shown that the microbiota in the GI tract is host specific and also region specific [12], [13], [14]. There has been a clear influence of diet on gut microbiota showing co-evolution between the diet, human and gut microbiota composition [15]. Studies using litter cross-fostering, embryo transplantation and mouse co-housing experiments and other stochastic and environmental factors show that they also contribute to microbiota composition [16]. One analysis of gut microbiota across human population showed three distinct clusters among microbiomes called enterotypes

[17]. Later it has been shown that two of three enterotypes could be driven by long term diet and it did not appear to correlate to host demography or health status [18], [19].

The interactions of host and their gut microbiota are of fundamental importance for host health and disease. Recent studies have shown that intestinal microbiota in humans can be correlated to inflammatory bowel disease [20], colon cancer [21], allergies [22], diabetes [23], neonatal necrotizing enterocolitis [24], metabolic syndrome [25] and obesity [26]. This awareness of health associated intestinal microbial communities underscores the importance of understanding the molecular basis and dynamics of host-microbe homoeostasis. Therefore it is necessary to have animal models that complement to human disease phenotype and allow us for targeted analysis of microbial, pathological and immunological aspects under controlled conditions. At the same time, genetically inbred mouse lines in a controlled environment help us understand the roles of host-genetic association with specific gut microbiota composition [27]. Mutation, up and down regulation or inactivation of specific genes have been significantly associated with specific bacterial community changes and it is also linked to metabolic diseases like obesity, diabetes and metabolic syndrome [28]. But whether these associations may be mediated by alterations in microbial community structure is unknown [29]. Most of the microbial research in past years is mostly focused on gastrointestinal tract. In recent years, skin-microbiota related research has also gained attention, with advances made towards identifying specific or microbial communities molecules involved in host skin physiology.

1.1 Skin microbiota

Skin being the largest organ of the body serves as a critical barrier between the host and the environment. It also provides a range of specialized niches that contain microbial communities comprising bacteria, viruses, mites and fungi, as a whole referred to as the skin microbiome (**Figure 1.1**). The mammalian skin flora is mainly dominated by diverse bacterial communities and

mostly found in skin structures such as hair follicles, eccrine and sebaceous glands (**Figure 1.1**). The composition of skin bacterial species is found to be mainly driven by local environmental factors e.g. local chemistry, nutrient availability, exposure to external environment, dry and wet areas [30]. Mammalian hosts and their microbiota have developed an evolutionary relationship that is vital for co-existence including host defense against pathogens, metabolism and development of host immune system. Therefore, it is very important to characterize these microbes in an appropriate way to fully understand their role in human disease and health status.

1.1.1 Skin microbiota characterization

In early studies using culture-based approach it has been shown that there is an association between a number of skin infections and microbes [31], [32]. Further procedures like gram staining [33] and other biochemical methods emerged in past decades (such as Oxidase, Catalase, Oxidation/Fermentation, Coagulase, Acid and Gas from Lactose tests etc.) were used to identify and characterize bacterial organisms. These traditional methods of identifying bacteria suffer from two main drawbacks. Firstly, most are not cultivable. Secondly, biochemical tests which look for common patterns cannot be used to identify all bacterial strains. There are some strains, which exhibit unique biochemical characteristics that cannot be found using traditional methods.

A revolutionary article in 1977 PNAS by two biologists Carl Woese and George Fox, phylogenetic analysis based on the 16S ribosomal RNA sequence revealed three distinct kingdoms: Eubacteria (Bacteria), Archaebacteria (Archaea) and Urkaryotes (Eucarya). 16S ribosomal RNA is a component of the 30S subunit of prokaryotic ribosomes and its use has revolutionized our understanding of microbial diversity [34], [35], [36]. The regions of 16S appear largely independent of ecological diversification and there are regions which evolve faster (variable regions) and slower (conserved regions), and can thus be used to find the relationships among different bacterial taxa at different phylogenetic depths. Based on this, several culture independent techniques have emerged

in the past 20 years. Microarrays and Real time PCR are widely used molecular techniques for identification and quantification of bacteria even at the species level. Real time PCR in particular is useful in quick detection of low amounts of bacterial DNA.
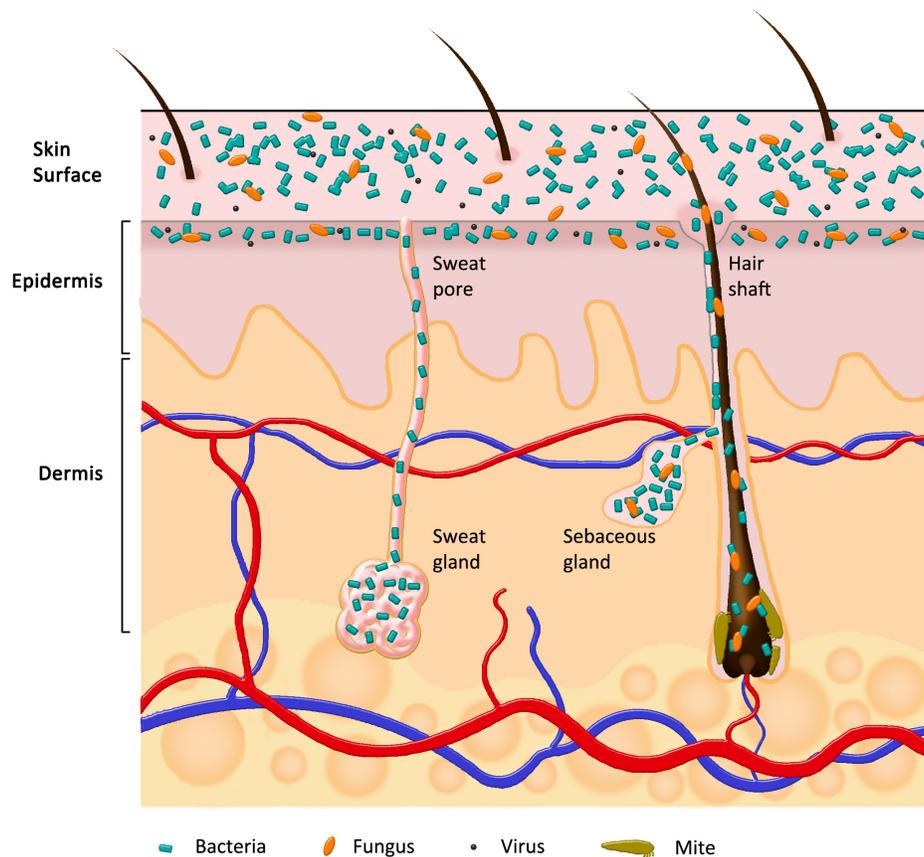
Figure 1.1: Cross section view of skin histology.
Microbes such as bacteria, fungi and viruses are found in the skin surface and also reside in the hair shaft, sweat pore, sweat and sebaceous glands. Staphylococcus spp. dominates the skin bacterial species population. Viruses are found to be living freely and also residing along with the bacterial cells. The figure is modified from Grice and Segre (2011) [37].

The recent introduction of low cost Next Generation Sequencing (NGS) technique has revolutionized the way in which diverse bacterial communities are characterized. Researchers use the 16S rRNA gene (also known as 16s rDNA) sequences to study bacterial phylogeny and taxonomy with greater accuracy than before. There are databases, including SILVA [38],
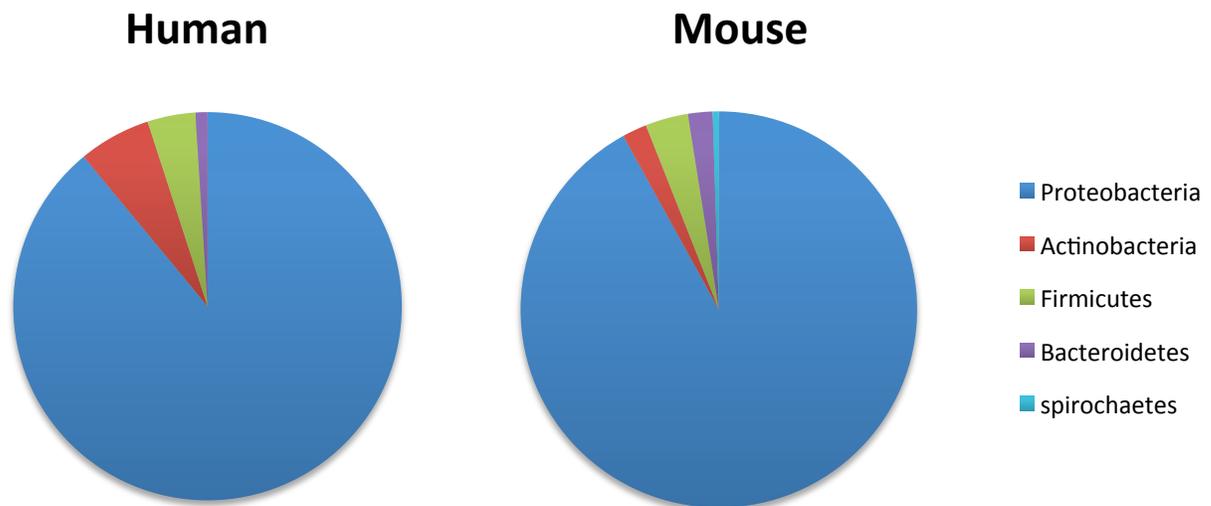
Figure 1.2: Comparison of relative abundance of bacteria phyla in mouse and human skin.
16s rDNA sequences are shown at phyla level. Figure was generated on publicly available data published by Grice *et al*. (2008) **[6]**.

Greengenes [39] and the Ribosomal Database Project [40] databases dedicated to 16S rRNA gene, that help classify a greater proportion of environmental sequences. Using this technique, researchers have found changes in microbial composition with skin conditions such as atopic dermatitis [41], psoriatic lesions [42] and acne [43] using three sampling methods: scrape, swap, and punch biopsy. These three sampling methods were first used in a single study to compare many layers and structures of human and mouse skin [6]. They also surveyed the bacterial composition in human antecubital fossa skin and mouse ear skin and identified striking parallels between them (**Figure 1.2**). They also reported that *Proteobacteria* phylum containing the genera

Pseudomonas and *Janthinobacterium* in mouse ear skin is identical to human antecubital fossae. These reports are very encouraging to take advantage of mouse disease models that could be useful to explore the roles of skin microbiota and its influence on disease.

1.1.2 Factors influencing skin microbiota composition

The microbial colonization of humans and other mammals begins right after birth, and the mode of delivery has a huge influence in the initial colonization. It has been reported that babies born by C-section have microbiota similar to their mother's skin while vaginally delivered babies have bacterial composition similar to vaginal microbiota [44]. It is noted that the maternal environment has a huge influence on microbiota composition. Several studies using mouse models have shown that genetically identical litters in the same cage have more similar intestinal microbiota than litters growing up in different cages [23], [45], [46].

Skin microbiota composition is also determined by demographic characteristics on skin conditions such as host's age, gender, hormones, and ethnicity [47], [48], [49], [50]. Local environmental characteristics like temperature, pH, moisture and exposure to ultraviolet radiation can influence the microbial composition in skin. These differences can be seen at different areas within the host which can lead to different skin microbial composition [30]. In skin, epithelial surfaces contain T lymphocytes, keratinocytes, mast cell and Langerhans cells, which express Toll like receptors and produce antimicrobial products such as cytokines, chemokines, β-definsins etc. These antimicrobial products are used by the innate immune system to regulate the skin microbiota. Thus, genetic variation within the host genome that can influence or alter the functions of immune system may also result in altering the microbial composition. To measure the variation caused by host genetics on microbial diversity, it is important to understand other factors influencing the variation. In order to effectively measure the host genotype effect, one has to control environmental factors to avoid noise. Benson and colleagues [27] showed that host genotype indeed controls significant variation

of gut microbiota. It is yet to be proven whether such alterations could lead to any changes in health status of the host.

1.1.3 Skin microbiota association with health and disease

Many common skin diseases such as atopic dermatitis [41], psoriatic lesions [42] and acne [43] are known to have microbial involvement [49]. This can be explained in two mutually exclusive events: a microbial infection that is promoting the disease, or microbes protects or prevents the host from getting the disease, and both scenarios are possible [51]. In atopic dermatitis (AD) patients are found densely colonized by *Staphylococcus species* (mainly *Staphylococcus aureus)* on their skin compared to healthy people [52]. It has also been found that AD patients have reduced amount of an antimicrobial peptide called dermicidin [53], which is found in skin as a part of epithelial innate defense system [54]. Dermicidin induces killing mechanism against *Staphylococcus aureus* [55] and prevent the infection by limiting the growth of *S. aureus* and/or its colonization and indirectly promoting other commensals. Dekio and colleagues speculates that the presence of *Dietzia Maris* bacterial species in healthy skin compared to that of AD patients points towards a very important role in maintaining healthy skin [41].

Microbial products from skin bacterial commensals are known to have immunoregulatory effects [56]. Accordingly, lipoteichoic acid (LTA) produced by Staphylococcal species has a unique anti-inflammatory effect on keratinocytes and has the opposite response when reacting with other immune cells [56]. These conclusions show the potential role of certain species of Staphylococcus inhabiting the skin. Another study clearly underscores the importance of skin commensals to tune the functions of local T cells for optimal skin immune fitness [57]. A recent landmark study of the mouse gut microbiota using a quantitative trait locus (QTL) approach unequivocally demonstrates the role of host genetics in shaping diversity between individuals [27]. It is now interesting to study the relationship between different bacterial strains and their interactions with the host genetics to

understand the role of skin commensals in maintaining healthy tissue function. This is still largely unexplored area of research and the results hold the key to develop targeting therapeutic disease specific treatments for (autoimmune) skin diseases.

1.2 Autoimmunity and Autoimmune diseases

One of the main functions of our immune system is to mount inflammatory responses to non-self while avoiding any harm to self-tissues. This non-self could be bacteria, viruses, parasites or any other foreign molecules. In the case of autoimmune diseases, the host immune system turns on against its own cells, tissues and organs. It can no longer differentiate between self and non-self; thus leading to autoimmune diseases (AD). Autoimmunity, on the other hand, refers merely to the presence or production of autoantibodies against self antigens, which does not lead to any tissue damage or inflammation *per se*, however, it can lead to false positive autoantibody tests in many cases [58]. For example T-lymphocytes reacting with self-antigens are found in all individuals [59].

1.2.1 Why study Autoimmune diseases?

Currently there are more than 80 known autoimmune diseases, affecting more than 100 million people worldwide [60], which has a great global economic and social impact globally. A recent American Autoimmune Related Disease Association (AARDA, source from www.aarda.org) report reveals that there are currently 50 million (~20% of the total population) Americans living with AD and it is disproportionately affecting women. Approximately every 3 in 4 affected individuals is a woman. AD is responsible for more than $100 billion in direct health care costs annually in America. A detailed research in this area not only potentially helps those directly affected by it but also helps us figure out how to modulate the immune system so that we can help cancer patients [61], it has the potential to revolutionize organ transplant [62], it improve treatment of AIDS patients as well as other infectious diseases [63].

1.2.2 Autoimmune skin blistering diseases

Autoimmune skin blistering diseases (ASBD) are a heterogeneous group of disorders characterized by autoantibodies that are mistakenly directed against self-specific structural proteins causing blistering lesions in skin. ASBD have different clinical manifestations not only ranging from blisters and erosions on the skin to surface mucous membranes but also share underlying common immunological mechanisms. ASBD can be divided into pemphigus disease and subepidermal bullous disease [64]. In the case of pemphigus disease, the autoantibodies are directed against intercellular structures called desmosomes, whereas in the case of subepidermal bullous disease, the autoantibodies are targeted against adhesion molecules of basal membrane called hemidesmosomes [64]. The pathophysiology of different skin blistering diseases (belonging to subepidermal bullous disease type) characterized by autoantibodies that target different subregions of epidermal basement membrane proteins is shown in detail in **Figure 1.3**.

In Germany, roughly 2000 new cases of ASBD affected patients are reported every year, with an overall prevalence of about 12000 cases at present. The incidence of ASBD in Germany has doubled over the past 12 years and the recent records show 25 new cases per million every year [64]. The treatment of ASBD is still challenging because of a very few prospective therapeutic trials and a lack of standard treatment guidelines available in Germany. Thus, it is a timely endeavor to now to focus more in this area of clinical research.
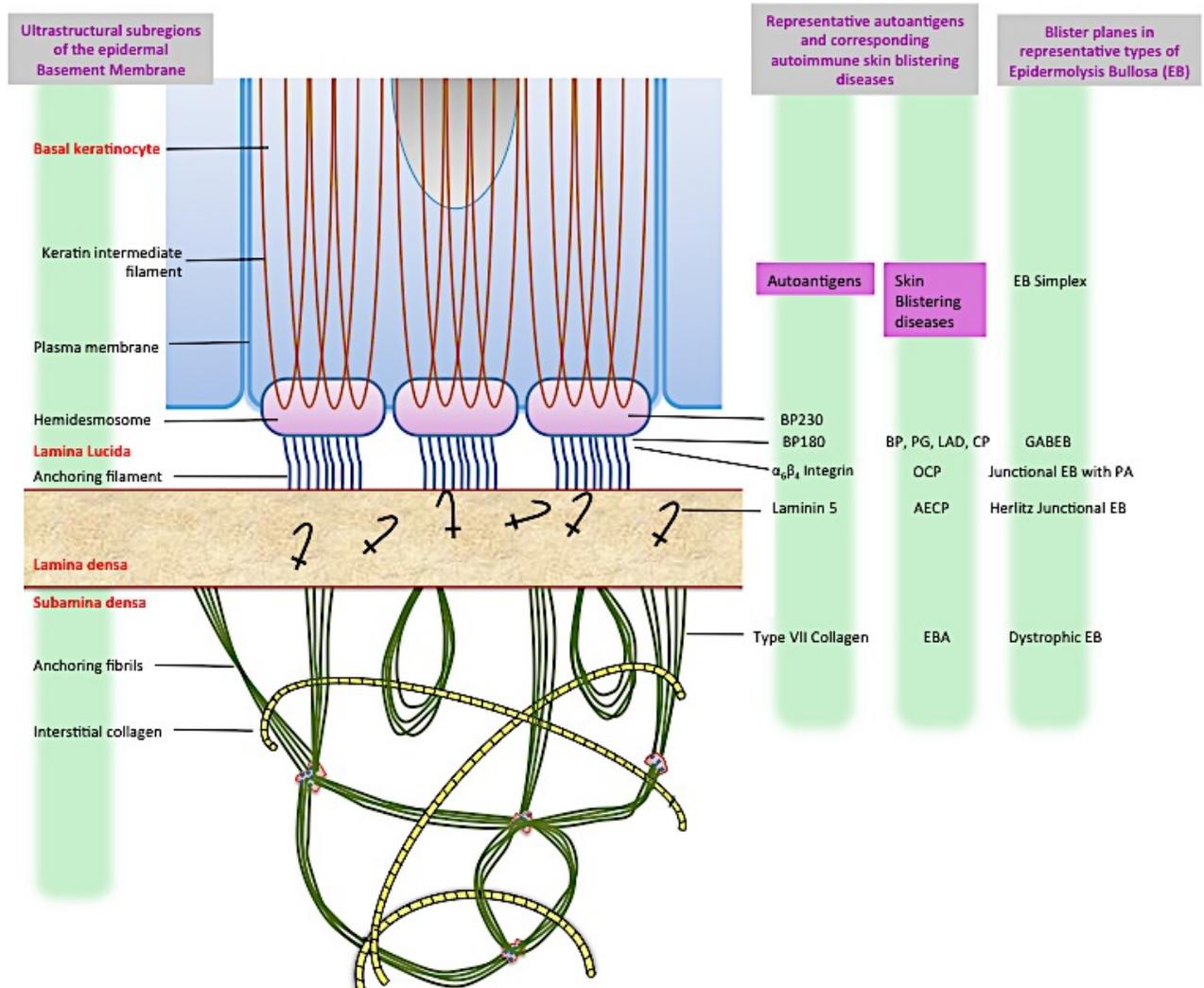
Figure 1.3: Schematic model of the epidermal basement membrane (BM).
The details of BM structural proteins that are targeted by autoantibodies in different autoimmune skin blistering diseases (ABSD) are shown**.** The major subregions of epidermal BM are depicted in the context of autoimmune and genetic blistering diseases. Proteins found as targets of autoantibodies are shown: AECP, Anti-Epiligrin Cicatricial Pemphigoid; CP, Cicatricial Pemphigoid; EBA, Epidermolysis Bullosa Acquisita; IB, Immunobullous; LAD, Linear IgA Dermatosis; OCP, Ocular Cicatricial Pemphigoid; GABEB, Generalized Atrophic Benign Epidermolysis Bullosa; PA, Pyloric Atresia. Figure is based on the idea from Yancey (2005) [65]**.**

1.3 Epidermolysis Bullosa Acquisita

Epidermolysis bullosa acquisita (EBA) is a chronic mucocutaneous and subepidermal autoimmune skin blistering disease characterized by autoantibodies to type VII collagen (COL7),

which is the major component of anchoring fibrils that attach the lamina densa of the basement membrane to the underlying dermis (**Figure 1.3**) [66], [67], [68]. Type VII collagen is made up of 3 identical α chains, each with a molecular weight of 290 kDa and each α chain is composed of an N-terminus non-collagen domain or NC1, Helical rod shaped collagenous domain in the center and a non-collagen globular domain or NC2, at the C-terminus [65]. The autoantibodies in EBA patients' sera were mapped to the NC1 domain of type VII collagen epitopes where interfering with the adhesion function of type VII collagen and the loss of anchoring fibrils eventually lead to the formation of subepidermal blisters [69], [70], [71], [72], [73] (**Figure 1.4**). In EBA both complement and neutrophilic granulocytes are involved along with activity of specific T cells, which is very similar to pemphigoid disease. Currently there are no specific drugs available to treat EBA patients, only combinations of immunosuppressant drugs that are used to treat pemphigus are also used to treat EBA patients [64].

1.3.1 Mouse models to study EBA

Disease models are used to understand the causes and the mechanisms of the disease, which can ultimately leads to the development of new treatments. Among other mammals, mouse is ideal to study human diseases for a number of reasons. Firstly, it has a short generation time, female yield an average of 5-10 pups, and their small in size (30-40g) makes them easy to handle. There are widely available techniques to manipulate the mouse genome and study genetic effects on disease phenotypes. For example, a germ free mouse model can be developed in a different genetic background, which could then be used to study the environmental and genetics effects on different phenotypes. Even differences in both development and metabolism exists between mice and humans [74], the mouse is still the most widely used experimental model animal used to study human disease in detail.

A spontaneously occurring sub epidermal blistering disease in mouse is rare, thus, induced models are necessary to reproduce the disease. Sitaru and colleagues showed that EBA blister formation could be produced in mice by passively transferring antibodies against type VII collagen from rabbit and injecting them into mice [75]. This model is called the passive EBA mouse model. This model suggests that EBA is an antibody mediated autoimmune disease [76]. Woodley and colleagues demonstrated that passive EBA mouse models can also be developed by injecting human EBA patient autoantibodies [77]. These antibody transferring passive EBA mouse models allow researchers to investigate mechanisms involved in autoantibody-induced tissue damage like complement activation [78] by alternative pathway [79], [80], neutrophils [81] and phagocyte-derived reactive oxygen species [81]. Another mouse model termed the "Active EBA mouse model" is made by injecting an immunogenic peptide from murine COL7 [82].
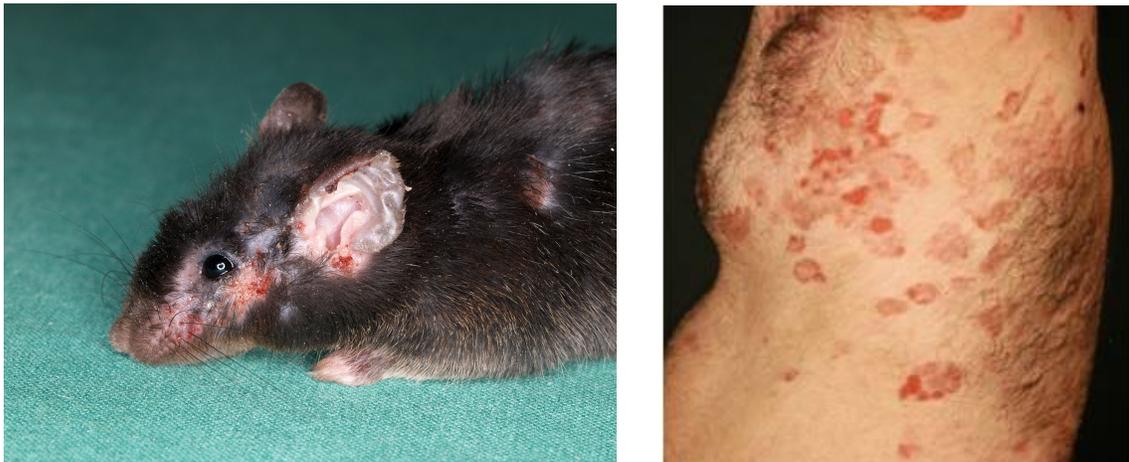
Figure 1.4: Active EBA mouse model and EBA disease in humans.
Blisters, erosions with crusts are seen in an EBA patient (photo was taken with patient's consent). Active EBA mouse model reproduces the autoimmune response type VII collagen where the active blistering with crusts can be seen in ears, near eyes and nose in the above mouse.

Different clinical phenotypes of EBA share common immunological features including autoantibodies to type VII collagen. This may or may not lead to a loss of anchoring fibrils depending on the mouse strain background [78], [82]. The active mouse model is comparable to the

human EBA phenotype (**Figure 1.4**) [78]. The active and passive mouse models are established in our lab, which allows researchers to efficiently study EBA pathogenesis and search for molecular clues in genetics, environmental factors as well as molecular and cellular pathways. Pros and cons of each EBA mouse model are explained in detail by Sitaru (2007) [76].

1.3.2 Genetics of EBA

Autoimmune diseases are mostly polygenic in origin (a few exceptions and a more detailed classification of autoimmune diseases are explained by McGonagle (2006) [83]), with major contributions from three main factors of genetics, environmental factors and immune dysregulation [84]. The interaction between genes and the environment determine susceptibility to complex diseases such as autoimmune and chronic inflammatory disorders is known [85], [86], [87], [88], [89]. A strong genetic link exists between certain kinds of haplotypes of major histocompatibility complexes (MHCs) and autoimmune diseases such as multiple sclerosis [90], [91], [92], [93] and rheumatoid arthritis [94], [95], [96]. The MHC region encodes genes for cell surface molecules that display peptides for immune recognition [97]. In the case of EBA using data from 29 patients, an association with HLA-DR2 has been shown [98]. This result varied between different populations because the previous study is based on a small number of patients.

Ludwig and associates [78] showed that EBA disease induction is strongly associated with the H2s mouse MHC haplotype by using an experimental EBA model inducing several inbred mouse strains. They also compared EBA disease induction in C57Bl/10.q and C57Bl/10.s mice where they differ only by H2q and H2s MHC haplotypes respectively. They found that C57Bl/10.q mice are completely resistant, whereas H2s developed mild, transient disease. By investigating different mouse strains they reported that 75% of mice carrying H2s developed EBA. Further questions are raised within the same study. The authors noted that the disease induction and severity in H2s MHC haplotype animals such as in MRL/MpJ or SJL/J is higher compared to same haplotype

in different genetic background such as C57Bl/10.s. This strongly suggested that there are non-MHC dependent genes found within the genome that also contribute to EBA disease induction or severity.
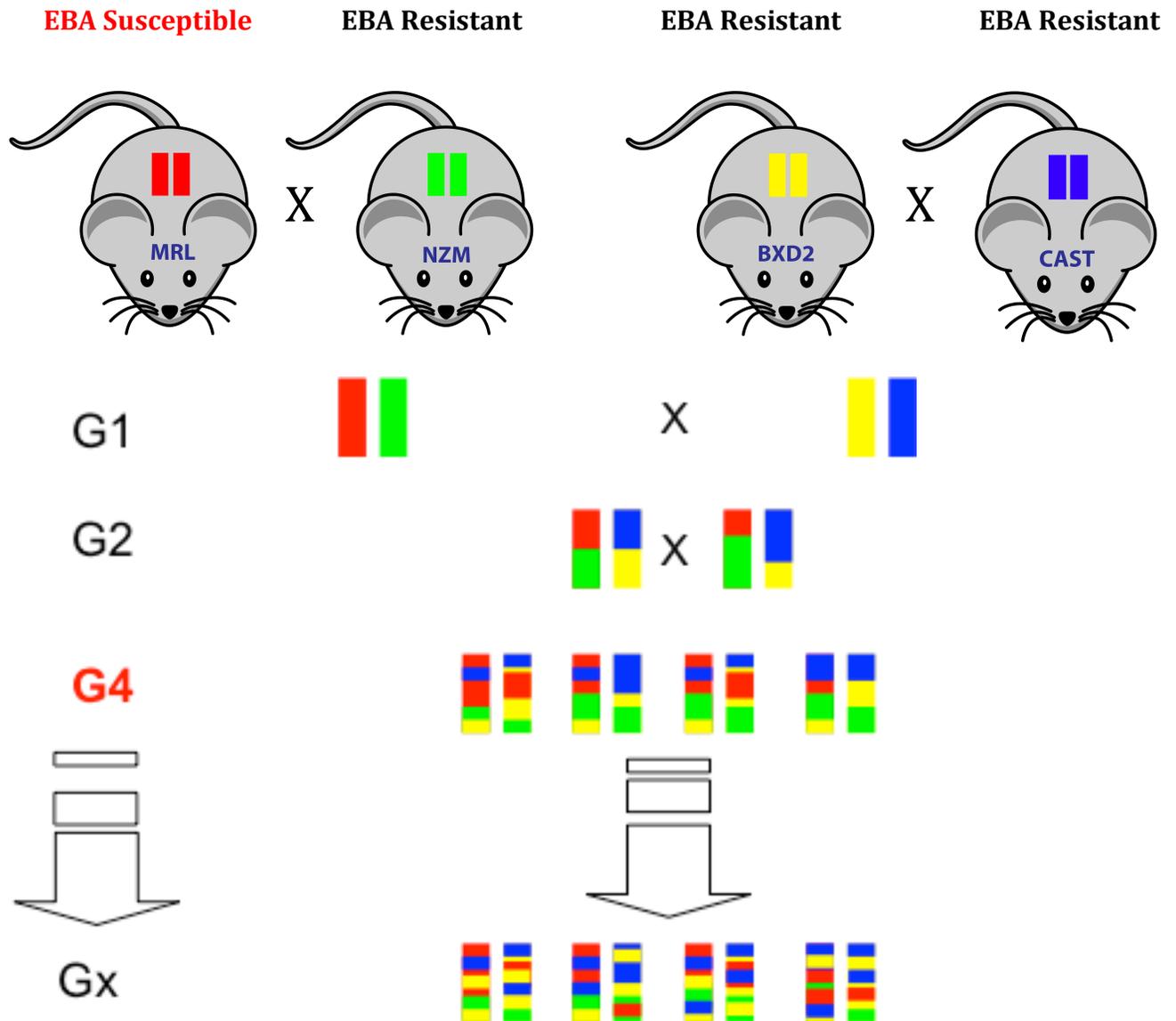
Figure 1.5: A schematic diagram of generation of four way autoimmune prone-advanced intercross mouse lines.
Genetic diversity of mice was obtained by intercrossing each generation of mice. Mice from generation four (G4 marked in red) were used in this study as well as by Ludwig *et.al.* (2012) EBA QTL study [99]. Mice generated in this population lines were susceptible to immunization induced EBA model. (Mouse cartoon used in this figure is modified from original; Illustration by Picsburg, source: http://animecartoondrawing.com)

The contribution of non-MHC genes to the susceptibility of other autoimmune disease is well known [100], [101]. This is commonly referred to as genetic heterogeneity, where multiple combinations of genes within the genome lead to a similar disease phenotype [101]. In order to find non-MHC genes contributing to EBA disease phenotype, Ludwig and colleagues used a four-way autoimmune-prone advanced intercross line between an EBA susceptible strain (MRL/MpJ) and three EBA resistant strains (BXD2/TyJ, NZM2410/J and CAST) as shown in **Figure 1.5**. They revealed six quantitative trait loci (QTLs) involved in controlling susceptibility to immunization induced EBA [99]. SJL/J mice carry the H2s haplotype, and immunization of these mice leads to anti-COL7 IgG production in all mice (Ellbrecht *et al*. unpublished). Interestingly, 20% of these mice remain healthy; i.e. no blister formation is observed after immunization, despite the presence of IgG2 antibodies [78], [82]. Presence of autoantibody production and absence of skin blistering in genetically identical mice held under same environmental conditions cannot be explained by the current understanding of EBA pathogenesis (Ellbrecht *et al*. unpublished). This results points to the direction that there are environmental factors along with host immune factors playing a significant role in the pathogenesis of EBA.

1.4 Scope of the thesis

The main aim of this study is to evaluate the interaction between host genetics and the skin microbiota in the context of Epidermolysis bullosa acquisita (EBA), an autoimmune skin blistering disease (**Figure 1.6**). This is the first study to investigate the contribution of host genetic control of the skin microbiota using such an approach (**Figure 1.5**). A further goal is to characterize the identified genetic loci influencing skin microbial community structure on EBA disease susceptibility.
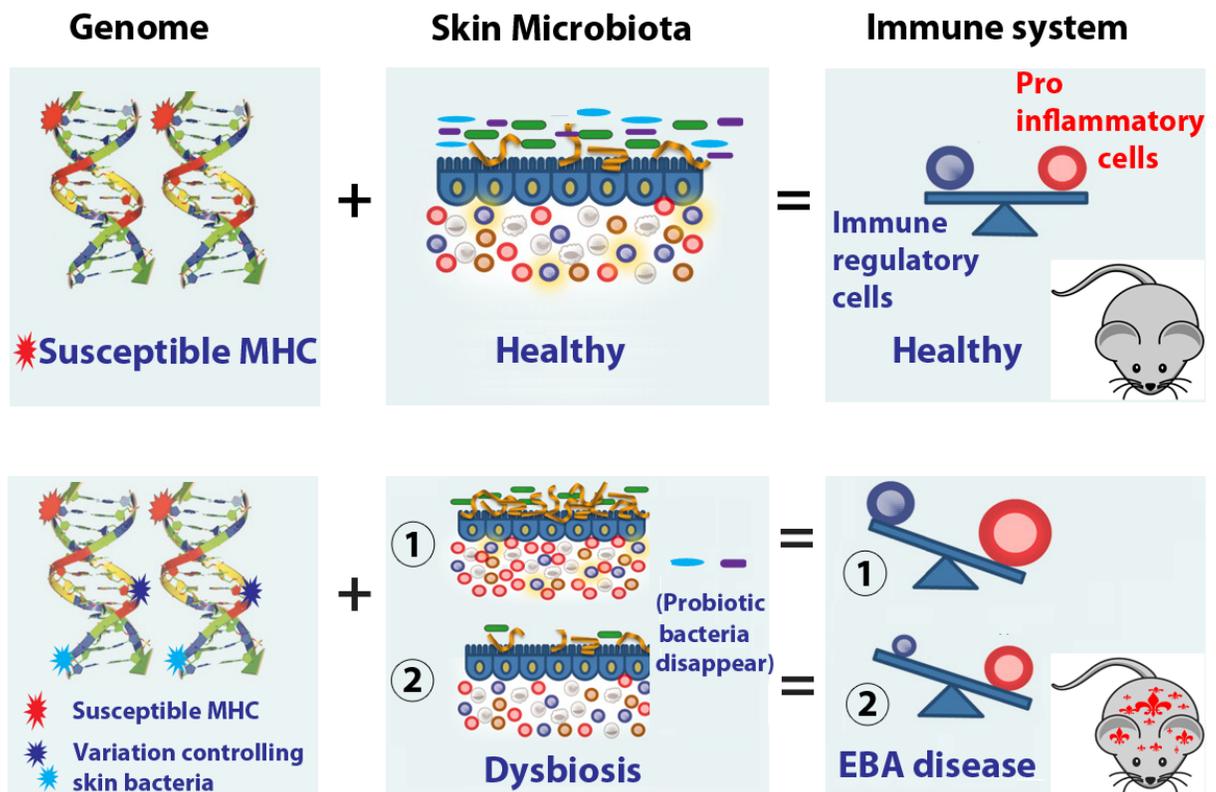
Figure 1.6: A simplified model showing the hypothesis of this study: skin microbiota and host genetics interaction leading to EBA disease susceptibility.
The unnatural shift of healthy skin microbiota is known as dysbiosis. The figure is modified from Lee YK *et.al.* (2010) [102].

2. Another aim of this study is to determine what extent changes in the skin bacterial composition influences the onset of the Epidermolysis bullosa acquisita (EBA), an autoimmune skin blistering disease. To test this hypothesis, genetically identical mice belonging to SJL/J mice carry the H2s haplotype (a susceptible MHC to EBA) were held under same environmental conditions and their ear samples were taken before immunization for characterization of skin bacterial communities **(Figure 1.7)**. Immunization induced EBA in the SJL/J mouse strain show incomplete susceptibility and this model could be further used to investigate disease-modulating mechanisms in genetically susceptible mice that harbor same environmental conditions. This model is an excellent tool to

investigate the presence of skin microbiome signatures that could be used to predict the onset of the EBA disease as microbiota phenotyping could be observed before disease induction. The results from this study are an important contribution to the ongoing debate about whether the changes in microbial communities are a cause or consequence of the disease.
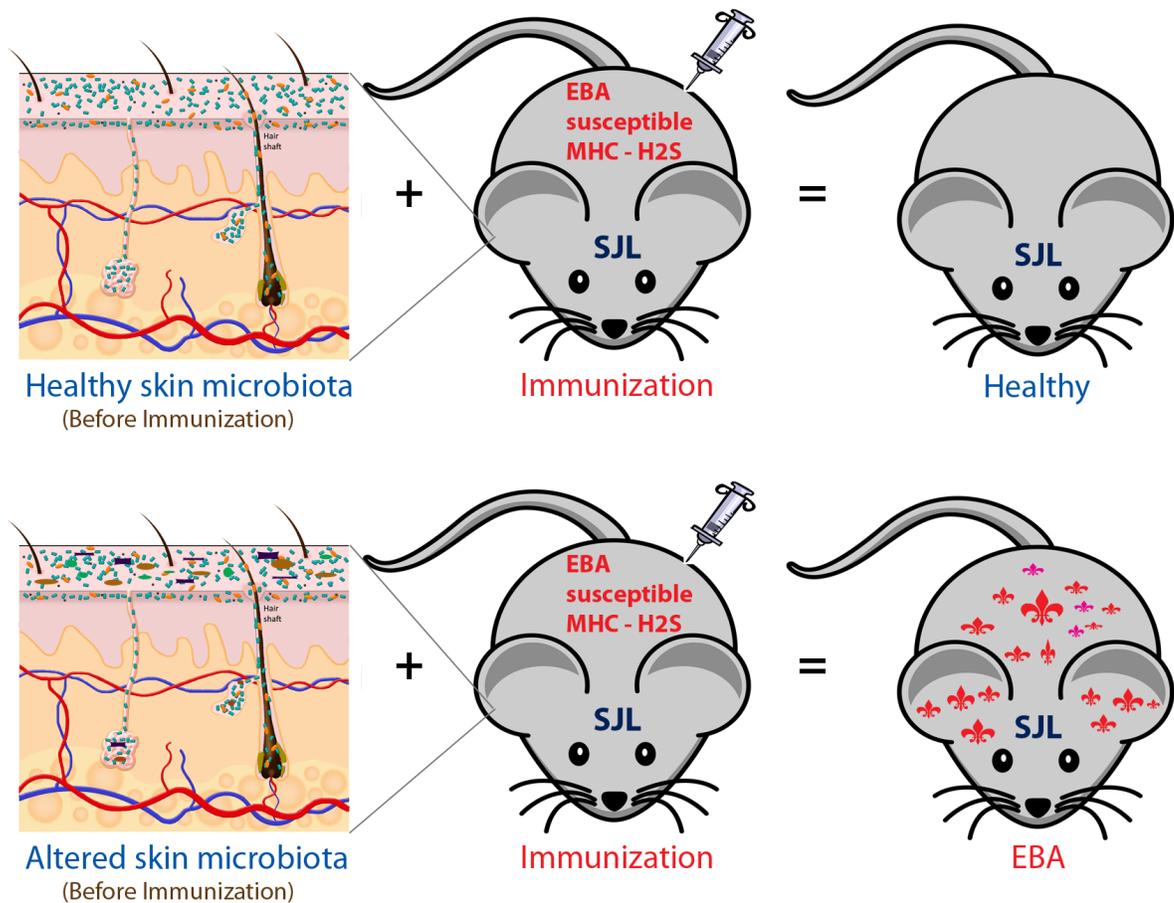
Figure 1.7: A schematic model explaining the importance of skin microbiota role in EBA disease susceptibility in SJL background mice carrying H2S haplotype.
H2s haplotype is susceptible to EBA by immunization method.

2 Materials and Methods

2.1 Generation of a four way advanced intercross lines

Parental strains (MRL/MpJ, NZM2410/J, BXD2/TyJ, Cast) for generating a heterogeneous intercross line [75] were purchased from the Jackson laboratory (Maine, USA) through European distributor Charles River Laboratories, Germany. Briefly, strains were intercrossed at an equal strain and sex distribution to generate a genetically diverse mouse line. First generation (G1) offspring mice were then mated considering their parental origin to maintain an equal distribution of parental alleles for successive generations by maintaining at least 50 breeding pairs per generation. Male and female offspring used in the study were transferred to separate cages according to sex after weaning. This procedure was also followed for intercrossing and producing next generation of mice lines (**Figure 1.5**). Animals were coded with unique three-digit identification number and cages are identified using cage numbers also recorded according to standard lab protocol. Details regarding each inter-cross were also maintained as per standard lab protocol. In each generation a greater genetic diversity between animals was reflected by the different morphological traits like weight, tail length, fur or coat color and it was re-coded in detail. Animals were held under specific pathogen free conditions at a 12-hour light/dark cycle with food and water *ad libidum*. All 261 animals in this study were taken from fourth generation of this advanced intercross line (AIL). All animal experiments were approved by the state of Schleswig-Holstein, Germany.

2.2 SJL/J mice

Female SJL/J mice were obtained from Charles River Laboratories (Sulzfeld, Germany). At the beginning of the experiments mice were 8-10 weeks old. Housing facilities were according to FELASA recommendations with a dark: light cycle of 12:12 hours, an ambient temperature of $23\pm1°C$ and a humidity of $65\pm5\%$. Housing of the mice was performed under conventional

husbandry practices in open-box cages on metal racks. SPF-conditions were ensured throughout the experiments without any detection of before unknown microbes. Mice were fed acidified drinking water and chow ad libitum without changing the supplier during experiments. All protocols were approved by the Animal Rights Commission of the Ministry of Agriculture and Environment of Schleswig-Holstein and performed by certified personnel.

2.3 Recombinant peptides

A GST fusion protein of the immunodominant mCOL7C epitope of the murine NC1 domain (amino acids 757-967) of COL7 in a prokaryotic expression system was generated. This fusion protein was purified by glutathione-affinity chromatography as described [104]. Subcloning the mCOL7C fragment into pQE40 (Qiagen, Hilden, Germany), using Bam HI and Hind III restriction sites, produced a His-tagged mCol7C. The subcloned protein was expressed by E. coli and purified using Talon®-immobilized cobalt affinity chromatography (Clontech, Saint-Germain-en-Laye, France).

2.4 Induction of experimental EBA and observation protocol

EBA was induced by immunization with an immunodominant peptide within the murine NC1 domain of type VII collagen (GST-mCOL7C) and this protocol was slightly modified from previously published paper [82]. In brief 60μg GST-mCOL7C emulsified in 60μl adjuvant (TiterMax, Alexix, Lörrach, Germany) was injected subcutaneously in to mice footpads and tail base. After immunization mice were screened for skin inflammation every 4th week for a period of 12 weeks, after which the ears were taken for analysis. In this study 183 mice were immunized using above protocol and 78 non-immunized mice were also taken from fourth generation of advanced intercross lines (G4). The extent of disease was determined by the percentage of body surface area covered by lesions. This was evaluated 4, 6, 8, and 10 weeks after immunization. The extent of EBA skin disease was scored as follows: 0 to 5, corresponding to 0%, <1%, ≥1% to <5%, ≥5% to <10%,

≥10% to <20%, and ≥20%, respectively. From every mouse, ear skin samples were obtained at the end of the 10-week observation period. Ear skin samples were fixed in 4% buffered formalin and snap-frozen at -80°C. The observed scores were later classified in to Low (0>x<5), Moderate (5>x<15) and Severe EBA (x>15) categories, where x is the percentage of the EBA disease covered. For SJL/J mice, disease incidence was evaluated by clinical examination for presence of erosions, blisters, crusts, alopecia and skin necrosis every week after immunization. Disease severity was calculated as percentage of body surface area affected by skin lesions. An affected maximum body surface less than 1.5% was regarded as clinically healthy, as this is comparable to unspecific lesions in controls. Two weeks after immunization, one of the popliteal, draining lymph nodes was surgically removed. As control group served mice that were injected with mCOL7C/TiterMax, but that were not biopsied.

2.5 Genomic DNA extraction for genotyping

In the genotyping protocol, genomic DNA was isolated from tail clippings was isolated by incubation in 500μl 50mM NaOH at 95°C for 2 hours (h), and posterior addition of 50 μl 1M Tris-HCl (pH 8.0). The DNA exact protocol was followed as in DNeasy Blood & Tissue Kit (Qiagen, Germany) according to manufacturer's instructions. The extracted genomic DNA content was later quantified using Nanodrop spectrophotometer. Extracted DNA sample from different mice were normalized to 50ng/μl in TE for genotyping. A typical protocol for making 1X TE buffer is: 10 mM Tris (bring it to pH 8.0 with HCl and 1 mM EDTA). Agarose gel electrophoresis was done for each sample for quality control purpose.

2.6 G4 population genotyping

Illumina mouse medium density linkage panel with 1449 SNPs were used for genotyping and it was processed at the Centre for Applied Genomics, Toronto, Canada. In order to avoid potential false positive association the quality control filtering in genotyping data is essential. SNPs with

insufficient genotyping quality were removed. Mainly genotyping quality is based on the Hardy-Weinberg equilibrium principle, minor allele frequency and missing proportion. Around 17% of SNPs were excluded due to low quality, only 1199 SNPs were found to be informative across 261 mice. These SNPs were further used in the downstream analysis in this study.

2.7 Bacterial DNA extraction and 16S rRNA gene pyrosequencing

Each sample mouse ear was split using mortar and pestle in the presence of liquid nitrogen to maintain low temperature. Approximately one third of an ear was transferred to the Power Bead tubes containing 60µl of C1 solution (from manufacturer's powerSoil® Kit from MoBio, Carlsbad, CA) and 20µl of 20mg/ml Proteinase K. Samples were incubated at 50°C for 2h at 850 rpm and the remaining steps were performed according to the manufacturer's instructions. The hypervariable V1 and V2 region of the 16S rRNA gene was amplified using the composite forward (5´-*CTATGCGCCTTGCCAGCCCGC*TCAGTC<u>AGAGTTTGATCCTGGCTCAG</u>-3´) and reverse (5´*CGTATCGCCTCCCTCGCGCCA*TCAGXXXXXXXXXXCA<u>TGCTGCCTCCCGTAGGAGT</u>-3´) primers. These primers included the 454 Life Sciences Adaptor A (for reverse primer) and B (for forward primer) – donated by italics. Additionally, to the reverse primers were added barcodes of 10 bp (designated as XXXXXXXXXX). The underlined sequences represent the broadly conserved bacterial primers 27F and 338R. A unique 10 base multiplex identifier (MID; designated as XXXXXXXXXX) was added to the reverse primer to tag each PCR product. 100 ng of template DNA was added to 25 µl PCR reactions performed using Phusion® Hot Start DNA Polymerase (Finnzymes, Espoo, Finland). The cycling conditions were as follows: initial denaturation for 30 sec at 98°C; 30 cycles of 9 sec at 98°C, 30 sec at 55°C, and 30 sec at 72°C; final extension for 10 min at 72°C. All reactions were performed in duplicate and combined after PCR. PCR products were extracted with the Qiagen MiniElute Gel Extraction Kit and quantified with the Quant-iT™ dsDNA BR Assay Kit on a nanodrop 3300 fluorometer (Thermo Scientific, US). Equimolar amounts of

purified PCR product were pooled and further purified using Ampure Beads (Agencourt). A sample of each library was run on an Agilent Bioanalyzer prior to emulsion PCR and sequencing as recommended by Roche. Amplicon libraries were subsequently sequenced according to the manufacturer's instructions on a Roche 454 GS-FLX using Titanium sequencing chemistry.

2.8 454 pyrosequencing data analysis

The following steps were followed in processing of 454 pyrosequencing data analysis and it is explained more in detail. Python scripts were used to process the sequence data and they are part of quantitative insights into Microbial Ecology (QIIME) package [105].

2.8.1 Pre-processing steps

Using the amplicon processing software on the standard 454 FLX, each region of the sequencing plate yield a FASTA file with .fna as extension. This file contains sequences. Simultaneously, it also generates a quality score file, which contains a score for each base in each sequence included in the FASTA file. A perl script using the Smith-Waterman algorithm [106] was written to match the forward primer and barcode allowing no insertions or deletions. Sequences were required to have a length between 290 to 370 nucleotides, an average quality score ≥ 20 and contain no ambiguous bases. Sequencing belonging to individual mouse is identified by matching the barcode and all sequences were separated accordingly. An average of 5732 reads (or sequences) per sample was obtained for 261 animals.

Software	Version	Task
sffinfo sfffile	Both software are part of Roche 454 GS-FLX system Version 1.3.3	To convert SFF files (raw files) into FASTA (sequence data) and QUAL (quality scores) files.

barcode_primer_trim.pl quality_sequence_check.pl barcode_primer_trim.pl qiime_sample_preparation.pl (available DVD)	Perl Version 5.5 for Linux	To check sequence quality as specified in pre-processing steps. Simultaneously separate sequences in to individual animal using barcode as unique identifier.
greengenes_blast_tab_converstion.pl (available DVD)		To convert the blast output to tab delimited format.

2.8.2 OTU determination

Trimmed sequences were grouped in to operational taxonomic unit (OTU). OTUs are clusters of sequences intended to represent some degree of taxonomic relatedness. Sequences clustered at 97% sequence similarity are typically thought of representing species level. This method is flawed and it is still an active area of research to determine how OTUs should be defined and at what sequence similarity cutoff should be used to represent each taxonomy depth? In this study OTUs were picked using 97% sequence similarity against a reference 16s rDNA dataset for V1-V2 region (This region is trimmed from full length 16s rDNA sequence from Greengenes database downloaded from http://greengenes.ibl.gov/). UCLUST [107] algorithm was used to generate clusters. The default parameters of the algorithm were used for optimal cluster formation. There are also chances that the candidate sequences are outside of the similarity threshold to any reference sequence in the reference database; so new clusters are allowed to form in such cases. Since each OTU clusters are made up of many related sequences, a representative sequence is picked to represent each OTU cluster. This representative sequence for each OTU is used in the subsequent analysis. In this study, the representative sequence for an OTU cluster is chosen as the most abundant sequence showing up in

that OTU. Each representative sequence from OTU clusters was given an arbitrary unique identifier. For those representative sequences from OTU clusters formed without having similarity threshold to any reference data set were identified using "N" followed by arbitrary OTU unique identifier.

Software	Version	Task
pick_otus.py (available under QIIME)	Python version 2.6.5 for Linux QIIME version 1.4.0 uclust version 1.2.22q usearch version 5.2.236	The OTU picking step assigns similar sequences to operational taxonomic units clusters by clustering sequences at 97% sequence similarity threshold.
pick_rep_set.py (available under QIIME)		A representative sequence will be chosen as the most abundant sequence in the OTU.
otu_table_rv_conversion.pl otu_table_log_conversion.pl (available in DVD)	Perl Version 5.5 for Linux	Values in OTU tables are converted to relative values and log values by respective scripts.

2.8.3 Sequence alignment

Representative OTU sequences were aligned using PyNAST [108], a python implementation of the NAST alignment algorithm [109]. This algorithm aligns each candidate sequence to the best matching sequence in a pre-aligned database sequences or template sequences, it is a core set of full length 16s rDNA sequences obtained from Greengenes website (http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/core_set_aligned.fasta.imputed). Default parameters of NAST alignment tool was chosen for optical alignment of the sequences. ChimeraSlayer perl script (http://microbiomeutil.sourceforge.net/) was used to remove chimeras with reference to chimera free gold standard database set published by DeSantis *et al.* (2006). After removing chimera sequences it resulted in 34624 species-level OTUs. In order to examine changes in

community composition within (alpha diversity) and between (beta diversity) compositions across samples based on health status, species-level binned OTUs were analyzed by different diversity indices.

Software	Version	Task
align_seqs.py (available under QIIME)	Python version 2.6.5 for Linux QIIME version 1.4.0 PyNAST version 1.1 PyCogent version 1.5.1 NumPy version 1.5.1 Perl Version 5.5 for Linux	To align sequences using the PyNAST method.
filter_alignment.py (available under QIIME)		This step will remove positions, which are gaps in every sequence. Typically, this will differentiate between non-conserved positions, which are uninformative for tree building, and conserved positions, which are informative for tree building.
ChimeraSlayer.pl (available online and link provided, please see in method description)	Perl Version 5.5 for Linux ChimeraSlayer Version 1.22.0	To remove chimeric sequences.
split_align_products.pl (available in DVD)	Perl Version 5.5 for Linux	To split alignment files according to PCR products for PCR chimera check.

2.8.4 Normalization using rarefaction method

The difference in the total number of sequences in each sample can lead to over or under representation of certain OTUs in samples. So it is very important to normalize the data across the samples. Rarefaction procedure was implemented. It creates a series of subsampled OTU tables by random sampling procedure without replacement from the initial original species level OTU table.

These tables are referred as rarified OTU tables and the analysis is called rarefaction analysis. In this study, multiple rarefaction analysis was performed on all samples at various sequencing depths. 10 iterations at each sampling depth of 50 beginning with 50 sequences/sample through 2500 sequences/sample were generated. The samples from some individuals did not yield more than 2500 sequences. These rarefaction OTU tables are not only used for normalization but also to construct rarefaction graphs that are widely used by researchers to estimate the sequence depth for each sample. One of the samples from SJL/J mice did not yield more than 1000 sequences. The rarefactions OTU tables are normalized at 1000 sequences/individual.

Software	Version	Task
multiple_rarefaction.py (available under QIIME)	Python version 2.6.5 for Linux QIIME version 1.4.0 PyNAST version 1.1 PyCogent version 1.5.1 NumPy version 1.5.1	This step creates a series of subsampled OTU tables by random sampling (without replacement) of the input OTU table. The pseudo-random number generator used for rarefaction by subsampling is NumPy's default.
Multiple_rarefaction_even_depth.py (available under QIIME)		

2.8.5 Alpha diversity analysis

In this estimation, the diversity is assessed within a sample sometimes also referred as within sample diversity. A series of scripts were written along with QIIME scripts to estimate the alpha diversity within the samples in this study. Molecular ecologists have used wide range of metric to measure alpha diversity and every metric has different limitations and strengths. These estimators are focused either on estimating different types of species (qualitative species based measurement) or their abundances at a given scale (quantitative based measurement). However, this is still a major drawback of these estimators. In this study four estimators were used. The Shannon index (based on

quantitative) is used to calculate species richness within the sample using the formula $E = e^{(H/S)}$, where S is the number of taxa in that group (S = species/phylotypes) and H is Shannon diversity index. The Shannon diversity index $H = \Sigma\ P_i\ ln(P_i)$ where P_i is the proportion of the *i*th OTU [110]. Chao1 [111] is focused on qualitative based species measurements. Chao1 index is slightly modified version of Chao and it is calculated using the formula (corrected for bias refer this book [112] to Chapter 4 for more details) $Choa1 = S_{obs} + (f_1(f_1-1))/2(\ f_2+1)$ where S_{obs} is the total number of species observed in a sample, f_1 is the number of singleton species, f_2 is the number of doubleton species. Phylogenetic Diversity (PD) is qualitative divergence based measure which sums the total branch length in a phylogenetic tree that leads to each member of a community [113]. It has a special property in which adding a new species will always increase the PD index value. Hence this measurement is highly sensitive to sampling efforts. The Phylogenetic tree was built using FastTree tool [114] and default parameters were used for optimal results. In order estimate the OTU richness within the sample observed species measure was chosen. All these measurements are applied for collection of rarefaction OTU species level tables generated by rarefaction method.

Software	Version	Task
alpha_diversity.py (available under QIIME)	Python version 2.6.5 for Linux	This step calculates alpha diversity, or within-sample diversity, using an otu rarefaction tables based on different metrics.
make_phylogeny.py (available under QIIME)	FastTree version 2.1 QIIME version 1.4.0	This step produces a tree from a multiple sequence alignment. Trees are constructed with a set of sequences representative of the OTUs using FastTree software program.

collate_alpha.py (available under QIIME)	PyNAST version 1.1 PyCogent version 1.5.1 NumPy version 1.5.1	The result of alpha_diversity.py comprises many files from rarefaction OTU tables, which need to be concatenated into a single file for generating rarefaction curves. This script joins those files.
make_rarefaction_plots.py (available under QIIME)		This script generates rarefaction plots based on the supplied collated alpha-diversity files.

2.8.6 Beta diversity analysis

Beta diversity represents the explicit comparison of microbial communities based on their composition between samples. This can explicitly explain the differences of the microbial communities found in different samples. This diversity analysis is based on a square matrix where a similarity or dissimilarity distance is calculated between every pair of samples. This matrix will reflect the dissimilarity between those samples. Many visualization techniques such as Principal component (or Coordinate) Analysis (PCoA), redundancy analysis, Constrained Principal component analysis and hierarchical clustering could be used. Rarefaction OTU table at 2500 sequences/sample is used in this analysis. This cutoff was used because all 261 samples have at least 2500 sequences. Jaccard Distance and Bray-Curtis distance between all 261 samples were compared using the R package 'vegan' [115]. If we denote the number of species shared between two sites as '*a*' and the numbers of the unique species as '*b*' and '*c*', then $S = a+b+c$ and $\alpha = (2a+b+c)/2$. Bray-Curtis dissimilarity measure = $(b+c)/(2a+b+c)$. This measure is also known as Sørensen dissimilarity. Jaccard dissimilarity measure = $(b+c)/(a+b+c)$. This measure ranges from 0 (no species in common) to 1 (share identical species lists). In this study additionally another measurement called UniFrac is also used. This measurement is based on the

phylogenetic information. Weighted UniFrac and unweighted UniFrac is calculated using Fast UniFrac tool[116]. Phylogenetic tree built using FastTree tool was used as input for weighted and unweighted UniFrac distance matrix. Constrained PCoA with healthy and disease status as a constraint. This analysis is very similar to a redundancy analysis but additionally it allows non-Euclidian dissimilarity indices. Constrained analysis of principal coordinates (CAP) were calculated using the Vegan package in R [115]. Statistical significance for CAP was determined by an ANOVA-like permutation test function with 1000 permutations (anova.cca function) in Vegan was used. Adonis multivariate statistics were applied using Vegan [115] package in R. R version 2.15.2 for Linux was used for all analysis (*R* Development Core Team (2012)).

Software	Version	Task
beta_diversity.py (available under QIIME)	Python version 2.6.5 for Linux QIIME version 1.4.0 Fast UniFrac (a new version of UniFrac that is specifically designed to handle very large datasets)	This step calculates beta diversity, or between-sample diversity, using an otu rarefaction tables based on different metrics.
Vegan (R package available in CRAN)	Vegan version 2.0-7 R version 2.15.2 for Linux	To perform beta diversity statistics.
vegan_data_prepartion.pl (available in DVD)	Perl Version 5.5 for Linux	Automatically converts all file formats to vegan file format for data processing.

2.8.7 Indicator species analysis

Indicator species analysis was performed using R package "indicspecies". This package provides a set of specific functions to assess the species relationship between categories. Indicator and biserial

correlation values of each species among healthy and EBA samples were calculated using "Indval.g" and "r.g" functions within the packages. Species level OTU measurements in relative values were used as input. In order to obtain high statistical confidence permutation scheme implemented and this function is available within the package. Permutation was done using 10000 permutations. To correct for multiple testing "Benjamini and Hochberg" correction was applied using "p.adjust" function in R.

Software	Version	Task
indicspecies (R package available in CRAN)	indicspecies version 1.6.7 R version 2.15.2 for Linux	This package provides a set of functions to assess the statistical significance of the relationship between species occurrence/abundance and groups of sites

2.8.8 Taxonomy classification

RDP classifier [117] (RDP Multi-Classifier version 1.0 available from http://sourceforge.net/projects/rdp-classifier/files/MultiClassifier/) was applied to assign taxonomy to the genus level using 0.80 as minimum confidence. The "fixrank" option was used to assign taxonomy from kingdom to genus level. Sequences unclassifiable at a given taxonomic rank were classified at the next possible higher rank while genus being the lowest. Since RDP classifier can assign sequences till genus level it is important to look for alternative taxonomy assignment at species level. BLAST taxonomy assignment method was implement. Species level taxonomy was obtained by BLAST [118] against the Greengenes reference database at species level OTUs (http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/Caporaso_Reference_OTUs/g g_otus_4feb2011.tgz) with an E-value cutoff of 0.001. For sequences that did not yield any species name is assigned as "*Unclassified Species*".

Software	Version	Task
MultiClassifier.jar (available online and link provided, please see above)	RDP multi classfier version 2.0 Java Version 7 for Linux	To assign bacterial 16S rRNA sequences to the new phylogenetically consistent higher-order bacterial taxonomy.

2.9 Data preparation for QTL analysis

The proportion values were calculated at each taxonomic level using the following equation where proportion values equals to ratio of number of reads for a given taxon to total sequence reads for a given animal for each taxonomic level. Proportion values were then log10 transformed and for animals which no sequence counts were obtained for a given taxon, proportion value was calculated using 0.5/total reads for that taxon and then log10 transformed. These equations were applied accordingly and resulted in calculated log10 transformed proportional value across 261 mice to 34624 species-level OTUs, 863 genera, 376 families, 218 orders, 93 classes and 51 phyla. These taxonomic depths contained at least one assigned sequence in their taxa. Out of 34624 species-level OTUs, 19443 covers nearly 99% of the total sequences, and 762 species OTUs represents nearly 90% of the total sequences.

2.10 Core measurement microbiota

To determine quantitatively measurable microbial traits across 261 samples from each taxonomic depth a core measurement microbiota is defined. 16S rDNA from five different samples were amplified with two different sets of barcoded primers. These samples were filtered and processed in similar fashion as explained above in this thesis. Sequence counts for each taxonomic level was log10 transformed. A scatter plot was generated from all pairwise combinations of the two repeat for each five different sample. A threshold of greater than 20 reads (or sequences) per bin (or taxon) leads to a correlation above 0.97 was observed. Thus, the Core Measurable Microbiota (CMM) taxa

were defined as taxonomic bins containing more than 20 reads in at least 20 animals. The resulting

CMM of 131 species-level OTUs represent nearly 80% of the total sequences.

Software	Version	Task
cor.test (function available in R)	R version 2.15.2 for Linux	This function is used to calculate correlation analysis.

2.11 QTL analysis

The underlying haplotype structure of each mouse was inferred for linkage disequilibrium

mapping of QTLs using HAPPY software version 2.1 (available at

http://www.well.ox.ac.uk/happy/happyR.shtml) [119]. HAPPY is originally written in C and uses R

interface for QTL mapping. It is based on hidden Markov model to infer the haplotype descent from

each mouse and also uses multipoint analysis that offers a significant improvement in statistical

power to detect QTLs compared to single marker association. It utilizes the known founder genotype

information along with recombination distances to provide a probabilistic estimate measure of

haplotype descent at each marker interval for each mouse. For mouse i at marker interval m, HAPPY

software computes a vector $g_i(m)$ which contains the expected proportion of genetic material

descended from each of the 4 possible pairs of founder haplotypes from AIL. Later this vector is

used to characterize variation at the given locus to test for phenotype association. QTL mapping

experiments in crosses of inbred mouse strains in this study were taken from generation four

population. The distances between informative recombinants are larger and limited. It is also known

that these populations are prone to uneven genetic relatedness among individuals thus require proper

model selection to avoid false positive associations from family (sibship) and environment (cage).

The strong association of family and cage in defining the microbial phenotype was also previously

described in Andrew K. Benson *et. al.* [27]. These factors were taken in to account during the QTL

model selection procedure and it is described in detail below.

Software	Version	Task
HAPPY (now called as happy.hbrem) (available online and link provided, please see above)	Happy version 2.1 R version 2.15.2 for Linux	To map QTL in Heterogeneous Stocks (HS), i.e. populations founded from known inbred lines, which have interbred over many generations without pedigree information.
compare_QTL.pl (available in DVD)	Perl Version 5.5 for Linux	To compare the QTL detected under different categories.
innate_immune_gene_overlap_sqtl.pl (available in DVD)		To find innate immune genes within the QTL.
parse_happy_QTL_output.pl (available in DVD)		To automatically parse all happy QTL output file and calculate confidence interval.
parse_happy_covariate_QTL_output.pl (available in DVD)		To automatically parse all covariate Happy QTL output file and calculate confidence interval
compare_covariate_QTLs.pl (available in DVD)		To compare the covariate QTL and EBA QTL (null model)
hfit.R (available in DVD)	Happy version 2.1 R version 2.15.2 for Linux	Modified R code to account for cage and family simultaneously during mapping.

2.11.1 Model selection

Each microbial trait from CMM at chosen taxonomic depth is treated as an individual phenotype trait in this model and the data were log10 transformed for normalization because HAPPY uses Analysis of Variance F statistics to test for linkage. To estimate the contribution of various factors such as cage, family (sibship), age, and sex in defining CMM, a linear mixed model analysis was performed using 'lmer function' in 'lme4 R package'. Mice with common parents are defined as one family. Family consists of genetic as well as environmental components. Within sibling from same family variation is explained by environmental variation and between families is explained by genetic variation. Same sex siblings from one or more family, depending on the number of mice, are kept in the same cage during the weaning stage. A model for a QTL at locus m on the Phenotype trait (microbial) was found using the following equations considering those factors that are significant from linear mixed models.

Null model: Equation 1

Microbial Trait $\sim$ sibship $_{k[i]}$ + cage $_{k[i]}$

Alternative model: Equation 2

Microbial Trait $\sim$ sibship $_{k[i]}$ + cage $_{k[i]}$ + Q_m

Where Q_m is the design matrix to fit a QTL at locus m. The design matrix Q_m contains information related to the strain probabilities of each genotype markers. The significance of QTL at each locus m is found by comparing the fit of a null model and an alternative model, via ANOVA test. To add covariates to the model, additional column was added to the additional design matrix. The HAPPY package allows the inclusion of covariate matrices and allows to test to see if they significantly improve the fit conditional upon the presence of the covariates.

Software	Version	Task
lme4 (R package available in CRAN)	lme4 version 0.999999-2 matrix version 1.0-12 lattice version 0.20-15 R version 2.15.2 for Linux	To perform linear mixed model analysis.

2.11.2 QTL mapping

HAPPY requires two input text files for QTL mapping. One is called alleles files (or marker files). This file describes the state of alleles for each marker in the founder populations and also contains the marker position in centimorgan (cM). Another file is called a data file that contains the phenotype and genotypes for the individuals. If there are K markers and N individuals, then the data file contains N rows and 2*K+2 columns. For each taxon bin a separate data file was generated. For every marker pair position, HAPPY calculates a LOD score. The above model equations (equation 1 and equation 2) were tested under additive model, which assumes there are independent QTLs and there are no interactions between alleles within each locus. HAPPY uses analysis of Variance F statistics (F-test) to look for association between locus and phenotypes using linear fit model. Let us assume the overall simulation over m loci is defined as Significance $S_i = -\log_{10} P_i$, where P_i is the P value from F-test at locus i [120]. Maximum over all loci is given by the equation $S_{max} = \max\{S_1,...,S_m\}$. The reported anova P value is highly anticonservative [120]. Since the Bonferroni correction is too conservative to use for the correction and it also ignores the linkage disequilibrium (LD) between markers. The distribution of most significant log P is found using genome wide scan with no QTL (equation 1) and compared it with each phenotype by permutation of phenotypes between animals while fixing the genotypes. This permutation scheme took LD in to account and also the phenotype distribution [121]. For each of 1000 permutations the genome wide scan was repeated and the most significant log P value was recorded. The quantiles form the extreme value

distribution from 1000 permutation were fitted and then used to estimate the genome wide significance for each phenotype [120].
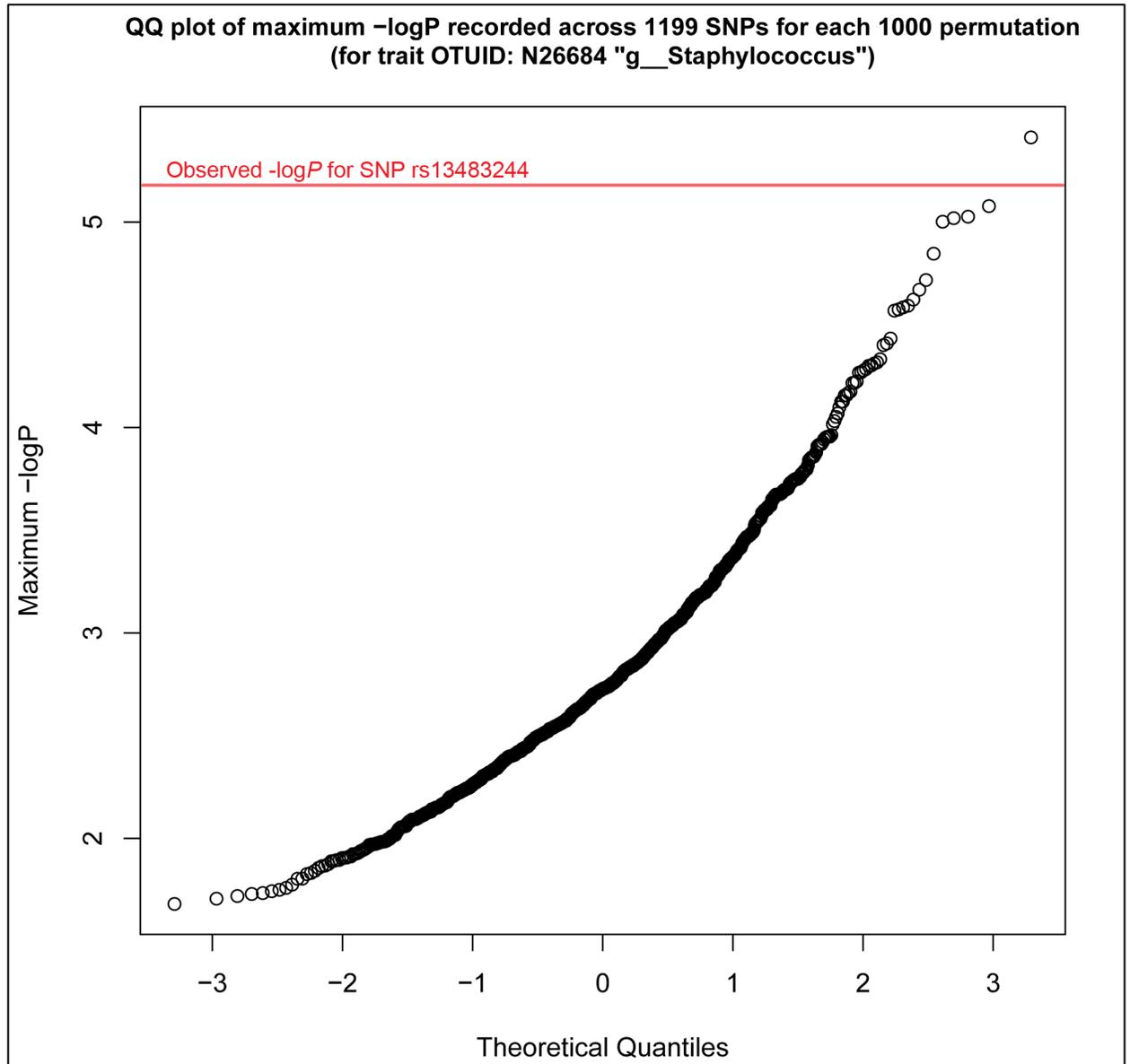
Figure 2.1: Distribution of maximum –log *P* scores recorded across 1199 SNPs for 1000 random permutations of the phenotype scores.
(a.k.a. "QQ plot", in this case for OTUID: N26684 "g_Staphylococcus"). This is used to estimate the genome wide significance threshold, or "*E* value". For example, the *E* value for SNP rs13483244 would be 1/1000 from the above plot (observed –log *P* value shown in red for trait OTUID: N26684). In other words, the probability of the observed –log *P* score for SNP rs13483244 (which is 5.24) to occur by chance is 1 in 1000.

To illustrate the procedure used to obtain the *E* value for a given trait, a "QQ-plot" was produced (**Figure 2.1**), which shows the distribution of maximum $-\log_{10} P$ scores recorded across 1199 SNPs for 1000 random permutations of the phenotype scores (for OTUID: N26684). The genome wide significance threshold was set at 5%, which is equivalent to one false positive per 50 genome wide scans. Overall genome wide significance threshold for all phenotypes were calculated and found that global significance cutoff at α = 0.05, a corresponding ANOVA $-\log P \geq 4.39$ for all QTLs and for α = 0.1, ANOVA $-\log P \geq 4.1$. The corresponding Anova $-\log P$ value to 5% global significance value between phenotypes varies only slightly and the same procedure has been widely implemented [121]. Confidence intervals were determined manually by a drop of 1.5 in ANOVA $-\log_{10} P$ score [122].The probability of overlap for QTLs was calculated as described by Graham *et.al.* [123] and the equation is written as $p_i = 1 - \frac{(1-W_i)^2}{2} - \frac{(1-W_T)^2}{2}$, where p_i is the probability of QTL overlap, W_i is the fraction of the confidence interval of the QTL to the overall physical size of the mouse genome and W_T is the ratio of the combined length of the confidence interval of QTL to the overall physical size of the mouse genome. 2500 Mb is used as the physical size of the mouse genome for the QTL overlapping calculation.

2.12 Data deposition

The microbiota sequence data has been submitted to the European Nucleotide Archive (ENA) under accession number ERP002614. The data will be released upon publication (revised and under review) and/or request by the Reviewer.

3. Results

3.1 Skin bacterial diversity

Skin samples were collected from ear of 261 mice from the fourth generation of an advanced intercross line (G_4 AIL) mapping population. This mapping population is a cross from four major strains (MRL/MpJ, NZM2410/J, BXD2/TyJ, Cast). Out of these 261 mice there are 119 healthy, 64 EBA, and 78 non immunized mice samples. Since the induction of EBA disease is by immunization procedure it is very important to control for immunization. This is the reason why this study also includes 78 non-immunized mice ear samples. DNA was extracted from each sample and microbial "phenotyping" was carried out using barcoded 454 pyrosequencing of the bacterial 16S rRNA gene. In total 1.5 million sequences were obtained form 261 animals. Each sequence was checked for sequence quality and assigned to respective samples by matching unique barcodes.

After quality control checks, sequences were clustered in to species-level operation taxonomic units using ≥ 97% sequence similarity. Operational taxonomic units (OTU) are used widely in ecological studies to classify bacterial sequences at different taxonomic level depths. Each taxonomic level distinction is made using respective sequence similarity threshold. The sequence similarity threshold is an arbitrary number and it is an important research focus area. Since current methods for defining species levels of bacterial diversity that are being covered new are incapable and inadequate, an OTU approach will allow us to classify newly found bacterial species using sequence similarity. For species level OTUs ≥ 97% sequence similarity is widely accepted cutoff and also followed in this study [124]. The representative sequences for each OTU bins were chosen. After controlling for chimeric sequences, there were 34624 species-level OTUs bins. A heat map is generated to visualize the abundances of OTUs distribution among 261 mice (**Appendix A Figure A.1**). This figure shows that there are only few species OTUs which are high in abundant across samples. In order to estimate the distribution of sequences among species OTUs precisely, another visualization technique was

used (**Appendix A Figure A.2**). This figure shows that 90% of the sequences from all samples are spread among 762 species OTUs. The rest of these OTU bins have at least one sequence present in one mouse sample. Using high quality curated 16s Greengenes database [39], (see **Method**) each of these representatives from OTU bins was classified till species level using BLAST method [118] with E-value threshold up to 0.001. Alternatively RDP multiclassifier [117] was also used to classify all sequences using 0.8 confidence threshold. RDP classifier could classify the sequences only till genus level. The independent classification of sequences helps us to find if there is any misclassification. After comparing the classification of sequences at each taxonomic depth by both classifiers resulted in no conflicting assignments. At the phyla level, the Firmicutes were most abundant (54%), followed by Proteobacteria (21%), Actinobacteria (12%) and Bacteroidetes (6%) (**Figure 3.1 a**), revealing communities similar to those observed in previous studies of the skin [6], [30], [47], [125]. The other phyla such as *Deferribacters, Deinococcus-Thermus, Fusobacteria,* and *Spirochaetes* were found less abundance in mouse ears (**Figure 3.1 b**). At the genus level, *Staphylococcus* (~36%), *Corynebacterium* (~9%), and *Ralstonia* (~8%) were most abundant (**Figure 3.2**). The majority of species level OTUs belonged to *Firmicutes* (**Figure 3.3**). The largest single OTU, a member of the division *Firmicutes* phylum belonging to *Staphylococcus equorum* species, which covers 27% of total sequence. The second largest OTU member belongs to *Actinobacteria* phylum and to the genus *Corynebacterium*. Followed by another OTU that belongs to *Proteobacteria* phylum and to the genus *Ralstonia*. Other common OTUs belonging to *Firmicutes* phylum division were *Staphylococcus* genus, *Aeroccocus* genus, *Lachnospiraceae* family, *Alicyclobacillus* genus, *Streptococcus Thremophilus* species, *Lactobacillus Delbruekii* species, *Bifidobacterium Longum* species, *Streptococcus* genus and *Lactobacillus* genus. Order *Bacteroidales* was the only species level OTU belonging to phylum division *Bacteroidetes* found among top ten abundance species OTUs classification. There were other OTUs classified to *Proteobacterium*

phylum found to be under *Enterobactriaceae* family, *Helicobacter Hepaticus*, *Enterobacter Cowaii* species and *Alphaproteobacteria* family.

a.
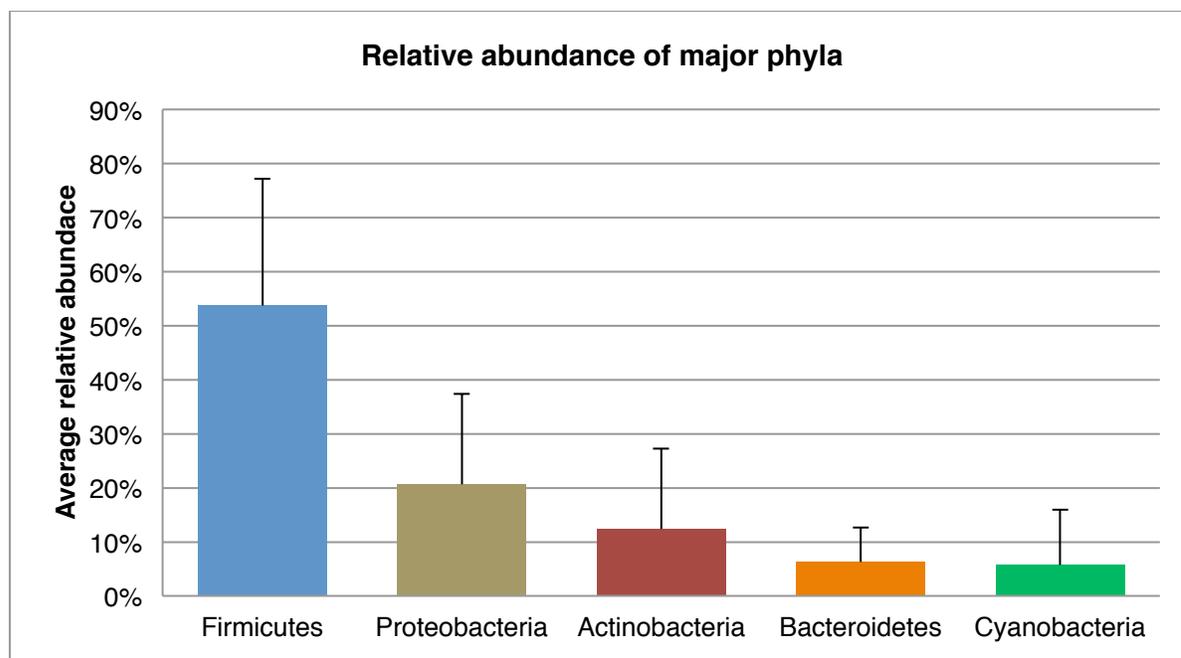
b.
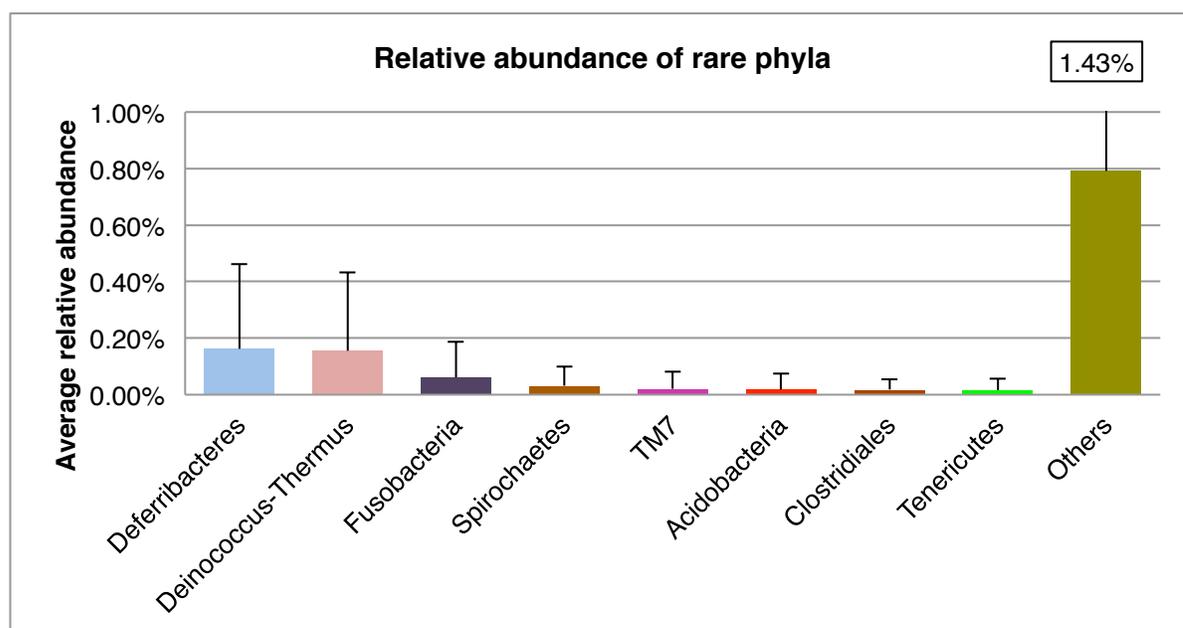
Figure 3.1: (a) Relative abundance of major and (b) minor phyla of mouse skin
Error bars indicate 1 standard deviation (SD).
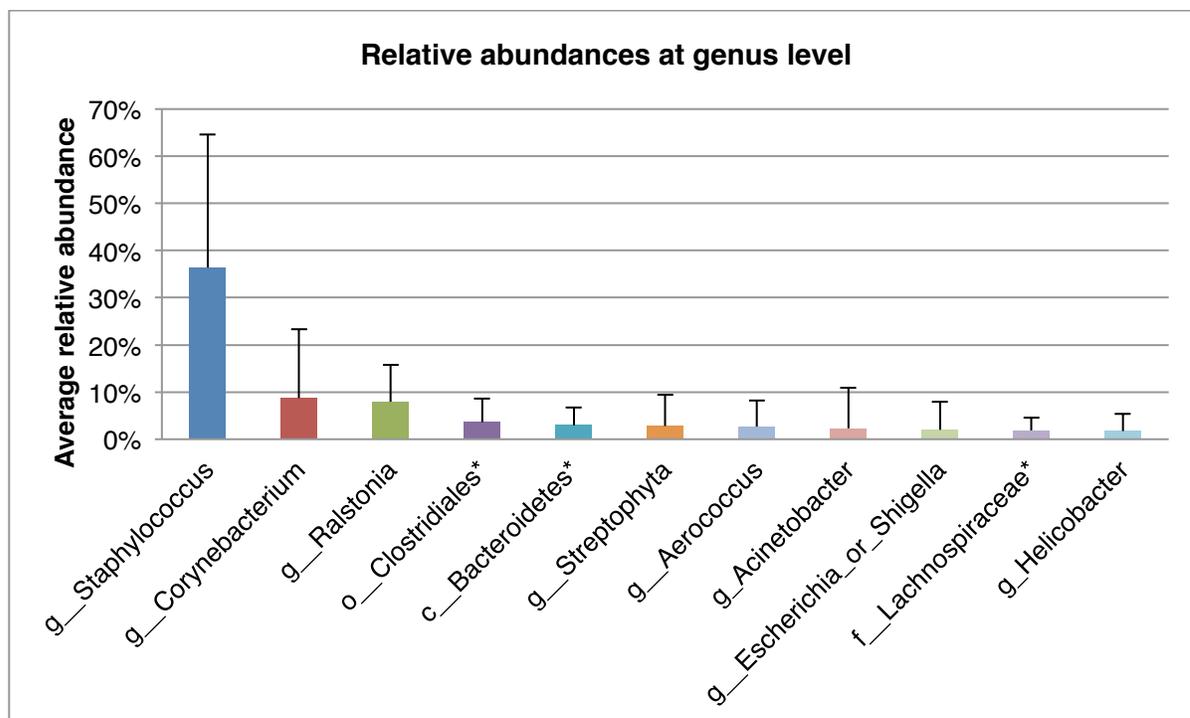
Figure 3.2: Ten most abundant genera in the mouse skin microbiota (n=261)
For those unclassified at the genus level, the next highest taxonomic level is shown (as marked in asterisk *). Error bars indicate 1 SD.
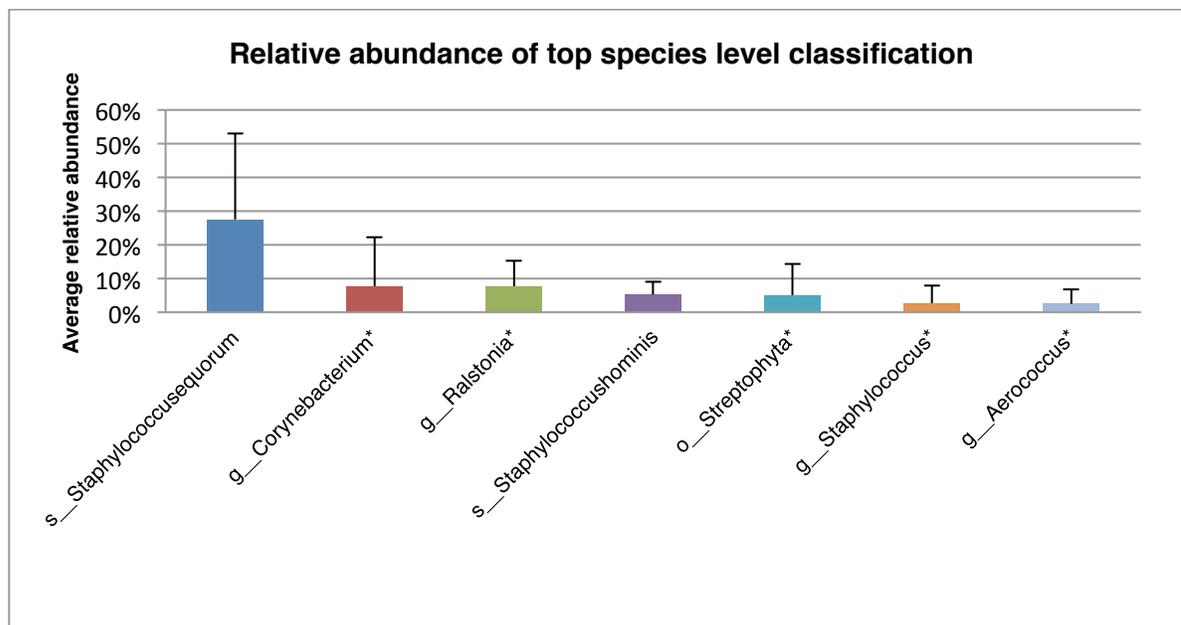
Figure 3.3: Relative abundance of major bacterial species of mouse skin those with common classification species OTUs were grouped.
For those unclassified at the species level, the next highest taxonomic level is shown (as marked in asterisk *). The taxonomic level of classification is indicated by k, p, c, o, f and g for kingdom, phylum, class, order, family and genus, respectively. Error bars indicate 1 SD.
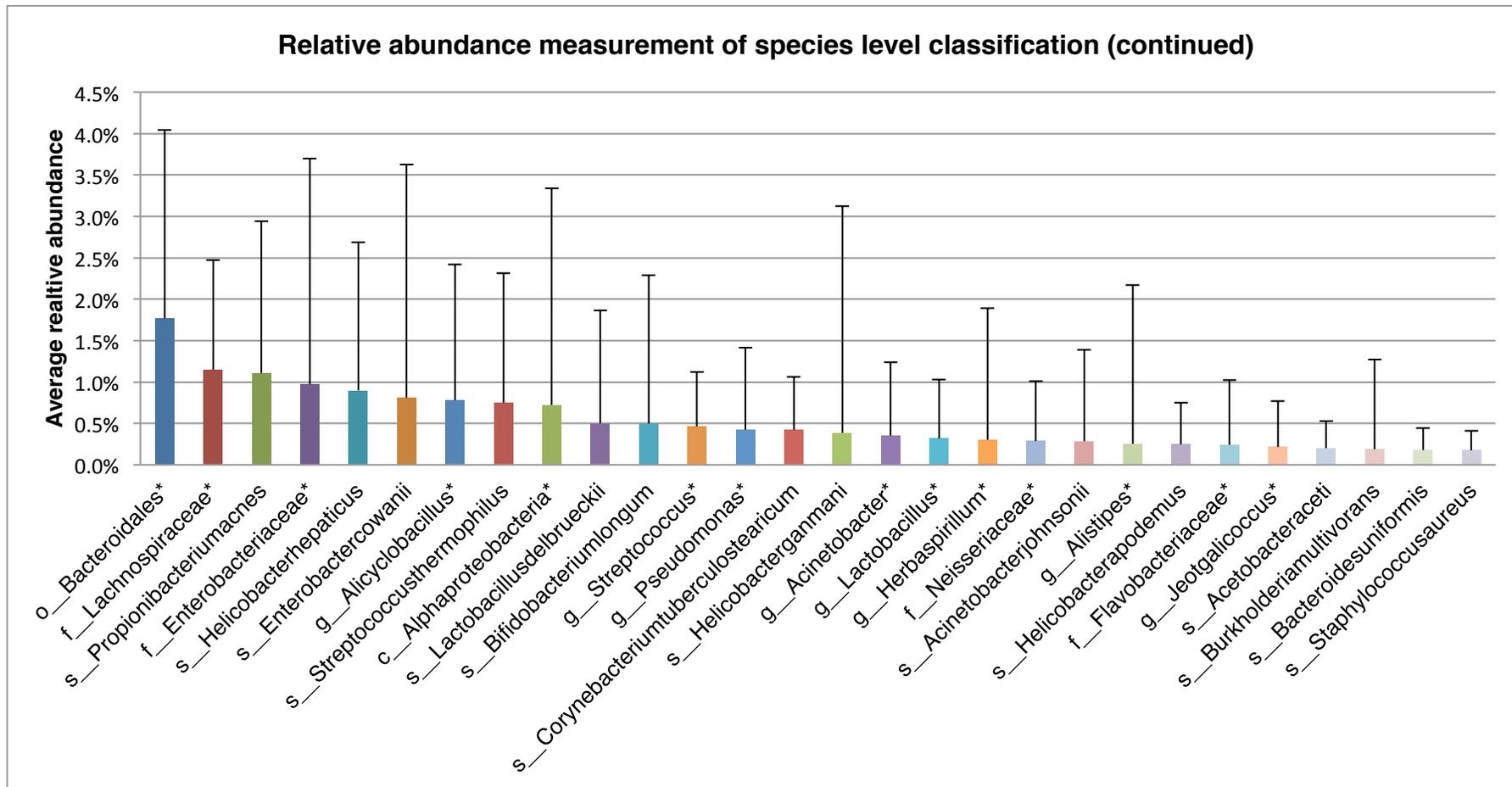
Figure 3.4: Relative abundance of bacterial species those with common classification species OTUs were grouped.
Relative abundance of species OTUs less than 0.1% is not shown here. For those unclassified at the species level, the next highest taxonomic level is shown (as marked in asterisk *). The taxonomic level of classification is indicated by k, p, c, o, f and g for kingdom, phylum, class, order, family and genus, respectively. Error bars indicate 1 SD.

3.1.1 Microbial communities in healthy and EBA afflicted individuals

Mice from G_4 AIL in these experiments contain genetic material from parental strains that are susceptible to an immunization-induced model of EBA [82], of which a total of 183 out of 261 were autoimmunized. Thus, in order to gain insight into the role of the skin microbiota in the pathogenesis of this inflammatory disorder, the bacterial diversity was compared within and between healthy mice (n = 119) and those that developed EBA (n = 64). At phylum level, three major phyla including the *Proteobacteria* (22% Healthy; 17% EBA; P-value = 0.01, ANOVA F value = 10.463), *Actinobacteria* (12% Healthy; 8% EBA; P-value = 0.02, ANOVA F value = 9.7776) and *Bacteroidetes* (7% Healthy; 6% EBA; P-value = 0.001, ANOVA F value = 10.926) were more abundant in healthy samples compared to those that developed EBA (**Figure 3.5**). In *Firmicutes* phyla division extreme outliers of samples belonging to EBA category drive the overall abundance greater than healthy mice and also the one-way ANOVA test was not statistically significant at P<0.05.
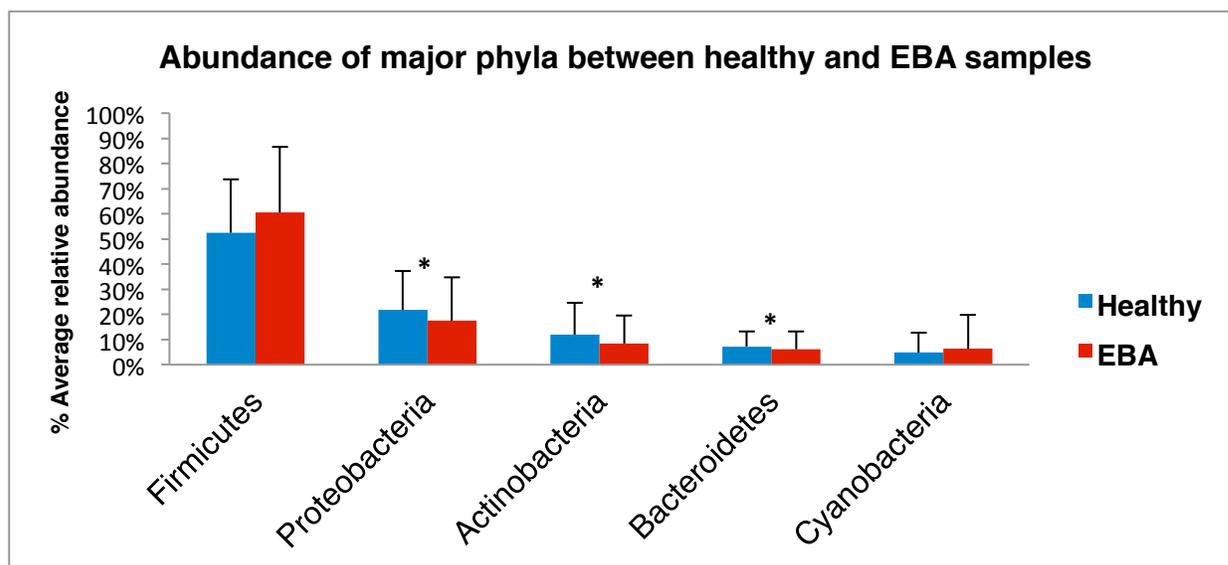
Figure 3.5: Relative abundance of major phyla between healthy and EBA samples.
Asterisk represent statistically significant P-values < 0.05 using one-way ANOVA test. Error bars indicate1 SD.
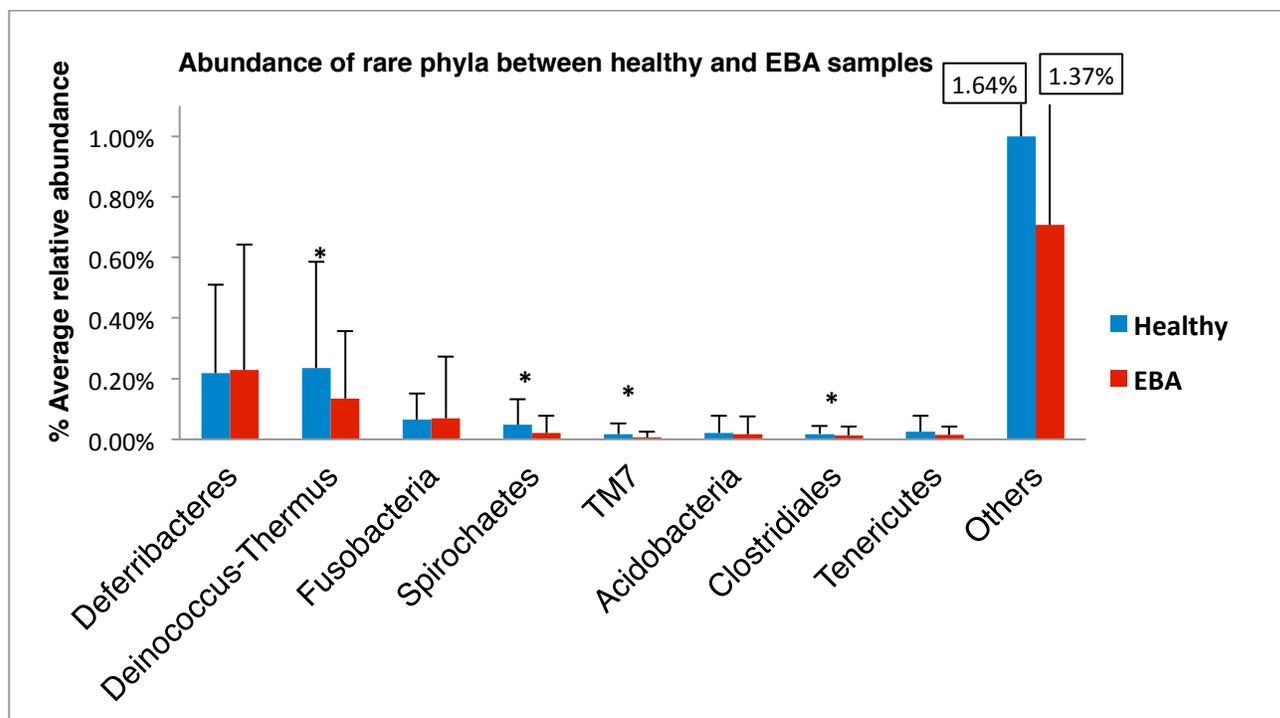
Figure 3.6: Relative abundance of rare phyla between healthy and EBA samples.
Asterisk represent statistically significant P-values < 0.05 using one-way ANOVA test. Error bars
indicate 1 SD.

Four other phyla that have the relative abundance less than 1% showed significant (P-value < 0.05,

one-way ANOVA test) abundance difference between healthy and EBA samples and they are

Deinococcus-Thermus, *Spriochaetes*, *TM7*, and *clostridiales* (**Figure 3.6**). Other Phyla that have the

relative abundance less than 0.01% were not included in the analysis but their abundances were

clubbed together to show the differences between the healthy and EBA mice.

3.1.1.1 Alpha diversity

To characterize the level and pattern of diversity within individuals, different measures of alpha

diversity was applied, which focus on species richness, evenness and abundance. These alpha

diversity measurements were applied to species level OTUs distributions. Each measurement

calculation was described in detail in the method section of this thesis. Rarefaction analysis was

simultaneously performed and the average within diversity measurements at each level was plotted.

These individual plots for different measurements not only reveal the highest detectable species richness and diversity based on sequencing depth but also show the measurement difference within the healthy and EBA samples. The Chao1 index was higher in the healthy individuals compared to those afflicted with EBA (**Figure 3.7**) (Wilcoxon rank sum test, W = 170, P-value = 0.005). The same pattern was also observed for Faith's[14] Phylogenetic Diversity index (PD whole tree) (Wilcoxon rank sum test, W = 176, P-value = 0.008) shown in **Appendix A Figure A.1**, the observed number of species (Wilcoxon rank sum test, W = 212, P-value = 0.05) as shown in **Appendix A Figure A.2** and the Shannon evenness measure (Wilcoxon signed-rank Test, Z = -4.3726, P-value < 0.001) (**Figure 3.8**). All four alpha diversity estimations revealed that the healthy samples contained the most diverse microbial community compared to the EBA samples. The rarefaction curves from 'Chao1', 'PD whole tree' and 'Observed species' strongly suggesting that more sampling depth would reveal a greater degree of species diversity and richness in the samples. The estimate shows that there is a minimum of 20-25 more species OTUs (mostly rare species) for every addition of 1000 sequences per animal to be discovered. There is no single study that claims to have sampled to completion and this study has covered large members of bacterial species across 261 samples.

The observed species measurement was later applied to samples belonging to different EBA disease severity scores. These scores were divided in to 3 categories i.e., low, moderate, and severe EBA. To know the detailed EBA scores for each disease severity category refer method section. Disease samples based on the EBA disease score were separated and measured for within species diversity measurement among the samples. The results revealed that bacterial diversity decreases with disease progression from healthy, low, moderate, and severe EBA categories respectively (Wilcoxon test for each pair, Healthy vs. Low: W = 357, P-value = 0.4; Healthy vs. Moderate: W = 446, P-value = 0.009; Healthy vs. High: W = 482, P-value = 0.0007) (**Figure 3.9**).
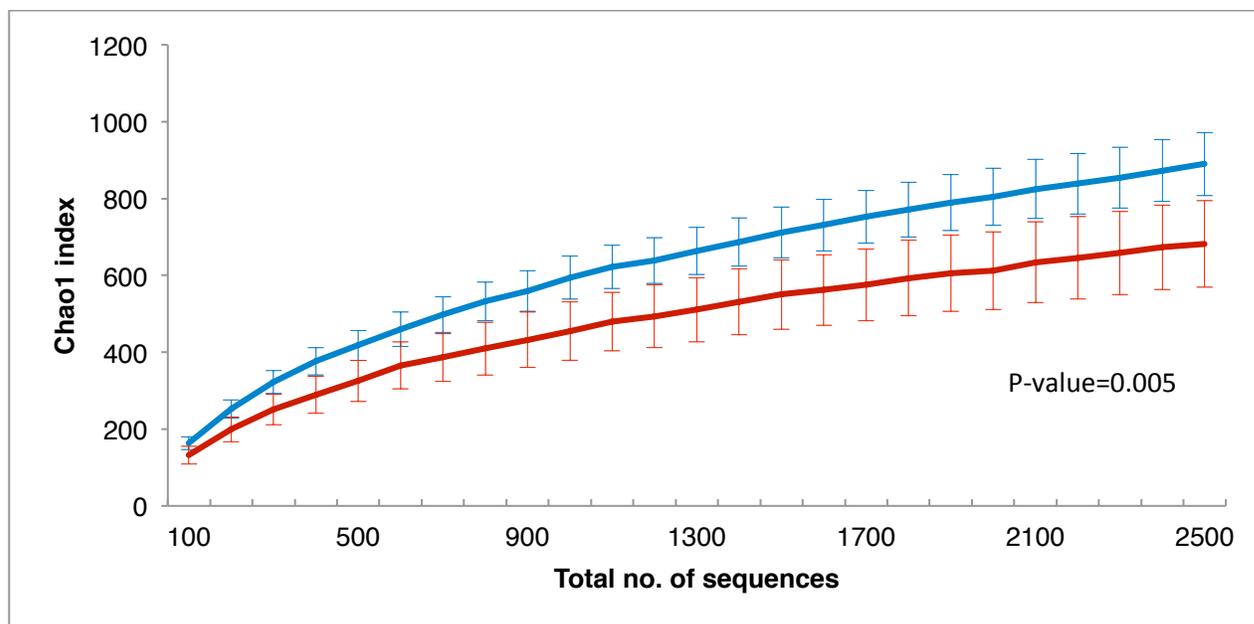
Figure 3.7: Chao1 alpha diversity measurement for species level OTUs in healthy (n = 119 in blue) and EBA (n = 64 in red) samples
Error bars represent the 95% confidence interval. P-value was determined by the Wilcoxon rank sum test in R.
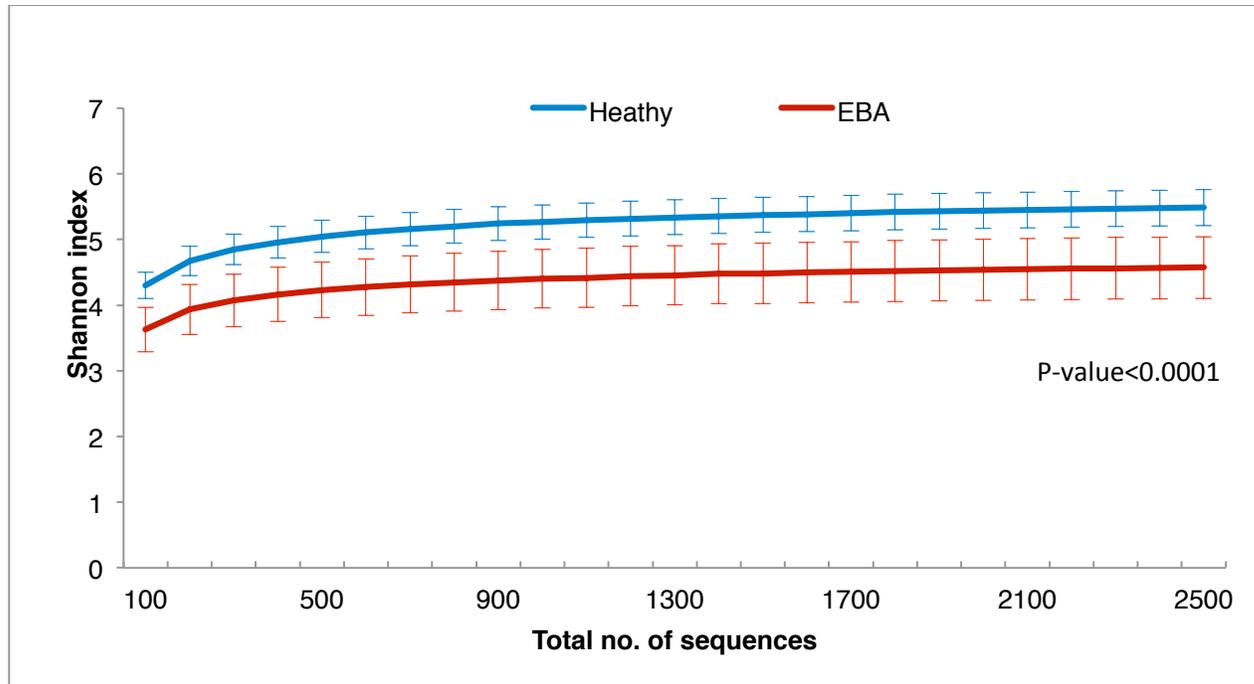
Figure 3.8: Shannon index alpha diversity measurement for species level OTUs in healthy (n = 119 in blue) and EBA (n = 64 in red) samples
Error bars represent the 95% confidence interval. P-value was determined by the Wilcoxon signed rank test in R.
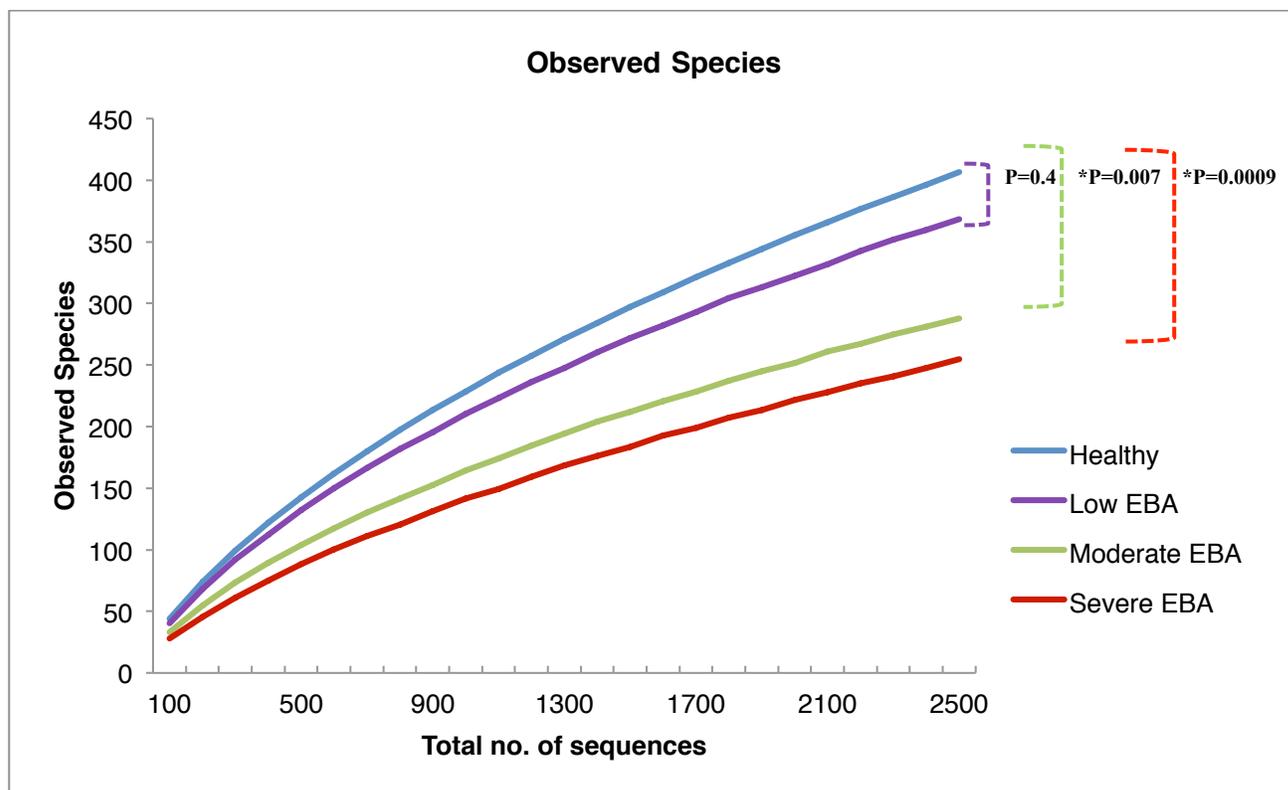
Figure 3.9: Observed species rarefaction curves for healthy (n = 119) and varying EBA clinical scores (Low EBA n = 43; Moderate EBA n = 9; Severe EBA n = 12).
P-values are calculated using the Wilcoxon rank sum test and significant P-values<0.05 are marked using asterisk (*).

3.1.1.2 Beta diversity

To analyze bacterial community composition and structure *between* individuals (*i.e.* beta diversity,) the weighted and unweighted Unifrac metric [116], [126] were used. This analysis was done on species level OTUs. Both the metrics are phylogenetic based measure weighted by taxon abundance and based on presence-absence information, respectively. In this diversity measurement a pairwise UniFrac distances are calculated between samples. UniFrac distances are defined as the fraction of branch length in a phylogenetic tree that is unique to sample pairs. Low UniFrac scores are result of common phylogenetic tree and dissimilar microbial communities results in high UniFrac scores. The calculated UniFrac scores were projected in a distance matrix and later the distance

matrix was visualized in a three-dimensional space using constrained Principal component (or Coordinate) Analysis (PCoA). Constrained PCoA was analyzed using healthy and disease status as a constraint factor (see **Method** section for details). This revealed significant variation between bacterial communities in healthy and EBA samples (Adonis, weighted UniFrac: $R^2 = 0.01922$, P-value = 0.008; unweighted UniFrac: $R^2 = 0.018$, P-value = 0.001). Constrained PCoA using EBA status as an explanatory variable and the weighted Unifrac metric as a response variable also revealed a small, but significant effect (P-value = 0.015), with the first principal coordinate axis explaining 1.794% of the variation between individuals (**Appendix A Figure A.3**). Similarly, the first principal coordinate axis of the unweighted Unifrac metric explains 2% of the significant variation (P-value = 0.005, **Figure 3.10**).

Analysis of beta diversity using OTU-based approaches yielded very similar results. In this study two non-phylogenetic based measurements were implied. Bray-Curtis distance matrix is based on the quantitative abundance, which yielded similar results to weighted UniFrac metric. Adonis on Bray-Curtis index, $R^2 = 0.01848$, P-value = 0.004, with first principal component axis explaining 1.791% of the variation (P-value = 0.015) (**Appendix A Figure A.4**). Another measurement Jaccard index is based on the presence-absence of OTU species. Adonis on Jaccard index, $R^2 = 0.02544$, P-value = 0.001. First principal component axis of contrained PCoA using Jaccard distance matrix explains 1% of the variation (P-value = 0.001) (**Appendix A Figure A.5**). All beta diversity metric plots on the basis of species OTU taxonomic composition showed high correlation to health factor. Therefore, skin microbial communities have considerable role in health status and this results further motivated to look for individual species OTU driving these signals.
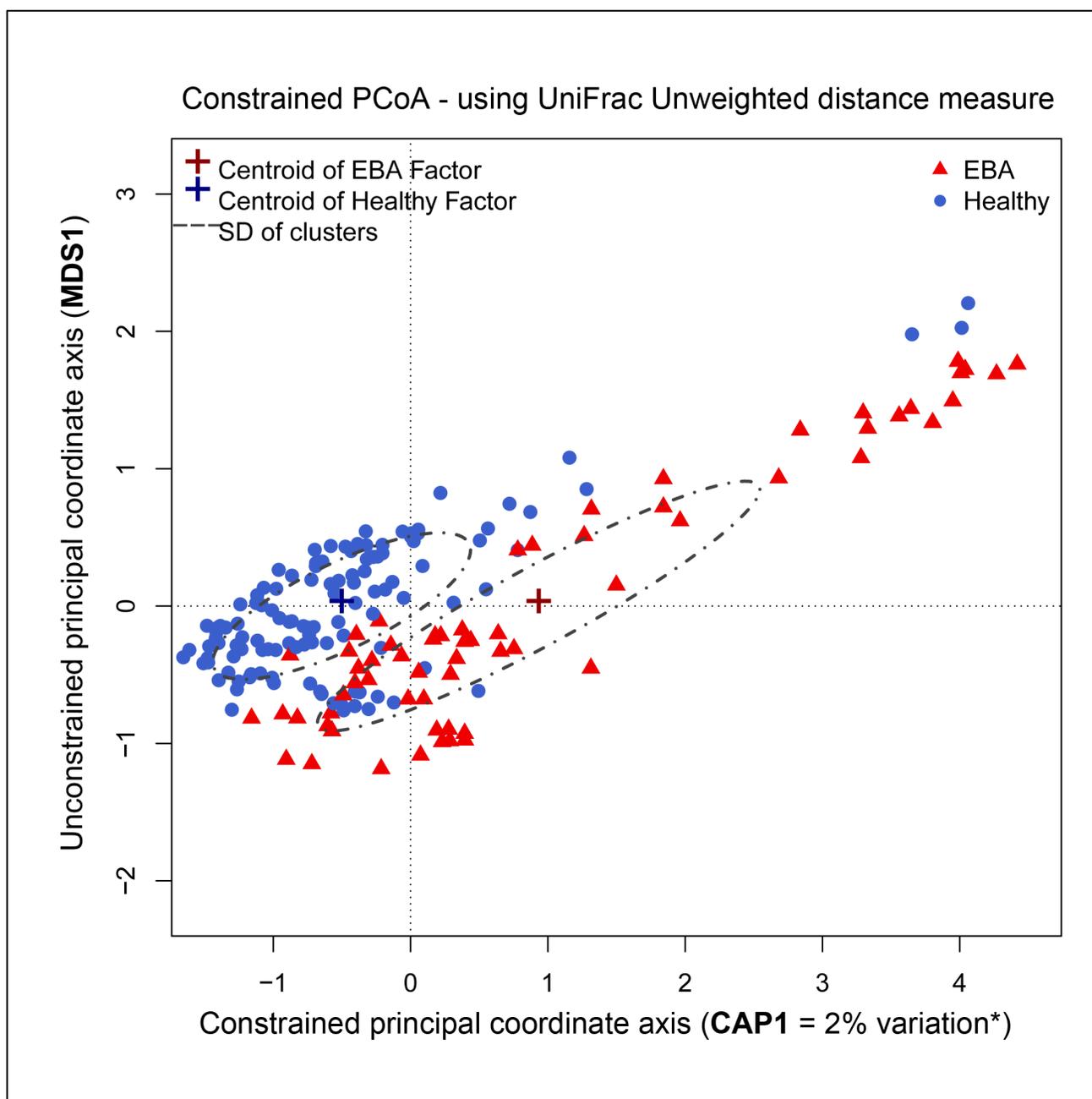
Figure 3.10: Constrained analysis of principal coordinates of the unweighted UniFrac metric using disease status as a constrained factor.
*P-value = 0.005

3.1.1.3 Indicator species analysis

The alpha and beta diversity analysis indicated that there are different community composition that is distinct from the healthy site to that of disease site group. Now it is further important to check

if individual species and/or species combinations are involved in driving the community composition among two different sites. Since there are many species involved in skin microbiota composition, it is highly important to account for multiple testing errors and validate the indicators.

In order to identify specific species OTUs that reliably distinguish between healthy and EBA-afflicted individuals, an "indicator species" analysis was performed. This method combines information of species OTU abundances in a particular category and it also calculates the probability of occurences by chance of a species OTU in a particular category. The assignment by this method is highly reliable as it is specifically designed to handle such data analysis. Interestingly, this revealed 39 OTUs more abundant in healthy individuals (Benjamini and Hochberg-adjusted [127] P-value $\leq$ 0.05, 1000 permutation), yet only a single OTU more abundant in EBA-afflicted individuals (**Appendix B Table B.1**). A rondomly chosen 10 out of 39 OTUs abundances are plotted between healthy and diseased individuals in **Figure 3.11** for visualization. Most of the OTUs associated with healthy individuals belong to the genera *Corynebacterium* and *Staphylococcus*, as well as several classified only to the order *Bacteroidales*. The single OTU associated with EBA also belonged to the genus *Staphylococcus* and most closely matched to *S. equorum*.
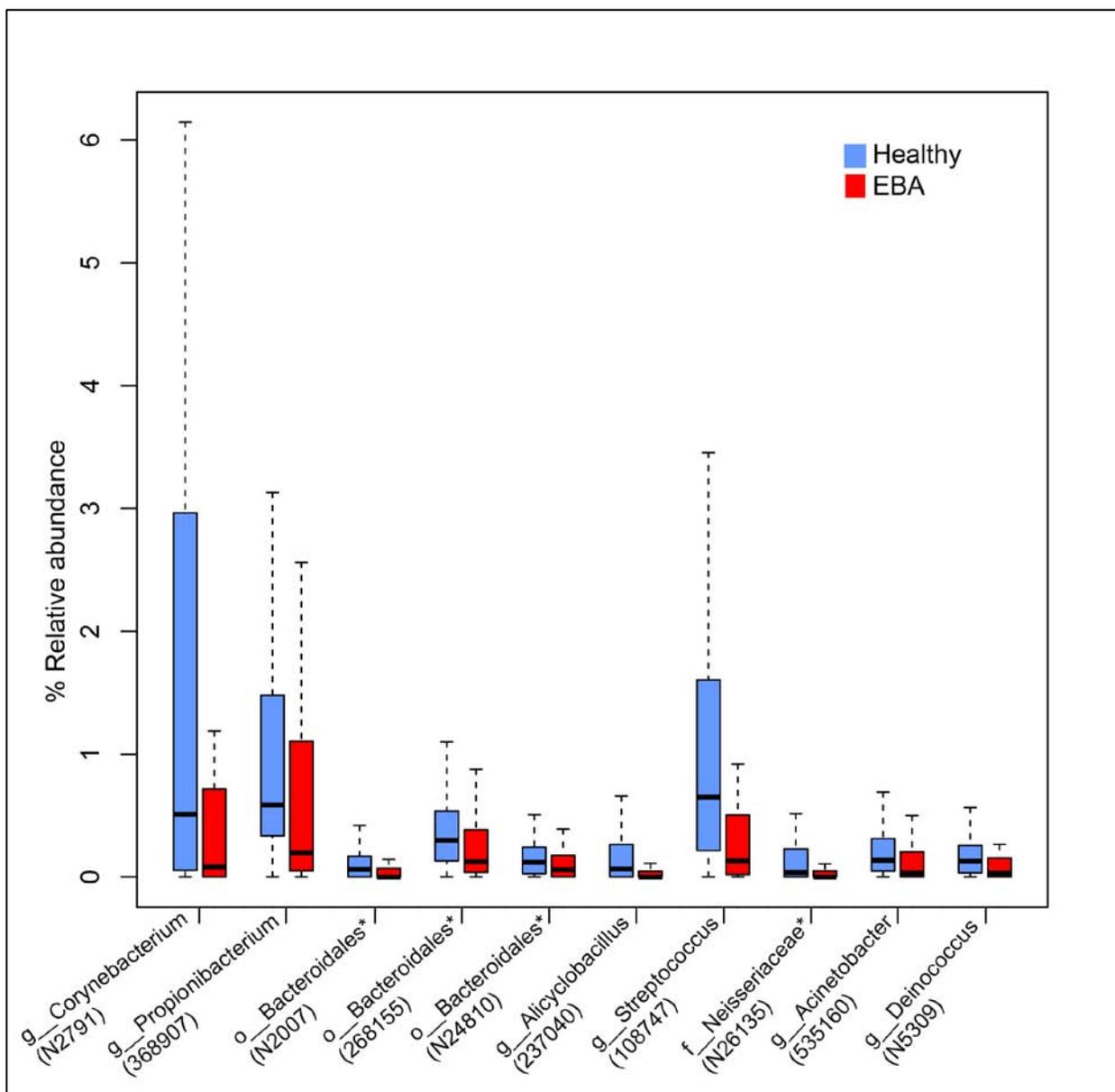
Figure 3.11: Relative abundances of randomly selected 10 species level OTUs out of 39 OTUs identified by indicator species analysis with respect to disease status. Further OTUs, P-values and details are shown in Appendix B Table B.1. For those OTUs unclassified at the genus level, the next highest taxonomic for which classification was possible is indicated by an asterisk. The taxonomic level of classification is indicated by k, p, c, o, f and g for kingdom, phylum, class, order, family and genus, respectively. OTU Ids are indicated in the parentheses. Box plots display the median, upper and lower quartiles, with whiskers denoting the maximum and minimum values within 1.5× of the interquantile range (IQR). Outliers are not shown.

3.2 Skin microbiota role in EBA susceptibility

Since cutaneous inflammation has been reported to be influenced by skin commensal microbiota [57], the skin microbiota was characterized to study its influence in immunization induced EBA disease. For this purpose, skin biopsies for microbial community profiling were collected before immunization of SJL/J mice. A classical barcoded 454 pyrosequencing of 16S rRNA gene approach was used to quantify the bacterial composition in 20 randomly selected mice, representing the 20/80% distribution of clinical health and overt blistering. In total, 16 phyla level bacteria in skin were found, which is dominated by *Proteobacteria*, *Firmicutes* and *Actinobacteria* (**Figure 3.12**) and at genus level by *Acinetobacter*, *Ralstonia*, *Streptophyta*, *Streptococcus* and *Staphylococcus* bacteria (**Figure 3.13**).

Additionally classical ecological measurements were applied to evaluate potential microbial compositional difference between the ultimately healthy and diseased mice before immunization. By applying alpha diversity measurements including the Chao1 and Shannon indices, greater bacterial species richness and evenness were found in the mice that did not develop disease symptoms compared to those that did display EBA symptoms after immunization (P-value = 0.03 for Chao1using Wilcoxon Rank sum test and P-value = 0.01 for Shannon using Wilcoxon signed rank test) (**Figure. 3.14** and **Figure 3.15**). Beta diversity measurements including the weighted UniFrac, unweighted UniFrac, Bray-Curtis and Jaccard distances were analyzed using constrained analysis of principle coordinates with the presence/absence of disease manifestation as the factor tested. However, no significant separation was identified with respect to this variable (**Figure 3.16**).
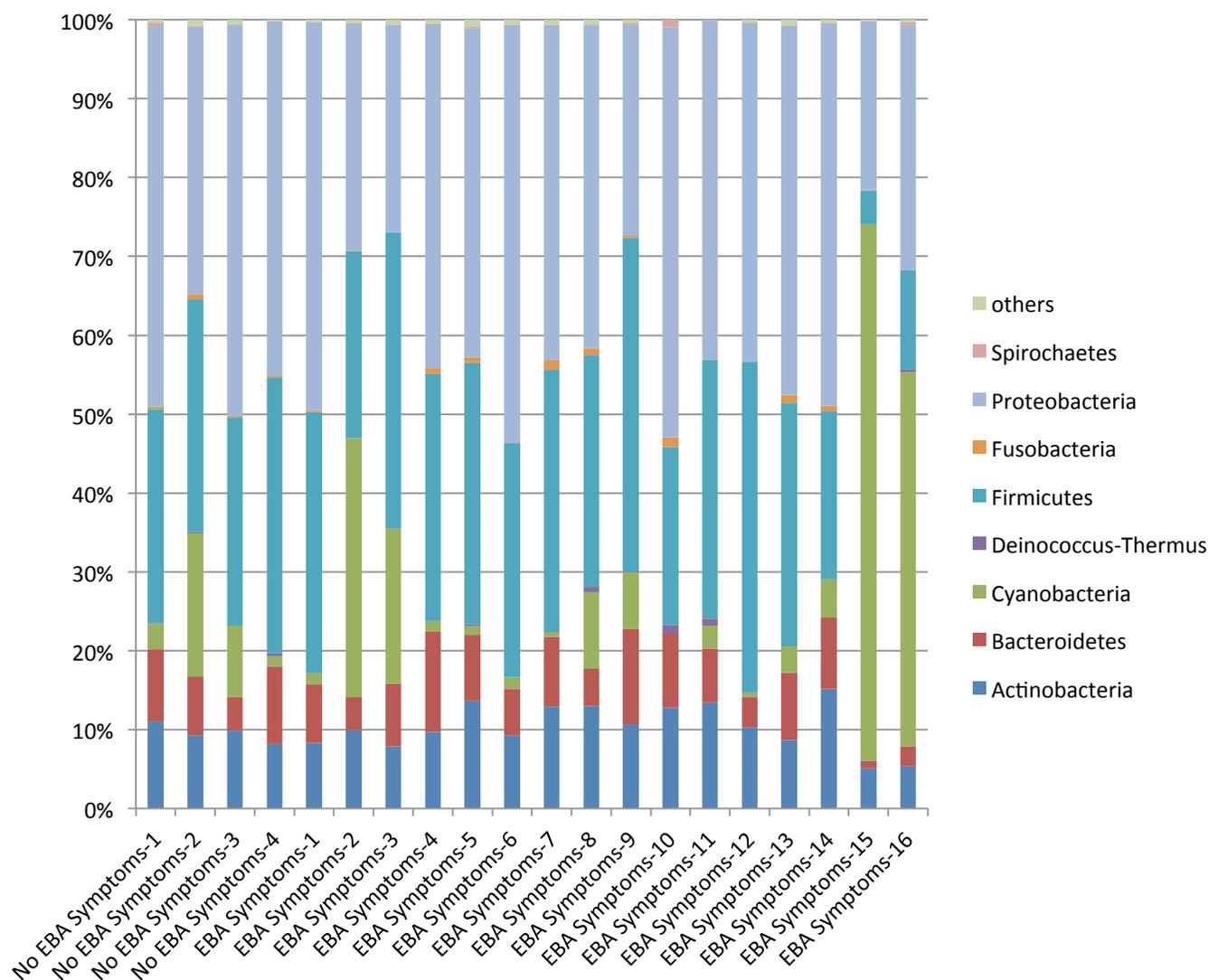
Figure 3.12: Major bacterial phyla in the mouse skin microbiota (SJL/J mouse strain)
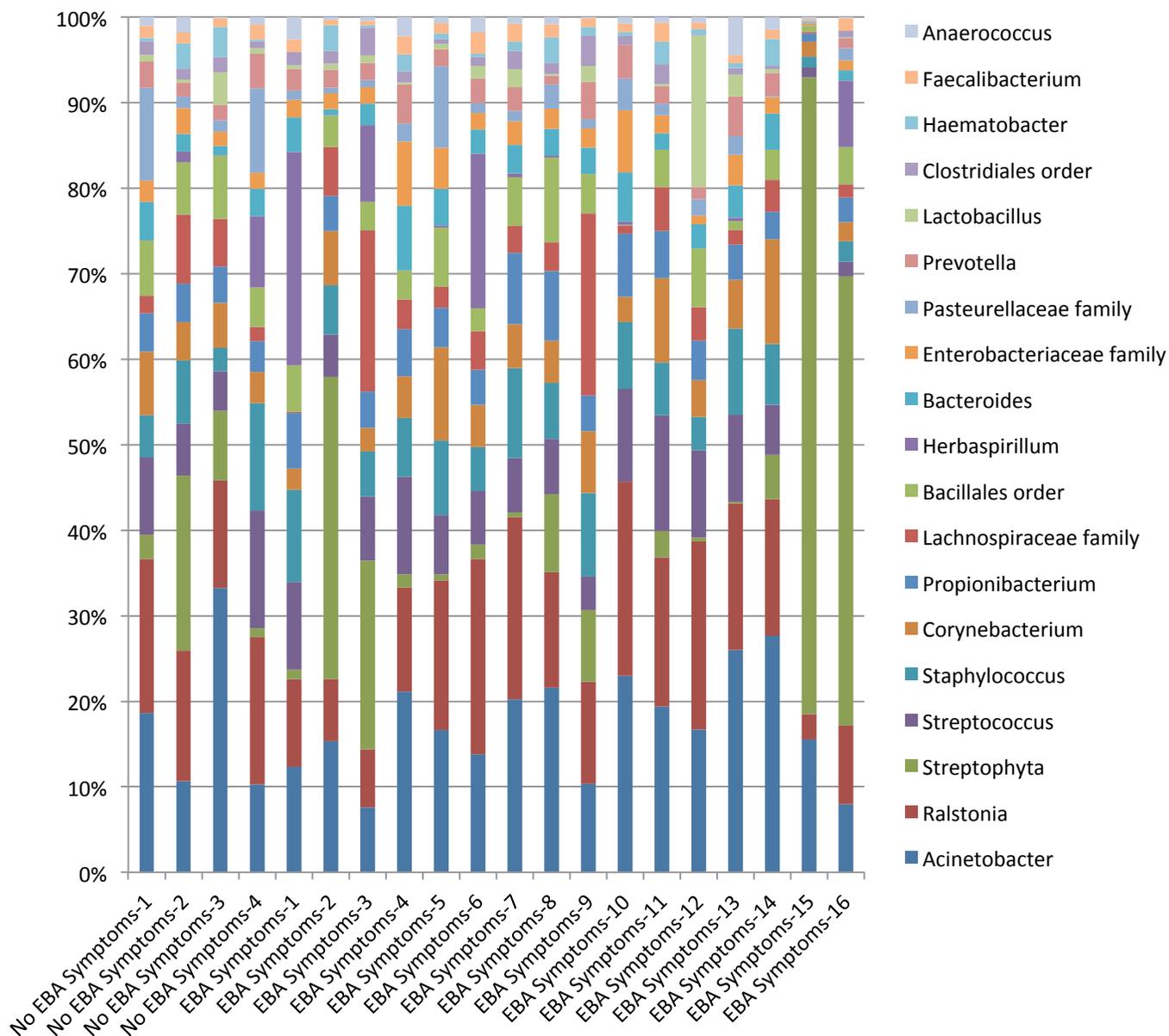
Figure 3.13: Major bacterial genera in the mouse skin microbiota (SJL/J mouse strain)
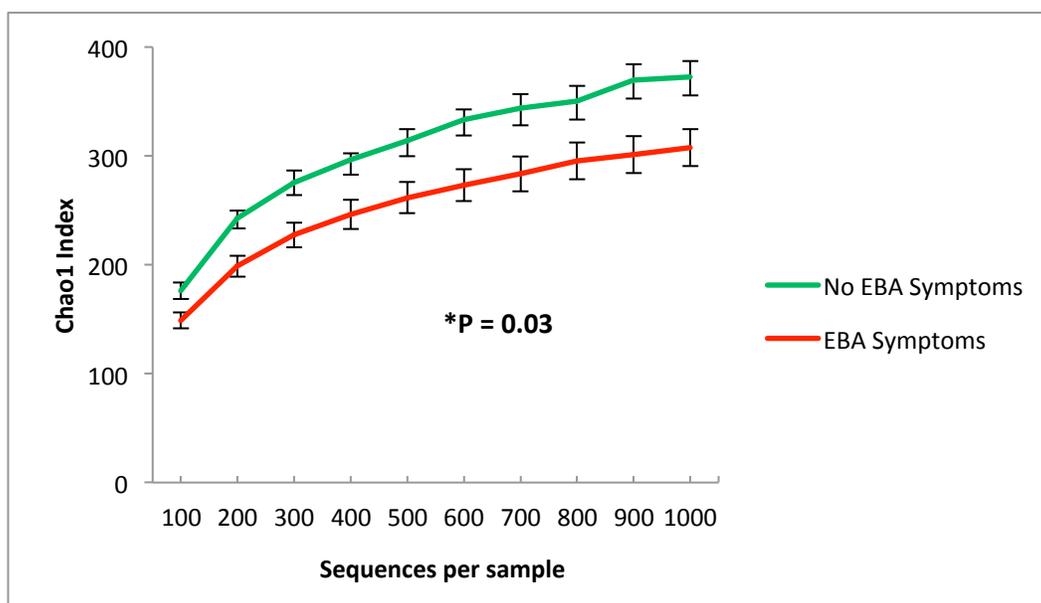
Figure 3.14: Chao1 alpha diversity measurement for species level OTUs in "No EBA Symptoms" (n = 4 in green) and "EBA Symptoms" (n = 16 in red) samples.
Error bars represent the 95% confidence interval. P-value was determined by the Wilcoxon rank sum test in R.
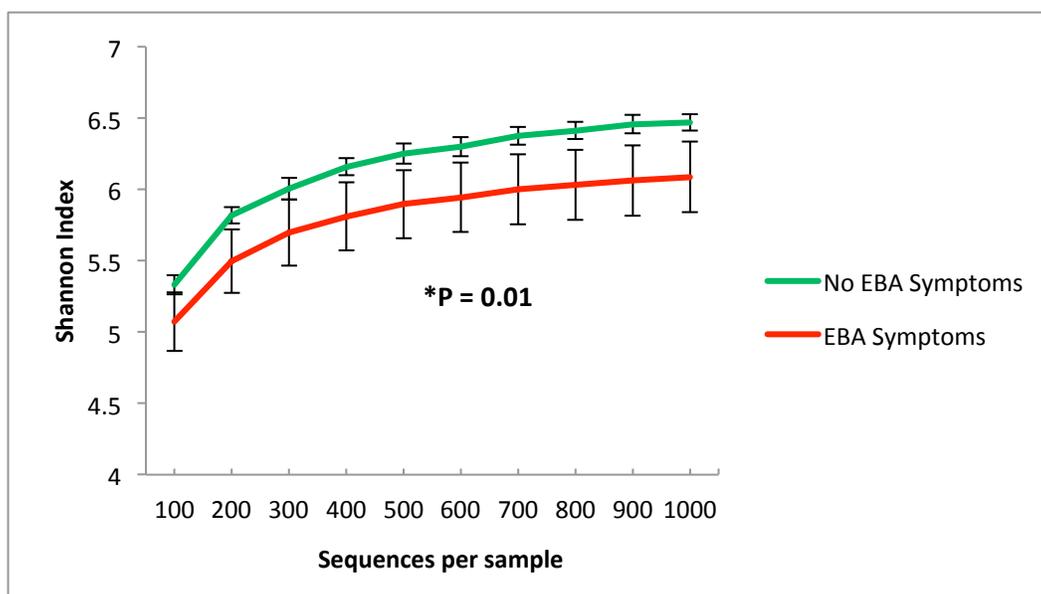
Figure 3.15: Shannon alpha diversity measurement for species level OTUs in "No EBA Symptoms" (n = 4 in green) and "EBA Symptoms" (n = 16 in red) samples.
Error bars represent the 95% confidence interval. P-value was determined by the Wilcoxon signed rank test in R.
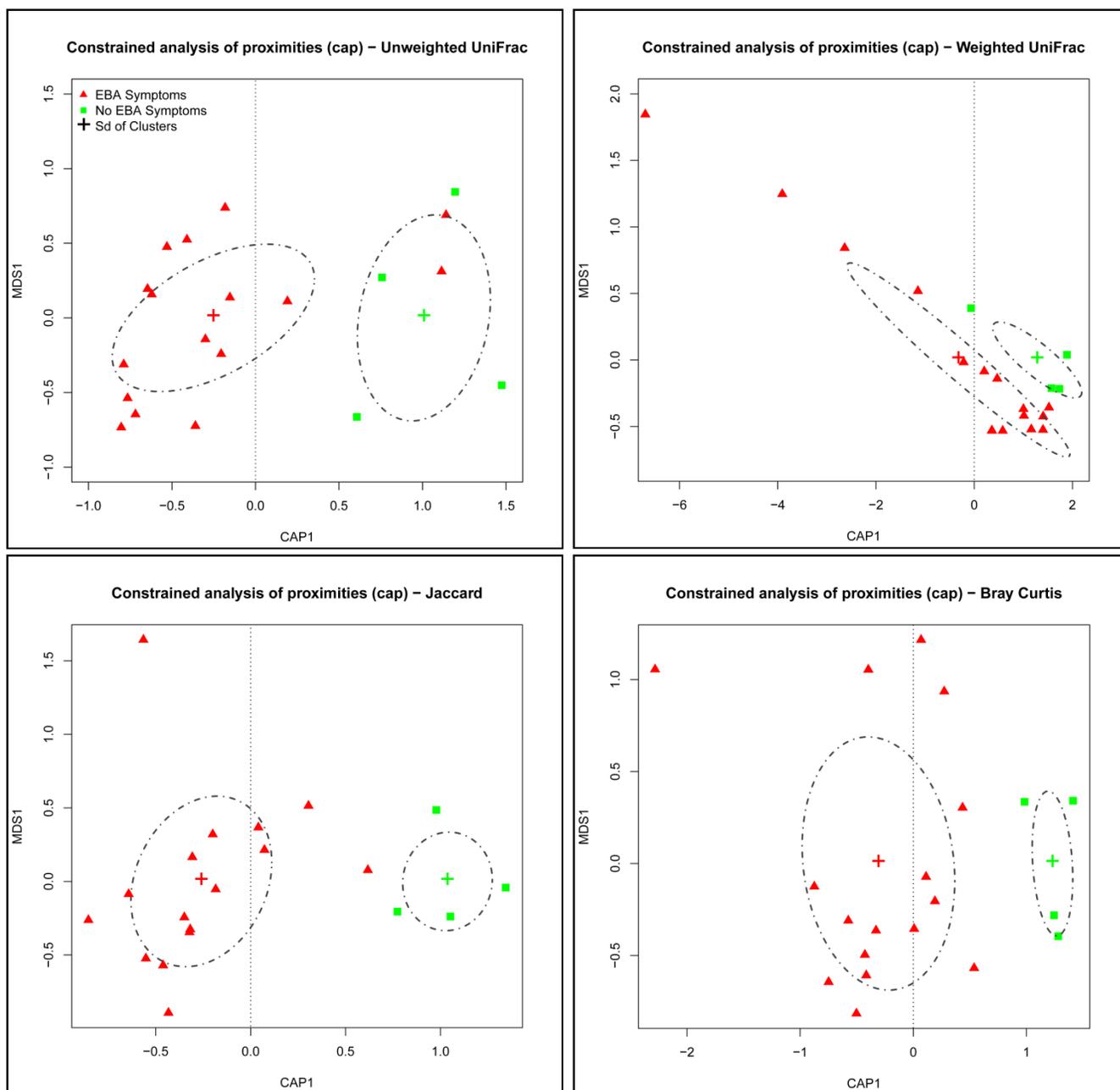
Figure 3.16: Constrained analysis of principle coordinates analysis of beta diversity metrics
No EBA Symptoms" (green) and "EBA Symptoms" (red) were used as constrained factors. None of
the first constrained (or CAP1) axes for any of the four distance measures is significant.

3.3 Factors influencing skin microbiota

In order to measure the contribution of host genetics on skin microbial diversity, it is very important to understand the various factors that can influence variation in the skin microbiota. There are very high chances that these environmental factors potentially constitute the noise that can mask host genetic effects. Many factors including cages where mice were kept, family, sex, age, and weight were carefully noted for each generation in AIL. Linear mixed model statistics were implemented to test the influence of each factor among skin microbiota traits and a statistical model was developed to efficient map the loci contributing to microbiota trait variation simultaneously controlling for environmental factors. From **Appendix A Figure A.1**, it can be seen that there are huge differences in abundances among 261 mice across species OTUs. For further analysis, 131 OTUs belonging to the "core measurable microbiota" (CMM) were identified (**Appendix A Figure A.8**) and it was determined in a manner similar to that of a paper published by Benson and associates[27]. In other words these 131 OTUs are indeed quantitatively reproducible data and this experiment also directly controls for PCR, sequencing, primer and adapter errors/influence in this study (**Appendix A Figure A.9**). It can be also seen from **Appendix A Figure A.2** that 131 CMM species OTUs cover roughly 80% of total sequence from all 261 animals and hence, with high confidence it is now possible to claim that these 131 CMM species OTUs are indeed a major representation of skin microbiota composition. Mixed model analysis showed that only two factors (cage and family) out of measured factors significantly influenced the variation among skin bacterial communities at species OTU level.

3.3.1 Cage

One of the common factors that can have a deep influence on the skin microbiota variation is the cage environment. Mice sharing common cages also share common microbiota compare to that of mice from different cages [27], [128]. 261 animals from AIL were separated in to 104 cages. A

linear mixed model was used to determine to which extent the cage environment causes the skin bacterial phenotype variation in each of 131 CMM traits (see **Methods**). The result showed that the cage directly accounting for 28% of overall variation in CMM species abundance. The individual cage variation between taxa is shown in detail in **Appendix B Table B.2**.

3.3.2 Family

Littermates reared within the same cage at weaning can influence skin microbiota community composition. It also important to account for the family structure in segregating populations created using a multiple generation breeding methods for Quantitative Trait Loci (QTL) mapping. In this study siblings of the same sex were cohoused together in separate cages after weaning. Also depending on the number per cage, females from multiple families were also cohoused together. 261 mice in this study originate from 41 different families. Accounting for the family structure on each CMM trait using linear mixed model showed an overall variation of 3 percent. A detailed variation by family factor on each CMM trait variation is shown in detail in **Appendix B Table B.2**.

3.4 QTL analysis of skin microbiota

To measure the genetic contribution, CMM abundances (131 species OTUs and 149 phylum-genus level traits) were tested for co-segregation against 1199 informative SNP markers after accounting for cage and family effects (see **Methods)**. This revealed host genetics to have significant control over members of the skin microbiota, which can be seen in **Figure 3.17**. Nine out of 131 CMM OTUs were associated with three significant and six suggestive (see **Methods**) species-level OTU QTLs, hereafter referred to as "spQTLs" (**Appendix B Table B.3**). The phenotypic variance explained by the spQTLs ranged from 2.8 to 11.31% (**Appendix B Table B.3**). The highest ANOVA –log P of 5.24 was given by an OTU belonging to the genus *Staphylococcus* on chromosome 18 (spQTL 7), with a peak at SNP rs13483244 (position 21 Mb), which accounted for 5.44% of the phenotypic variance.
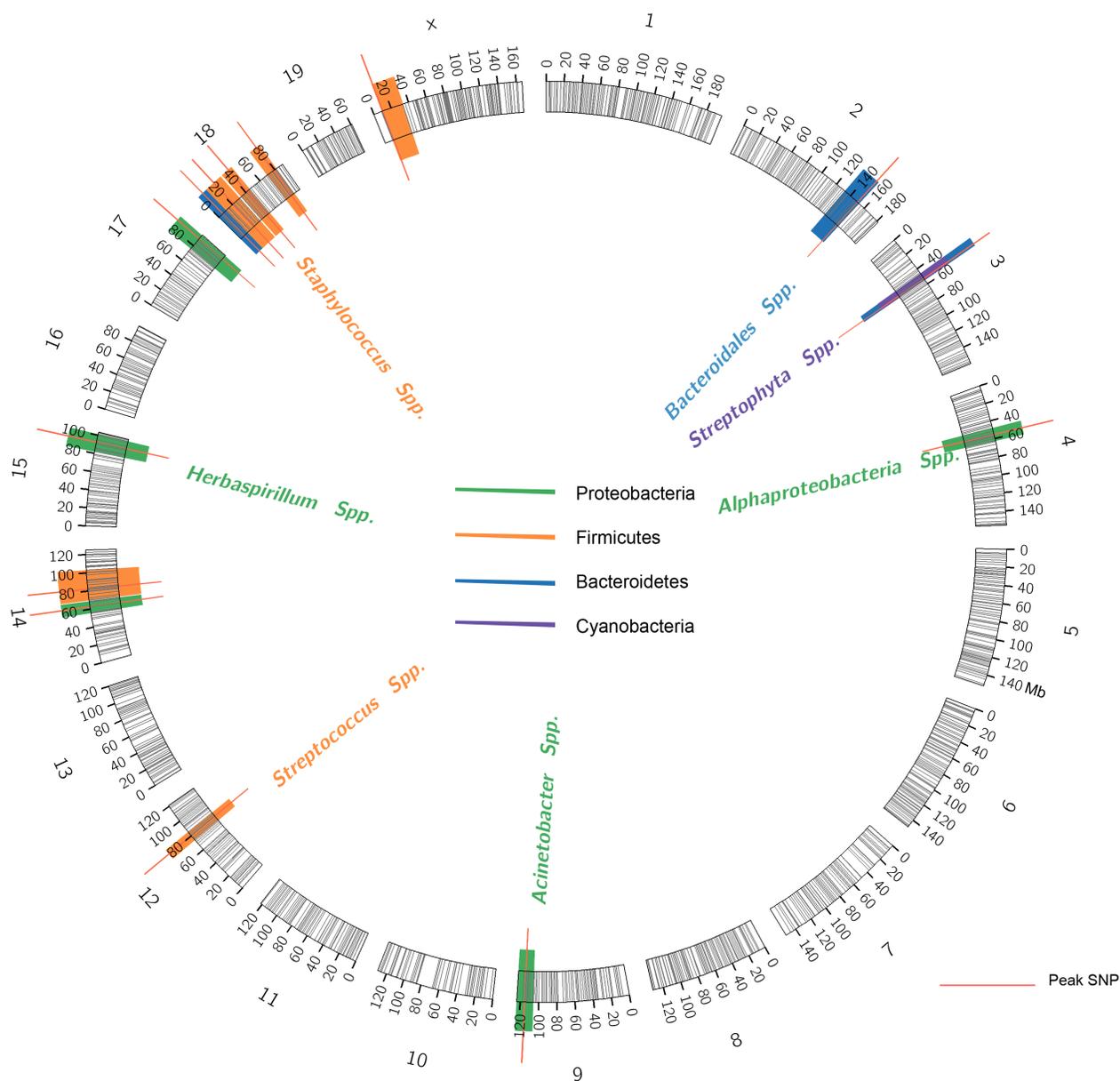
Figure 3.17: QTL mapping of skin microbiota
19 mouse autosomes and X chromosome are scaled along with 1199 SNPs indicated in black lines. QTL representing at species level OTUs are color coded according to their respective phyla classification and highest taxonomic classification of that OTU is written next to QTL color shade. Other genus to phylum level QTLs are detected according to their phylum classification and color shaded respectively.

Mapping at higher taxonomic levels including phylum, class, order, family and genus revealed a total of six QTLs, including three with one or more significant associations and three with suggestive associations, hereafter referred to as "gpQTLs" (**Appendix B Table B.4**). Two out of the nine spQTLs are contained within gpQTLs, thus, in total thirteen unique QTLs were identified (**Figure 3.17**). To gain further insight we compared our results to previously published QTL studies of the gut microbiota [27], [128] and revealed evidence of overlap greater than expected by chance (**Figure 3.18; Methods**). Interestingly, the confidence intervals of our spQTLs and gpQTLs contain five and four genes related to innate immunity, respectively (see **Discussion**, **Appendix B Table B.5**).

3.4.1 Effects of immunization on QTL mapping

Since the model of EBA used in this study is immunization-based, we also included 78 non-immunized mice to control for the effect of immunization in the QTL mapping. Accordingly, we analyzed a subset where both EBA-afflicted and non-immunized individuals were removed (*i.e.* only the 119 healthy, autoimmunized samples). Despite decreasing the sample size from 261 to 119, two out of nine spQTLs and two out of six gpQTLs were still detected (**Appendix B Table B.3**). Next, we analyzed a subset where the EBA afflicted mice were removed (*i.e.* including 119 healthy autoimmunized and 78 non-immunized samples). One out of nine spQTLs and two out of six gpQTLs were still detectable despite lowering the sample size from 261 to 197 (**Appendix B Table B.3 − B.4**). Thus, the presence of QTLs among subsamples not influenced by differences in disease/autoimmunization status supports the presence of true genetic effects. In order to further confirm this, immunization factor was included in the QTL mapping model and the mapping procedure was repeated for 280 traits. All previously found QTLs stayed significant. This strongly suggests that the QTLs found in this study for skin microbiota traits are indeed independent of the immunization.
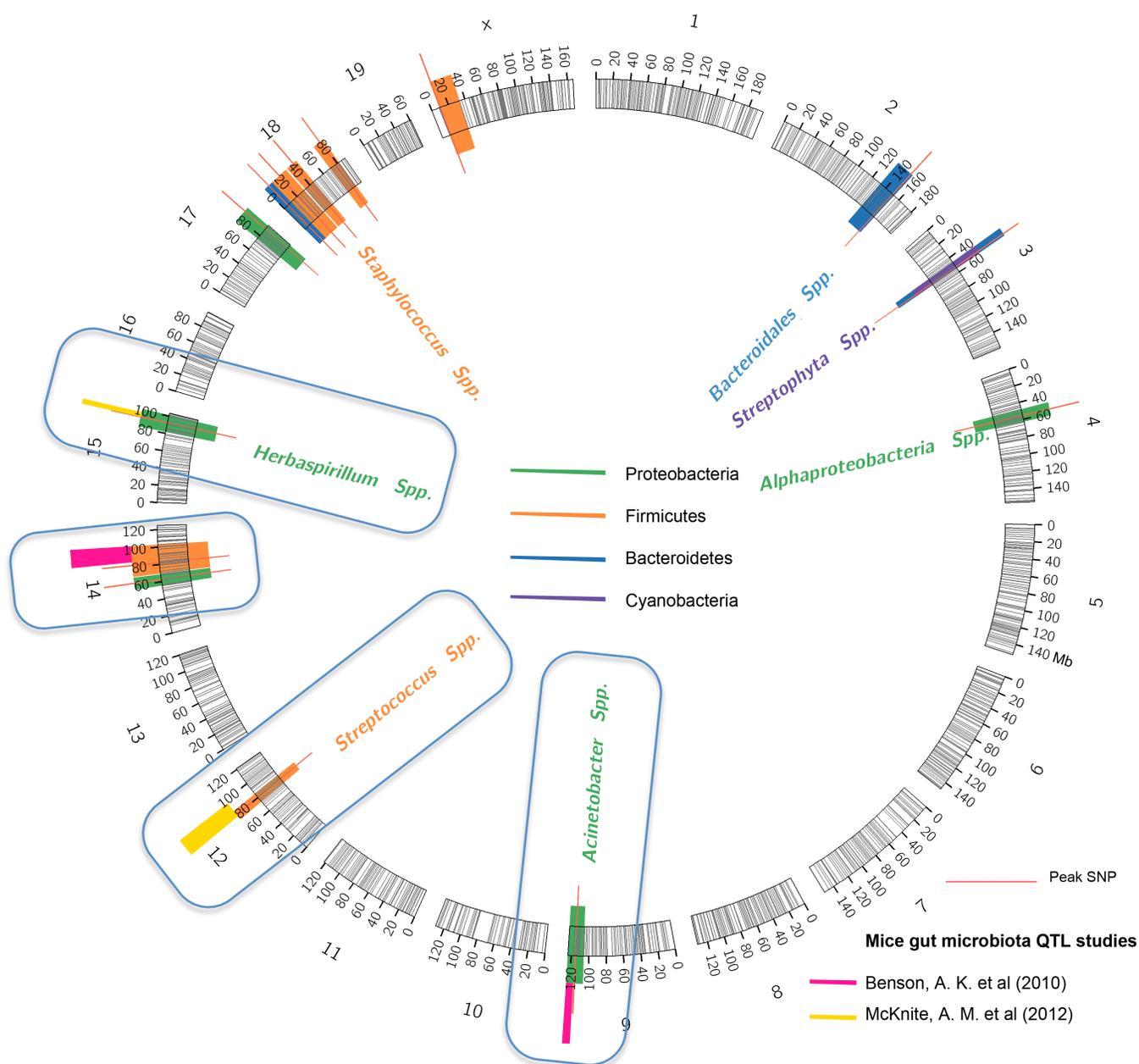
Figure 3.18: Overlapping of skin microbiota QTLs on published gut microbiota QTLs
Only those gut QTLs overlap on skin QTLs are shown above and color-coded respectively. 19 mouse autosomes and X chromosome are scaled along with 1199 SNPs indicated in black lines. QTL representing at species level OTUs are color coded according to their respective phyla classification and highest taxonomic classification of that OTU is written next to QTL color shade. Other genus to phylum level QTLs are detected according to their phylum classification and color shaded respectively.

3.5 Genetics association of EBA

To investigate the potential role of host genetically-based variation for EBA disease, first the subset of 183 immunized mice common to this and our previous study on EBA was re-analyzed [99]. This revealed no significant QTL for EBA presence/absence at an *E* value cutoff of < 0.1 (see **Methods**), likely due to the reduced number of animals (**Figure 3.19**). This previously published study of EBA by Ludwig and associates [99] did not look in to presence/absence disease as phenotypic trait.
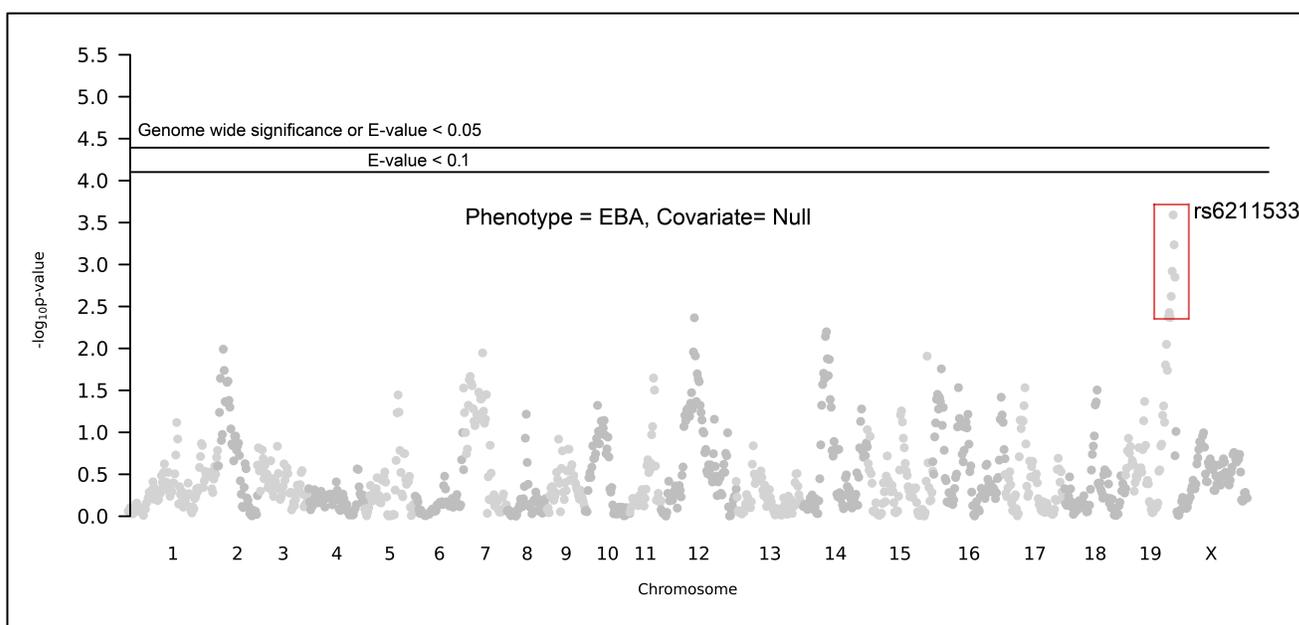
Figure 3.19: Manhattan plot of –log *P* values for each SNP tested against EBA disease phenotype (presence/absence)
Each dot represents 1199 SNPs in X-axis according to their position on each chromosome along with respective ANOVA –log *P* values in Y-axis.

3.6 Genetics and skin microbiota interaction

The results in this study reveal that the bacterial taxon abundance from mouse skin does display a clear genetic component. Covariate QTL analysis was applied in order to evaluate potential

interactions between bacterial species and disease susceptibility between each of the 131 CMM species abundances (as covariates) and EBA disease susceptibility (presence/absence of disease as primary phenotype) (**Figure 3.20**).
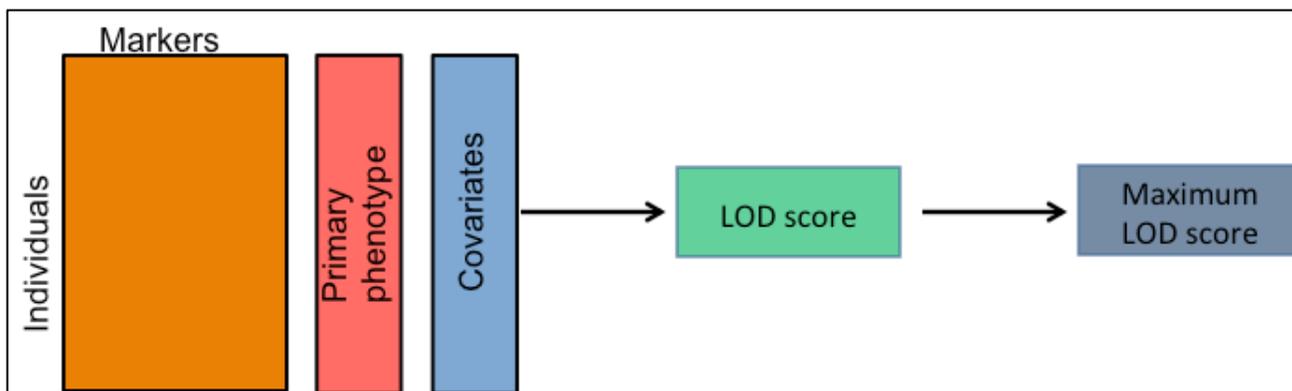
Figure 3.20: A graphical representation of covariate QTL mapping procedure.
Addition of covariates in the QTL analysis will result in reduced residual variation, which directly improves the power of QTL detection. This analysis directly reflects the extent of QTL x covariate interaction.
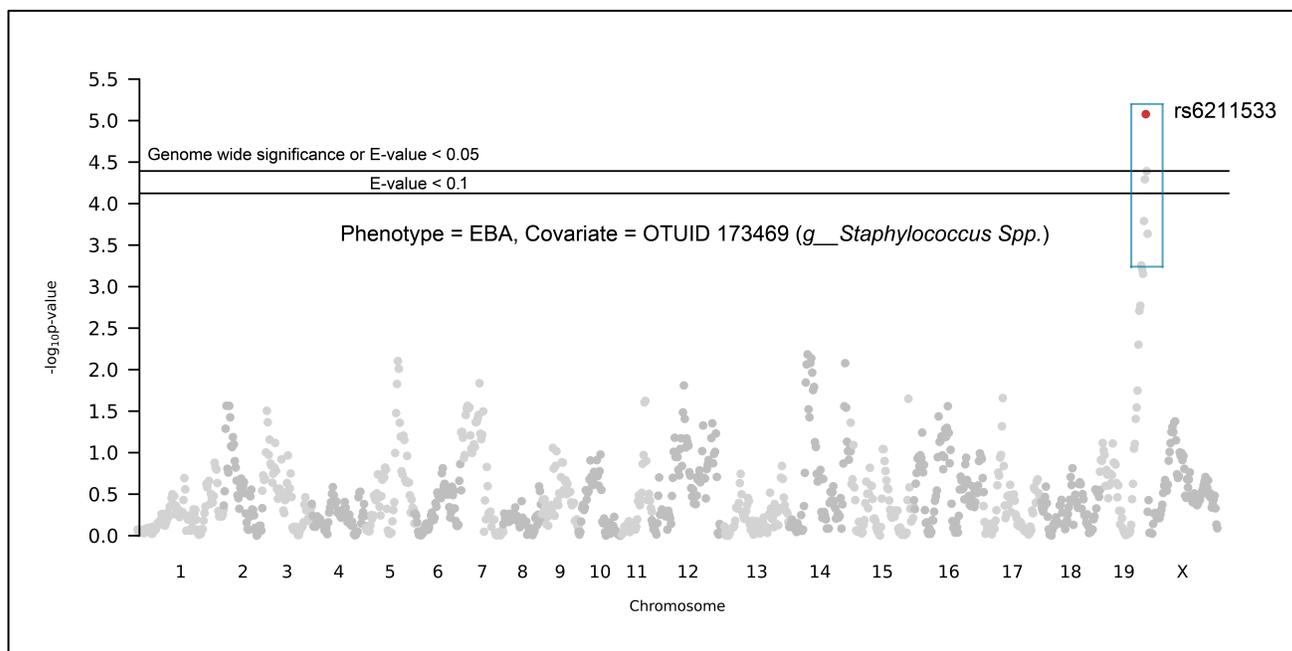
Figure 3.21: Manhattan plot showing –log *P* values for each SNP tested against EBA including *Staphylococcus spp.* (OTUID 173469) abundance as a covariate
SNPs with an E-value (Genome wide significance) < 0.05 are shown in red.

The covariate analysis revealed a significant covariation (E value < 0.1) involving 10 out of 131 taxa, which, intriguingly, increased the power of detecting EBA QTLs, as a novel locus (covariate QTL) (Chr.19, CI 53 – 60, peak at 57 Mb) was detected (**Figure 3.21**, **Appendix B Table B.6**). Two OTUs belonging to the genus *Staphylococcus* clearly display a gene-bacterial interaction (E value < 0.05) (**Figure 3.22, Appendix B Table B.6**).
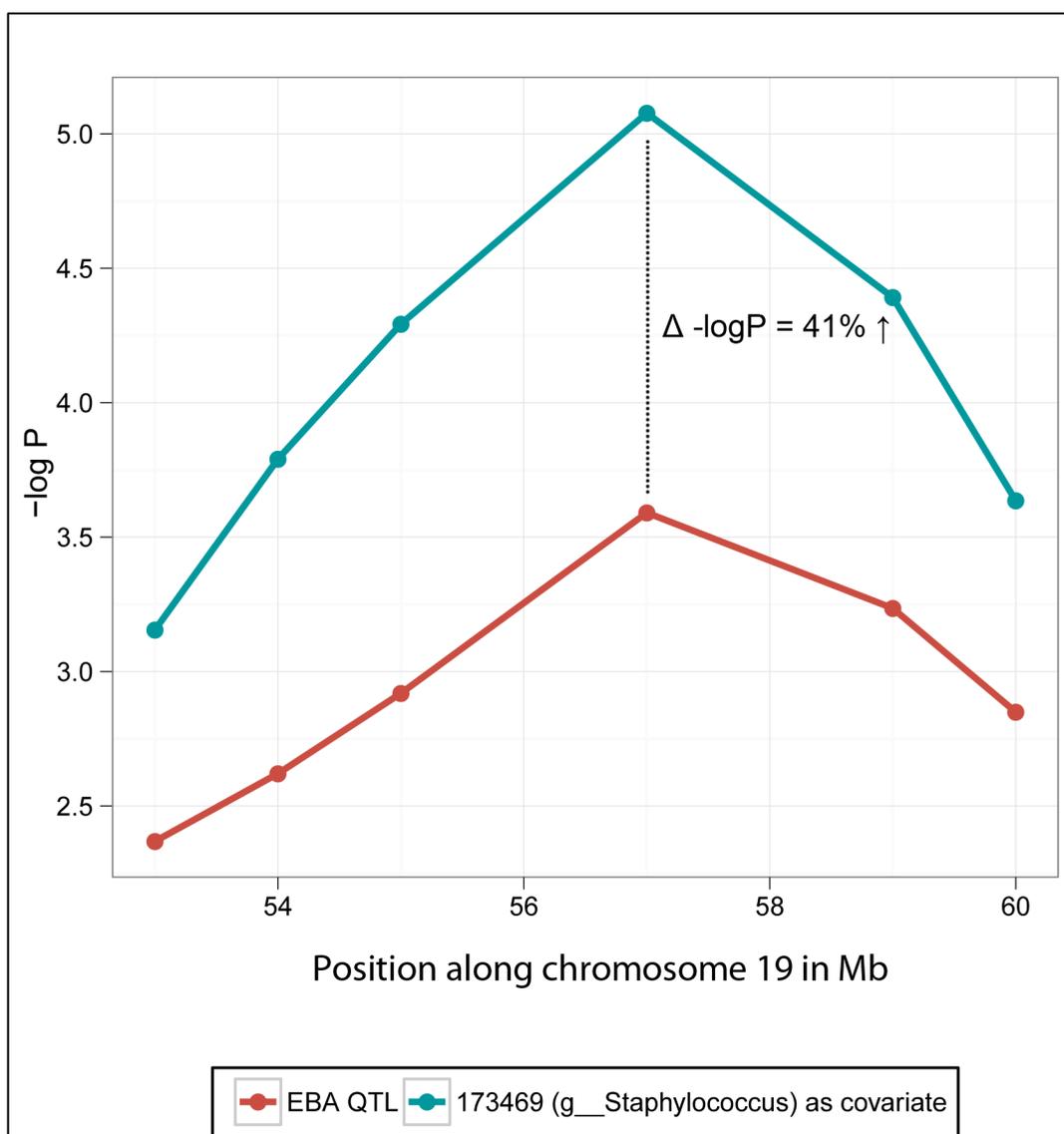
Figure 3.22: Portion of chromosome 19 containing the covariate QTL with a peak at SNP rs6211533
SNPs shown in this figure in red and blue lines are those SNPs marked in Figure 3.19 and Figure 3.21 highlighted in respective colors within respective figures.

To further characterize the nature of the identified covariate QTL, individuals were divided arbitrarily into "high" (top 50%) and "low" (bottom 50%) groups with respect to their individual OTU abundances and analyzed the proportion of individuals developing EBA with respect to host genotype. This revealed that for most cases the proportion of animals developing EBA was higher among the low OTU abundance category (n = 10; one of which was also significant by Fisher's exact test between these defined abundance categories (**Figure 3.23**); It also noted however, that all 10 taxa display significant covariation), suggesting a probiotic role (**Appendix B Table B.7**). While community-level alterations of the skin microbiota in the context of EBA are present (*e.g.* **Appendix A Figure A.3-A.5, Figure 3.11, and Figure 3.12**), it can be note that the putative probiotic covariate taxa identified here do not vary in abundance simply according to disease status. Namely, both healthy and diseased individuals are found among the low abundance categories, thus, low abundance of *e.g. Staphylococcus spp.* is not a simple byproduct of disease, but increases the probability of developing symptoms.

The large number of covariate bacterial taxa interacting with a single host locus suggests that individual bacterial taxa may not be acting independently. Thus, to identify potential interactions among covariate taxa we performed a pairwise correlation analysis (**Figure 3.24**). Indeed, this revealed significant positive correlations (Pearson correlation; P-value ≤ 0.05 after correction for multiple testing (Benjamini-Hochberg [127])) between many taxa, suggesting interactions between the host locus and bacterial species assemblages or individual driver species.
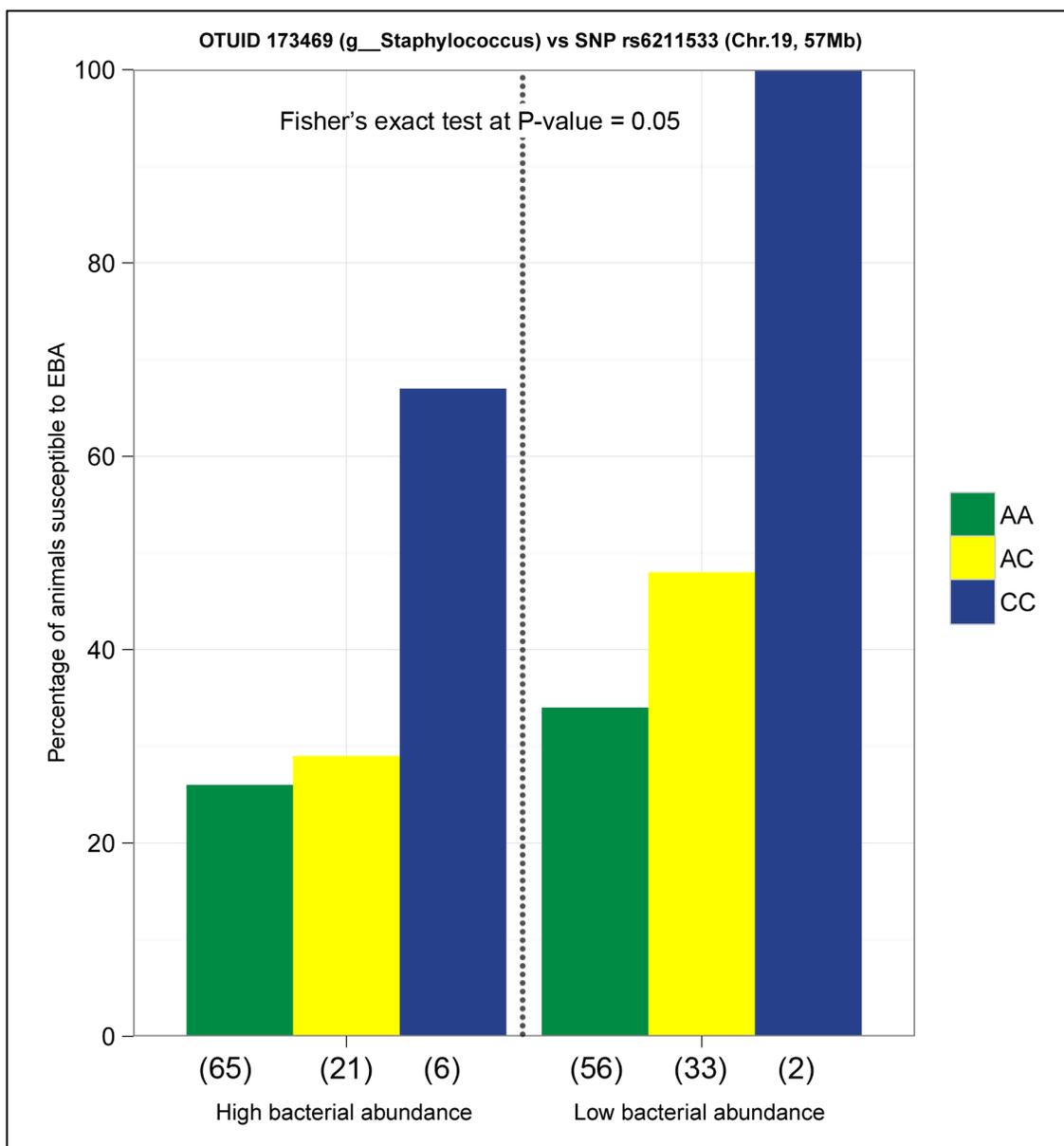
Figure 3.23: Percentage of animals developing EBA among high (top 50%) and low (bottom 50%) *Staphylococcus spp*. (OTUID 173469) abundance categories with respect to host genotype at rs6211533, represented by green (AA), yellow (AC) and blue (CC). Numbers in parentheses indicate the sample size within each genotype category.
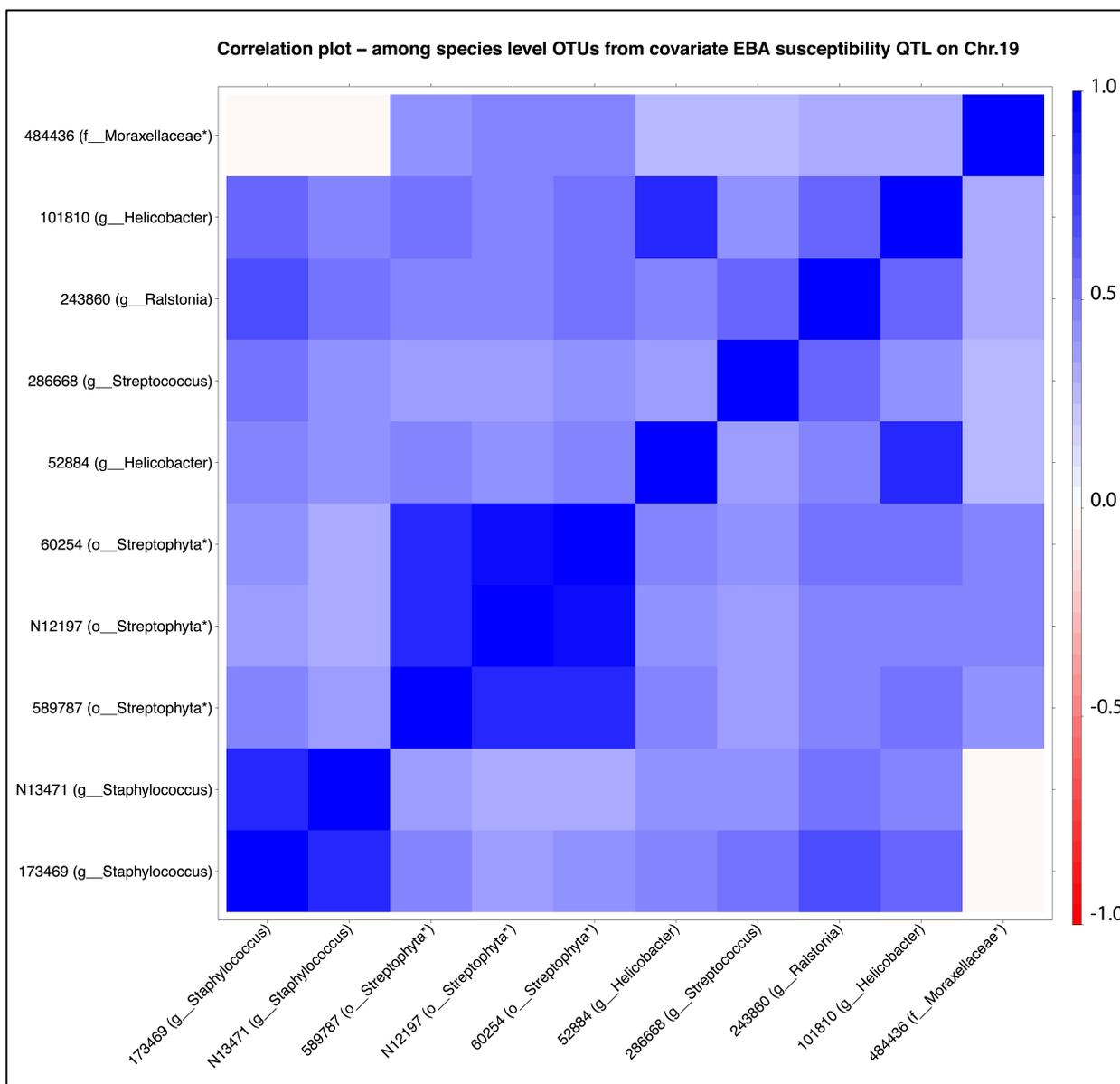
Figure 3.24: Pearson correlation matrix of 10 species OTUs that significantly co-vary with EBA susceptibility locus on chromosome 19.
For OTUs not classified at the genus level, the next highest taxonomic level is shown (as marked in asterisk *). The taxonomic level of classification is indicated by k, p, c, o, f and g for kingdom, phylum, class, order, family and genus, respectively. Only values significantly differing from zero after correction for multiple testing[127] are shown by either blue (positive correlation) or red (negative correlation) squares.

4. Discussion

Mammalian hosts are essentially sterile at birth but quickly become colonized by microbes [129], [130]. The microbial communities on human skin are primarily driven by the mode of delivery. It has been reported that babies born by C-section have microbiota similar to their mother's skin while vaginally delivered babies have bacterial composition similar to vaginal microbiota [44]. Understanding the complex population structure of skin microbiota in the context of host-symbiont and host-pathogen relationship would indeed greatly advance the field mainly focused on the role of the environment in the development and progression of the autoimmune disease. Unlike the microbiota studies in gastrointestinal tract, a need exists to study the role of skin microbiota especially in the context of health and disease. Current results on abundance differences among skin communities in skin diseases such as atopic dermatitis [41], psoriatic lesions [42] and acne [43] fail to address whether the change is due to causes or consequences of the disease. In the context of autoimmune skin blistering disease, no study to date has focused on the skin microbiota. This study makes the first attempt to simultaneously study interactions between skin microbiota and host-genetics and its influence on EBA disease susceptibility, using a large number of animals (261 mice) from fourth generation of AIL population (see **Methods**). Animals in this generation include genetically susceptible individuals (~33%) as well as those resistant to immunization-dependent EBA disease. Non-immunized mice were also included in the study to control for immunization effect (see **Methods and Results**).

Ears from 261 mice were specifically chosen for skin microbiota phenotyping because it is the site of the first disease manifestations of EBA clinical symptoms. The EBA clinical symptoms include redness, erythema, blistering, erosions, lesions, etc. High throughput sequencing of variable regions (V1-V2) of the bacterial 16s rRNA gene was used to phenotype the skin bacterial composition from mouse ears. The skin bacterial composition at the bacterial phyla level observed in this study is comparable to that of previously published studies on skin microbial communities [6],

71

[30], [47], [125] and hence confirms that the skin bacterial composition findings in this study are indeed a true representation of skin microflora. The analysis results of diversity "within" (Alpha diversity analysis) and "between" (Beta diversity analysis) individuals revealed significant differences between microbial community composition among healthy and diseased individuals. This further pointed to the direction towards the existence of important bacterial species candidates contributing to these patterns. At the species level, indicator species analysis showed that most OTUs displayed a strong association with healthy individuals compared to diseased individuals. This change could be a result of both causes as well as consequences of the disease. Increase in the bacterial diversity and evenness among healthy individuals supports the well-known hygiene hypothesis. This hypothesis suggests that the regulation of immune system response is compromised by diminished exposure to pathogens early in life because of changes in life style factors such as overuse of antibiotics, vaccines and improve hygiene methods. This immediately raises the question of what exactly constitutes hygiene microflora hypothesis? This theory is mainly supported by the observation or circumstantial evidences resulting from various microbiota studies on the context of disease state. Scientists are still trying to figure out the ways to understand this complex puzzle on how commensal bacteria are involved in changing the host health states.

Recent clinical studies have provided valuable information about the different types and abundance of skin microbes, but they fail to address their function [49], [50], [125], [131], [132]. There are three different categories of relationship existing between host and microbial flora and they are parasitism, commensalism or mutualism [133]. These relationships also exist among microbial communities found within or on the host bodies making it more complex to disentangle their individual functions to host. They might act as an individual bacterial species or group of driving species contributing for specific host function at a given host environment or in the presence of certain unknown factors. In this study there are 38 bacterial species OTUs found in high abundance among healthy individuals compared to EBA diseased mice (**Appendix B Table B.1**).

Out of them eight bacterial OTUs belong to the genus *Corynebacterium* and also seven belong to *Staphylococcus*. These two genus level bacterial populations are common members of the skin flora. Metagenomics studies of different skin areas of mouse as well as humans have revealed that *Corynebacterium* and *Staphylococcus* are the most abundant organisms colonizing moist areas of the skin [30], [48]. Coryneforms of bacteria are Gram-positive, nonmotile facultative anaerobic bacteria that they are found in skin regions such as hair follicles, eccrine and sebaceous glands. There are 17 different bacterial species belonging to the genus *Corynebacterium* and not all are present on mammalian skin [133]. Among these there are two bacterial species *Corynebacterium dipththeriae* [134], [135], [136], [137] and *Corynebacterium jeikeium*, which are known to be linked to different disease states in humans [138], [139], [140]. *C.jeikeium* is also found within the normal skin flora of most humans and is commonly found in hospitalized patients [140], [141]. This species produces bacteriocin-like compounds which might be used to ward off potential microbial pathogens and competitors, a possible way by which this bacteria species gives protection to host [133]. This particular bacterial species looks as if it exhibits both pathogenesis and benefits to the host. It remains to be explored what makes this bacterium turn pathogenic.

Staphylococcus epidermidis and *Staphylococcus aureus* are two other species widely found in skin flora. They are Gram-positive coccus found in clusters. *S. aureus* has been usually classified as transient pathogen but it is also found in the nasal microbiota of healthy individuals [142]. Certain strains of S. *aureus* have been shown to produce bacteriocins, which inhibit the growth of other *S. aureus* strains [143]. It has also been shown that *S.epidermidis* produces a variety of bacteriocins [144], [145], [146]. In particular, it produces a toxic peptide that directly regulate other microorganisms such as *S.aureus* and group A *Streptococcus* (*S. pyogenes*) [147]. In this way *S.epidermidis* provides the host with additional protection against common pathogens. This "cross inhibition" mechanism may also be used by the host to regulate other commensal bacteria populations as well. A recent study demonstrated that the colonization of *S.epidermidis* as a single

bacterial species on germ free mouse skin is sufficient to rescue Interleukin-17 (IL-17) production in the skin, and this result shows that the local resident skin bacteria is indeed necessary to drive T cell function in the skin to main cutaneous immune homeostasis [57]. It has been further shown that the presence of *S.epidermidis* in mouse skin provides immunity to mice against the protozoan parasite *Leishmania major* [57]. It may be now possible to modulate various inflammatory cells by commensals as a tool to restore host fitness. But most of these studies are performed on C57BL/6 (B6), a widely used laboratory mouse strain. It is of greater scientific importance to focus on more studies on different genetic backgrounds and derive germ free mouse strains from different genetic backgrounds to make available for scientists to perform similar experiments. Understanding the cross talk between the host immune system, host genetics and skin microbiota might help us in understanding the etiology and pathology of complex diseases such as autoimmune skin blistering diseases. Furthermore, it would be very exciting to study and explore how these microbes from different areas within host may affect physiological, metabolic and immunological balance.

It can be seen that shifts in normal microbial community composition (also known as dysbiosis) make the host more susceptible to diseases. The association of host genetic variation in maintaining healthy skin microbiota or involvement in dysbiosis remains largely unexplored. Results from this study provide strong evidence on how host genetically based variation contributes to differences in the bacterial communities observed in the skin (**Appendix B Table B.3 – B.4**). This finding is consistent with the previous observation by Benson and associates [27]. They identified that the host genetics indeed controls the variation of the mouse gut community. These two findings clearly show a glimpse of the host genetic association over microbial communities found in gut and skin communities. In a previous QTL analysis of the mouse fecal community, Benson *et al*. [27] reported 13 significant and 5 additional suggestive QTLs for 26 out of 64 taxonomic groups tested from CMM. However, their analysis was not extended beyond the level of bacterial genera. Despite a more inclusive set of phenotypic traits extending to the bacterial species (OTU) level, this study

detected nine significant QTLs for nine out of 131 species level traits (**Appendix B Table B.3**). Interestingly, despite differences in experimental setup and the obvious distinction between the two environments, a strong evidence of overlap between studies was found. The confidence intervals of two out of 18 QTLs controlling bacterial abundance in murine feces contain the peak SNP of a skin QTL, which overlaps more than expected by chance (*P* value < 0.05; **Figure 3.18**; **Methods**). One of these QTLs is consistent at the phyla level (Firmicutes, Chr. 14), while the other is at the order level (*Pseudomonadales*, Chr. 9). Similarly, an additional two of our skin spQTLs (OTU N31208 belonging to *Streptococcus* on Chr. 12 and OTU 130241 belonging to *Herbaspirillum* on Chr. 15) overlap with fecal QTLs from another recently published study [128] although the taxonomic assignments do not agree at even the phylum level. Given the evidence of overlap of skin QTLs with previous studies of the gut, certainly a portion of these variants will have influences wider reaching than the ear. This speculation could be true as there are high chances that the host-coevolution of skin and gut bacteria communities could have evolved at the same time and the host has indeed found a common mechanism of tolerance to allow certain commensal bacterial taxon to reside in gut and skin simultaneously for its fitness. It further requires fine mapping within these QTL regions to find out whether the same genes are indeed involved in controlling the same or different bacterial taxa in the gut and skin regions and how this could directly contribute to the fitness of specific local regions of the host. However, these fine-mapped gene functions should be verified from gene knockout studies to further confirm the claim. It is indeed interesting to study whether the same genes are involved in improving the fitness of the host by manipulating the gut as well as skin communities simultaneously. Future research may enhance our understanding of the evolutionary forces that shaped the host-genes influencing certain microbial communities.

In addition to exogenous factors, genetic variation of innate immune system genes also appears to shape the composition and structure of commensal bacteria. IgA-deficient mice become highly colonized by segmented filamentous bacteria (SFB) [148]. It has been shown that SFB colonization

in the gut promotes IL-17 production through TH17 cells (Interleukin -17 (IL-17) producing helper T cells) [149]. Toll-like receptor (TLRs) signaling is the main part of the innate immune system and is widely studied in the context of its influence over microbiota composition in the mammalian gut. A recent study on TLR4 knockouts (KO) clearly showed that gut microbial products regulate gastrointestinal motility in mice through TLR4 signaling [150]. Another study simultaneously compared bacterial taxa that are affected by TLR signaling in the ileum or cecum from colonies of MyD88, TLR2, TLR3, TLR4, TLR5, or TLR9 KO mice and their respective wild type (WT) healthy controls [151]. In the results they showed that there is high possibility of specific bacterial taxa getting affected by defects in the TLR signaling pathways [151]. These results further motivated us to look for innate immune genes within the QTLs found in this study. The confidence intervals of skin microbiota QTLs found in this study contain nine genes known to be involved in the functioning of the innate immune system (**Appendix B Table B.5**). Interleukin-1 receptor-associated kinase (IRAK)-4 is an interesting candidate found within the confidence interval of spQTL6, which modulates an OTU (ID 130241) belonging to the genus *Herbaspirillum*. Deficiencies of this gene in humans lead to increased susceptibility to pyogenic bacterial infections including *Staphylococcus aureus* [152], and its interaction with the MYD88 adapter protein is used by several Toll-like receptor (TLR) pathways in host defence [153], as well as being involved in controlling commensal bacteria [154]. Another gene coding for CD14 antigen is found within spQTL8 on chromosome 18, which modulates an OTU (ID N10459) belonging to the genus *Staphylococcus*. Increasing CD14 expression enhances TLR2 activation in skin in the presence of vitamin D_3—1,25-dihydroxyvitamin D_3 (1,25D3) [155], which in turn influences the skins sensitivity to microbial challenge. Furthermore, several studies have shown that components of *Staphyloccus aureus* (LTA and peptidoglycan) interact with the CD14 molecule [156], [157], [158]. Finally, by treating bacterial abundances as covariates with the presence/absence of EBA, an additional significant EBA QTL on chromosome 19 was identified (Fig. 4). One potential candidate gene lying within this chromosomal

interval (53-60 Mb) is caspase-7 (casp7), a member of the cytosolic cysteine protease family known to be involved with inflammatory disorders [159], [160] and defense against pathogens [161].

The results from this study confirmed previous findings that immunization of SJL/J mice with COL7 leads to autoantibody production in all mice, while clinically overt blistering is observed in 80% of the mice. Interestingly, the remaining 20% of mice do not develop skin blistering, despite the presence of autoantibody production. The ultimate reason, why genetically identical mice housed under same environmental conditions, react differently to the same attempt of disease induction, can be explained by sub-clinical inflammatory events, which are obviously triggered by environmental factors in genetically prone individuals [57], [162], [163], [164]. In order to test whether exogenous factors lead to this altered state of subclinical inflammation, the microbial communities of the skin present before disease induction were characterized. Mice which do not develop symptoms after disease induction showed an increased microbial diversity compared to individuals that developed EBA symptoms, identifying bacterial diversity as a key element of protection against autoimmune diseases [165]. This result is in consistent with similar analyses done on the gut microbiota of identical twin humans, where healthy individuals showed greater bacterial diversity compared to individuals with Crohn's disease [164]. The microbiome along with sub-clinical inflammation should be further evaluated as tools for surveillance of mouse experiments to minimize the bias by subclinical differences between the mice. Additionally, these two factors may serve as a target in prevention and therapy of autoimmune disease symptoms. Moreover, the results of this study show possible inflammatory and environmental events which enable further investigation of gene-environment interactions in the pathogenesis of autoimmunity, as KC and TNFα have both been reported to be induced in the skin and other organs by the commensal flora, thus leading to inflammation and regulating autoimmune responses [166], [167].

Similar to previous studies of chronic inflammatory skin diseases, findings in this study support a role of resident microbial communities in disease pathogenesis. The differences in community composition and structure between mice with and without EBA symptoms are akin to shifts in the skin microbiota associated with atopic dermatitis disease flares and treatment [168] or between psoriatic lesions and both unaffected skin in patients and healthy controls [169]. A rough three-fold increase over the last 30 years of atopic dermatitis in industrialized countries suggests more complex environmental influences, possibly mediated by changes in microbial communities. By investigating disease provocation in a large mouse mapping population under controlled environmental conditions, this study is able to identify individual, genotype-dependent microbial risk factors among a core set of taxa inhabiting the skin of both healthy and diseased mice, more closely resembling a disease-modifying effect. Validation and characterization of these interactions await more intensive experimental interrogation in gnotobiotic animals, for example. Thus, the further identification and functional analysis of host genetic and probiotic bacterial factors represent promising avenues for research in preventative and therapeutic treatment development.

5. Bibliography

[1]   A. Lauer, M. A. Simon, J. L. Banning, B. A. Lam, and R. N. Harris, "Diversity of cutaneous bacteria with antifungal activity isolated from female four-toed salamanders," *The ISME Journal*, vol. 2, no. 2, pp. 145–157, Dec. 2007.

[2]   J. C. Clemente, L. K. Ursell, L. W. Parfrey, and R. Knight, "The impact of the gut microbiota on human health: an integrative view," *Cell*, vol. 148, no. 6, pp. 1258–1270, Mar. 2012.

[3]   P. J. Turnbaugh, M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight, and J. I. Gordon, "A core gut microbiome in obese and lean twins," *Nature*, vol. 457, no. 7228, pp. 480–484, Jan. 2009.

[4]   F. Bäckhed, H. Ding, T. Wang, L. V. Hooper, G. Y. Koh, A. Nagy, C. F. Semenkovich, and J. I. Gordon, "The gut microbiota as an environmental factor that regulates fat storage," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 44, pp. 15718–15723, Nov. 2004.

[5]   A. L. Kau, P. P. Ahern, N. W. Griffin, A. L. Goodman, and J. I. Gordon, "Human nutrition, the gut microbiome and the immune system," *Nature*, vol. 474, no. 7351, pp. 327–336, Jun. 2011.

[6]   E. A. Grice, H. H. Kong, G. Renaud, A. C. Young, G. G. Bouffard, R. W. Blakesley, T. G. Wolfsberg, M. L. Turner, and J. A. Segre, "A diversity profile of the human skin microbiota," *Genome Res.*, vol. 18, no. 7, pp. 1043–1050, Jul. 2008.

[7]   F. Bäckhed, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon, "Host-bacterial mutualism in the human intestine," *Science*, vol. 307, no. 5717, pp. 1915–1920, Mar. 2005.

[8]   G. B. Rogers, M. P. Carroll, D. J. Serisier, P. M. Hockey, G. Jones, and K. D. Bruce, "Characterization of Bacterial Community Diversity in Cystic Fibrosis Lung Infections by Use of 16S Ribosomal DNA Terminal Restriction Fragment Length Polymorphism Profiling," *Journal of Clinical Microbiology*, vol. 42, no. 11, pp. 5176–5183, Nov. 2004.

[9]   W. B. Whitman, D. C. Coleman, and W. J. Wiebe, "Prokaryotes: the unseen majority," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, no. 12, pp. 6578–6583, Jun. 1998.

[10]   R. E. Ley, D. A. Peterson, and J. I. Gordon, "Ecological and evolutionary forces shaping microbial diversity in the human intestine," *Cell*, vol. 124, no. 4, pp. 837–848, Feb. 2006.

[11]   P. B. Eckburg, "Diversity of the Human Intestinal Microbial Flora," *Science*, vol. 308, no. 5728, pp. 1635–1638, Jun. 2005.

[12]   Erwin G. Zoetendal, Antoon D. L. Ak, "The Host Genotype Affects the Bacterial Community in the Human Gastronintestinal Tract," *Microbial Ecology in Health and Disease*, vol. 13, no. 3, pp. 129–134, Jan. 2001.

[13]   J. Jalanka-Tuovinen, A. Salonen, J. Nikkilä, O. Immonen, R. Kekkonen, L. Lahti, A. Palva, and W. M. de Vos, "Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms," *PLoS ONE*, vol. 6, no. 7, p. e23035, 2011.

[14]   M. Rajilić-Stojanović, H. G. H. J. Heilig, D. Molenaar, K. Kajander, A. Surakka, H. Smidt, and W. M. de Vos, "Development and application of the human intestinal tract chip, a

phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults," *Environ. Microbiol.*, vol. 11, no. 7, pp. 1736–1751, Jul. 2009.

[15] J. Walter and R. Ley, "The human gut microbiome: ecology and recent evolutionary changes," *Annu. Rev. Microbiol.*, vol. 65, pp. 411–429, 2011.

[16] M. K. Friswell, H. Gika, I. J. Stratford, G. Theodoridis, B. Telfer, I. D. Wilson, and A. J. McBain, "Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice," *PLoS ONE*, vol. 5, no. 1, p. e8584, 2010.

[17] M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J.-M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E. G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. Brunak, J. Doré, M. Antolín, F. Artiguenave, H. M. Blottiere, M. Almeida, C. Brechot, C. Cara, C. Chervaux, A. Cultrone, C. Delorme, G. Denariaz, R. Dervyn, K. U. Foerstner, C. Friss, M. van de Guchte, E. Guedon, F. Haimet, W. Huber, J. van Hylckama-Vlieg, A. Jamet, C. Juste, G. Kaci, J. Knol, O. Lakhdari, S. Layec, K. Le Roux, E. Maguin, A. Mérieux, R. Melo Minardi, C. M'rini, J. Muller, R. Oozeer, J. Parkhill, P. Renault, M. Rescigno, N. Sanchez, S. Sunagawa, A. Torrejon, K. Turner, G. Vandemeulebrouck, E. Varela, Y. Winogradsky, G. Zeller, J. Weissenbach, S. D. Ehrlich, and P. Bork, "Enterotypes of the human gut microbiome," *Nature*, vol. 473, no. 7346, pp. 174–180, May 2011.

[18] G. D. Wu, J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters, R. Knight, R. Sinha, E. Gilroy, K. Gupta, R. Baldassano, L. Nessel, H. Li, F. D. Bushman, and J. D. Lewis, "Linking long-term dietary patterns with gut microbial enterotypes," *Science*, vol. 334, no. 6052, pp. 105–108, Oct. 2011.

[19] S. M. Huse, Y. Ye, Y. Zhou, and A. A. Fodor, "A core human microbiome as viewed through 16S rRNA sequence clusters," *PLoS ONE*, vol. 7, no. 6, p. e34242, 2012.

[20] E. Louis, C. Van Kemseke, P. Latour, J. Belaiche, and C. Reenaers, "[Genetics and environment in chronic inflammatory bowel diseases]," *Rev Med Liege*, vol. 67, no. 5–6, pp. 298–304, Jun. 2012.

[21] W. E. Moore and L. H. Moore, "Intestinal floras of populations that have a high risk of colon cancer," *Appl. Environ. Microbiol.*, vol. 61, no. 9, pp. 3202–3207, Sep. 1995.

[22] M. Noval Rivas, O. T. Burton, P. Wise, Y. Zhang, S. A. Hobson, M. Garcia Lloret, C. Chehoud, J. Kuczynski, T. DeSantis, J. Warrington, E. R. Hyde, J. F. Petrosino, G. K. Gerber, L. Bry, H. C. Oettgen, S. K. Mazmanian, and T. A. Chatila, "A microbiota signature associated with experimental food allergy promotes allergic sensitization and anaphylaxis," *J. Allergy Clin. Immunol.*, vol. 131, no. 1, pp. 201–212, Jan. 2013.

[23] L. Wen, R. E. Ley, P. Y. Volchkov, P. B. Stranges, L. Avanesyan, A. C. Stonebraker, C. Hu, F. S. Wong, G. L. Szot, J. A. Bluestone, J. I. Gordon, and A. V. Chervonsky, "Innate immunity and intestinal microbiota in the development of Type 1 diabetes," *Nature*, vol. 455, no. 7216, pp. 1109–1113, Oct. 2008.

[24] Y. Wang, J. D. Hoenig, K. J. Malin, S. Qamar, E. O. Petrof, J. Sun, D. A. Antonopoulos, E. B. Chang, and E. C. Claud, "16S rRNA gene-based analysis of fecal microbiota from

preterm infants with and without necrotizing enterocolitis," *The ISME Journal*, vol. 3, no. 8, pp. 944–954, Apr. 2009.

[25] C. Zhang, M. Zhang, S. Wang, R. Han, Y. Cao, W. Hua, Y. Mao, X. Zhang, X. Pang, C. Wei, G. Zhao, Y. Chen, and L. Zhao, "Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice," *The ISME Journal*, vol. 4, no. 2, pp. 232–241, Oct. 2009.

[26] R. E. Ley, P. J. Turnbaugh, S. Klein, and J. I. Gordon, "Microbial ecology: human gut microbes associated with obesity," *Nature*, vol. 444, no. 7122, pp. 1022–1023, Dec. 2006.

[27] A. K. Benson, S. A. Kelly, R. Legge, F. Ma, S. J. Low, J. Kim, M. Zhang, P. L. Oh, D. Nehrenberg, K. Hua, S. D. Kachman, E. N. Moriyama, J. Walter, D. A. Peterson, and D. Pomp, "Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 44, pp. 18933–18938, Nov. 2010.

[28] A. Spor, O. Koren, and R. Ley, "Unravelling the effects of the environment and host genotype on the gut microbiome," *Nature Reviews Microbiology*, vol. 9, no. 4, pp. 279–290, Apr. 2011.

[29] A. M. Bowcock and W. O. C. M. Cookson, "The genetics of psoriasis, psoriatic arthritis and atopic dermatitis," *Hum. Mol. Genet.*, vol. 13 Spec No 1, pp. R43–55, Apr. 2004.

[30] E. K. Costello, C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight, "Bacterial community variation in human body habitats across space and time," *Science*, vol. 326, no. 5960, pp. 1694–1697, Dec. 2009.

[31] J. Masenga, C. Garbe, J. Wagner, and C. E. Orfanos, "Staphylococcus aureus in atopic dermatitis and in nonatopic dermatitis," *Int. J. Dermatol.*, vol. 29, no. 8, pp. 579–582, Oct. 1990.

[32] A. Nakabayashi, Y. Sei, and J. Guillot, "Identification of Malassezia species isolated from patients with seborrhoeic dermatitis, atopic dermatitis, pityriasis versicolor and normal subjects," *Med. Mycol.*, vol. 38, no. 5, pp. 337–341, Oct. 2000.

[33] R. Austrian, "The Gram stain and the etiology of lobar pneumonia, an historical note," *Bacteriol Rev*, vol. 24, no. 3, pp. 261–265, Sep. 1960.

[34] C. R. Woese, "Interpreting the universal phylogenetic tree," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, no. 15, pp. 8392–8396, Jul. 2000.

[35] N. Iwabe, K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata, "Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 86, no. 23, pp. 9355–9359, Dec. 1989.

[36] N. R. Pace, "A molecular view of microbial diversity and the biosphere," *Science*, vol. 276, no. 5313, pp. 734–740, May 1997.

[37] E. A. Grice and J. A. Segre, "The skin microbiome," *Nature Reviews Microbiology*, vol. 9, no. 4, pp. 244–253, Apr. 2011.

[38] E. Pruesse, C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner, "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB," *Nucleic Acids Res.*, vol. 35, no. 21, pp. 7188–7196, 2007.

[39] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen, "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB," *Appl. Environ. Microbiol.*, vol. 72, no. 7, pp. 5069–5072, Jul. 2006.

[40] J. R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje, "The Ribosomal Database Project: improved alignments and new tools for rRNA analysis," *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D141–145, Jan. 2009.

[41] I. Dekio, M. Sakamoto, H. Hayashi, M. Amagai, M. Suematsu, and Y. Benno, "Characterization of skin microbiota in patients with atopic dermatitis and in normal subjects using 16S rRNA gene-based comprehensive analysis," *J. Med. Microbiol.*, vol. 56, no. Pt 12, pp. 1675–1683, Dec. 2007.

[42] Z. Gao, C. Tseng, B. E. Strober, Z. Pei, and M. J. Blaser, "Substantial alterations of the cutaneous bacterial biota in psoriatic lesions," *PLoS ONE*, vol. 3, no. 7, p. e2719, 2008.

[43] M. Bek-Thomsen, H. B. Lomholt, and M. Kilian, "Acne is not associated with yet-uncultured bacteria," *J. Clin. Microbiol.*, vol. 46, no. 10, pp. 3355–3360, Oct. 2008.

[44] M. G. Dominguez-Bello, E. K. Costello, M. Contreras, M. Magris, G. Hidalgo, N. Fierer, and R. Knight, "Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 26, pp. 11971–11975, Jun. 2010.

[45] R. E. Ley, F. Bäckhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, and J. I. Gordon, "Obesity alters gut microbial ecology," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 31, pp. 11070–11075, Aug. 2005.

[46] A. K. Benson, S. A. Kelly, R. Legge, F. Ma, S. J. Low, J. Kim, M. Zhang, P. L. Oh, D. Nehrenberg, K. Hua, S. D. Kachman, E. N. Moriyama, J. Walter, D. A. Peterson, and D. Pomp, "Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors," *Proceedings of the National Academy of Sciences*, vol. 107, no. 44, pp. 18933–18938, Oct. 2010.

[47] N. Fierer, M. Hamady, C. L. Lauber, and R. Knight, "The influence of sex, handedness, and washing on the diversity of hand surface bacteria," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, no. 46, pp. 17994–17999, Nov. 2008.

[48] E. A. Grice, H. H. Kong, S. Conlan, C. B. Deming, J. Davis, A. C. Young, G. G. Bouffard, R. W. Blakesley, P. R. Murray, E. D. Green, M. L. Turner, and J. A. Segre, "Topographical and temporal diversity of the human skin microbiome," *Science*, vol. 324, no. 5931, pp. 1190–1192, May 2009.

[49] D. N. Fredricks, "Microbial ecology of human skin in health and disease," *J. Investig. Dermatol. Symp. Proc.*, vol. 6, no. 3, pp. 167–169, Dec. 2001.

[50] R. R. Roth and W. D. James, "Microbial ecology of the skin," *Annu. Rev. Microbiol.*, vol. 42, pp. 441–464, 1988.

[51] A. V. Chervonsky, "Influence of microbial environment on autoimmunity," *Nature Immunology*, vol. 11, no. 1, pp. 28–35, Dec. 2009.

[52] M. Gloor, G. Peters, and D. Stoika, "On the resident aerobic bacterial skin flora in unaffected skin of patients with atopic dermatitis and in healthy controls," *Dermatologica*, vol. 164, no. 4, pp. 258–265, Apr. 1982.

[53] S. Rieg, H. Steffen, S. Seeber, A. Humeny, H. Kalbacher, K. Dietz, C. Garbe, and B. Schittek, "Deficiency of dermcidin-derived antimicrobial peptides in sweat of patients with atopic dermatitis correlates with an impaired innate defense of human skin in vivo," *J. Immunol.*, vol. 174, no. 12, pp. 8003–8010, Jun. 2005.

[54] B. Schittek, R. Hipfel, B. Sauer, J. Bauer, H. Kalbacher, S. Stevanovic, M. Schirle, K. Schroeder, N. Blin, F. Meier, G. Rassner, and C. Garbe, "Dermcidin: a novel human antibiotic peptide secreted by sweat glands," *Nat. Immunol.*, vol. 2, no. 12, pp. 1133–1137, Dec. 2001.

[55] I. Senyürek, M. Paulmann, T. Sinnberg, H. Kalbacher, M. Deeg, T. Gutsmann, M. Hermes, T. Kohler, F. Götz, C. Wolz, A. Peschel, and B. Schittek, "Dermcidin-derived peptides show a different mode of action than the cathelicidin LL-37 against Staphylococcus aureus," *Antimicrob. Agents Chemother.*, vol. 53, no. 6, pp. 2499–2509, Jun. 2009.

[56] Y. Lai, A. Di Nardo, T. Nakatsuji, A. Leichtle, Y. Yang, A. L. Cogen, Z.-R. Wu, L. V. Hooper, R. R. Schmidt, S. von Aulock, K. A. Radek, C.-M. Huang, A. F. Ryan, and R. L. Gallo, "Commensal bacteria regulate Toll-like receptor 3-dependent inflammation after skin injury," *Nat. Med.*, vol. 15, no. 12, pp. 1377–1382, Dec. 2009.

[57] S. Naik, N. Bouladoux, C. Wilhelm, M. J. Molloy, R. Salcedo, W. Kastenmuller, C. Deming, M. Quinones, L. Koo, S. Conlan, S. Spencer, J. A. Hall, A. Dzutsev, H. Kong, D. J. Campbell, G. Trinchieri, J. A. Segre, and Y. Belkaid, "Compartmentalized Control of Skin Immunity by Resident Commensals," *Science*, vol. 337, no. 6098, pp. 1115–1119, Jul. 2012.

[58] P. E. Lipsky, "Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity," *Nat. Immunol.*, vol. 2, no. 9, pp. 764–766, Sep. 2001.

[59] C. D. Surh, "Homeostatic T Cell Proliferation: How Far Can T Cells Be Activated to Self-Ligands?," *Journal of Experimental Medicine*, vol. 192, no. 4, p. 9F–14, Aug. 2000.

[60] J. Lis, A. Jarząb, and D. Witkowska, "Molecular mimicry in the etiology of autoimmune diseases," *Postepy Hig Med Dosw (Online)*, vol. 66, pp. 475–491, 2012.

[61] K. Hemminki, X. Liu, J. Ji, J. Sundquist, and K. Sundquist, "Autoimmune disease and subsequent digestive tract cancer by histology," *Annals of Oncology*, vol. 23, no. 4, pp. 927–933, Aug. 2011.

[62] B. Afzali, G. Lombardi, R. I. Lechler, and G. M. Lord, "The role of T helper 17 (Th17) and regulatory T cells (Treg) in human organ transplantation and autoimmune disease," *Clinical & Experimental Immunology*, vol. 148, no. 1, pp. 32–46, Feb. 2007.

[63] G. R. Yannam, T. Gutti, and L. Y. Poluektova, "IL-23 in Infections, Inflammation, Autoimmunity and Cancer: Possible Role in HIV-1 and AIDS," *Journal of Neuroimmune Pharmacology*, vol. 7, no. 1, pp. 95–112, Sep. 2011.

[64] E. Schmidt and D. Zillikens, "The Diagnosis and Treatment of Autoimmune Blistering Skin Disease," *Dtsch Arztebl Int.*, no. 108(23), pp. 399–405, 2011.

[65] K. B. Yancey, "The pathophysiology of autoimmune blistering diseases," *Journal of Clinical Investigation*, vol. 115, no. 4, pp. 825–828, Apr. 2005.

[66]   E. Schmidt and D. Zillikens, "Autoimmune and inherited subepidermal blistering diseases: advances in the clinic and the laboratory," *Adv Dermatol*, vol. 16, pp. 113–157; discussion 158, 2000.

[67]   D. T. Woodley, R. A. Briggaman, E. J. O'Keefe, A. O. Inman, L. L. Queen, and W. R. Gammon, "Identification of the skin basement-membrane autoantigen in epidermolysis bullosa acquisita," *N. Engl. J. Med.*, vol. 310, no. 16, pp. 1007–1013, Apr. 1984.

[68]   J. R. Stanley, N. Rubinstein, and V. Klaus-Kovtun, "Epidermolysis bullosa acquisita antigen is synthesized by both human keratinocytes and human dermal fibroblasts," *J. Invest. Dermatol.*, vol. 85, no. 6, pp. 542–545, Dec. 1985.

[69]   D. T. Woodley, R. E. Burgeson, G. Lunstrum, L. Bruckner-Tuderman, M. J. Reese, and R. A. Briggaman, "Epidermolysis bullosa acquisita antigen is the globular carboxyl terminus of type VII procollagen," *J. Clin. Invest.*, vol. 81, no. 3, pp. 683–687, Mar. 1988.

[70]   J. C. Lapiere, D. T. Woodley, M. G. Parente, T. Iwasaki, K. C. Wynn, A. M. Christiano, and J. Uitto, "Epitope mapping of type VII collagen. Identification of discrete peptide sequences recognized by sera from patients with acquired epidermolysis bullosa," *J. Clin. Invest.*, vol. 92, no. 4, pp. 1831–1839, Oct. 1993.

[71]   W. R. Gammon, D. F. Murrell, M. W. Jenison, K. M. Padilla, P. S. Prisayanh, D. A. Jones, R. A. Briggaman, and S. W. Hunt 3rd, "Autoantibodies to type VII collagen recognize epitopes in a fibronectin-like region of the noncollagenous (NC1) domain," *J. Invest. Dermatol.*, vol. 100, no. 5, pp. 618–622, May 1993.

[72]   M. Chen, D. R. Keene, F. K. Costa, S. H. Tahk, and D. T. Woodley, "The carboxyl terminus of type VII collagen mediates antiparallel dimer formation and constitutes a new antigenic epitope for epidermolysis Bullosa acquisita autoantibodies," *J. Biol. Chem.*, vol. 276, no. 24, pp. 21649–21655, Jun. 2001.

[73]   T. Tanaka, F. Furukawa, and S. Imamura, "Epitope mapping for epidermolysis bullosa acquisita autoantibody by molecularly cloned cDNA for type VII collagen," *J. Invest. Dermatol.*, vol. 102, no. 5, pp. 706–709, May 1994.

[74]   R. P. Erickson, "Why isn't a mouse more like a man?," *Trends Genet.*, vol. 5, no. 1, pp. 1–3, Jan. 1989.

[75]   C. Sitaru, S. Mihai, C. Otto, M. T. Chiriac, I. Hausser, B. Dotterweich, H. Saito, C. Rose, A. Ishiko, and D. Zillikens, "Induction of dermal-epidermal separation in mice by passive transfer of antibodies specific to type VII collagen," *J. Clin. Invest.*, vol. 115, no. 4, pp. 870–878, Apr. 2005.

[76]   C. Sitaru, "Experimental models of epidermolysis bullosa acquisita," *Exp. Dermatol.*, vol. 16, no. 6, pp. 520–531, Jun. 2007.

[77]   D. T. Woodley, R. Ram, A. Doostan, P. Bandyopadhyay, Y. Huang, J. Remington, Y. Hou, D. R. Keene, Z. Liu, and M. Chen, "Induction of Epidermolysis Bullosa Acquisita in Mice by Passive Transfer of Autoantibodies from Patients," *Journal of Investigative Dermatology*, vol. 126, no. 6, pp. 1323–1330, Mar. 2006.

[78]   R. J. Ludwig, A. Recke, K. Bieber, S. Müller, A. de C. Marques, D. Banczyk, M. Hirose, M. Kasperkiewicz, N. Ishii, E. Schmidt, J. Westermann, D. Zillikens, and S. M. Ibrahim, "Generation of antibodies of distinct subclasses and specificity is linked to H2s in an active mouse model of epidermolysis bullosa acquisita," *J. Invest. Dermatol.*, vol. 131, no. 1, pp. 167–176, Jan. 2011.

[79]   S. Mihai, M. T. Chiriac, K. Takahashi, J. M. Thurman, V. M. Holers, D. Zillikens, M. Botto, and C. Sitaru, "The alternative pathway of complement activation is critical for blister induction in experimental epidermolysis bullosa acquisita," *J. Immunol.*, vol. 178, no. 10, pp. 6514–6521, May 2007.

[80]   A. Sesarman, S. Mihai, M. T. Chiriac, F. Olaru, A. G. Sitaru, J. M. Thurman, D. Zillikens, and C. Sitaru, "Binding of avian IgY to type VII collagen does not activate complement and leucocytes and fails to induce subepidermal blistering in mice," *Br. J. Dermatol.*, vol. 158, no. 3, pp. 463–471, Mar. 2008.

[81]   M. T. Chiriac, J. Roesler, A. Sindrilaru, K. Scharffetter-Kochanek, D. Zillikens, and C. Sitaru, "NADPH oxidase is required for neutrophil-dependent autoantibody-induced tissue damage," *J. Pathol.*, vol. 212, no. 1, pp. 56–65, May 2007.

[82]   C. Sitaru, M. T. Chiriac, S. Mihai, J. Büning, A. Gebert, A. Ishiko, and D. Zillikens, "Induction of complement-fixing autoantibodies against type VII collagen results in subepidermal blistering in mice," *J. Immunol.*, vol. 177, no. 5, pp. 3461–3468, Sep. 2006.

[83]   D. McGonagle, "A Proposed Classification of the Immunological Diseases," *PLoS Medicine*, vol. 3, no. 8, p. e297, 2006.

[84]   J. Ermann and C. G. Fathman, "Autoimmune diseases: genes, bugs and failed regulation," *Nat. Immunol.*, vol. 2, no. 9, pp. 759–761, Sep. 2001.

[85]   G. Cooper and F. Miller, "Chapter 25: Environmental influences on autoimmunity and autoimmune diseases , in Immunopharmacology and Immunotoxicology," 3rd edition., Luebke R,ed., CRC Press (New York), 2007, pp. 437–454.

[86]   L. Naldi, L. Chatenoud, D. Linder, A. Belloni Fortina, A. Peserico, A. R. Virgili, P. L. Bruni, V. Ingordo, G. Lo Scocco, C. Solaroli, D. Schena, A. Barba, A. Di Landro, E. Pezzarossa, F. Arcangeli, C. Gianni, R. Betti, P. Carli, A. Farris, G. F. Barabino, and C. La Vecchia, "Cigarette smoking, body mass index, and stressful life events as risk factors for psoriasis: results from an Italian case-control study," *J. Invest. Dermatol.*, vol. 125, no. 1, pp. 61–67, Jul. 2005.

[87]   J.-W. Han, H.-F. Zheng, Y. Cui, L.-D. Sun, D.-Q. Ye, Z. Hu, J.-H. Xu, Z.-M. Cai, W. Huang, G.-P. Zhao, H.-F. Xie, H. Fang, Q.-J. Lu, J.-H. Xu, X.-P. Li, Y.-F. Pan, D.-Q. Deng, F.-Q. Zeng, Z.-Z. Ye, X.-Y. Zhang, Q.-W. Wang, F. Hao, L. Ma, X.-B. Zuo, F.-S. Zhou, W.-H. Du, Y.-L. Cheng, J.-Q. Yang, S.-K. Shen, J. Li, Y.-J. Sheng, X.-X. Zuo, W.-F. Zhu, F. Gao, P.-L. Zhang, Q. Guo, B. Li, M. Gao, F.-L. Xiao, C. Quan, C. Zhang, Z. Zhang, K.-J. Zhu, Y. Li, D.-Y. Hu, W.-S. Lu, J.-L. Huang, S.-X. Liu, H. Li, Y.-Q. Ren, Z.-X. Wang, C.-J. Yang, P.-G. Wang, W.-M. Zhou, Y.-M. Lv, A.-P. Zhang, S.-Q. Zhang, D. Lin, Y. Li, H. Q. Low, M. Shen, Z.-F. Zhai, Y. Wang, F.-Y. Zhang, S. Yang, J.-J. Liu, and X.-J. Zhang, "Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus," *Nat. Genet.*, vol. 41, no. 11, pp. 1234–1237, Nov. 2009.

[88]   A. Franke, T. Balschun, C. Sina, D. Ellinghaus, R. Häsler, G. Mayr, M. Albrecht, M. Wittig, E. Buchert, S. Nikolaus, C. Gieger, H. E. Wichmann, J. Sventoraityte, L. Kupcinskas, C. M. Onnie, M. Gazouli, N. P. Anagnou, D. Strachan, W. L. McArdle, C. G. Mathew, P. Rutgeerts, S. Vermeire, M. H. Vatn, M. Krawczak, P. Rosenstiel, T. H. Karlsen, and S. Schreiber, "Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL)," *Nat. Genet.*, vol. 42, no. 4, pp. 292–294, Apr. 2010.

[89] H. Renz, E. von Mutius, P. Brandtzaeg, W. O. Cookson, I. B. Autenrieth, and D. Haller, "Gene-environment interactions in chronic inflammatory disease," *Nat. Immunol.*, vol. 12, no. 4, pp. 273–277, Apr. 2011.

[90] S. Sawcer, H. B. Jones, R. Feakes, J. Gray, N. Smaldon, J. Chataway, N. Robertson, D. Clayton, P. N. Goodfellow, and A. Compston, "A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22," *Nat. Genet.*, vol. 13, no. 4, pp. 464–468, Aug. 1996.

[91] J. L. Haines, M. Ter-Minassian, A. Bazyk, J. F. Gusella, D. J. Kim, H. Terwedow, M. A. Pericak-Vance, J. B. Rimmler, C. S. Haynes, A. D. Roses, A. Lee, B. Shaner, M. Menold, E. Seboun, R. P. Fitoussi, C. Gartioux, C. Reyes, F. Ribierre, G. Gyapay, J. Weissenbach, S. L. Hauser, D. E. Goodkin, R. Lincoln, K. Usuku, and J. R. Oksenberg, "A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group," *Nat. Genet.*, vol. 13, no. 4, pp. 469–471, Aug. 1996.

[92] G. C. Ebers, K. Kukay, D. E. Bulman, A. D. Sadovnick, G. Rice, C. Anderson, H. Armstrong, K. Cousin, R. B. Bell, W. Hader, D. W. Paty, S. Hashimoto, J. Oger, P. Duquette, S. Warren, T. Gray, P. O'Connor, A. Nath, A. Auty, L. Metz, G. Francis, J. E. Paulseth, T. J. Murray, W. Pryse-Phillips, R. Nelson, M. Freedman, D. Brunet, J. P. Bouchard, D. Hinds, and N. Risch, "A full genome search in multiple sclerosis," *Nat. Genet.*, vol. 13, no. 4, pp. 472–476, Aug. 1996.

[93] L. Steinman, "Multiple sclerosis: a two-stage disease," *Nat. Immunol.*, vol. 2, no. 9, pp. 762–764, Sep. 2001.

[94] A. S. Rigby, A. J. Silman, L. Voelm, J. C. Gregory, W. E. Ollier, M. A. Khan, G. T. Nepom, and G. Thomson, "Investigating the HLA component in rheumatoid arthritis: an additive (dominant) mode of inheritance is rejected, a recessive mode is preferred," *Genet. Epidemiol.*, vol. 8, no. 3, pp. 153–175, 1991.

[95] M. Feldmann, "Pathogenesis of arthritis: recent research progress," *Nat. Immunol.*, vol. 2, no. 9, pp. 771–773, Sep. 2001.

[96] A. J. Silman, A. J. MacGregor, W. Thomson, S. Holligan, D. Carthy, A. Farhan, and W. E. Ollier, "Twin concordance rates for rheumatoid arthritis: results from a nationwide study," *Br. J. Rheumatol.*, vol. 32, no. 10, pp. 903–907, Oct. 1993.

[97] A. Svejgaard, P. Platz, and L. P. Ryder, "HLA and disease 1982--a survey," *Immunol. Rev.*, vol. 70, pp. 193–218, 1983.

[98] W. R. Gammon, E. R. Heise, W. A. Burke, J. D. Fine, D. T. Woodley, and R. A. Briggaman, "Increased frequency of HLA-DR2 in patients with autoantibodies to epidermolysis bullosa acquisita antigen: evidence that the expression of autoimmunity to type VII collagen is HLA class II allele associated," *J. Invest. Dermatol.*, vol. 91, no. 3, pp. 228–232, Sep. 1988.

[99] R. J. Ludwig, S. Müller, A. D. C. Marques, A. Recke, E. Schmidt, D. Zillikens, S. Möller, and S. M. Ibrahim, "Identification of quantitative trait Loci in experimental epidermolysis bullosa acquisita," *J. Invest. Dermatol.*, vol. 132, no. 5, pp. 1409–1415, May 2012.

[100] E. F. Remmers, R. E. Longman, Y. Du, A. O'Hare, G. W. Cannon, M. M. Griffiths, and R. L. Wilder, "A genome scan localizes five non–MHC loci controlling collagen–induced arthritis in rats," *Nature Genetics*, vol. 14, no. 1, pp. 82–85, Sep. 1996.

[101] A. Wandstrat and E. Wakeland, "The genetics of complex autoimmune diseases: non-MHC susceptibility genes," *Nat. Immunol.*, vol. 2, no. 9, pp. 802–809, Sep. 2001.

[102] Y. K. Lee and S. K. Mazmanian, "Has the Microbiota Played a Critical Role in the Evolution of the Adaptive Immune System?," *Science*, vol. 330, no. 6012, pp. 1768–1773, Dec. 2010.

[103] F. Asghari, B. Fitzner, S.-A. Holzhuter, H. Nizze, A. de Castro Marques, S. Muller, S. Moller, S. M. Ibrahim, and R. Jaster, "Identification of quantitative trait loci for murine autoimmune pancreatitis," *Journal of Medical Genetics*, vol. 48, no. 8, pp. 557–562, Jun. 2011.

[104] C. Sitaru, S. Mihai, C. Otto, M. T. Chiriac, I. Hausser, B. Dotterweich, H. Saito, C. Rose, A. Ishiko, and D. Zillikens, "Induction of dermal-epidermal separation in mice by passive transfer of antibodies specific to type VII collagen.," *J Clin Invest*, vol. 115, pp. 870–878, 2005.

[105] J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight, "QIIME allows analysis of high-throughput community sequencing data," *Nature Methods*, vol. 7, no. 5, pp. 335–336, Apr. 2010.

[106] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *J. Mol. Biol.*, vol. 147, no. 1, pp. 195–197, Mar. 1981.

[107] R. C. Edgar, "Search and clustering orders of magnitude faster than BLAST," *Bioinformatics*, vol. 26, no. 19, pp. 2460–2461, Aug. 2010.

[108] J. G. Caporaso, K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen, and R. Knight, "PyNAST: a flexible tool for aligning sequences to a template alignment," *Bioinformatics*, vol. 26, no. 2, pp. 266–267, Nov. 2009.

[109] T. Z. DeSantis, P. Hugenholtz, K. Keller, E. L. Brodie, N. Larsen, Y. M. Piceno, R. Phan, and G. L. Andersen, "NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes," *Nucleic Acids Research*, vol. 34, no. Web Server, pp. W394–W399, Jul. 2006.

[110] C. Shannon, "The mathematical theory of communicaton," *University of Illinois Press; Urbana*, vol. Il, 1949.

[111] A. Chao, "Non-parametric estimation of the number of classes in a population.," *Scand J Stat.*, no. 11, pp. 265–270, 1984.

[112] *Biological diversity: frontiers in measurement and assessment*. Oxford ; New York: Oxford University Press, 2011.

[113] D. P. Faith and A. M. Baker, "Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges," *Evol. Bioinform. Online*, vol. 2, pp. 121–128, 2006.

[114] M. N. Price, P. S. Dehal, and A. P. Arkin, "FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments," *PLoS ONE*, vol. 5, no. 3, p. e9490, Mar. 2010.

[115] P. Dixon, "VEGAN, a package of R functions for community ecology," *Journal of Vegetation Science*, vol. 14, no. 6, pp. 927–930, Dec. 2003.

[116] M. Hamady, C. Lozupone, and R. Knight, "Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data," *ISME J*, vol. 4, no. 1, pp. 17–27, Jan. 2010.

[117] Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole, "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy," *Appl. Environ. Microbiol.*, vol. 73, no. 16, pp. 5261–5267, Aug. 2007.

[118] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.

[119] R. Mott, C. J. Talbot, M. G. Turri, A. C. Collins, and J. Flint, "A method for fine mapping quantitative trait loci in outbred animal stocks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, no. 23, pp. 12649–12654, Nov. 2000.

[120] W. Valdar, "Simulating the Collaborative Cross: Power of Quantitative Trait Loci Detection and Mapping Resolution in Large Sets of Recombinant Inbred Strains of Mice," *Genetics*, vol. 172, no. 3, pp. 1783–1797, Dec. 2005.

[121] W. Valdar, L. C. Solberg, D. Gauguier, S. Burnett, P. Klenerman, W. O. Cookson, M. S. Taylor, J. N. P. Rawlins, R. Mott, and J. Flint, "Genome-wide genetic association of complex traits in heterogeneous stock mice," *Nature Genetics*, vol. 38, no. 8, pp. 879–887, Jul. 2006.

[122] J. Dupuis and D. Siegmund, "Statistical methods for mapping quantitative trait loci from a dense set of markers," *Genetics*, vol. 151, no. 1, pp. 373–386, Jan. 1999.

[123] A. M. Graham, M. D. Munday, O. Kaftanoglu, R. E. Page Jr, G. V. Amdam, and O. Rueppell, "Support for the reproductive ground plan hypothesis of social evolution and major QTL for ovary traits of Africanized worker honey bees (Apis mellifera L.)," *BMC Evol. Biol.*, vol. 11, p. 95, 2011.

[124] D. Gevers, F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. V. de Peer, P. Vandamme, F. L. Thompson, and J. Swings, "Opinion: Re-evaluating prokaryotic species," *Nature Reviews Microbiology*, vol. 3, no. 9, pp. 733–739, Aug. 2005.

[125] Z. Gao, C. Tseng, Z. Pei, and M. J. Blaser, "Molecular analysis of human forearm superficial skin bacterial biota," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, no. 8, pp. 2927–2932, Feb. 2007.

[126] C. Lozupone and R. Knight, "UniFrac: a new phylogenetic method for comparing microbial communities," *Appl. Environ. Microbiol.*, vol. 71, no. 12, pp. 8228–8235, Dec. 2005.

[127] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing.," *Journal of the Royal Statistical Society*, vol. Ser. B 57, pp. 289–289, 1995.

[128] A. M. McKnite, M. E. Perez-Munoz, L. Lu, E. G. Williams, S. Brewer, P. A. Andreux, J. W. M. Bastiaansen, X. Wang, S. D. Kachman, J. Auwerx, R. W. Williams, A. K. Benson, D. A. Peterson, and D. C. Ciobanu, "Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits," *PLoS ONE*, vol. 7, no. 6, p. e39191, Jun. 2012.

[129] B. Stecher and W.-D. Hardt, "The role of microbiota in infectious disease," *Trends Microbiol.*, vol. 16, no. 3, pp. 107–114, Mar. 2008.

[130] T. Hrncir, R. Stepankova, H. Kozakova, T. Hudcovic, and H. Tlaskalova-Hogenova, "Gut microbiota and lipopolysaccharide content of the diet influence development of regulatory T cells: studies in germ-free mice," *BMC Immunol.*, vol. 9, p. 65, 2008.

[131] I. Dekio, H. Hayashi, M. Sakamoto, M. Kitahara, T. Nishikawa, M. Suematsu, and Y. Benno, "Detection of potentially novel bacterial components of the human skin microbiota using culture-independent molecular profiling," *J. Med. Microbiol.*, vol. 54, no. Pt 12, pp. 1231–1238, Dec. 2005.

[132] L. C. Hadaway, "Skin flora and infection," *J Infus Nurs*, vol. 26, no. 1, pp. 44–48, Feb. 2003.

[133] A. L. Cogen, V. Nizet, and R. L. Gallo, "Skin microbiota: a source of disease or defence?," *British Journal of Dermatology*, vol. 158, no. 3, pp. 442–455, Feb. 2008.

[134] G. E. Austin and E. O. Hill, "Endocarditis due to Corynebacterium CDC group G2," *J. Infect. Dis.*, vol. 147, no. 6, p. 1106, Jun. 1983.

[135] V. Barakett, G. Morel, D. Lesage, and J. C. Petit, "Septic arthritis due to a nontoxigenic strain of Corynebacterium diphtheriae subspecies mitis," *Clin. Infect. Dis.*, vol. 17, no. 3, pp. 520–521, Sep. 1993.

[136] I. Poilane, F. Fawaz, M. Nathanson, P. Cruaud, T. Martin, A. Collignon, and J. Gaudelus, "Corynebacterium diphtheriae osteomyelitis in an immunocompetent child: a case report," *Eur. J. Pediatr.*, vol. 154, no. 5, pp. 381–383, May 1995.

[137] M. B. Coyle, N. B. Groman, J. Q. Russell, J. P. Harnisch, M. Rabin, and K. K. Holmes, "The molecular epidemiology of three biotypes of Corynebacterium diphtheriae in the Seattle outbreak, 1972-1982," *J. Infect. Dis.*, vol. 159, no. 4, pp. 670–679, Apr. 1989.

[138] A. Jucglà, G. Sais, J. Carratala, A. Moreno, A. Fernandez-Sevilla, and J. Peyri, "A papular eruption secondary to infection with Corynebacterium jeikeium, with histopathological features mimicking botryomycosis," *Br. J. Dermatol.*, vol. 133, no. 5, pp. 801–804, Nov. 1995.

[139] H. van der Lelie, M. Leverstein-Van Hall, M. Mertens, H. C. van Zaanen, R. H. van Oers, B. L. Thomas, A. E. von dem Borne, and E. J. Kuijper, "Corynebacterium CDC group JK (Corynebacterium jeikeium) sepsis in haematological patients: a report of three cases and a systematic literature review," *Scand. J. Infect. Dis.*, vol. 27, no. 6, pp. 581–584, 1995.

[140] M. B. Coyle and B. A. Lipsky, "Coryneform bacteria in infectious diseases: clinical and laboratory aspects," *Clin. Microbiol. Rev.*, vol. 3, no. 3, pp. 227–246, Jul. 1990.

[141] E. L. Larson, K. J. McGinley, J. J. Leyden, M. E. Cooley, and G. H. Talbot, "Skin colonization with antibiotic-resistant (JK group) and antibiotic-sensitive lipophilic diphtheroids in hospitalized and normal adults," *J. Infect. Dis.*, vol. 153, no. 4, pp. 701–706, Apr. 1986.

[142] C. von Eiff, K. Becker, K. Machka, H. Stammer, and G. Peters, "Nasal carriage as a source of Staphylococcus aureus bacteremia. Study Group," *N. Engl. J. Med.*, vol. 344, no. 1, pp. 11–16, Jan. 2001.

[143] E. M. Hale and R. D. Hinsdill, "Biological activity of staphylococcin 162: bacteriocin from Staphylococcus aureus," *Antimicrob. Agents Chemother.*, vol. 7, no. 1, pp. 74–81, Jan. 1975.

[144] G. Bierbaum, F. Götz, A. Peschel, T. Kupke, M. van de Kamp, and H. G. Sahl, "The biosynthesis of the lantibiotics epidermin, gallidermin, Pep5 and epilancin K7," *Antonie Van Leeuwenhoek*, vol. 69, no. 2, pp. 119–127, Feb. 1996.

[145] M. B. Ekkelenkamp, M. Hanssen, S.-T. Danny Hsu, A. de Jong, D. Milatovic, J. Verhoef, and N. A. J. van Nuland, "Isolation and structural characterization of epilancin 15X, a novel lantibiotic from a clinical strain of Staphylococcus epidermidis," *FEBS Lett.*, vol. 579, no. 9, pp. 1917–1922, Mar. 2005.

[146] H. G. Sahl, "Staphylococcin 1580 is identical to the lantibiotic epidermin: implications for the nature of bacteriocins from gram-positive bacteria," *Appl. Environ. Microbiol.*, vol. 60, no. 2, pp. 752–755, Feb. 1994.

[147] A. L. Cogen, K. Yamasaki, K. M. Sanchez, R. A. Dorschner, Y. Lai, D. T. MacLeod, J. W. Torpey, M. Otto, V. Nizet, J. E. Kim, and R. L. Gallo, "Selective antimicrobial action is provided by phenol-soluble modulins derived from Staphylococcus epidermidis, a normal resident of the skin," *J. Invest. Dermatol.*, vol. 130, no. 1, pp. 192–200, Jan. 2010.

[148] K. Suzuki, B. Meek, Y. Doi, M. Muramatsu, T. Chiba, T. Honjo, and S. Fagarasan, "Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 7, pp. 1981–1986, Feb. 2004.

[149] Y. K. Lee, J. S. Menezes, Y. Umesaki, and S. K. Mazmanian, "Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108 Suppl 1, pp. 4615–4622, Mar. 2011.

[150] M. Anitha, M. Vijay-Kumar, S. V. Sitaraman, A. T. Gewirtz, and S. Srinivasan, "Gut microbial products regulate murine gastrointestinal motility via Toll-like receptor 4 signaling," *Gastroenterology*, vol. 143, no. 4, pp. 1006–1016.e4, Oct. 2012.

[151] C. Ubeda, L. Lipuma, A. Gobourne, A. Viale, I. Leiner, M. Equinda, R. Khanin, and E. G. Pamer, "Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice," *Journal of Experimental Medicine*, vol. 209, no. 8, pp. 1445–1456, Jul. 2012.

[152] C. Picard, A. Puel, M. Bonnet, C.-L. Ku, J. Bustamante, K. Yang, C. Soudais, S. Dupuis, J. Feinberg, C. Fieschi, C. Elbim, R. Hitchcock, D. Lammas, G. Davies, A. Al-Ghonaium, H. Al-Rayes, S. Al-Jumaah, S. Al-Hajjar, I. Z. Al-Mohsen, H. H. Frayha, R. Rucker, T. R. Hawn, A. Aderem, H. Tufenkeji, S. Haraguchi, N. K. Day, R. A. Good, M.-A. Gougerot-Pocidalo, A. Ozinsky, and J.-L. Casanova, "Pyogenic bacterial infections in humans with IRAK-4 deficiency," *Science*, vol. 299, no. 5615, pp. 2076–2079, Mar. 2003.

[153] S. Akira, S. Uematsu, and O. Takeuchi, "Pathogen Recognition and Innate Immunity," *Cell*, vol. 124, no. 4, pp. 783–801, Feb. 2006.

[154] S. Rakoff-Nahoum, J. Paglino, F. Eslami-Varzaneh, S. Edberg, and R. Medzhitov, "Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis," *Cell*, vol. 118, no. 2, pp. 229–241, Jul. 2004.

[155] J. Schauber, Y. Oda, A. S. Büchau, Q.-C. Yun, A. Steinmeyer, U. Zügel, D. D. Bikle, and R. L. Gallo, "Histone Acetylation in Keratinocytes Enables Control of the Expression of Cathelicidin and CD14 by 1,25-Dihydroxyvitamin D3," *Journal of Investigative Dermatology*, vol. 128, no. 4, pp. 816–824, Oct. 2007.

[156] S. H. Han, J. H. Kim, M. Martin, S. M. Michalek, and M. H. Nahm, "Pneumococcal lipoteichoic acid (LTA) is not as potent as staphylococcal LTA in stimulating Toll-like receptor 2," *Infect. Immun.*, vol. 71, no. 10, pp. 5541–5548, Oct. 2003.

[157] T. Kusunoki, E. Hailman, T. S. Juan, H. S. Lichenstein, and S. D. Wright, "Molecules from Staphylococcus aureus that bind CD14 and stimulate innate immune responses," *J. Exp. Med.*, vol. 182, no. 6, pp. 1673–1682, Dec. 1995.

[158] N. W. J. Schröder, S. Morath, C. Alexander, L. Hamann, T. Hartung, U. Zähringer, U. B. Göbel, J. R. Weber, and R. R. Schumann, "Lipoteichoic acid (LTA) of Streptococcus pneumoniae and Staphylococcus aureus activates immune cells via Toll-like receptor (TLR)-2, lipopolysaccharide-binding protein (LBP), and CD14, whereas TLR-4 and MD-2 are not involved," *J. Biol. Chem.*, vol. 278, no. 18, pp. 15587–15594, May 2003.

[159] J. R. García-Lozano, B. Torres, O. Fernández, G. Orozco, A. Alvarez-Márquez, A. García, M. A. González-Gay, A. García, A. Núñez-Roldán, J. Martín, and M. F. González-Escribano, "Caspase 7 influences susceptibility to rheumatoid arthritis," *Rheumatology (Oxford)*, vol. 46, no. 8, pp. 1243–1247, Aug. 2007.

[160] V. H. Teixeira, L. Jacq, S. Lasbleiz, P. Hilliquin, C. R. Oliveira, F. Cornelis, and E. Petit-Teixeira, "Genetic and expression analysis of CASP7 gene in a European Caucasian population with rheumatoid arthritis," *J. Rheumatol.*, vol. 35, no. 10, pp. 1912–1918, Oct. 2008.

[161] A. Akhter, M. A. Gavrilin, L. Frantz, S. Washington, C. Ditty, D. Limoli, C. Day, A. Sarkar, C. Newland, J. Butchar, C. B. Marsh, M. D. Wewers, S. Tridandapani, T.-D. Kanneganti, and A. O. Amer, "Caspase-7 activation by the Nlrc4/Ipaf inflammasome restricts Legionella pneumophila infection," *PLoS Pathog.*, vol. 5, no. 4, p. e1000361, Apr. 2009.

[162] S. Conlan, H. H. Kong, and J. A. Segre, "Species-level analysis of DNA sequence data from the NIH Human Microbiome Project," *PLoS ONE*, vol. 7, no. 10, p. e47075, 2012.

[163] A. Fahlén, L. Engstrand, B. S. Baker, A. Powles, and L. Fry, "Comparison of bacterial microbiota in skin biopsies from normal and psoriatic skin," *Arch. Dermatol. Res.*, vol. 304, no. 1, pp. 15–22, Jan. 2012.

[164] J. Dicksved, J. Halfvarson, M. Rosenquist, G. Järnerot, C. Tysk, J. Apajalahti, L. Engstrand, and J. K. Jansson, "Molecular analysis of the gut microbiota of identical twins with Crohn's disease," *ISME J*, vol. 2, no. 7, pp. 716–727, Jul. 2008.

[165] T. R. Abrahamsson, H. E. Jakobsson, A. F. Andersson, B. Björkstén, L. Engstrand, and M. C. Jenmalm, "Low diversity of the gut microbiota in infants with atopic eczema," *J. Allergy Clin. Immunol.*, vol. 129, no. 2, pp. 434–440, 440.e1–2, Feb. 2012.

[166] F. A. Amaral, D. Sachs, V. V Costa, C. T. Fagundes, D. Cisalpino, T. M. Cunha, S. H. Ferreira, F. Q. Cunha, T. A. Silva, J. R. Nicoli, L. Q. Vieira, D. G. Souza, and M. M. Teixeira, "Commensal microbiota is fundamental for the development of inflammatory pain.," *Proc Natl Acad Sci U S A*, vol. 105, pp. 2193–2197, 2008.

[167] N. A. Johnston, R. A. Trammell, S. Ball-Kell, S. Verhulst, and L. A. Toth, "Assessment of immune activation in mice before and after eradication of mite infestation.," *J Am Assoc Lab Anim Sci*, vol. 48, pp. 371–377, 2009.

[168] H. H. Kong, J. Oh, C. Deming, S. Conlan, E. A. Grice, M. A. Beatson, E. Nomicos, E. C. Polley, H. D. Komarow, P. R. Murray, M. L. Turner, and J. A. Segre, "Temporal shifts in

the skin microbiome associated with disease flares and treatment in children with atopic dermatitis," *Genome Res.*, vol. 22, no. 5, pp. 850–859, May 2012.

[169] L. Fry, B. S. Baker, A. V. Powles, A. Fahlen, and L. Engstrand, "Is chronic plaque psoriasis triggered by microbiota in the skin?," *Br. J. Dermatol.*, Mar. 2013.

6. Appendices

6.1 Appendix A – Additional figures

Figure A.1: Heat map represents the abundances of the species OTUs of skin microbiota across 261 mice from generation four of AIL population.
Abundance levels are gradient color coded where yellow being high abundance and black indicates absent or low abundance taxa.
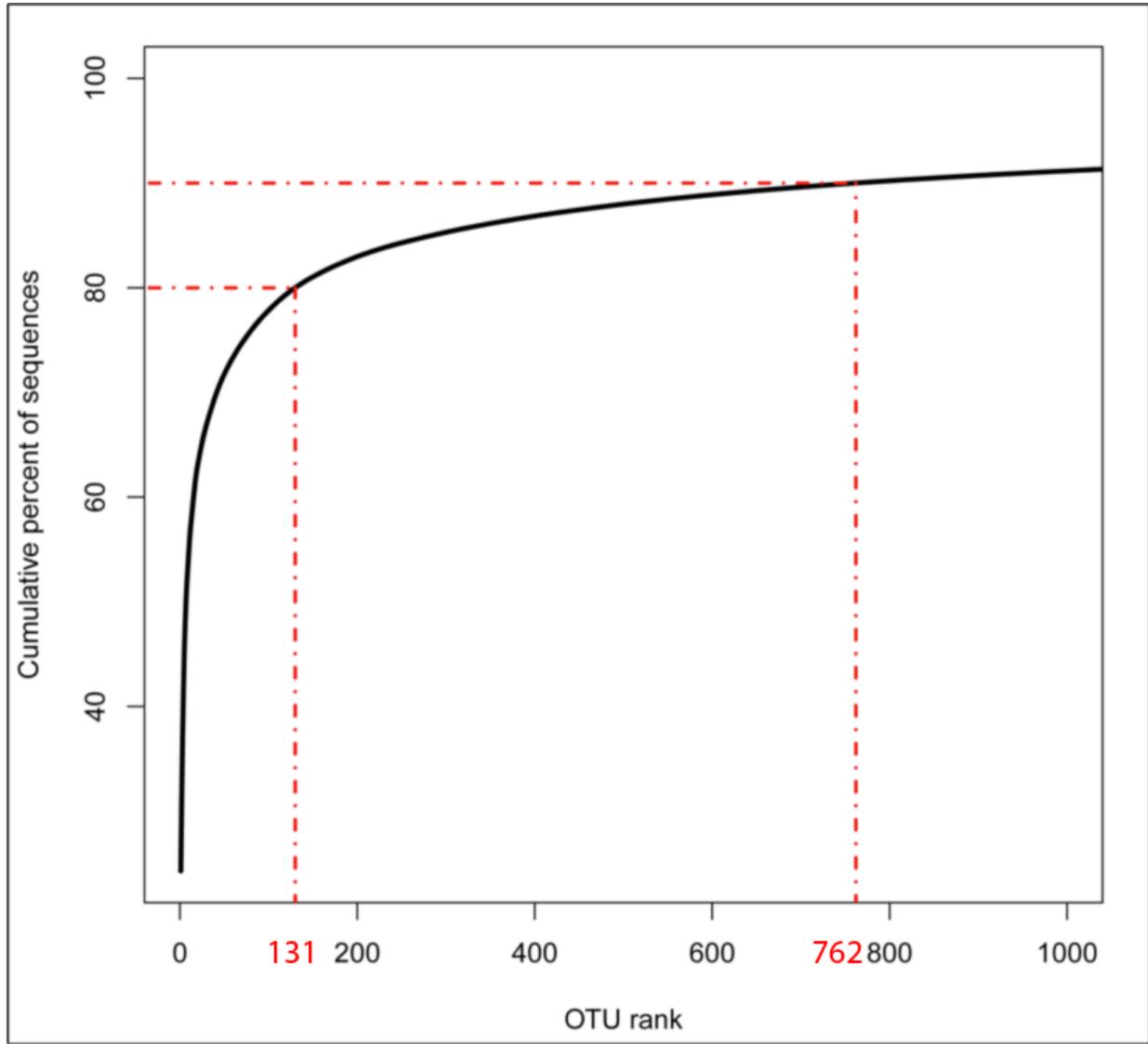
Figure A.2: Distribution of sequences among the 1000 most abundant species level OTUs
The X-axis indicates the individual OTUs ranked according to their relative abundance from high to low. The Y-axis indicates the cumulative percent of the total number of sequences. The dotted lines represent 80% and 90% of the total sequence reads from 261 samples, corresponding to the 131 and 762 most abundant OTUs, respectively.
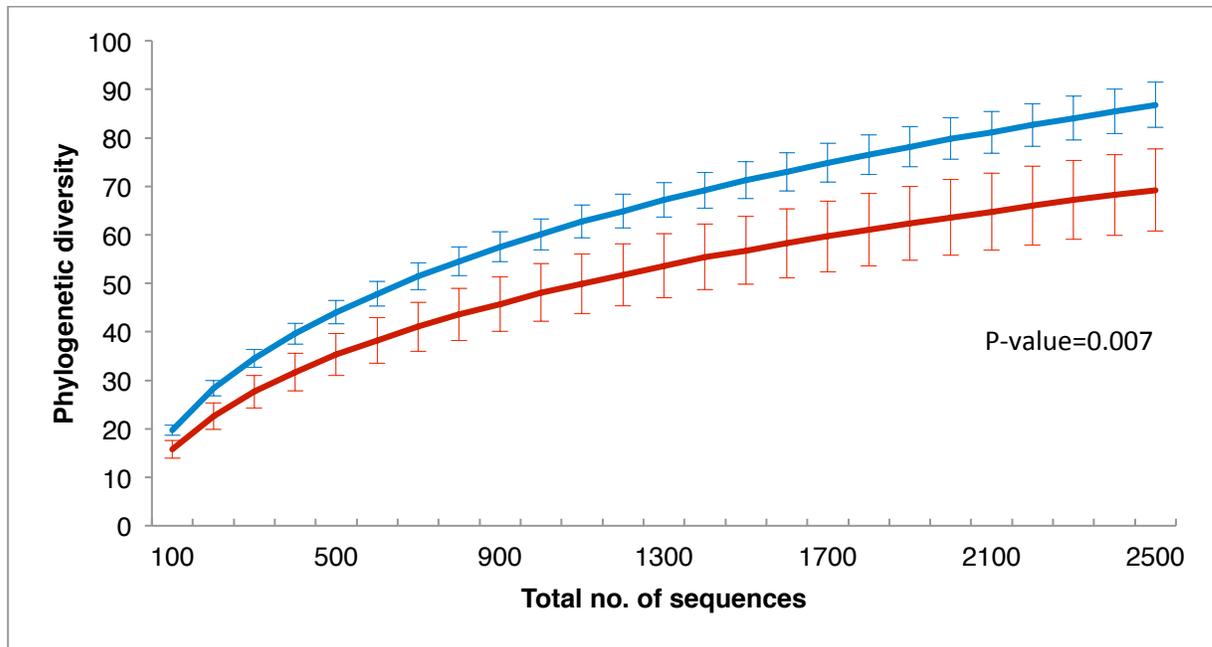
Figure A.3: Faiths Phylogenetic diversity alpha diversity measurement for species level OTUs in healthy (n = 119 in blue) and EBA (n = 64 in red) samples
Error bars represent the 95% confidence interval. P-value was determined by the Wilcoxon rank sum test in R.
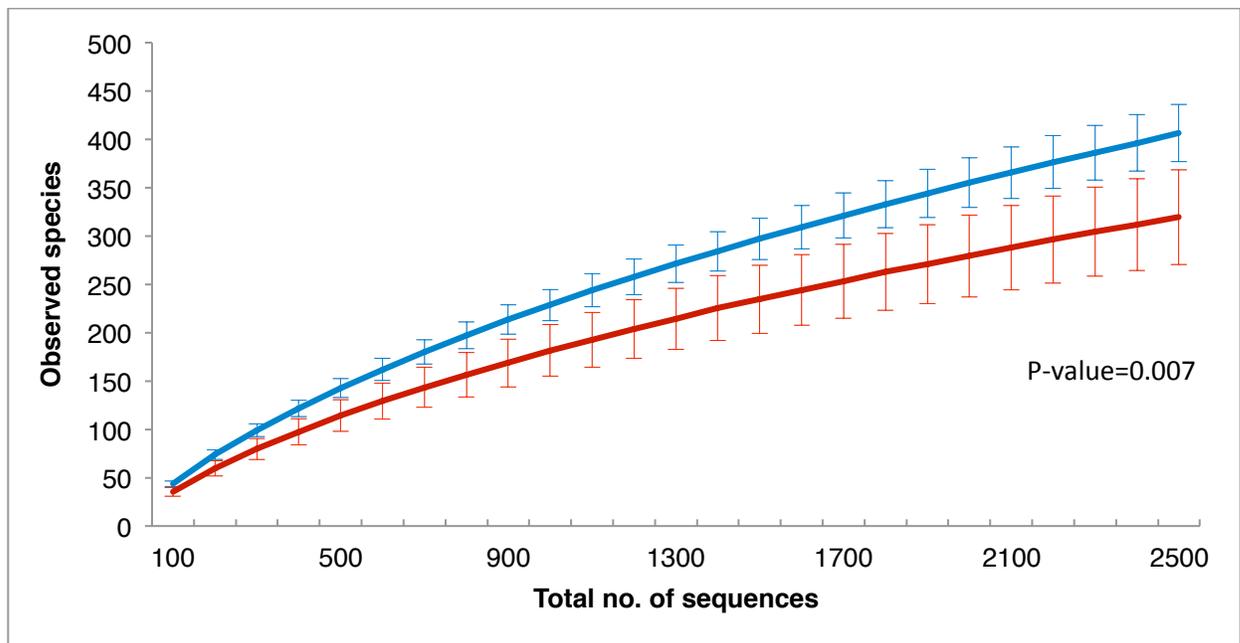
Figure A.4: Observed species alpha diversity measurement for species level OTUs in healthy (n = 119 in blue) and EBA (n = 64 in red) samples
Error bars represent the 95% confidence interval. P-value was determined by the Wilcoxon rank sum test in R.
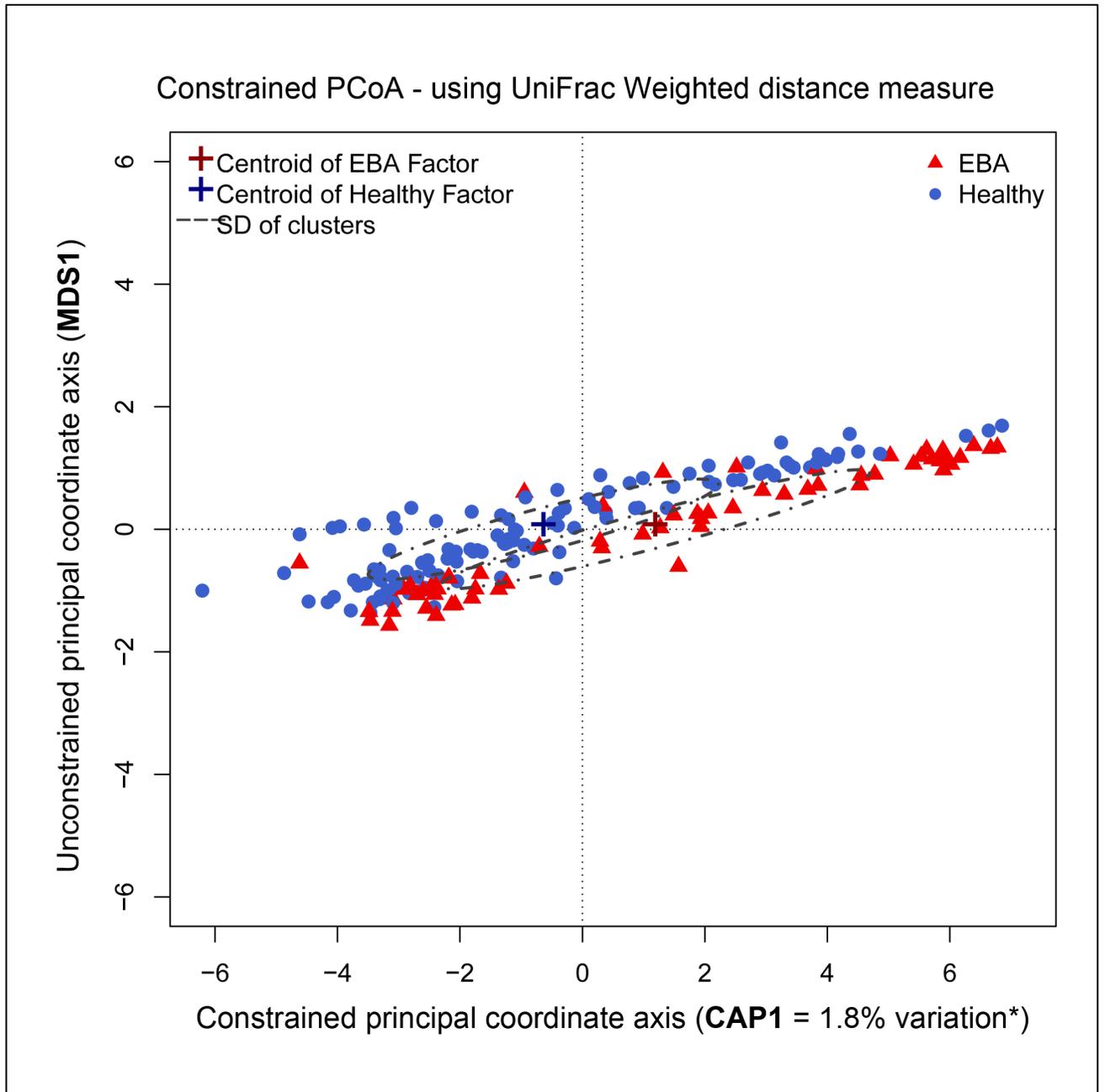
Figure A.5: Constrained analysis of principal coordinates of the weighted UniFrac metric using disease status as a constrained factor
*P-value = 0.015
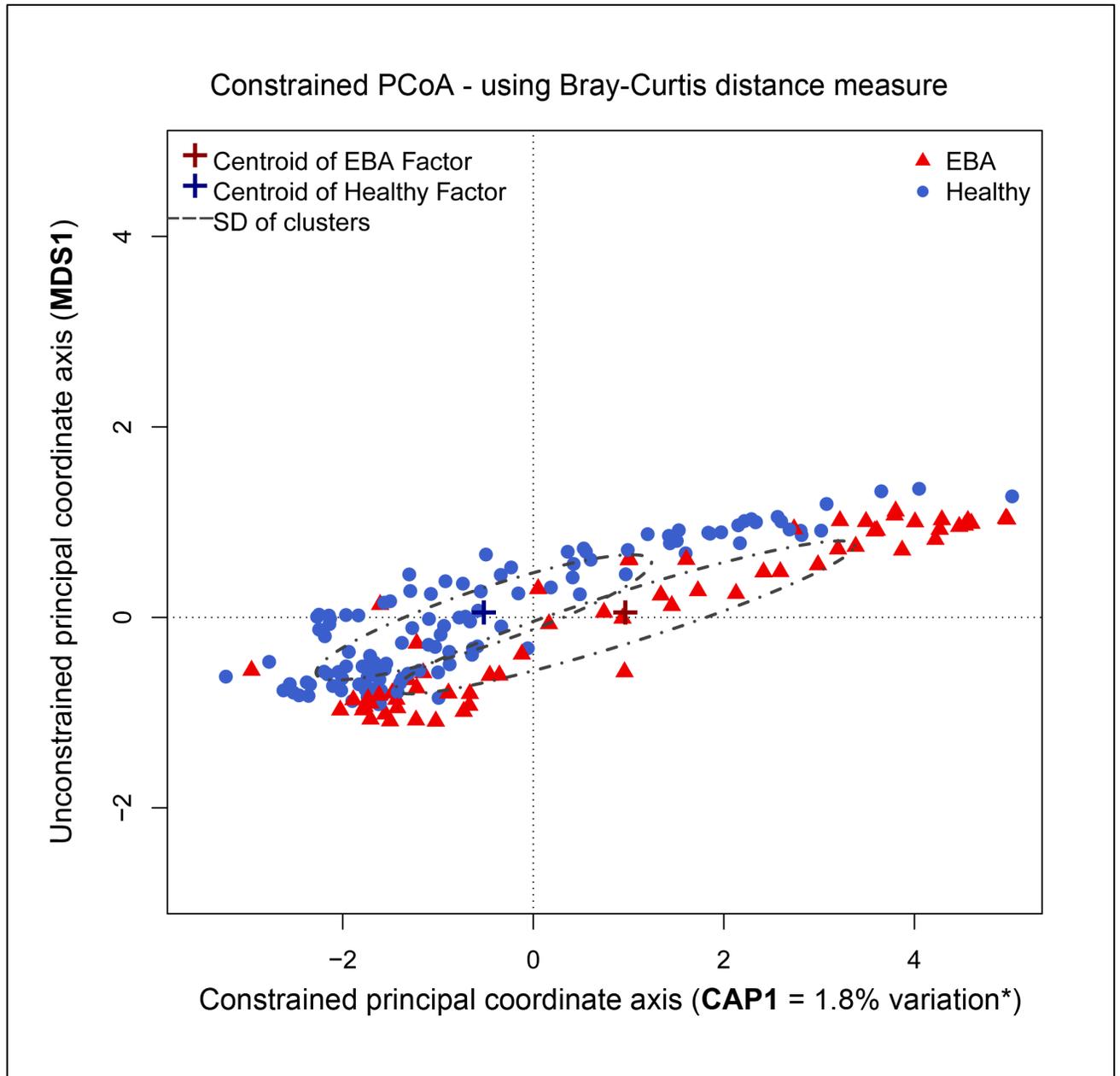
Figure A.6: Constrained analysis of principal coordinates of the Bray-Curtis index using disease status as a constrained factor
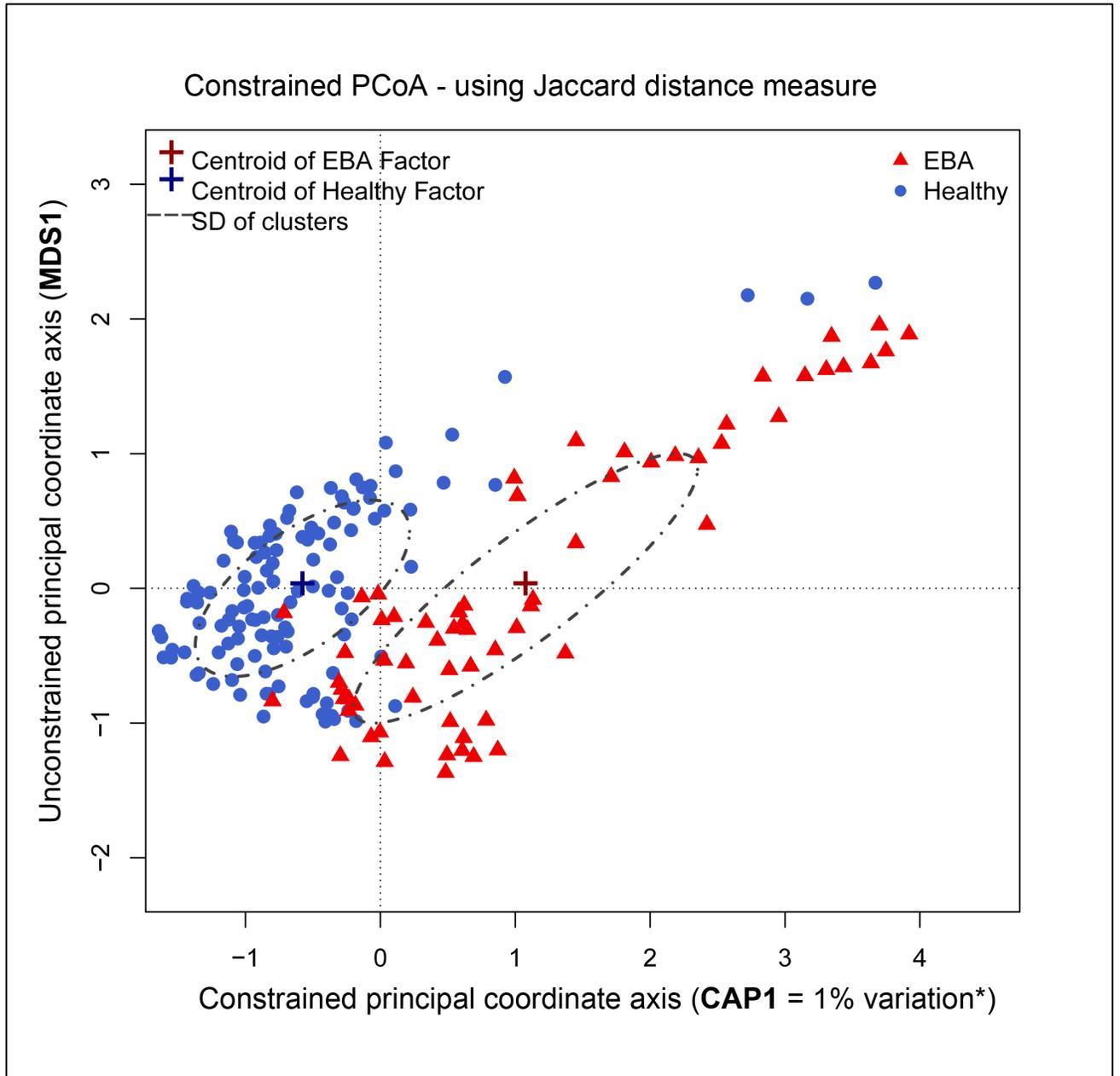
*P-value = 0.015

Figure A.7: Constrained analysis of principal coordinates of the Jaccard index using disease status as a constrained factor
*P-value = 0.005
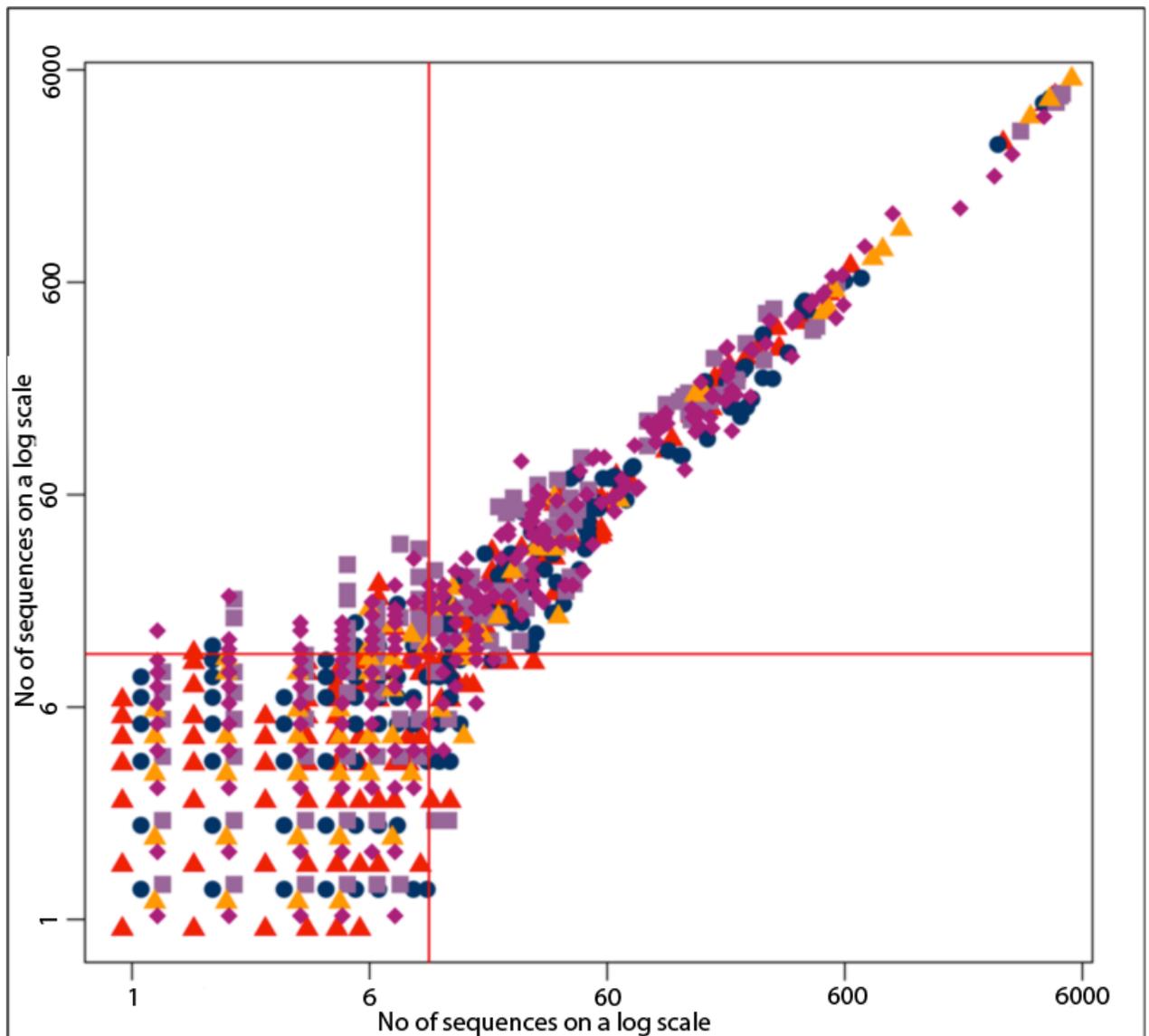
Figure A.8: Scatterplot from pairwise comparisons of taxonomic bins from technical repeats performed on five different samples

Each sample is displayed by a different color/symbol. The V1-V2 region of the 16s rRNA gene from each sample was amplified with two different sets of barcoded primers. The sequence data was processed and taxonomic assignments were performed (see Methods). Sequence counts for each taxonomic bin were log-transformed and plotted for all pairwise comparisons of the two technical repeats for each sample. Taxonomic bins falling above the red lines indicate those with a correlation >0.97 between replicates and correspond to having at least 20 reads per bin. Thus, OTUs with at least 20 reads per bin and occurring in at least 20 animals were chosen to comprise the "Core Measurable Microbiota" (CMM).
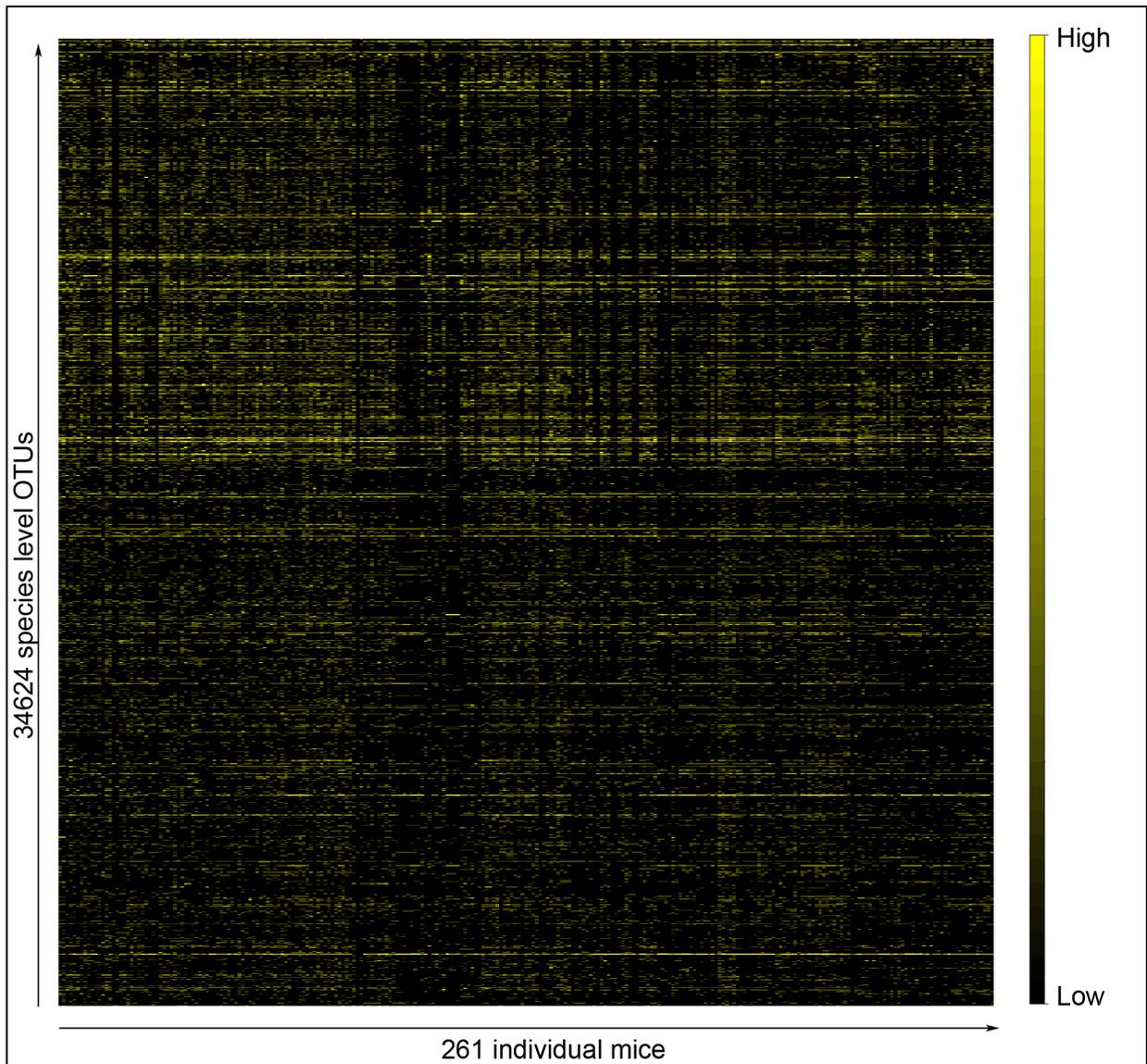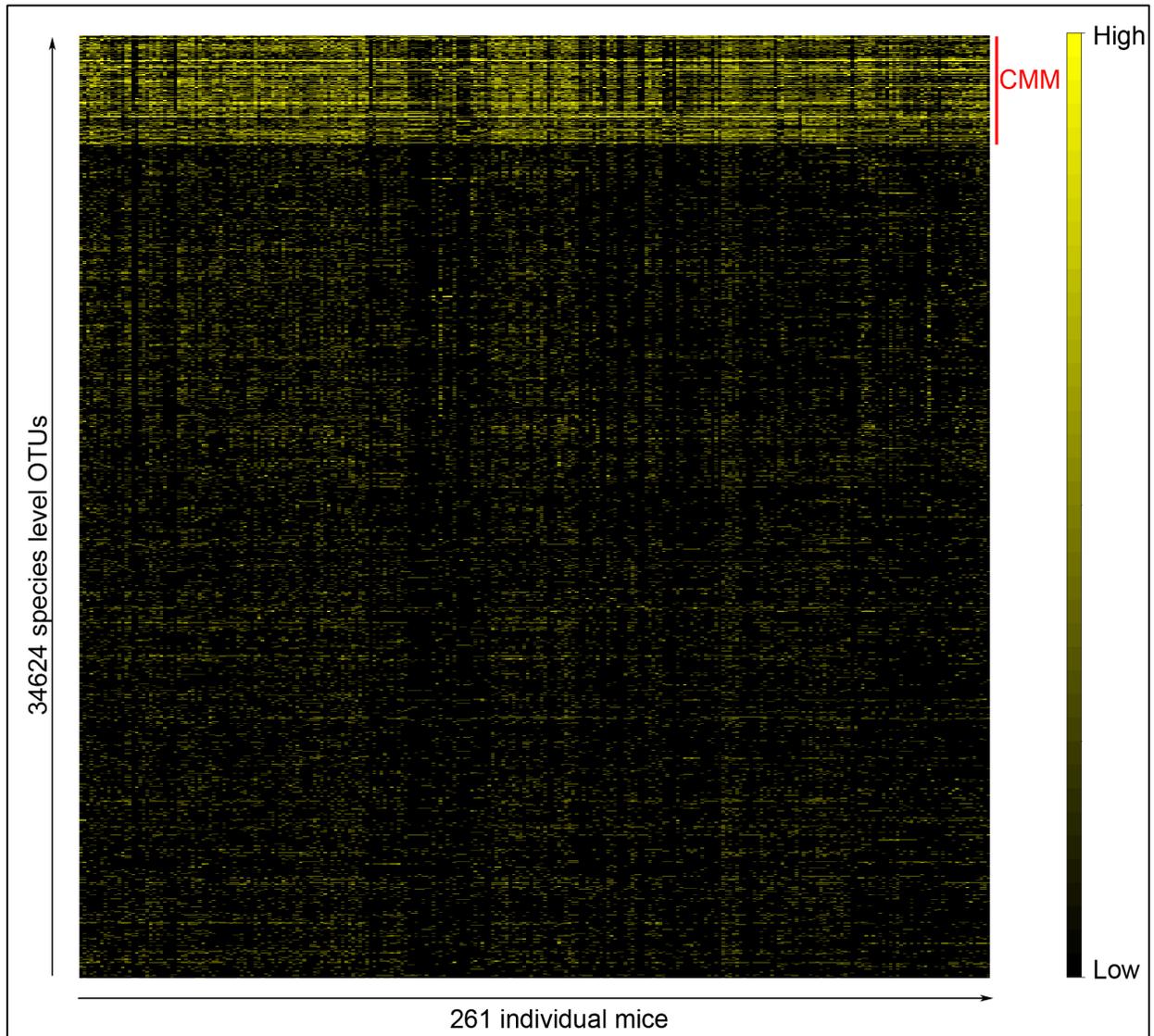
Figure A.9: Heat map represents the abundances of the CMM species OTUs of skin microbiota across 261 mice from generation four of AIL population
Abundance levels are gradient color coded where yellow being high abundance and black indicates absent or low abundance taxa. Most abundant quantitatively reproducible 131 CMM species OTUs across the AIL population is indicated by red line.

6.2 Appendix B – Additional tables

Table B.1: Indicator species analysis of EBA vs. healthy samples

| S.No	OTUID[‡]	Taxonomic Classification[δ]	Closely related species[°]
1	N2791	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
2	264597	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
3	N19632	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
4	N15384	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
5	N31433	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
6	530894	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
7	82696	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
8	470219	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*Corynebacterium tuberculostearicum*
9	273465	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Microbacterium	*unclassified species*
10	226812	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Nocardioidaceae;g__Propionicimonas	*Propionicimonas paludicola*
11	368907	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;g__Propionibacterium	*Propionibacterium acnes*
12	N2007	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
13	268155	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
14	N24810	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
15	N10167	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
16	182382	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
17	325622	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
18	174809	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides	*unclassified species*
19	237040	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus	*unclassified species*
20	N10292	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
21	538741	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
22	N19020	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
23	92651	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
24	157775	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus aureus*
25	N13471	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
26	173469	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
27	**N22308**	**k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus**	***Staphylococcus equorum***
28	18951	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*Streptococcus pneumoniae*
29	108747	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*Streptococcus thermophilus*
30	N26397	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria	*unclassified species*
31	294146	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria	*unclassified species*
32	293030	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Ralstonia	*unclassified species*
33	243860	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Ralstonia	*unclassified species*
34	N26135	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae	*unclassified species*
35	269548	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	*Helicobacter hepaticus*
36	535160	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter	*Acinetobacter johnsonii*
37	66816	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas	*unclassified species*
38	181076	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas	*unclassified species*
39	N5309	k__Bacteria;p__Thermi;c__Deinococci;o__Deinococcales;f__Deinococcaceae;g__Deinococcus	*unclassified species*

Table B.1 (continued)

S.No	OTUID[‡]	Healthy samples (%Average relative abundance)	EBA samples (%Average relative abundance)	Standard Deviation of Healthy samples	Standard Deviation of EBA samples	P-value (without correction)	P-value (Benjamini and Hochberg-adjusted)	More abundance direction
1	N2791	5.02	2.26	11.73	7.98	0.001	0.01	Healthy ↑
2	264597	0.14	0.07	0.16	0.12	0.002	0.013	Healthy ↑
3	N19632	0.08	0.03	0.12	0.17	0.002	0.013	Healthy ↑
4	N15384	0.07	0.04	0.11	0.08	0.003	0.016	Healthy ↑
5	N31433	0.17	0.07	0.22	0.39	0.002	0.013	Healthy ↑
6	530894	0.13	0.1	0.19	0.19	0.007	0.028	Healthy ↑
7	82696	0.11	0.08	0.16	0.18	0.012	0.041	Healthy ↑
8	470219	0.52	0.46	1.53	1.66	0.009	0.034	Healthy ↑
9	273465	0.04	0.01	0.05	0.16	0.005	0.022	Healthy ↑
10	226812	0.1	0.02	0.47	0.09	0.001	0.01	Healthy ↑
11	368907	1.03	0.75	1.04	1.13	0.001	0.01	Healthy ↑
12	N2007	0.13	0.08	0.18	0.17	0.001	0.01	Healthy ↑
13	268155	0.47	0.27	0.54	0.33	0.001	0.01	Healthy ↑
14	N24810	0.2	0.1	0.46	0.13	0.001	0.01	Healthy ↑
15	N10167	0.22	0.15	0.26	0.26	0.004	0.019	Healthy ↑
16	182382	0.16	0.09	0.24	0.11	0.009	0.034	Healthy ↑
17	325622	0.13	0.09	0.18	0.18	0.013	0.044	Healthy ↑
18	174809	0.12	0.07	0.18	0.11	0.002	0.013	Healthy ↑
19	237040	0.26	0.09	0.5	0.34	0.001	0.01	Healthy ↑
20	N10292	0.12	0.09	0.25	0.33	0.002	0.013	Healthy ↑
21	538741	0.31	0.29	0.41	0.64	0.004	0.019	Healthy ↑
22	N19020	0.14	0.08	0.18	0.12	0.003	0.016	Healthy ↑
23	92651	1.98	1.77	3.74	3.45	0.014	0.046	Healthy ↑
24	157775	0.27	0.18	0.46	0.35	0.002	0.013	Healthy ↑
25	N13471	0.13	0.1	0.14	0.17	0.005	0.022	Healthy ↑
26	173469	1.37	1.03	2.26	1.34	0.004	0.019	Healthy ↑
27	**N22308**	**0.17**	**0.27**	**0.18**	**0.26**	**0.016**	**0.05**	**EBA ↑**
28	18951	0.21	0.16	0.28	0.25	0.003	0.016	Healthy ↑
29	108747	1.21	0.73	1.79	1.77	0.001	0.01	Healthy ↑

S.No	OTUID[‡]	Healthy samples (%Average relative abundance)	EBA samples (%Average relative abundance)	Standard Deviation of Healthy samples	Standard Deviation of EBA samples	P-value (without correction)	P-value (Benjamini and Hochberg-adjusted)	More abundance direction
30	N26397	0.6	0.26	1.6	0.43	0.002	0.013	Healthy ↑
31	294146	0.7	0.26	2.17	0.46	0.003	0.016	Healthy ↑
32	293030	0.97	0.92	0.85	1.19	0.003	0.016	Healthy ↑
33	243860	6.92	6.1	6.31	6.95	0.01	0.036	Healthy ↑
34	N26135	0.41	0.09	1.56	0.37	0.001	0.01	Healthy ↑
35	269548	1.35	1.03	2.01	2.09	0.006	0.025	Healthy ↑
36	535160	0.28	0.16	0.4	0.31	0.001	0.01	Healthy ↑
37	66816	0.31	0.23	0.5	0.38	0.007	0.028	Healthy ↑
38	181076	0.11	0.1	0.16	0.22	0.001	0.01	Healthy ↑
39	N5309	0.22	0.13	0.35	0.22	0.001	0.01	Healthy ↑

‡ OTU IDs. The Ids for clusters displaying no match to the reference Greengenes database start with an "N"
δ Taxonomic classification from the RDP database and its prefixes k, p, c, o, f and g for kingdom, phylum, class, order, family and genus bacterial taxonomic   level, respectively. For OTUs that were unclassified, the next highest taxonomic level to which they could be classified is shown.
φ Closely related species were identified using the Greengenes database (see Methods) and were assigned using BLAST.
Species OTU having more abundance in EBA sample is shown in bold text.

Table B.2: Estimation of cage effect and family variances using linear mixed model method of CMM species level OTUs

S.No	OTUID[‡]	Taxonomic Classification[δ]	Closely related species[*]
1	N19632	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
2	N2791	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
3	N31433	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
4	530894	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
5	N16209	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
6	13405	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
7	N15384	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
8	82696	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	*unclassified species*
9	273465	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Microbacterium	*unclassified species*
10	226812	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Nocardioidaceae;g__Propionicimonas	*Propionicimonas paludicola*
11	591285	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium	*Bifidobacterium longum*
12	325622	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
13	N10167	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
14	N79	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
15	398943	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
16	264534	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
17	N24810	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
18	177453	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
19	182382	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
20	206324	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
21	469709	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides	*Bacteroides dorei*
22	356164	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides	*Bacteroides acidifaciens*
23	469832	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides	*Bacteroides uniformis*
24	190522	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella	*unclassified species*
25	N27927	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae	*unclassified species*
26	N10366	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes	*unclassified species*
27	309188	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes	*unclassified species*
28	167498	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes	*unclassified species*
29	263010	k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae	*unclassified species*
30	N14520	k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Chryseobacterium	*unclassified species*
31	560209	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	*unclassified species*
32	60254	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	*unclassified species*
33	N12197	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	*unclassified species*
34	N8891	k__Bacteria;p__Deferribacteres;c__Deferribacteres;o__Deferribacterales;f__Deferribacteraceae;g__Mucispirillum	*Mucispirillum schaedleri*
35	237040	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus	*unclassified species*
36	N18531	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Jeotgalicoccus	*unclassified species*
37	N6868	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*
38	N13471	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
39	N12039	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*
40	N29011	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
41	538741	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*

S.No	OTUID[‡]	Taxonomic Classification[δ]	Closely related species
42	92651	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
43	N2179	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
44	N15687	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*
45	N13661	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
46	N19525	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
47	120648	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*
48	173469	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
49	N2632	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*
50	N19020	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
51	157775	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus aureus*
52	N10292	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
53	N1906	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*
54	102524	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
55	N10459	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
56	276600	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Aerococcus	*unclassified species*
57	N4110	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Aerococcus	*unclassified species*
58	160031	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus	*Lactobacillus delbrueckii*
59	252321	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus	*Lactobacillus helveticus*
60	567604	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus	*unclassified species*
61	286668	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*unclassified species*
62	18951	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*Streptococcus pneumoniae*
63	108747	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*Streptococcus thermophilus*
64	N31208	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*unclassified species*
65	346400	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae	*unclassified species*
66	195445	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae	*unclassified species*
67	331878	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae	*unclassified species*
68	339014	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae	*unclassified species*
69	N19958	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae	*unclassified species*
70	328536	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Clostridium	*unclassified species*
71	381715	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae	*unclassified species*
72	N26397	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria	*unclassified species*
73	294146	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria	*unclassified species*
74	549374	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Acetobacteraceae;g__Acetobacter	*Acetobacter aceti*
75	243860	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Ralstonia	*unclassified species*
76	130241	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Herbaspirillum	*unclassified species*
77	N26135	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae	*unclassified species*
78	52884	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	*Helicobacter apodemus*
79	101810	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	*Helicobacter ganmani*
80	269548	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	*Helicobacter hepaticus*
81	143131	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae	*unclassified species*
82	288912	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter	*unclassified species*
83	66816	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas	*unclassified species*
84	308167	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas	*unclassified species*
85	231154	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas	*Pseudomonas geniculate*

S.No	OTUID[‡]	Taxonomic Classification[δ]	Closely related species[®]
86	470139	k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Clostridium	*Clostridium ramosum*
87	15711	k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Clostridium	*Clostridium cocleatum*
88	N5309	k__Bacteria;p__Thermi;c__Deinococci;o__Deinococcales;f__Deinococcaceae;g__Deinococcus	*unclassified species*

Table B.2 (continued)

S.No	OTUID[‡]	Variance			% Total variance explained*		
		Family	Cage	Residue	Family	Cage	Residue
1	N19632	0	0.1267	0.1387	0%	47.74%	52.26%
2	N2791	0.0001	0.4634	0.3696	0.01%	55.62%	44.36%
3	N31433	0	0.1662	0.1525	0%	52.15%	47.85%
4	530894	0.0227	0.0512	0.3019	6.04%	13.62%	80.34%
5	N16209	0.0003	0.0808	0.1891	0.11%	29.90%	69.99%
6	13405	0	0.1301	0.3132	0%	29.35%	70.65%
7	N15384	0	0.0975	0.1168	0%	45.50%	54.50%
8	82696	0	0.0809	0.2299	0%	26.03%	73.97%
9	273465	0.0152	0.0604	0.2006	5.50%	21.87%	72.63%
10	226812	0.0302	0.0624	0.2134	9.87%	20.39%	69.74%
11	591285	0	0.1791	0.3662	0%	32.84%	67.16%
12	325622	0	0.052	0.2816	0%	15.59%	84.41%
13	N10167	0	0.1401	0.2344	0%	37.41%	62.59%
14	N79	0	0.0952	0.2306	0%	29.22%	70.78%
15	398943	0.0077	0.0514	0.2183	2.78%	18.53%	78.70%
16	264534	0	0.0848	0.1961	0%	30.19%	69.81%
17	N24810	0.0568	0.0496	0.276	14.85%	12.97%	72.18%
18	177453	0.0029	0.0976	0.4144	0.56%	18.96%	80.48%
19	182382	0	0.0865	0.2235	0%	27.90%	72.10%
20	206324	0.0017	0.0416	0.2069	0.68%	16.63%	82.69%
21	469709	0.0353	0.0509	0.231	11.13%	16.05%	72.82%
22	356164	0.0642	0.0332	0.1855	22.69%	11.74%	65.57%
23	469832	0	0.1076	0.2346	0%	31.44%	68.56%
24	190522	0.0403	0.024	0.2131	14.53%	8.65%	76.82%
25	N27927	0.0263	0.0409	0.1713	11.03%	17.15%	71.82%
26	N10366	0.0388	0.0354	0.2213	13.13%	11.98%	74.89%
27	309188	0.0064	0.0723	0.2255	2.10%	23.77%	74.13%
28	167498	0	0.0577	0.3246	0%	15.09%	84.91%
29	263010	0.0406	0.0689	0.2902	10.16%	17.24%	72.60%
30	N14520	0.0007	0.0937	0.1592	0.28%	36.95%	62.78%
31	560209	0.0582	0.0375	0.2205	18.41%	11.86%	69.73%
32	60254	0.046	0.0652	0.269	12.10%	17.15%	70.75%
33	N12197	0	0.0727	0.2442	0%	22.94%	77.06%

S.No	OTUID[‡]	Variance			% Total variance explained*		
		Family	Cage	Residue	Family	Cage	Residue
34	N8891	0.0003	0.1197	0.2002	0.09%	37.38%	62.52%
35	237040	0.0001	0.2919	0.3206	0.02%	47.65%	52.33%
36	N18531	0	0.1379	0.1688	0%	44.96%	55.04%
37	N6868	0	0.1978	0.2407	0%	45.11%	54.89%
38	N13471	0	0.0959	0.1602	0%	37.45%	62.55%
39	N12039	0	0.1131	0.1603	0%	41.37%	58.63%
40	N29011	0.0004	0.1362	0.2516	0.10%	35.09%	64.81%
41	538741	0	0.1034	0.2704	0%	27.66%	72.34%
42	92651	0	0.1498	0.3397	0%	30.60%	69.40%
43	N2179	0.0001	0.048	0.1246	0.06%	27.79%	72.15%
44	N15687	0	0.0891	0.1535	0%	36.73%	63.27%
45	N13661	0	0.1294	0.1681	0%	43.50%	56.50%
46	N19525	0	0.0986	0.1841	0%	34.88%	65.12%
47	120648	0	0.1504	0.1977	0%	43.21%	56.79%
48	173469	0	0.1301	0.3254	0%	28.56%	71.44%
49	N2632	0	0.0526	0.1297	0%	28.85%	71.15%
50	N19020	0	0.1	0.1499	0%	40.02%	59.98%
51	157775	0	0.1449	0.2712	0%	34.82%	65.18%
52	N10292	0	0.077	0.2042	0%	27.38%	72.62%
53	N1906	0	0.0617	0.2143	0%	22.36%	77.64%
54	102524	0	0.1531	0.2797	0%	35.37%	64.63%
55	N10459	0	0.0838	0.1449	0%	36.64%	63.36%
56	276600	0	0.2862	0.2458	0%	53.80%	46.20%
57	N4110	0	0.0807	0.1084	0%	42.68%	57.32%
58	160031	0.0868	0.0881	0.4521	13.84%	14.05%	72.11%
59	252321	0.0384	0.0732	0.252	10.56%	20.13%	69.31%
60	567604	0	0.0944	0.2788	0%	25.29%	74.71%
61	286668	0.0093	0.0636	0.3263	2.33%	15.93%	81.74%
62	18951	0.089	0.0445	0.2638	22.40%	11.20%	66.40%
63	108747	0	0.1926	0.376	0%	33.87%	66.13%
64	N31208	0.0163	0.0507	0.2109	5.87%	18.24%	75.89%
65	346400	0	0.1126	0.2918	0%	27.84%	72.16%
66	195445	0.0029	0.039	0.2025	1.19%	15.96%	82.86%
67	331878	0	0.1772	0.2814	0%	38.64%	61.36%
68	339014	0	0.0932	0.2281	0%	29.01%	70.99%
69	N19958	0.0005	0.044	0.2256	0.19%	16.29%	83.52%
70	328536	0	0.0874	0.2283	0%	27.68%	72.32%
71	381715	0	0.0899	0.1782	0%	33.53%	66.47%
72	N26397	0	0.2232	0.2094	0%	51.60%	48.40%
73	294146	0	0.2295	0.2507	0%	47.79%	52.21%
74	549374	0	0.1106	0.2204	0%	33.41%	66.59%
75	243860	0.0102	0.059	0.2635	3.07%	17.73%	79.20%
76	130241	0	0.0241	0.3727	0%	6.07%	93.93%

S.No	OTUID[‡]	Variance			% Total variance explained*		
		Family	Cage	Residue	Family	Cage	Residue
77	N26135	0	0.1387	0.2505	0%	35.64%	64.36%
78	52884	0.0003	0.1635	0.253	0.07%	39.23%	60.70%
79	101810	0	0.2396	0.2272	0%	51.33%	48.67%
80	269548	0	0.2234	0.3299	0%	40.38%	59.62%
81	143131	0.0072	0.0328	0.2122	2.85%	13.01%	84.14%
82	288912	0.0033	0.046	0.2685	1.04%	14.47%	84.49%
83	66816	0	0.11	0.2874	0%	27.68%	72.32%
84	308167	0	0.1062	0.2107	0%	33.51%	66.49%
85	231154	0.0041	0.0426	0.2111	1.59%	16.52%	81.89%
86	470139	0	0.0292	0.2111	0%	12.15%	87.85%
87	15711	0.0001	0.0675	0.165	0.04%	29.02%	70.94%
88	N5309	0.0091	0.0624	0.3072	2.40%	16.48%	81.12%

‡ **OTU IDs. The Ids for clusters displaying no match to the reference Greengenes database start with an "N"**
δ **Taxonomic classification from the RDP database and its prefixes k, p, c, o, f and g for kingdom, phylum, class, order, family and genus bacterial taxonomic   level, respectively. For OTUs that were unclassified, the next highest taxonomic level to which they could be classified is shown.**
φ **Closely related species were identified using the Greengenes database (see Methods) and were assigned using BLAST.**
Species OTU having more abundance in EBA sample is shown in bold text.
***Percentage of total variance is the variance divided by the sum of the cage and residue variances**

Table B.3: QTLs detected for species level OTUs of the skin microbiota

spQTL list	OTUID[‡]	Taxonomic Classification[δ]	Closely related species[φ]
spQTL1	193038	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*
spQTL2	60254	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	*unclassified species*
spQTL3	294146	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria	*unclassified species*
spQTL4	314572	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter	*unclassified species*
spQTL5	N31208	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*unclassified species*
spQTL6	130241	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Herbaspirillum	*unclassified species*
spQTL7	N26684	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*
spQTL8	N10459	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*
spQTL9	N6868	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*

spQTL list	Nearest marker to peak	Chr	Peak position in Mb	Anova - log P score (LOD)	Genome wide significance	Confidence Interval (CI) in Mb using 1.5 - log P drop	% Phenotypic Variance[#]	QTL found using non-immunized plus immunized mice without clinical EBA mice (n=197)[¥]	QTL found only in immunized mice without clinical EBA mice (n=119)[¥]
spQTL1	rs6246342	2	150	4.23	E-value < 0.1	132-152	3.26	--	--
spQTL2	rs6241331	3	52	4.9	E-value < 0.05	46-55	2.8	yes	yes
spQTL3	rs3677770	4	55	4.23	E-value < 0.1	48-62	6.17	--	--
spQTL4	rs6320810	9	115	4.5	E-value < 0.05	106-	3.29	--	--
spQTL5	Gnf12.077.067	12	80	4.3	E-value < 0.1	74-85	7.54	--	--
spQTL6	rs13482712	15	92	4.28	E-value < 0.1	82-101	7.35	--	yes
spQTL7	rs13483244	18	21	5.24	E-value < 0.05	11-31	5.44	--	--
spQTL8	rs13483319	18	41	4.28	E-value < 0.1	33-46	2.88	--	--
spQTL9	rs3706601	18	78	4.14	E-value < 0.1	70-82	11.31	--	--

‡ OTU Ids. The Ids for clusters displaying no match to the reference Greengenes database start with an "N"

δ Taxonomic classification from the RDP database and its prefixes k, p, c, o, f and g for kingdom, phylum, class, order, family and genus bacterial taxonomic level, respectively. For OTUs that were unclassified, the next highest taxonomic level to which they could be classified is shown.

φ Closely related species were identified using the Greengenes database (see Methods) and were assigned using BLAST.

Percent of phenotypic variation of QTL was estimated using linear regression

¥ QTL found is determined by location of peak position within the confidence interval of spQTLs and it is not taxon specific

Table 1Table B.4: Skin microbiota QTLs detected from genus to phylum level

gpQTL list	Taxonomic Classification[δ]
gpQTL1	k___Bacteria;p__Bacteroidetes;c__Flavobacteria
	k___Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales
	k___Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae
	k___Bacteria;p__Cyanobacteria
	k___Bacteria;p__Cyanobacteria;c__Cyanobacteria
gpQTL2	k___Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Neisseria
gpQTL3	k___Bacteria;p__Firmicutes
gpQTL4	k___Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
gpQTL5	k___Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
gpQTL6	k___Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae
	k___Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
	k___Bacteria;p__Firmicutes;c__Clostridia
	k___Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales

Table B.4 (continued)

gpQTL list	Nearest marker to peak	Chr	Peak position in Mb	Anova - log *P* score (LOD)	Genome wide significance	Confidence Interval (CI) in Mb using 1.5 - log *P* drop	% Phenotypic Variance[#]	QTL seen in spQTL[φ]	Taxon specific QTL seen in spQTL[‡]	QTL found using non-immunized plus immunized mice without clinical EBA mice (n=197)[¥]	QTL found only in immunized mice without clinical EBA mice (n=119)[¥]
gpQTL1	rs6241331	3	52	4.1	E-value < 0.1	46-55	4.36	yes	--	--	--
	rs6241331	3	52	4.1	E-value < 0.1	46-55	4.36	yes	--	--	--
	rs6241331	3	52	4.28	E-value < 0.1	46-55	4.25	yes	--	--	--
	rs6241331	3	52	4.97	E-value < 0.05	46-55	3.59	yes	yes	--	--
	rs6241331	3	52	4.97	E-value < 0.05	46-55	2.73	yes	yes	--	--
gpQTL2	rs13482216	14	63	5.8	E-value < 0.05	56-69	4.13	--	--	yes	--
gpQTL3	CEL.14_73162771	14	82	4.14	E-value < 0.1	70-103	2.25	--	--	yes	--
gpQTL4	rs13483157	17	89	4.25	E-value < 0.1	76-95	2.15	--	--	--	--
gpQTL5	CEL.18_5565618	18	5	4.21	E-value < 0.1	-12	3.25	yes	--	--	yes
gpQTL6	CEL.X_8334947	X	10	4.47	E-value < 0.05	9 – 34	5.25	--	--	--	yes
	CEL.X_8334947	X	10	4.52	E-value < 0.05	9 – 34	5.22	--	--	--	yes
	CEL.X_8334947	X	10	4.2	E-value < 0.1	9 – 36	3.34	--	--	--	yes
	CEL.X_8334947	X	10	4.26	E-value < 0.1	9 – 36	3.32	--	--	--	yes

δ Taxonomic classification from the RDP database and its prefixes k, p, c, o, f and g for kingdom, phylum, class, order, family and genus bacterial taxonomic level, respectively. For OTUs that were unclassified, the next highest taxonomic level to which they could be classified is shown.

 # Percent of phenotypic variation of QTL was estimated using linear regression

φ QTL found is determined by location of QTL peak position within the confidence interval of spQTLs, but it is not taxon specific

‡ QTL found is determined by location of QTL peak position within the confidence interval of spQTLs. At least one of the taxonomic levels of the QTL matches with that of the spQTL.

¥ QTL found is determined by location of peak position within the confidence interval of gpQTLs, but it is not taxon specific

Table B.5: List of innate immunity related genes found within the confidence intervals of spQTLs

S.No	Gene related to innate immunity	Chr	Genome Coordinates NCBI Build 37 in Mb	cM	Gene name
1	Polr3f	2	144.35-144.37	71.02	polymerase (RNA) III (DNA directed) polypeptide F
2	Irak4	15	94.37-94.41	48.55	interleukin-1 receptor-associated kinase 4
3	Aqp4	18	15.55-15.56	8.74	aquaporin 4
4	Tmem173	18	35.89-35.90	19.23	transmembrane protein 173
5	Cd14	18	36.88-36.89	19.46	CD14 antigen

List of innate immunity related genes found within the confidence intervals of gpQTLs

6	Polr3d	14	70.84-70.84	36.32	polymerase (RNA) III (DNA directed) polypeptide D
7	Prkce	17	86.57-87.06	56.74	protein kinase C, epsilon
8	Colec12	18	9.71-9.88	4.91	collectin sub-family member 12
9	Cfp	20	20.50-20.51	16.44	complement factor properdin

S.No	Gene related to innate immunity	OTUID[‡]	Taxonomic Classification[◊]	Closely related species[◊]	Nearest marker to peak	Chr	Peak position in Mb	Anova - log P score (LOD)	Confidence Interval (CI) in Mb using 1.5 - log P drop
1	Polr3f	193038	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	*unclassified species*	rs6246342	2	132	4.23	150-152
2	Irak4	130241	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Herbaspirillum	*unclassified species*	rs13482712	15	82	4.28	92-101
3	Aqp4	N26684	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus equorum*	rs13483244	18	11	5.24	21-31
4	Tmem173	N10459	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*	rs13483319	18	33	4.28	41-46
5	Cd14	N10459	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*unclassified species*	rs13483319	18	33	4.28	41-46

List of innate immunity related genes found within the confidence intervals of gpQTLs

Table B.5 (continued)

S.No	Gene related to innate immunity	OTUID[‡]	Taxonomic Classification[δ]	Closely related species[φ]	Nearest marker to peak	Chr	Peak position in Mb	Anova - log *P* score (LOD)	Confidence Interval (CI) in Mb using 1.5 - log *P* drop
6	Polr3d	--	k___Bacteria;p__Firmicutes	--	CEL.14_73162 771	14	82	4.14	70-103
7	Prkce	--	k___Bacteria;p__Proteobacteria; c__Gammaproteobacteria;o__Enterobacteriale; f__Enterobacteriaceae	--	rs13483157	17	89	4.25	76-95
8	Colec12	--	k___Bacteria;p__Bacteroidetes;c__Bacteroidia; o__Bacteroidales;f__Prevotellaceae; g__Prevotella	--	CEL.18_55656 18	18	5	4.21	-12
9	Cfp	--	k___Bacteria;p__Firmicutes;c__Bacilli; o__Lactobacillales;f__Streptococcaceae	--	CEL.X_833494 7	X	10	4.47	9 – 34
		--	k___Bacteria;p__Firmicutes;c__Bacilli; o__Lactobacillales;f__Streptococcaceae; g__Streptococcus	--	CEL.X_833494 7	X	10	4.52	9 – 34
		--	k___Bacteria;p__Firmicutes;c__Clostridia	--	CEL.X_833494 7	X	10	4.2	9 – 36
		--	k___Bacteria;p__Firmicutes;c__Clostridia; o__Clostridiales	--	CEL.X_833494 7	X	10	4.26	9 – 36

[‡] **OTU Ids. The Ids for clusters displaying no match to the reference Greengenes database start with an "N"**

[δ] **Taxonomic classification from the RDP database and its prefixes k, p, c, o, f and g for kingdom, phylum, class, order, family and genus bacterial taxonomic level, respectively. For OTUs that were unclassified, the next highest taxonomic level to which they could be classified is shown.**

[φ] **Closely related species were identified using the Greengenes database (see Methods) and were assigned using BLAST.**

Table B.6: Covariate analysis using EBA susceptibility as the primary phenotype and bacterial species OTUs as covariates

S.No	OTUID[‡]	Taxonomic Classification[δ]	Closely related species[φ]
1	173469	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales; f__Staphylococcaceae;g__Staphylococcus	Staphylococcus hominis
2	N13471	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales; f__Staphylococcaceae;g__Staphylococcus	Staphylococcus hominis
3	589787	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	unclassified species
4	N12197	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	unclassified species
5	60254	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	unclassified species
6	52884	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria; o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	Helicobacter apodemus
7	286668	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales; f__Streptococcaceae;g__Streptococcus	unclassified species
8	243860	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria; o__Burkholderiales;f__Burkholderiaceae;g__Ralstonia	unclassified species
9	101810	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria; o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	Helicobacter ganmani
10	484436	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria; o__Pseudomonadales;f__Moraxellaceae	unclassified species

S.No	OTUID[‡]	Nearest marker to peak	Chr	Peak position in Mb	Confidence Interval (CI) in Mb using 1.5 - log P drop	Covariate Anova -log P score	E value for covariate QTL *	EBA score Anova -log P score	E value for EBA QTL *	Difference in -log P score[γ]	Percent increase of -log P
1	173469	rs6211533	19	57	53-60	5.08	E-value < 0.05	3.59	ns	1.49	41.50%
2	N13471	rs6211533	19	57	53-60	4.62	E-value < 0.05	3.59	ns	1.03	28.69%
3	589787	rs6211533	19	57	53-60	4.35	E-value < 0.1	3.59	ns	0.76	21.17%
4	N12197	rs6211533	19	57	53-60	4.24	E-value < 0.1	3.59	ns	0.65	18.11%
5	60254	rs6211533	19	57	53-60	4.23	E-value < 0.1	3.59	ns	0.64	17.83%
6	52884	rs6211533	19	57	53-60	4.23	E-value < 0.1	3.59	ns	0.64	17.83%
7	286668	rs6211533	19	57	53-60	4.19	E-value < 0.1	3.59	ns	0.6	16.71%
8	243860	rs6211533	19	57	53-60	4.17	E-value < 0.1	3.59	ns	0.58	16.16%
9	101810	rs6211533	19	57	53-60	4.16	E-value < 0.1	3.59	ns	0.57	15.88%

Table B.6 (continued)

S.No	OTUID[‡]	Nearest marker to peak	Chr	Peak position in Mb	Confidence Interval (CI) in Mb using 1.5 - log *P* drop	Covariate Anova -log *P* score	*E* value for covariate QTL *	EBA score Anova -log *P* score	*E* value for EBA QTL *	Difference in -log *P* score[γ]	Percent increase of -log *P*
10	484436	rs6211533	19	57	53-60	4.09	E-value < 0.1	3.59	ns	0.5	13.93%

‡ OTU Ids. The Ids for clusters displaying no match to the reference Greengenes database start with an "N"

δ Taxonomic classification from the RDP database and its prefixes k, p, c, o, f and g for kingdom, phylum, class, order, family and genus bacterial taxonomic level, respectively. For OTUs that were unclassified, the next highest taxonomic level to which they could be classified is shown.

φ Closely related species were identified using the Greengenes database (see Methods) and were assigned using BLAST.

γ Difference in -log P score of the peak snp associated with EBA (presence/absence) and that of EBA along with an OTU as a covariate.

*Significance threshold was set to an *E* value of 0.05 and 0.1 (see Methods). "ns" stands for not significant at a 0.1 significance threshold

Table B.7: Percentages of animals developing EBA with respect to genotype and bacterial abundance class (high or low)

S.No	OTUID[‡]	Taxonomic Classification[δ]	Closely related species[ᵠ]
1	173469	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
2	N13471	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	*Staphylococcus hominis*
3	589787	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	*unclassified species*
4	N12197	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	*unclassified species*
5	60254	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	*unclassified species*
6	52884	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	*Helicobacter apodemus*
7	286668	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	*unclassified species*
8	243860	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Ralstonia	*unclassified species*
9	101810	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter	*Helicobacter ganmani*
10	484436	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae	*unclassified species*

Table B.7 (continued)

S.No	Nearest marker to peak	Chr	Peak position in Mb	Confidence Interval (CI) in Mb using 1.5 - log *P* drop	Covariate Anova - log *P* score	Peak SNP (allele information)	Percentage of animals developing EBA with respect to peak genotype and high bacterial abundance*			Percentage of animals developing EBA with respect to peak genotype and low bacterial abundance*			*P* value from Fisher exact test[R]	Direction
						(A/B)	AA (n)	AB (n)	BB (n)	AA (n)	AB (n)	BB (n)		
1	rs6211533	19	57	53-60	5.08	A/C	26% (65)	29% (21)	67% (6)	34% (56)	48% (33)	100% (2)	0.05	Probiotic
2	rs6211533	19	57	53-60	4.62	A/C	21% (63)	26% (23)	67% (6)	40% (58)	52% (31)	100% (2)	ns	Probiotic
3	rs6211533	19	57	53-60	4.35	A/C	27% (62)	23% (26)	75% (4)	32% (59)	57% (28)	75% (4)	ns	Probiotic
4	rs6211533	19	57	53-60	4.24	A/C	26% (58)	37% (30)	75% (4)	33% (63)	46% (24)	75% (4)	ns	Probiotic
5	rs6211533	19	57	53-60	4.23	A/C	25% (59)	37% (30)	67% (3)	34% (62)	46% (24)	80% (5)	ns	Probiotic
6	rs6211533	19	57	53-60	4.23	A/C	25% (59)	25% (28)	80% (5)	34% (62)	58% (26)	67% (3)	ns	Probiotic
7	rs6211533	19	57	53-60	4.19	A/C	27% (59)	43% (28)	60% (5)	32% (62)	38% (26)	100% (3)	ns	Probiotic
8	rs6211533	19	57	53-60	4.17	A/C	28% (61)	36% (25)	67% (6)	32% (60)	45% (29)	100% (2)	ns	Probiotic
9	rs6211533	19	57	46-60	4.16	A/C	32% (62)	24% (25)	80% (5)	27% (59)	55% (29)	67% (3)	ns	Probiotic
10	rs6211533	19	57	53-60	4.09	A/C	25% (61)	36% (28)	100% (3)	35% (60)	46% (26)	60% (5)	ns	Probiotic

‡ OTU Ids. The Ids for clusters displaying no match to the reference Greengenes database start with an "N"

δ Taxonomic classification from the RDP database and its prefixes k, p, c, o, f and g for kingdom, phylum, class, order, family and genus bacterial taxonomic level, respectively. For OTUs that were unclassified, the next highest taxonomic level to which they could be classified is shown.

φ Closely related species were identified using the Greengenes database (see Methods) and were assigned using BLAST.

* High and low bacterial abundance for each species level OTUs is defined by taking the median of relative values

R ns stands for not significant p value at alpha ≤ 0.05

7. Scientific achievements during doctoral research

7.1 Publication related to doctoral thesis

1. **G. Srinivas**, S. Möller, J. Wang, S. Künzel, D. Zillikens, J. F. Baines[*] and S. M. Ibrahim[*], Genome-wide mapping of gene–microbiota interactions in susceptibility to autoimmune skin blistering. *Nat. Commun.* 4:2462 doi: 10.1038/ncomms3462 (2013). ***** Contributed equally.

7.2 Workshops, ECTS credits and grades

Date	Course Name	Location	ECTS points	Grade/Prize
28.09. – 01.10.2010	EvoGen workshop 2010	IST, Vienna, Austria	4	Not applicable
03.10 – 04.10.2010	12th workshop about jobs for natural scientists	Frankfurt Messe, Frankfurt, Germany	Not applicable	Not applicable
25.02.2011	3rd Theoretical workshop	MPI Evolutionary Biology, Plön, Germany	Not applicable	Not applicable
28.06.2011	Grant proposal writing workshop	University of Lübeck, Germany	Not applicable	Won 100 Euro prize money (Won as a Team)
01.04. – 27.07.2012	Innovation Management and Marketing	Fachhochschule Lübeck, Germany	5	1.3

7.3 Presentation in scientific conferences

The following are the list of selected scientific conferences attended during my doctoral research to exchange scientific ideas and results of my doctoral work.

Date	Conference Name	Place, Country	Presentation type
01.03. – 03.03.2012	Arbeitsgemeinschaft Dermatologische Forschung (ADF)	Marburg, Germany	Talk
02.05. – 03.05.2012	Aquavit 2012 - Max Planck Institute for Evolutionary Biology	Plön, Germany	Talk
19.08. – 24.08.2012	14th International Symposium on Microbial Ecology (ISME14)	Copenhagen, Denmark	Poster
02.10. – 06.10.2012	25th Annual Mouse Molecular Genetics Conference	California, USA	Poster
23.11.2012	35th Symposium of the North-German Immunologists	Borstel, Germany	Talk
22.02. – 23.02.2013	Interdisciplinary Inflammation at Interfaces Symposium	Hamburg, Germany	Poster
06.05. – 07.05.2013	International pre IID 2013 Satellite Meeting on Autoimmune Bullous Diseases	Lübeck, Germany	Talk & Poster
06.06. – 07.06.2013	3rd Heidelberg Forum for Young Life Scientists	Heidelberg, Germany	Poster

8. Affidavit

I declare that the dissertation in form and content and except for advices given by my supervisors constitutes my own work. I only used those sources and resources referred to in the thesis, and that I have identified citations as such. This work has been undertaken in compliance with the German Research Foundations (Deutsche Forschungsgemeinschaft, DFG) rules of good academic practice.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Lübeck, 20th June 2013   Girish Srinivas

9. Copyright Statement

1.   Copyright for the text of this dissertation rests with the author. Copies of this work by any process, either in full, or as extracts, may be made only in accordance with instructions given by and the express permission of the author. Further details may be obtained from the appropriate Graduate Office. This page must form part of any such copies made. Further copies (by any process) made in accordance with such instructions should only be made with the permission (in writing) of the author.

2.   The ownership of any intellectual property rights arising from this dissertation is vested with the University of Lübeck, Germany, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

3.   Further information on the conditions under which disclosures and use may be permitted is available from the Dean of the Department of Natural Sciences, University of Lübeck. Lübeck, Germany.

10. Index