

Profiling skin-associated archaea and bacteria with 16S rRNA and *cpn60* genes

by

Alexander K. Umbach

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2021

© Alexander K. Umbach 2021

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Alexander K. Umbach was the sole author for Chapters 1 and 4, and the lead author of Chapters 2 and 3, all written under the supervision of Dr. Josh D. Neufeld. Chapters 2 and 3 involved external collaborations that are detailed below:

Chapter 2

Dr. Janet Hill of the University of Saskatchewan helped with experimental design. High-throughput *cpn60* sequencing of mammalian skin swab samples was completed by Champika Fernando under the guidance of Dr. Janet Hill at the University of Saskatchewan. Dr. Janet Hill was also involved in post-sequencing troubleshooting and guidance. Mammalian skin swab samples were provided by Ashley A. Stegelmeier as part of a prior project.

Chapter 3

This chapter consists of a manuscript accepted for publication by Alexander K. Umbach, co-authored by Ashley A. Stegelmeier of the University of Guelph and Josh D. Neufeld of the University of Waterloo. Both Ashley A. Stegelmeier and Josh D. Neufeld contributed to the conceptualization, methodology, and review/editing of the final manuscript. Mammalian skin swab samples were provided by Ashley A. Stegelmeier as part of a prior project.

Umbach AK, Stegelmeier AA, Neufeld JD. Archaea are rare and uncommon members of the mammalian skin microbiome. *mSystems* 2021.

DOI:10.1128/mSystems.00642-21

Abstract

As the largest organ of the mammalian body, skin is associated with commensal microorganisms that impact host health. Characterizing host-microbe associations is critical to our understanding of skin health, function, and disease, and the potential co-evolutionary relationships that have occurred throughout mammalian and prokaryotic evolution. The research within this thesis focused on profiling the bacteria and archaea that inhabit mammalian skin using two phylogenetic marker genes: *cpn60* and 16S rRNA. The *cpn60* gene was applied to mammalian skin swab samples to provide increased taxonomic resolution for microbial populations on mammalian skin. Datasets previously generated using the 16S rRNA gene were included to assess archaeal populations associated with the skin and skin-associated surfaces.

Previous research into the mammalian skin microbiome using the 16S rRNA gene identified evidence for phylosymbiosis within the Perissodactyla and Artiodactyla, as well as highlighted core taxa common to all sampled mammalian skin. The increased taxonomic resolution provided by the *cpn60* gene has the potential to reveal additional co-evolutionary patterns and can more thoroughly probe specific populations of the mammalian skin microbiome. Chapter 2 of this thesis describes a newly generated *cpn60* gene dataset sourced from the mammalian skin microbiome of Carnivora, Perissodactyla, Artiodactyla, and Primate hosts. Significant patterns of phylosymbiosis for Artiodactyla and Perissodactyla were confirmed when using weighted ($p = 4.43 \times 10^{-2}$) and unweighted ($p = 4.36 \times 10^{-2}$) UniFrac metrics, which are observations not made previously with a comparable 16S rRNA gene dataset. Using the *cpn60* gene, specific *Staphylococcaceae* communities were successfully delimited from their genus classifications, with improved species-level resolution for *Macrococcus*, *Staphylococcus*, and *Salinicoccus*, compared to the 16S rRNA gene dataset. Additionally, *Jeotgalicoccus halophilus* was detected broadly within mammalian skin microbiomes, representing a first report of widescale association of this species with mammalian skin. These results demonstrate associations between mammalian hosts and skin-associated taxa warrants further investigation. Future amplicon-based skin

microbiome studies focusing on host-microbe interactions would benefit from continued use of the *cpn60* gene given the increased taxonomic resolution that it provides.

Limited skin-related archaea research has not yet allowed for a consensus on the prevalence of skin-associated archaea. Recent studies suggest that archaea are consistently detected and relatively abundant on human skin, with skin “archaeomes” dominated by putative ammonia oxidizers of the *Nitrososphaeria* class (*Thermoproteota* phylum - formerly *Thaumarchaeota*). Chapter 3 evaluated new and existing 16S rRNA gene sequence data sourced from mammalian skin and skin-associated surfaces, generated with two commonly used universal prokaryotic primers sets, to assess archaeal prevalence, relative abundance, and taxonomic distributions. Archaeal 16S rRNA gene sequences were detected in only 17.5% of 1,688 sample high-throughput sequence data, with most of the archaea-positive samples associated with non-human mammalian skin. Only 5.9% of human-associated skin sample datasets contained sequences affiliated with archaeal 16S rRNA genes. When detected, the relative abundance of sequences affiliated with archaeal ASVs was less than 1% for most mammalian skin samples and did not exceed 2% for any samples. Although several computer keyboard microbial profiles were dominated by *Nitrososphaeria* sequences, all other skin microbiome datasets tested were primarily composed of sequences affiliated with *Methanobacteriota* and *Halobacteriota* phyla. Our findings revise downwards recent estimates of human skin archaeal distributions and relative abundances, especially those affiliated with the *Nitrososphaeria*, reflecting a limited and infrequent archaeal presence within the mammalian skin microbiome.

This work provides insight in the microbial communities of mammalian skin-associated bacteria and archaea, their relationships with mammalian hosts. Increased sequencing depth and mammalian host representation could reveal additional co-evolutionary patterns, and the species-level resolution provided by the *cpn60* gene could be used to target additional microbial populations of interest. Likewise, additional research into the mammalian skin archaeome would benefit from further comparisons of universal and archaea-specific primers and help elucidate potential differences in abundance and distribution caused by regionality.

Acknowledgements

Science is a collaborative process and the work detailed within my thesis is no exception. However, scientific writing does not always allow for the recognition of contributions “unrelated” to the specific work. Thus, I feel compelled to use this opportunity to thank those who made contributions to my work – and my life – in ways that are not as easily presented in publications but were nonetheless critically important to my success.

My supervisor, Dr. Josh Neufeld, is a person of great patience – or at least hides his frustrations well. I would like to thank Josh for providing me with a Biology 499 opportunity in his lab, then believing enough in my abilities to accept me as a graduate student. I like to think that during my masters I have made large strides in my development as a scientist and am incredibly appreciative of his support. Especially with academic writing, I have occasionally struggled with clarity and conciseness, and am prone to overly complicated sentences or word choices. I am grateful for how Josh edited initial manuscript drafts because I learned more through that experience about scientific communication than I did during my entire undergraduate degree. Thank you again, Josh, and I hope this thesis is a reflection of my development under your guidance. At the very least, I hope it is a cromulent read.

To my academic collaborators, Drs. Janet Hill and Ashley Stegelmeier, you have been duly acknowledged as integral components to the completion of this work. Beyond that, I extend my personal thanks for your willing involvement and engagement throughout my project and hope that our collaborations continue. I would also like to thank my committee members, Dr. Kirsten Müller and Dr. Andrew Doxey, for advice and support.

When I think of my colleagues in the Neufeld lab, I feel nothing but respect and gratitude. The lab has benefited from what seems like a perfect pairing of personalities and perspectives to create a collaborative, engaged, and fun group of people. All the members have in some way contributed to my success, either directly with advice and suggestions, or indirectly with emotional support and camaraderie. I would like to thank Katja Engel for her work in managing the lab and providing a strong foundation for all its members. Katja is a constant we can always return to and is a force that pushes the Neufeld lab onward. Thank

you to Emilie Spasov and Rachel Beaver for training me on specifics within the wet-lab and humoring my curiosity and frequent questions. Thank you to Jackson Tsuji, Michelle McKnight, and Daniel Kim for their experience with bioinformatics and the dry-lab side of things, and for reducing the trauma of having to try and learn those techniques independently. Thank you to Melody Vachon and Sarah Al-Ajeel for their general support, enthusiasm, energy, and fun conversations. To all others, thank you for contributing to the lab with your experience and unique perspectives and for making the lab such a welcoming environment.

A general “thank you” to friends and family, past and present, who have supported me through my masters. Thank you to my parents and siblings for supporting me and my journey in finding something I enjoy. Thank you to Richard Nguyen for his friendship, meaningful conversation, and compassionate support through the last couple of years. Thank you to Tim Shardlow – academic colleague, good friend, and fireteam member – for sharing in my elations and frustrations throughout my masters and for always being available for my complaints and brainstorming sessions. Together, we will defeat Savathûn! Lastly, thank you to my partner, Rachel Beaver, for her generosity and caring, and for bringing fulfillment and purpose to my life. She is the calibre of scientist I attempt to emulate and her engagement in my work is greatly appreciated.

Table of Contents

Author's Declaration.....	ii
Statement of Contributions	iii
Abstract.....	iv
Acknowledgements.....	vi
List of Figures.....	xi
List of Tables	xiii
List of Abbreviations	xiv
Chapter 1.....	1
1.1 Introduction.....	1
1.2 Mammalian skin.....	1
1.2.1 The physiology of mammalian skin	1
1.2.2 The mammalian skin environment is chemically diverse	2
1.3 The mammalian skin microbiome.....	3
1.3.1 Bacterial communities of mammalian skin	3
1.3.2 Archaeal communities on mammalian skin.....	4
1.4 Host-microbe interactions	5
1.4.1 Host-microbe evolution and phyllosymbiosis	5
1.4.2 The influence of microbes on skin health and disease	7
1.5 Profiling skin-associated microbial communities	8
1.5.1 The 16S rRNA gene phylogenetic marker	8
1.5.2 Chaperonin-60 as an alternative phylogenetic marker	9
1.6 Thesis structure and objectives	11
Chapter 2 Validating the <i>cpn60</i> gene for high-resolution profiling of the mammalian skin microbiome	13
2.1 Introduction.....	13
2.2 Methods.....	14
2.2.1 Sample selection, PCR amplification, and high-throughput sequencing	14
2.2.2 Processing of sequence reads	15

2.2.3 Mammalian COXI gene, 16S rRNA gene, and <i>cpn60</i> gene microbial dendrograms	16
2.2.4 Phylosymbiosis	16
2.3 Results	17
2.4 Discussion	38
2.4.1 Validating the <i>cpn60</i> gene for microbial profiling of mammalian skin and assessing phylosymbiosis	38
2.4.2 Using <i>cpn60</i> to define species within <i>Staphylococcaceae</i>	42
2.5 Conclusion	45
Chapter 3 Archaea are rare and infrequent members of mammalian skin microbiome	47
3.1 Archaea on mammalian skin	47
3.2 Skin archaea methods	50
3.2.1 Sample collection, selection, and processing	50
3.2.2 PCR and sequencing	51
3.2.3 Processing of sequencing reads	52
3.2.4 Tree generation	53
3.2.5 Assessment of primer coverage	53
3.3 Results	54
3.4 Discussion	69
3.4.1 Archaea on mammalian skin are rare and uncommon	69
3.4.2 Implications for previous studies of human skin archaea	72
3.4.3 Skin as a potential habitat for archaea	74
3.5 Conclusion	76
Chapter 4 Conclusion and recommendations	78
4.1 Summary	78
4.2 Validation of <i>cpn60</i> , observations of phylosymbiosis, and high-resolution profiling of the mammalian skin microbiota	79
4.3 Skin-associated archaea are infrequent and in low relative abundance on mammalian skin	82

Bibliography 84

List of Figures

Figure 2-1 The total number and distribution of <i>cpn60</i> gene reads produced following sequencing and processing of the reads through DADA2.	19
Figure 2-2 The total number and distribution of <i>cpn60</i> gene reads within the final 37-sample dataset, following processing through DADA2.....	21
Figure 2-3 Comparisons of sequence taxonomic affiliations provided by the naïve-Bayes and VSEARCH classifiers.....	24
Figure 2-4 The distribution, relative abundance, and taxonomic affiliation of <i>cpn60</i> gene sequences for selected mammalian hosts.....	25
Figure 2-5 Comparison of relative abundances and prevalence of genera within the <i>cpn60</i> (blue) and 16S rRNA (red) gene datasets.	26
Figure 2-6 PCoA plots generated using the <i>cpn60</i> gene show differentiation between microbial community composition of mammalian hosts.....	28
Figure 2-7 Procrustes analysis between ordinations produced by 16S rRNA (primary plot) and <i>cpn60</i> (secondary plot) gene profiles.	29
Figure 2-8 Assessment of phyllosymbiosis using Bray-Curtis microbial dendrograms for both the 16S rRNA (left) and <i>cpn60</i> (right) genes.....	30
Figure 2-9 Assessment of phyllosymbiosis using unweighted UniFrac microbial dendrograms for both the 16S rRNA (left) and <i>cpn60</i> (right) genes.	31
Figure 2-10 Assessment of phyllosymbiosis using weighted UniFrac microbial dendrograms for both the 16S rRNA (left) and <i>cpn60</i> (right) genes.	32
Figure 2-11 Abundance and distribution of <i>Staphylococcaceae</i> -associated <i>cpn60</i> reads between mammalian hosts.	35
Figure 2-12 The relative abundance of all <i>Staphylococcaceae</i> species detected (top panel) and the proportion of <i>Staphylococcaceae</i> reads for each mammalian hosts (bottom panel)..	36
Figure 2-13 Relative abundance and distribution of <i>Staphylococcaceae</i> species between mammalian hosts generated by amplification of the <i>cpn60</i> (A) and 16S rRNA (B) genes....	37

Figure 3-1 The effect of truncation and primer removal methods on archaeal taxonomic proportions following processing through DADA2.	60
Figure 3-2 The effect of truncation and primer removal methods on archaeal and bacterial read numbers following processing through DADA2.	61
Figure 3-3 Comparison of the relative abundance (A) and ASVs (B) of archaea on human and non-human mammalian skin.	63
Figure 3-4 Archaeal 16S rRNA gene relative abundances (A). The taxonomic proportion of the archaeome of the skin and skin-associated surfaces is separated by phyla or class (B). .	64
Figure 3-5 Distribution of all archaeal genera across skin and skin-associated surfaces.	66
Figure 3-6 The prevalence and overlap of archaeal ASVs on skin and skin-associated surfaces.	68

List of Tables

Table 2-1 Summary of the mammalian skin samples included within the study and the <i>cpn60</i> gene reads produced during sequencing.	18
Table 2-2 Summary of dataset with samples containing at least 1000 reads.....	20
Table 3-1 Summary of skin and skin-associated datasets.....	55
Table 3-2 Coverage comparison of the bacterial and archaeal coverage of universal prokaryotic and archaea-specific primers.	56
Table 3-3 Summary of archaeal reads and ASVs of the skin and skin-associated environment.	62

List of Abbreviations

°C	degrees Celsius
µg	microgram
µL	microliter
µM	micromolar
16S rRNA	16S ribosomal ribonucleic acid
AOA	ammonia-oxidizing archaea
ASV	amplicon sequence variant
COXI	cytochrome oxidase I
<i>cpn60</i>	chaperonin protein 60
cpnDB	chaperonin-60 database
DNA	deoxyribonucleic acid
DPANN	<i>Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota,</i> <i>Nanohaloarchaeota</i>
EMP	Earth Microbiome Project
ENA	European Nucleotide Archive
FISH	fluorescent in situ hybridization
FTIR-FPA	Fourier transform infrared focal plane array hyperspectral imaging
GTDB	Genome Taxonomy Database
GTR	general time reversible
HEPA	high efficiency particulate air
ISO	International Organization for Standardization
kDa	kilodalton
m/s	meters per second
mM	millimolar
NCBI	National Center for Biotechnology Information
ORE	Office of Research Ethics
OTU	operational taxonomic unit
PCoA	principal coordinates analysis
PCR	polymerase chain reaction

pM picomolar
QIIME Qualitative Insights into Microbial Ecology
qPCR quantitative polymerase chain reaction
Rel. ab. Relative abundance
SSU small subunit
TACK *Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota*
UPGMA Unweighted pair group method with arithmetic mean
UV..... ultraviolet

Chapter 1

1.1 Introduction

As the largest organ of mammalian bodies, skin serves as a protective barrier to separate temporal fluctuations of external conditions from the regulated internal environment of the host. Skin has evolved with its mammalian host [1] to endure stresses associated with specific host environments, with the purpose of maintaining skin barrier integrity. Because external forces experienced by the skin, such as exposure to extreme heat or cold, moisture, or mechanical forces can vary among hosts, differences among mammalian skin surfaces can be similarly variable [2]. Exploring such host-specific variation is essential for understanding the importance of skin for host function and health. However, mammalian skin does not exist in isolation, nor is skin health and function confined to and dependent on the intrinsic properties of the skin alone. Just as mammals interact with their local environments, so too do the microorganisms that inhabit mammalian skin. The interconnectivity between skin microbiota and host health is evolutionarily linked. Mammalian immune systems have adapted and evolved temporally with microbial populations – of pathogens and commensals alike – for hundreds of millions of years [1]. As such, researchers interested in mammalian skin health and function must incorporate a microbiological perspective in their work. By profiling the mammalian skin microbiome and establishing a baseline knowledge of the microorganisms associated with specific mammals, questions regarding mammalian skin health and function can be more thoroughly answered.

1.2 Mammalian skin

1.2.1 The physiology of mammalian skin

All skin research is dependent on and influenced by the physical structure of the skin. Skin physiology is complex but can be generalized into two primary layers: the epidermis and the dermis. Along with hair, glands, and nails, the epidermis and dermis represent the “integumentary system” of mammals [2, 3]. The epidermis consists of the top-most skin layers (i.e., between four and five layers), is without vasculature, and provides water resistance, protection from mechanical abrasion, and prevents microorganisms from

penetrating into the dermis and into soft tissues and the blood stream. The outer-most layers of the epidermis, the stratum corneum, stratum lucidum (i.e., an additional layer found in the “thick skin” of the soles, palms, and digits), and the stratum granulosum are comprised of heavily cornified keratinocytes that resist mechanical abrasion and mitigate immediate damage to the skin [2, 3]. It is within these first few layers where most skin-associated microbiota are found, although microbial communities can also extend further into the subepidermal compartments [4]. The dermis is comprised of several layers involved in producing keratinocytes and immune cells, houses various sensory cells, and provides the vasculature and protein-rich matrix that provides flexibility and tension, and supports the upper layers of the skin [2, 3]. The epidermis and dermis together represent the protective barrier that mammals depend on for survival. Despite variations in evolutionary history and such as the proportion and distribution of apocrine and eccrine glands within the mammalian order of Primates [5], the basic structure of mammalian skin is fundamentally shared among mammals.

1.2.2 The mammalian skin environment is chemically diverse

Skin is a chemically complex environment and the types and concentrations of skin-associated secretions varies by host and habitat. Intrinsic physiochemical characteristics of skin include pH, sebum production, and hydration; collectively, these can have a large impact on skin barrier function [6]. As well, lipid type and concentration can vary with host and with host features (i.e., age, physiology) [7–9]. Various “allelochemicals” that play an important role in mammalian behaviour modification (e.g., territory marking, hunting, mate selection) can be both localized or dispersed broadly across the host skin [10, 11]. Spatial distribution and abundance of the dermal apocrine and eccrine glands is host dependent and will result in differences in sweating [5] and the release of pheromones across the surface of the skin [3]. Human skin is unique because of hygiene practices, skin care products, and synthetic textiles, and thus has considerable chemical heterogeneity [12, 13].

1.3 The mammalian skin microbiome

Together, the physical and chemical characteristics of mammalian skin establish a habitat capable of supporting a diverse community of microorganisms. This community includes representatives from all the domains of life: *Bacteria*, *Archaea*, and *Eukarya*. However, the focus of this thesis will be on the prokaryotes alone, which include members of the *Bacteria* and *Archaea*.

1.3.1 Bacterial communities of mammalian skin

Prior to the modern discipline of microbial ecology, skin-associated bacteria were explored through the lens of disease. One of the earliest references to disease-causing skin bacteria is by Sir Arthur Clark, who in an 1821 essay regarding the fumigation treatment for skin disorders, remarked that "... others believe that ... the diseases originate from animalculae" [14]. Decades later, German "microscopical observers" would find stronger evidence of "animalciili" in acne punctata pustules [15], followed by the suggestion that healthy individuals also shared skin-associated microorganisms [16, 17]. The application of modern microbial techniques and ecological theory to mammalian skin microbiome research has greatly improved understanding of skin-associated microbiota and their relation to skin and host health.

Bacteria are the most abundant members of the mammalian skin microbiome and have thus been a dominant research focus. Because of this, bacterial communities of the skin have been thoroughly characterized. An established "core" skin microbiota consist primarily of taxa affiliated with the *Actinobacteria*, *Firmicutes*, *Proteobacteria*, and *Bacteroidetes* [18]. The proportions of these taxa vary based on host individuality [19], age [20], body region [21], cohabitation [22, 23], and the surrounding built environment [23, 24]. Features dependent on the mammalian host, such as hair/fur coverage and underlying skin physiology [25, 26], as well as the host itself [27], may also impact skin-associated bacterial communities. This is in addition to temporal variations, which are stronger determinants of the skin microbiome compared to the microbiota of the gut or mouth [28], although the skin microbiome can still retain temporal stability [29]. Additionally, skin topography can have a

large influence on microbiome spatiality [30] and, although the majority of bacteria live on the outer layers of the epidermis [31], others suggest that skin bacteria may extend into subepidermal tissue [4].

1.3.2 Archaeal communities on mammalian skin

The *Archaea* is the most recently discovered domain of life, having only been formalized in the late 1970s [32]. As such, knowledge of archaeal distributions and ecological roles is lacking, compared to bacteria, which have benefited from centuries of study. Archaea differ from their bacterial counterparts, often by containing unique metabolic pathways [33], such as methanogenesis [34], and distinct cell membranes composed of isoprenoid-glycerol ethers [35]. Although initially thought to be relegated to extreme environments only, archaea have seen a considerable level of discovery and taxonomic radiation, elevating them from their “extremophile” status to one of accepted ubiquity within global environments [36–38]. However, skin-associated archaea have been poorly studied and thus represent a new frontier in mammalian skin research.

In the last half decade, archaea have been reported as skin microbiota members of multiple individuals and body locations [39–42]. Probst and colleagues were the first to demonstrate that archaea can be detected as representatives of human skin microbiota, possibly representing as much as 8% of the total microbial community [43]. A subsequent study suggested that human skin archaeal communities could exceed 10% abundance, with such communities dominated by putative ammonia-oxidizing archaea (AOA) of the class *Nitrososphaeria* (previously phyla *Crenarchaeota/Thaumarchaeota*, now *Thermoproteota*) [40]. In contrast, extensive skin sampling of cohabitating couples revealed average archaeal sequence relative abundances of less than 0.5%, with archaea detected only for a few samples [22]. Furthermore, a large-scale human metagenome survey revealed that archaea accounted for approximately 1% of all human metagenome sequences, with the majority of these sequences affiliated with the *Methanobacteriota* phylum from gut or mucosal membranes [44]. A comprehensive analysis of mammalian skin concluded that less than 0.1% of all sequences were associated with archaea, with most sequences affiliated with the

Methanobacteriota and *Halobacteriota* phyla [27]. Because of such limited research, there is no consensus on the taxonomic representation and abundance of archaea on mammalian skin, nor an understanding of how the skin “archaeome” may contribute to host physiology or skin microbial ecology more broadly.

1.4 Host-microbe interactions

1.4.1 Host-microbe evolution and phylosymbiosis

Microorganisms and their mammalian hosts share a longstanding evolutionary connection [45]. Initial predatory and nutritional interactions between ancestors of modern bacteria and eukaryotes are thought to have led to multicellularity [46], and further developed to include complex metabolic symbioses [45] and vertebrate innate immune responses [47, 48]. Given variations in physiology [25], hair and fur coverage [25, 26], geographic origin and habitat features [49], and evolutionary histories and relatedness [27], the mammalian skin environment is ideal for host-specific microbial co-evolution. Because of the variations provided by the mammalian host, microbial community assemblage is thought to be deterministic (i.e., influenced by specific environmental or host factors) rather than stochastic (i.e., random assemblage, birth-death events) [50]. For mammals, evidence suggests that host phylogeny correlates with microbial community composition (i.e., phylosymbiosis) [27].

Phylosymbiosis is a pattern wherein the microbial community composition of a host reflects the host’s environmental and phylogenetic history [51–53], with more distantly related host species showing greater differences in microbial community composition compared to their more closely related members [27]. Phylosymbiosis can occur from both stochastic and deterministic processes with co-evolution as one mechanism that can lead to observations of phylosymbiosis [53]. Observed patterns of phylosymbiosis, as reflected in microbial community composition linked to host evolutionary history, is distinct from co-evolution, although phylosymbiosis can be influenced by or facilitate the development of co-evolutionary interactions between the host and associated microbiota. Co-evolution results from maintained close associations between organisms (e.g., mammalian host and skin microorganisms) wherein the associations result in reciprocal changes in the genotype of the

organisms over evolutionary time [52]. Phylosymbiosis caused by stochastic processes could foster close interactions between the host and associated microbiota over time, potentially leading to co-evolutionary relationships. Therefore, the study of phylosymbiosis of mammalian skin highlights key microbe-host associations that could indicate existing or the potential for future development of co-evolutionary relationships within class Mammalia. Furthermore, understanding host-microbe relationships can help inform and direct future research in managing microbial populations that have the potential to impact skin health and function.

Most phylosymbiosis research of both of invertebrates and vertebrates relates to the gut microbiome because of its functional importance and relative stability [54–58]. Gut phylosymbiosis studies of *Nasonia* wasps, mice, mosquitos [51], and hominids [56, 58] show increased diversity of their microbiome as host distance increases. Phylosymbiosis has also been linked to a decrease in fitness in wasps that received transplanted microbiomes from distant relatives [51]. In hominids, the gut microbiota is influenced by host phylogeny, an observation that was maintained despite masking confounding factors such as proximity, diet, or local environment [56]. As well, the use of *Bacteroidaceae*-specific PCR primers allowed for higher resolution species analysis that revealed evidence for co-evolution as a mechanism for phylosymbiosis, detected among hominid gut microbiota [56]. Conversely, chipmunk gut microbiomes were shown to be more influenced by environment change or other external factors rather than phylosymbiosis [59].

Relative to the gut microbiome, non-human mammalian skin has received less attention and is thus generally lacking microbial community comparisons and evidence for phylosymbiosis. Ross and colleagues showed first evidence of phylosymbiosis within the Perissodactyla and Artiodactyla orders, which constitute odd-toes and even-toed ungulates, respectively [27]. Their study also identified a core microbiome common to all sampled orders [27]. This core microbiome includes soil-associated bacteria, such as *Agrobacterium* and *Arthrobacter*, and taxa from the common skin bacterial genus *Staphylococcus* [27, 60, 61]. A core Primate axillary microbiome containing *Staphylococcus* was also identified in a study that showed support for skin phylosymbiosis among sampled primates [25]. As well,

the same study showed that the distribution of *Staphylococcus* was one of the main contributors to beta diversity among primates. However, observations of amphibian [62] and bat [63] skin microbiomes did not reveal significant evidence for phylosymbiosis. Thus, phylosymbiosis might not be universally relevant to mammalian skin, could be masked by external factors, or is not observable with the limited taxonomic resolution provided by 16S rRNA gene analyses.

1.4.2 The influence of microbes on skin health and disease

The assemblage of microorganisms on the skin has a direct impact on the health and function of the skin itself and therefore on the host. Host-microbe interactions on the skin are evolutionarily linked [48], and range from commensal (and arguably mutualistic) to parasitic or pathogenic [64]. Commensal microbes are involved in regulating immune responses and act as an extension of the innate immune system [65] while also occupying nutritional and spatial niches that prevent pathogenic bacterium from establishing themselves on the skin [47]. Recently, bacteria have been suggested to induce skin regeneration following damage [66]. However, these relationships are dynamic and prokaryotes that commonly maintain mutualistic relationships can become opportunistic pathogens when the integrity of the skin barrier is broken [67] or when microbial communities are disrupted [68, 69]. For example, *Staphylococcus epidermidis* is a commensal bacterium very commonly found on skin, has not evolved to cause disease, yet is the most frequent opportunistic pathogen and cause of nosocomial infections [70]. Whether commensal or pathogenic, skin-associated prokaryotes have a major impact on skin function.

Currently, all mammal-associated prokaryotic pathogens are of bacterial origin and there are no known archaeal pathogens belonging to mammalian hosts, on the skin or otherwise [71]. Although no causal relationships have been identified, halophilic archaea have been found in the biopsied tissues of inflammatory bowel disease patients [72] and methanogenic archaea are suggested to be a contributor to human obesity [73]. A lack of detected archaeal pathogens may relate to the metabolic requirements of archaea that are distinct from their mammalian hosts [74] such that they do not compete for the same

resources [75]. Additionally, Gill and Brinkman [75] hypothesize that the development of pathogenic traits is evolutionarily rare, and that the non-overlap between bacterial and archaeal phages and constrained lateral gene transfer has limited pathogen development in archaea. Regardless, the comparatively limited depth of skin-associated archaeal research has impeded exploration into the influence that skin-associated archaea might have on skin health.

1.5 Profiling skin-associated microbial communities

The tools and techniques used for profiling microbial communities have experienced great advances in recent decades. The advent of high-throughput sequencing methods and their associated bioinformatic tools, coupled with developments in both universal and tax-specific primers, have provided the means for both broad and specific characterizations of microbial communities. Of the many tools available, this thesis will focus on the widely used 16S rRNA gene and the more recently developed chaperonin 60 (*cpn60*) universal phylogenetic markers.

1.5.1 The 16S rRNA gene phylogenetic marker

The 16S ribosomal RNA (16S rRNA) is critical for protein synthesis because of its structural and functional role within ribosomes. As a result, the 16S rRNA gene is present in all prokaryotic genomes, often encoded by several copies within the same microorganism [76]. Although generally highly conserved, the 16S rRNA gene contains nine hypervariable regions that allow for mutability and sequence divergence over time. From a microbiology perspective, this universal dependence and combination of both conserved and variable regions is ideal for the identification and classification of microorganisms. Thus, the 16S rRNA gene is now the most widely used phylogenetic marker for amplicon-based studies of prokaryotic communities [76].

It was Carl Woese's early use of the 5S rRNA [77] gene and transition to the 16S rRNA gene that allowed for first proposal of the three-domain tree of life [32], which became solidified years later [78]. Woese's seminal contribution laid the framework for the development and use of 16S rRNA gene primers for profiling environmental microbiota [79].

Prior to short-read high-throughput sequencing, 16S rRNA gene sequencing was performed with by amplification, cloning, and sequencing of near-full length genes [80, 81]. Although full-length 16S rRNA genes provide higher taxonomic resolution and phylogenetic accuracy, the reduction in cost and time presented by short-read sequencing was a necessary compromise [82]. In response, 16S rRNA gene primers were developed to meet the size limitations imposed by sequencing technologies and the differences in phylogenetic resolution offered by the nine 16S rRNA hypervariable regions [83, 84]. Because of the variety of ways that 16S rRNA gene amplicon datasets can be generated, and the potential bias that those methods might introduce to taxonomic classification and analysis, efforts have been made to standardize 16S rRNA gene-based amplicon studies. One of the most prominent of these is the Earth Microbiome Project (EMP). Originally developed as a “collaborative effort to characterize microbial life on this planet” [85], the EMP also provides a sample collection and processing framework with the goal of standardizing 16S rRNA gene based microbial analysis [86]. Additionally, the EMP has also facilitated the widescale use of the 515F [87], 805R [88], and 926R [89] 16S rRNA gene primer sets. In combination with extensive 16S rRNA gene reference databases from the NCBI [90], the Genome Taxonomy Database [91, 92], and SILVA database [93, 94], 16S rRNA gene amplicon sequencing has been established as the standard for amplicon-based studies.

1.5.2 Chaperonin-60 as an alternative phylogenetic marker

Despite its usefulness and widespread support, the 16S rRNA gene has limitations. For example, although full 16S rRNA gene sequences are ideal for taxonomic classification [82], producing them is costly and time consuming. Additionally, although the 16S rRNA gene is often used in short-read high-throughput sequencing, the limited read length and high nucleotide conservation of the gene fragment limits phylogenetic resolution and thus taxonomic classification [95]. Thus, taxonomy derived from the 16S rRNA gene are commonly limited to a genus or species resolution [95]. Additionally, because prokaryotic genomes often contain multiple copies of the 16S rRNA gene, intragenomic heterogeneity and relative abundance bias is a concern for microbiome studies [96]. In response to these

limitations, several alternative universal phylogenetic gene markers have been proposed, such as *rpoB* [96, 97], *gyrB* [98–100], and *cpn60* [101].

Since first proposing *cpn60* as a universal phylogenetic marker for microbiome studies [102, 103], methodological development has led to establishment of a reference database [104], PCR primer development and validation [101, 105, 106], and increased adoption by the scientific community [107–112]. Several features of the *cpn60* gene make it a useful phylogenetic marker. The *cpn60* gene itself, alternatively known as *hsp60* [102], encodes for a 60kDa GroEL protein belonging to a family of type I chaperonins present in bacteria and chloroplasts, with equivalent type II chaperonins found in archaea and eukaryotes [113]. The structure of the GroEL protein consists of two stacked heptameric rings of seven identical subunits, which when combined with the complimentary GroES protein “lid”, assist in cellular protein folding [114, 115]. Like the 16S rRNA gene, the presence of *cpn60*, and function of GroEL, is essential for cell survival and thus present in all prokaryotic life [116]. However, unlike the 16S rRNA gene, the copy number of *cpn60* gene for most organisms is one, even though multiple copies exist in several taxa [101]. Additionally, the *cpn60* gene contains a much greater sequence diversity compared to the 16S rRNA gene, with a lower median nucleotide identity and greater barcode gap [101]. Furthermore, the primers used to amplify the 274 – 828 “universal target” region of the *cpn60* gene [101] produces an amplicon of approximately 554 bp in length. Although this length exceeds current short-read high-throughput sequencing technology, only 150 to 250 bases of the forward read is necessary to produce sufficient data for taxonomic classification to the species level [117], allowing it to be used on current sequencing platforms (e.g., MiSeq; Illumina). Because of intrinsic features of the *cpn60* gene, active development of primers and reference databases, and increased taxonomic resolution, the *cpn60* gene is a powerful complement to the 16S rRNA gene for use as a universal phylogenetic marker within the context of microbial community characterization and phylosymbiosis/co-evolution research.

1.6 Thesis structure and objectives

This thesis is separated into four chapters, with Chapters 1 and 4 providing introduction and conclusion context for the research chapters, respectively. The main research presented in this thesis is described in Chapters 2 and 3, both of which involve profiling microbial communities found on mammalian skin and skin-associated surfaces.

The first research chapter, Chapter 2, provides the first documented use of the *cpn60* phylogenetic marker for profiling mammalian skin microbial communities. The objective of Chapter 2 was to evaluate the *cpn60* gene as an alternative to the commonly used 16S rRNA gene through direct comparisons of mammalian skin microbial profiles. In addition, this chapter leveraged *cpn60*-based microbial profiles to assess evidence for phyllosymbiosis within class Mammalia. The *cpn60* gene data were also used to explore *Staphylococcaceae* taxa as core mammalian skin populations as potential indicators of specific host-microbe associations. Because of the increase in taxonomic resolution provided by the *cpn60* gene marker [108, 117], analysis using the *cpn60* gene has the potential to provide more confident species and strain-level classifications compared to the 16S rRNA gene when using amplicon-based methods, and might better reveal patterns of phyllosymbiosis or co-evolution. As well, because species and strain-level variation can have a considerable impact on disease state, the application of *cpn60*-based methods for skin microbiome analysis can provide microbial profiles with higher taxonomic resolution and species identification and has particular importance with respect to mammalian skin and host health.

The second research chapter, Chapter 3, reviews what is currently known about skin-associated archaea and contributes to the discussion surrounding the skin archaeome. The first objective was to evaluate mammalian skin and skin-associated datasets that were previously generated using the Pro341F/Pro805R primers for potential bias against archaeal 16S rRNA gene amplification by comparing them with newly generated profiles using primer pairs 515F-Y/926R. The second objective was to combine these datasets [22, 27, 118] with newly generated human skin (fingers) and skin-associated (keyboards) 16S rRNA gene amplicon sequence data and compare the relative abundance, sample prevalence, and taxonomic proportions of skin-associated archaea. Because of the limited research

surrounding skin-associated archaea, the work presented in Chapter 3 provides a substantial contribution to existing skin-associated archaea literature by suggesting limited distributions and abundances of archaea within the skin microbiome.

Chapter 2

Validating the *cpn60* gene for high-resolution profiling of the mammalian skin microbiome

2.1 Introduction

Performing universal prokaryotic profiling of microbial communities requires that the targeted phylogenetic marker has several important features. First, it must be universally present in all prokaryotes. Second, it should ideally be present in low copy number, or multiple copies should have high similarity. Third, it should contain sufficient nucleotide dissimilarity among species to be phylogenetically informative. The *cpn60* gene is ideal in that it is universally present, usually contains one copy per genome, and is taxonomically informative enough to provide confident species-level classification when compared to the commonly used 16S rRNA gene [101]. Applying the *cpn60* gene to the mammalian skin microbiome has the potential to generate prokaryotic microbial profiles with increased species-level taxonomic resolution that can be used to address questions regarding host-microbe patterns, such as phylosymbiosis and co-evolution.

Patterns of phylosymbiosis on mammalian skin were observed recently for Artiodactyla and Perissodactyla, using 16S rRNA genes clustered within 99% identity OTUs [27]. Also observed from the same study was the presence of a “core” mammalian skin microbiome consisting of taxa affiliated with *Staphylococcus* and other genera [27]. Although another study observed phylosymbiosis for primate skin microbiota, they identified a core primate skin microbial community where *Staphylococcus* members were differentially distributed among primate hosts [25]. Both studies produced microbial profiles based on the 16S rRNA gene and OTU-based clustering and are thus less capable of species-level classification [95]. By using the increased taxonomic resolution provided by the *cpn60* gene, in combination with amplicon sequence variants (ASVs) capable of identifying individual amplicon sequences [119], the mammalian skin microbiota could be more thoroughly characterized and evaluated for patterns such as phylosymbiosis and co-evolution. As well, specific microbial genera present across all mammalian skin, such as those included within

the family *Staphylococcaceae*, could be further defined to the species level to provide insight into species-level dynamics and host-microbe interactions. Finally, *cpn60*-based methods have yet to be applied to mammalian skin, although used previously for microbial profiling of the vagina [110, 111, 120] and pig feces [105]. Thus, profiling the mammalian skin microbiome using both the *cpn60* and 16S rRNA genes could help validate the *cpn60* gene for skin microbiome profiling and provide novel insight into mammalian skin microbial communities as they relate to host phylogeny.

2.2 Methods

2.2.1 Sample selection, PCR amplification, and high-throughput sequencing

Mammalian skin swabs were obtained and their genomic DNA were extracted as part of a previous 16S rRNA gene survey [27]. From these, a representative 95 sample subset was chosen for *cpn60*-based sequencing and microbial profiling. The PCR amplification and high-throughput sequencing of samples was completed at the University of Saskatchewan, in collaboration with Dr. Janet Hill, as previously described [121, 122] The *cpn60* gene was PCR amplified using a primer mix comprised of 100 μ M M279 (5' – GAIIIIIGCIGGIGAYGGIACIACIAC – 3'), M280 (5' – YKIYKITCICCRAAICCIGGIGC– 3'), M1612 (5' – GAIIIIIGCIGGYGACGGYACSAACSAC– 3'), and M1613 (5' – CGRCGRTRCCGAAGCCSGGIGCCTT– 3'). Primers were combined in a 1:3 molar ratio of M279 and M280 (3 μ L each) and M1612 and M1613 (9 μ L each) and diluted in 276 μ L of Ultrapure water for a total volume of 300 μ L. All PCR tubes, plates, and Ultrapure water used for PCR and sequencing were decontaminated prior to use by exposing to UV light for 20 minutes. A PCR master mix was prepared using 38.1 μ L of UV-treated Ultrapure water, 0.4 μ L of Invitrogen Platinum Taq (ThermoFisher Scientific), 5 μ L of 10X Thermopol buffer, 1 μ L of 10 mM dNTPs, and 1 μ L of the 1:3 primer cocktail for a total reaction volume of 50 μ L for 2 μ L of template. The amplification reaction conditions were 95°C initial denaturation for 5 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds, extension at 72°C for 30 seconds, and a final extension at 72°C for 2 minutes. All PCR amplifications were first visualized on a 1%

ethidium bromide gel then extracted from the gel and purified using NucleoMag beads (Macherey-Nagel) as previously described [122]. The purified amplicon library was sequenced using a 401x101 cycle using TG MiSeq Reagent Nano Kit v2 (Illumina Canada, MS-103-1003) on a MiSeq (Illumina).

2.2.2 Processing of sequence reads

Demultiplexed sequences were processed to generate ASVs using QIIME2 version 2019.10.0 [43]. Only the forward reads containing a 400-nucleotide long amplicon were imported into QIIME2. Forwards reads were truncated to 200 nucleotides using DADA2 version 2019.10.0 [123]. The trimmed reads were then denoised and chimeric sequences removed prior to ASV generation using with DADA2. Three taxonomic classification strategies were assessed prior to analysis. The first two used a naïve Bayes classifier, one using the cpnDB reference database [104], and the second using the cpnDB in combination with sequences retrieved from the NCBI using BLAST results (“cpnDB_B”). Representative sequences produced from DADA2 and nucleotide BLAST were used to query the NCBI non-redundant, cultured-only reference database with an e-value cutoff of $1e^{-6}$. The top three results for each sequence query were retained to create a database containing 16,624 sequences. Replicates and sequences below 180 nucleotides were removed (4,400) and the remaining sequences (12,224) were combined with the cpnDB (5,489) to create a final database of 17,713 reference sequences. The third strategy used VSEARCH [124] against the cpnDB_B as an alternative to the naïve Bayes classifier, a similar method to watered-BLAST [120]. All taxonomic information used in the databases originated from the NCBI [90] and were accessed through Entrez Direct [125].

All *cpn60* sequences generated for the current study were deposited in the European Nucleotide Archive (ENA) under project accession number PRJEB43503. All ASV and species-collapsed data tables have been made available at 10.6084/m9.figshare.14955753 (ASV table) and 10.6084/m9.figshare.14955744 (species-collapsed, *Staphylococcaceae*-only table).

2.2.3 Mammalian COXI gene, 16S rRNA gene, and *cpn60* gene microbial dendrograms

Cytochrome oxidase I (COXI) genes for the Cape eland, donkey, goat, horse, olive baboon, Przewalski's horse, sheep, spotted hyena, and Sumatran orangutan were obtained from a previous study [27] and used to construct a mammalian COXI-based phylogeny. The mammalian COXI gene sequences were aligned using ClustalW [126] in MEGA X version 10.1.8 [127], with trimming and gap removal as appropriate. The optimal nucleotide substitution model was determined using JModelTest2 version 2.1.10 [128, 129]. For the mammalian COXI gene, a maximum likelihood dendrogram was constructed using MEGA X with a GTR +G +I substitutional model and with a confidence assessment of 1000 bootstraps. The mammalian phylogeny was then compared with literature and confirmed for accuracy.

Microbial dendrograms were created using 16S rRNA gene sequences obtained from a previously generated amplicon dataset [27] and the *cpn60* gene sequences from this current study. For the 16S rRNA and *cpn60* gene microbial dendrograms, the respective ASV tables were sample-collapsed based on mammalian host, and the ASV read counts were summed for each category. Bray-Curtis, unweighted UniFrac, and weighted UniFrac distance matrices for the ASV tables were generated using the QIIME2 *diversity beta-rarefaction* command, rarefied to 1000 reads, and used to construct UPGMA dendrograms with a confidence assessment of 1000 bootstraps.

2.2.4 Phylosymbiosis

Mammalian COXI gene and microbial 16S rRNA and *cpn60* gene dendrograms were compared and evaluated for congruence using the *vegan* and *ape* packages in R as previously described [27]. Phylosymbiosis was assessed with Robinson-Foulds for both the 16S rRNA and *cpn60* gene dendrograms against the mammalian COXI dendrogram. The significance of the Robinson-Foulds metric was determined by evaluating the mammalian COXI gene dendrogram against 100,000 randomly generated trees containing identical terminal nodes (i.e., taxa). Congruency between dendrograms were measured with the normalized Robinson-Foulds score, ranging between 0 and 1, with 0 representing perfect congruency. Random dendrograms were considered significant if they obtained Robinson-Foulds scores equal to or

better than those obtained from comparisons between the mammalian phylogeny and the microbial gene (16S rRNA or *cpn60*) dendrograms.

2.3 Results

To assess microbial communities of mammalian skin using the *cpn60* gene, 95 representative skin swab samples already extracted for genomic DNA and sequenced from a previous project [27] were selected for further amplification and sequencing of the *cpn60* gene. This *cpn60* gene dataset contained samples from 19 unique mammals, with varying representation in terms of both number of samples and read proportions (Table 2-1). Of the 95 samples submitted for sequencing, 88 unique samples contained at least one read, with most samples associated with <500 total reads (Figure 2-1). The horse and Przewalski's horse samples represented 36.6% and 25.4% of all sequenced *cpn60* gene reads, respectively, followed by the olive baboon (15.8%) and Cape eland (10.0%). All other mammalian groups represented fewer than 5% of total *cpn60* gene reads. To avoid downstream analysis issues related to shallow sequencing depth, all samples with fewer than 1000 reads were removed, resulting in a final dataset of 37 samples (Table 2-2). These 37 samples (42.0% of the original dataset) represented 97.6% (118,645) of all the *cpn60* gene reads produced (121,622). Read loss was therefore minimal, although mammalian host representation in the dataset was reduced from 19 hosts to 9: the Cape eland, donkey, goat, horse, olive baboon, Przewalski's horse, sheep, spotted hyena, and the Sumatran orangutan (Figure 2-2).

Table 2-1 Summary of the mammalian skin samples included within the study and the *cpn60* gene reads produced during sequencing.

Mammalian host	Samples	Reads	Average per sample	Standard deviation	Proportion of dataset (%)
African lion	1	9	NA	NA	0.0
Alpaca	3	473	157.7	267.0	0.4
Arctic wolf	6	17	2.8	1.9	0.0
Beaver	1	1	NA	NA	0.0
Cape eland	5	12157	2431.4	1483.7	10.0
Cat	2	13	6.5	7.8	0.0
Cheetah	2	7	3.5	3.5	0.0
Donkey	1	2986	NA	NA	2.5
Giant panda	4	690	172.5	307.9	0.6
Goat	1	1785	NA	NA	1.5
Horse	18	44588	2477.1	2172.7	36.7
Olive baboon	9	19245	2138.3	1745.3	15.8
Przewalski's horse	12	30953	2579.4	2749.9	25.5
Rabbit	2	84	42.0	58.0	0.1
River otter	1	2	NA	NA	0.0
Sheep	1	5142	NA	NA	4.2
Spotted hyena	1	1441	NA	NA	1.2
Straw-coloured fruit bat	2	5	2.5	2.1	0.0
Sumatran orangutan	8	1699	212.4	404.7	1.4
Swamp wallaby	2	296	148.0	15.6	0.2
Two-toed sloth	2	7	3.5	0.7	0.0
White lion	4	22	5.5	9.0	0.0
Total	88	121622	NA	NA	NA

“NA” indicates values for which an average, standard deviation, and proportion could not be calculated because of singleton samples.

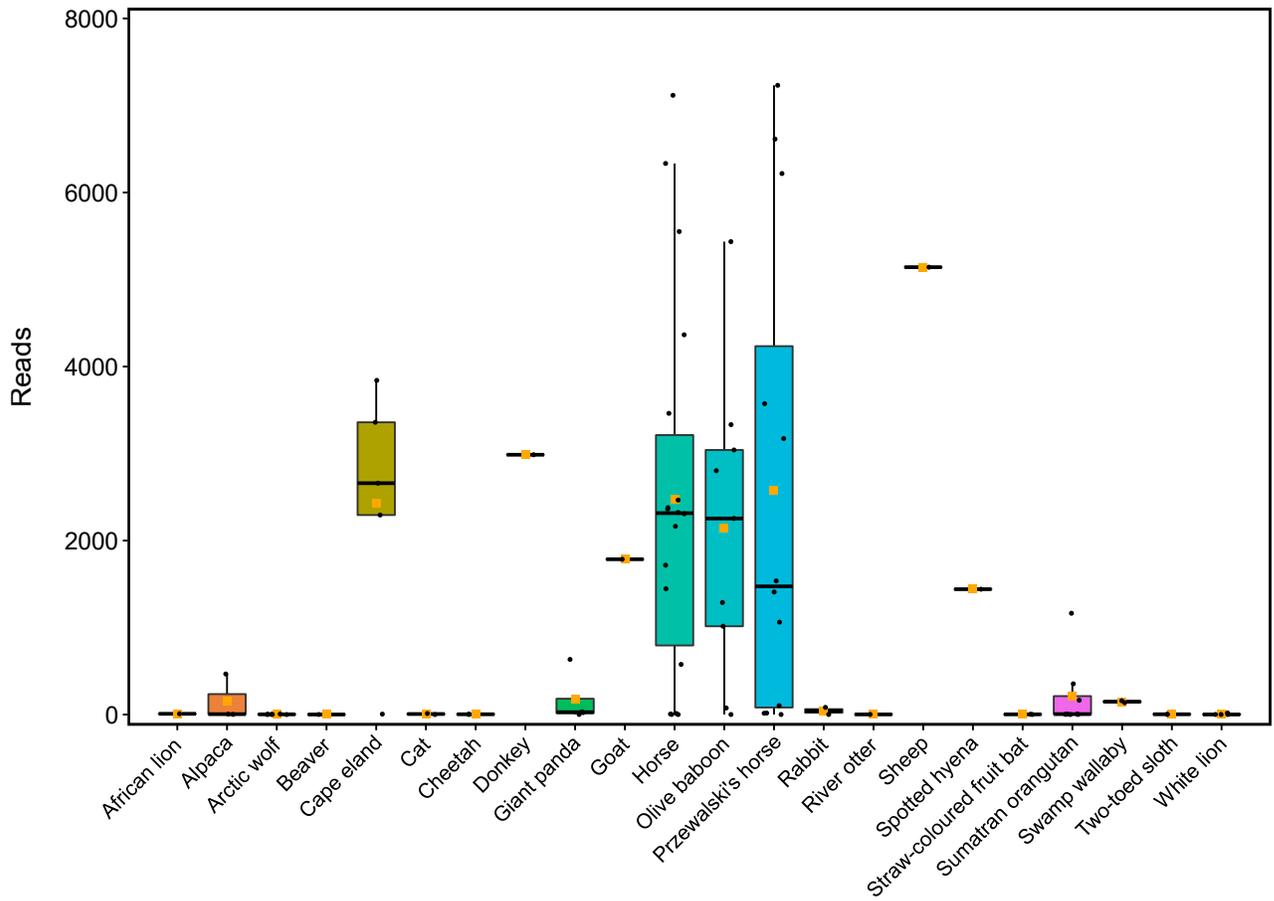


Figure 2-1 The total number and distribution of *cpn60* gene reads produced following sequencing and processing of the reads through DADA2. The average number of reads for each mammalian host is indicated with orange squares.

Table 2-2 Summary of dataset with samples containing at least 1000 reads.

Mammalian host	Samples	Reads	Average reads per sample	Standard deviation	Proportion of dataset (%)
Cape eland	4	12152	3038	694	10.2
Donkey	1	2986	NA	NA	2.5
Goat	1	1785	NA	NA	1.5
Horse	13	43990	3383	1860	37.1
Olive baboon	7	19168	2738	1473	16.2
Przewalski's horse	8	30817	3852	2515	26.0
Sheep	1	5142	NA	NA	4.3
Spotted hyena	1	1441	NA	NA	1.2
Sumatran orangutan	1	1164	NA	NA	1.0
Total	37	118645	NA	NA	NA

“NA” indicates values for which an average, standard deviation, and proportion could not be calculated because of singleton samples.

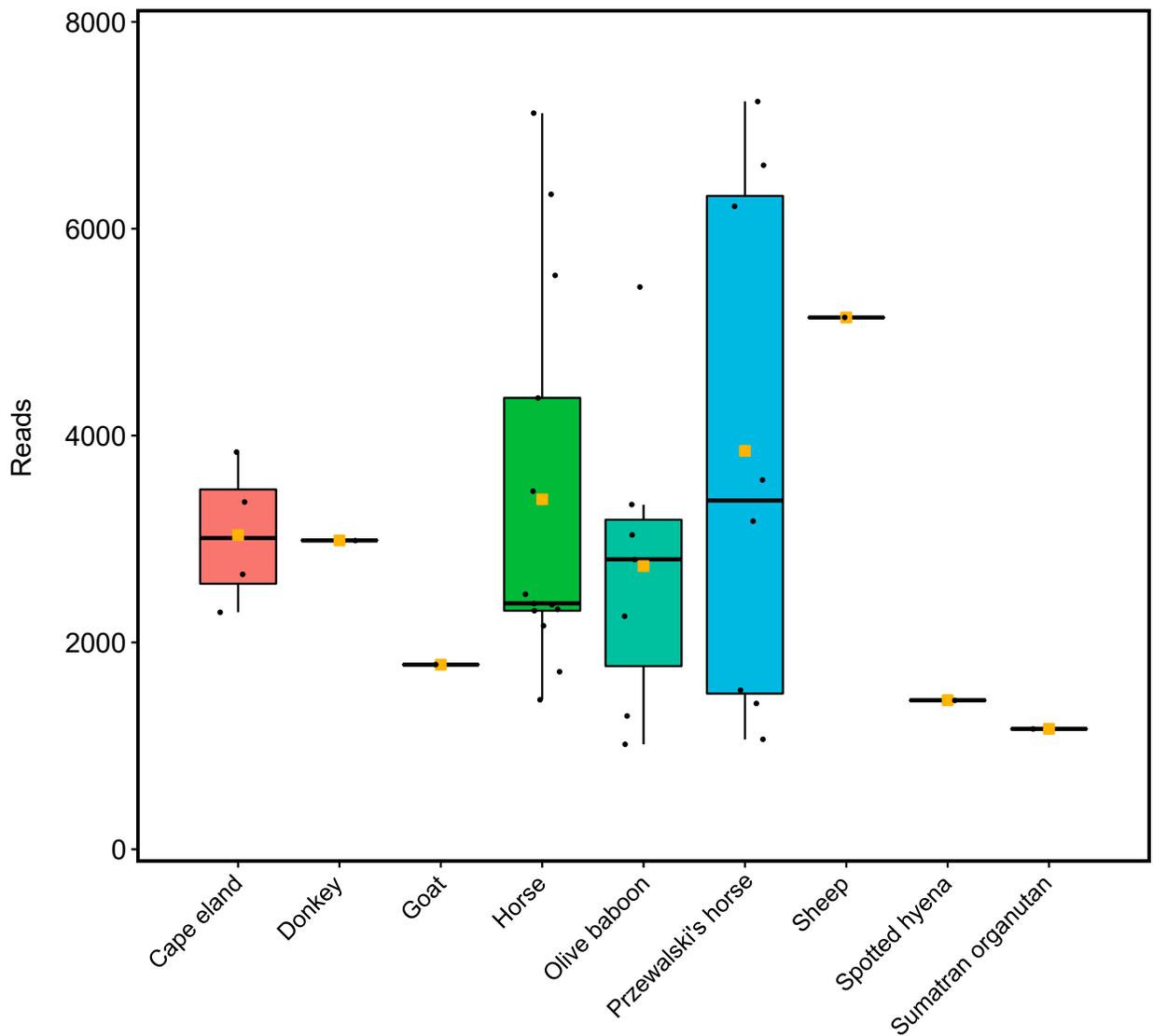


Figure 2-2 The total number and distribution of *cpn60* gene reads within the final 37-sample dataset, following processing through DADA2. The average number of reads per mammalian host is indicated with an orange box.

Three taxonomic classification strategies were assessed prior to taxa-based analyses (Figure 2-3). Both naïve Bayes based classifiers (cpnDB and cpnDB_B) produced similar taxonomic profiles. The VSEARCH classifier performed poorly, failing to classify most of the sequences beyond a domain or phylum level. The *cpn60* gene amplicons used for analysis were assigned taxonomy using the cpnDB_B strategy. Mammalian skin samples generated sequences with both genus-resolved and unresolved/unclassified taxa and were shown to be distinct based on mammalian host (Figure 2-4). The olive baboon samples had the most visually distinct and consistent microbial profiles of any other mammalian host. These profiles were represented by *Prevotella*, *Prophyromonas*, and *Butyrivibrio*, which were almost exclusive to the olive baboons. The Cape eland, horse, and Przewalski's horse samples had less consistent microbial profiles among samples. The Cape eland samples contained sequences associated with *Jeotgalicoccus* that were shared sparingly between the goat, horse, and sheep. For horses, microbial community profiles were more inconsistent between samples and contained horse-exclusive sequences associated with *Moraxella* and *Salinicoccus* genera. The microbial profiles of the Przewalski's horse samples were more consistent and contained exclusive sequences affiliated with *Planomicrobium* and *Macrococcus*. Of the singleton samples, the Sumatran orangutan microbial profiles included more unique genera (nine) than the donkey (four), goat (two), sheep (zero), and spotted hyena (one). Across all samples, *Corynebacterium* was the most prevalent and represented in moderate relative abundance (>5%) within the goat, horse, olive baboon, Przewalski's horse, and sheep, though in cases reached as high as 75% of the total community. Sequences affiliated with *Acidobacteria* were also present on the donkey, horse, olive baboon and Przewalski's horse at relative abundances ranging from 4% to 38%. The most observed sequences belonged to unclassified bacteria (*Bacteria_394*) which were present in 35 of the 37 samples in high abundance, as well as an unresolved *Proteobacteria* (*Proteobacteria_383*) that was present in 30 of the 37 samples. Other unresolved taxa include *Staphylococcaceae* (*Staphylococcaceae_138*), and *Clostridiales* (*Clostridiales_177*), which were also shared among several samples.

Comparisons made between the new *cpn60* gene microbial profiles and those generated previously using the 16S rRNA gene showed minimal overlap in their detected

taxa, although similar patterns in microbial community composition and differentiation based on mammalian host. Samples from the *cpn60* and 16S rRNA gene datasets contained both unique and overlapping genera at >3% relative abundance within their respective sample pairs (Figure 2-5). Overall, a large proportion of taxa were unable to be resolved to a genus level in the *cpn60* gene dataset. Both the horse and Przewalski's horse samples detected *Corynebacterium* in the *cpn60* and 16S rRNA gene datasets, as well as *Salinicoccus* in the horse samples, and *Massilia* in the Przewalski's horses. The goat samples included sequences associated with the genera *Massilia*, and *Sphingomonas* in high relative abundance, for both the *cpn60* and 16S rRNA gene datasets, though shared no additional genera.

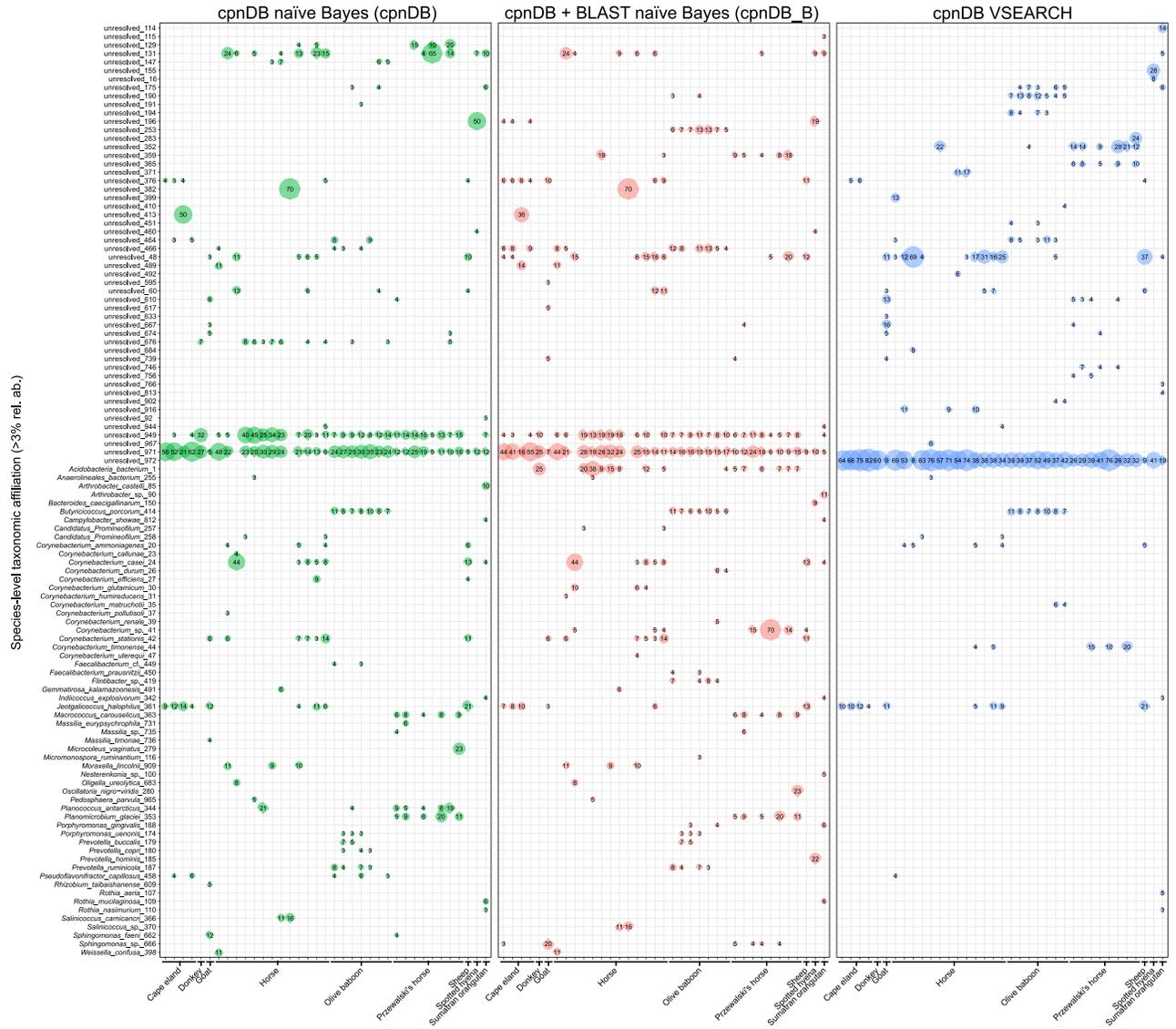


Figure 2-3 Comparisons of sequence taxonomic affiliations provided by the naïve-Bayes and VSEARCH classifiers. The *cpn60* gene ASV tables were collapsed to a species level for each classifier method. Bubble sizes and associated numeric values represent the relative abundances of taxa for a given sample.

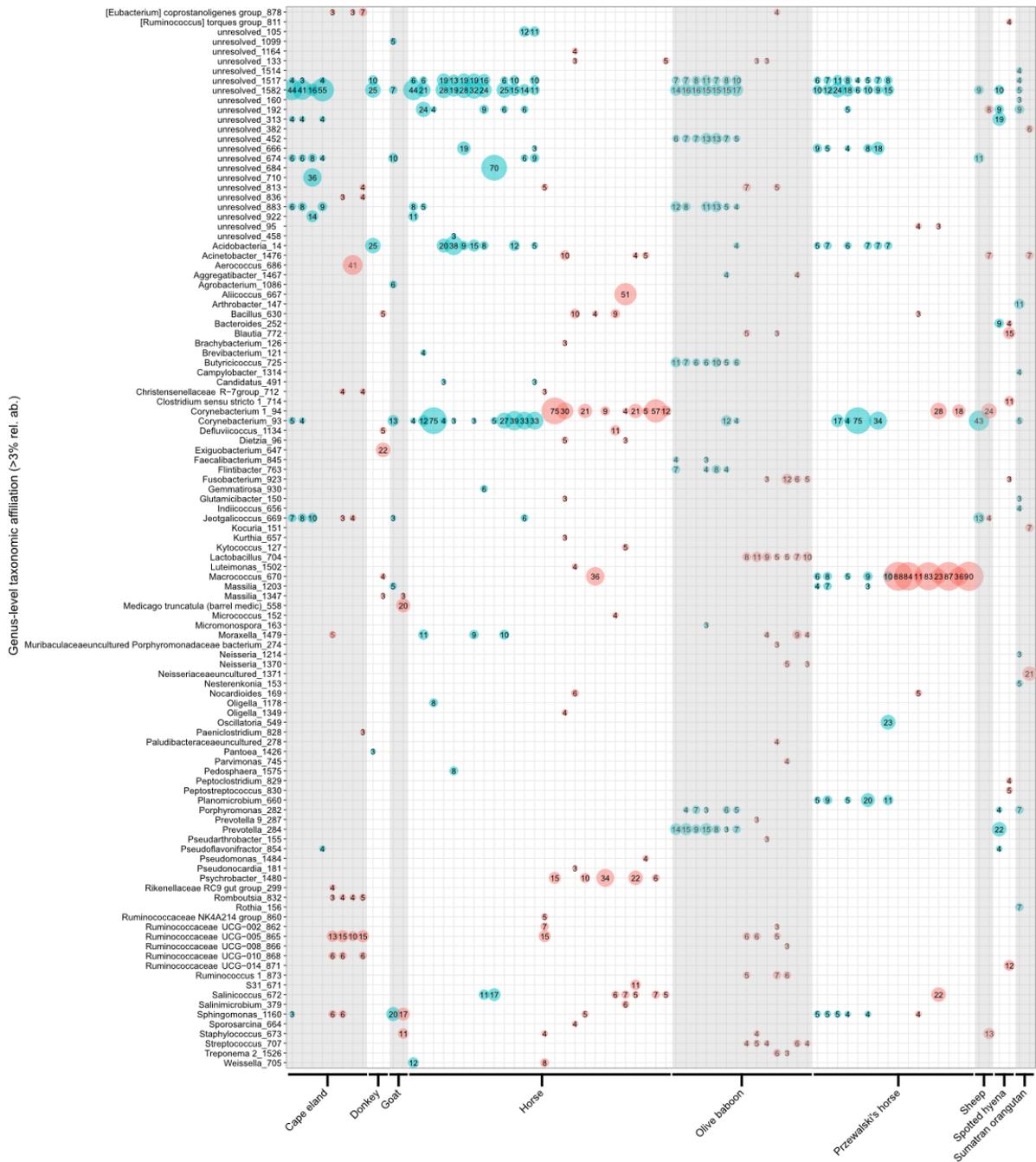


Figure 2-5 Comparison of relative abundances and prevalence of genera within the *cpn60* (blue) and 16S rRNA (red) gene datasets. The ASVs were collapsed to a genus level and filtered for taxa greater than 3% relative abundance. Bubble sizes and associated numeric values represent the relative abundances of taxa for a given sample.

In addition to direct taxonomic comparisons among samples, principal coordinate analyses (PCoA) can be used to compare microbial community composition similarity/dissimilarity independent of taxonomic classifications. Using such an approach, the results showed how olive baboon sample profiles were distinct, with samples grouping separately for each tested metric (Figure 2-6). The Przewalski's horse and Cape eland samples had similarly strong and distinct associations with each other and for most metrics. All other mammalian host samples grouped homogeneously within the horse samples for each diversity metric. An analysis of community composition between the *cpn60* and 16S rRNA gene datasets showed congruence (Figure 2-7). Procrustes analysis of ordinations made with the *cpn60* and 16S rRNA genes demonstrated significant correlations ($p < 0.05\%$) with correlation coefficients of 0.91, 0.69, and 0.66 for Bray-Curtis, weighted UniFrac, and unweighted UniFrac, respectively.

Evidence of phylosymbiosis within the *cpn60* and 16S rRNA gene microbial profiles was assessed through comparing microbial community composition dendrograms against a COXI mammalian dendrogram representing mammalian phylogenetic history. Significant ($p = 6.73 \times 10^{-3}$) patterns of phylosymbiosis were observed in the 16S rRNA gene Bray-Curtis microbial dendrogram for clades containing the Cape eland, goat, and sheep (Artiodactyla) and the donkey, Przewalski's horse, and horse (Perissodactyla), although was not significant for the *cpn60*-based microbial dendrogram (Figure 2-8). Conversely, significant results were observed for the *cpn60* gene unweighted ($p = 4.36 \times 10^{-2}$) (Figure 2-9) and weighted (Figure 2-10) ($p = 4.43 \times 10^{-2}$) UniFrac microbial dendrograms for Artiodactyla (Cape eland excluded) and Perissodactyla, although not within the 16S rRNA gene-based dendrograms. Neither the Primates (olive baboon and Sumatran orangutan) nor Carnivora (spotted hyena) displayed patterns of phylosymbiosis in either the *cpn60* or 16S rRNA gene microbial dendrograms using any of the diversity metrics.

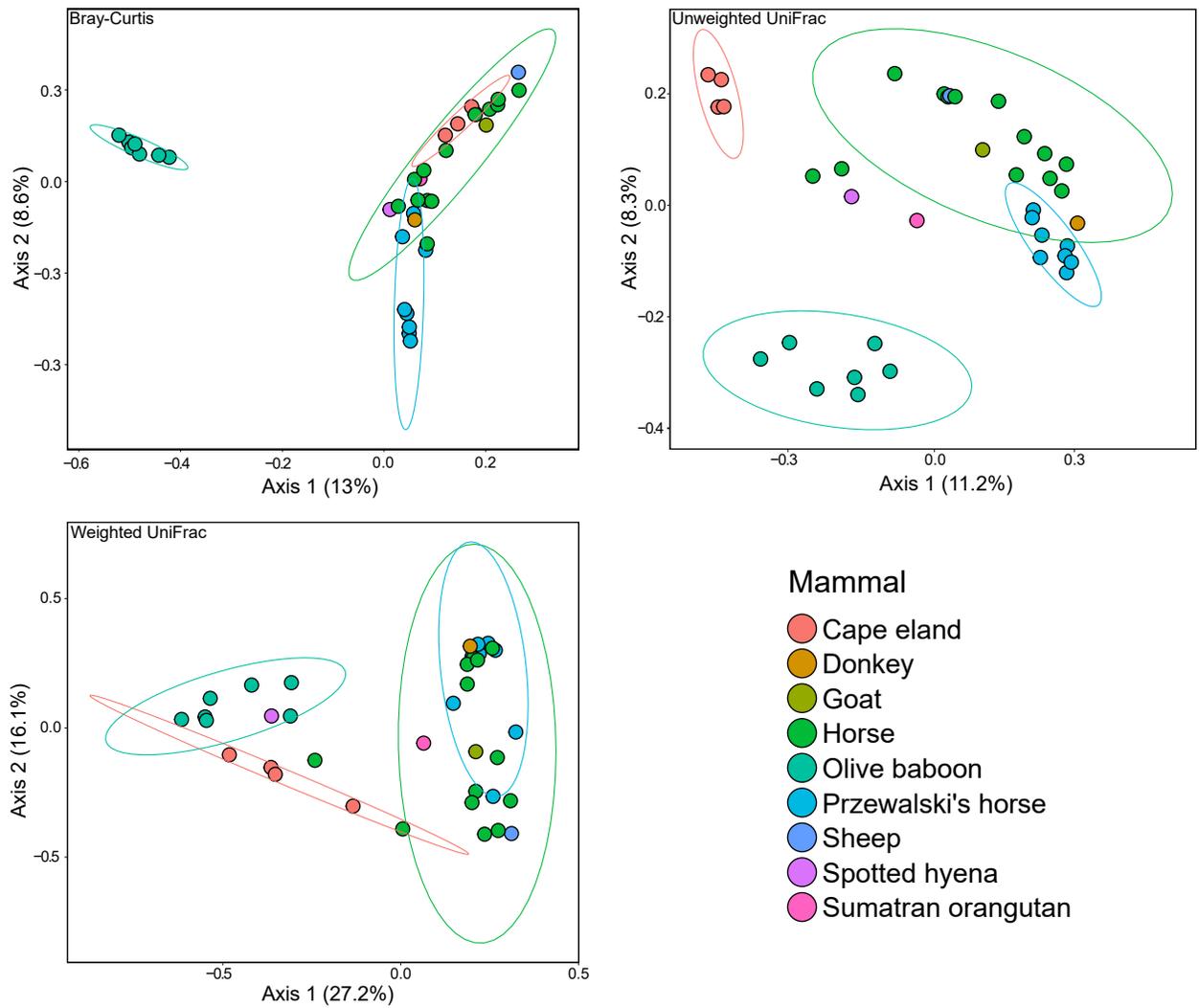


Figure 2-6 PCoA plots generated using the *cpn60* gene show differentiation between microbial community composition of mammalian hosts. Samples were rarefied to 1000 reads for each metric.

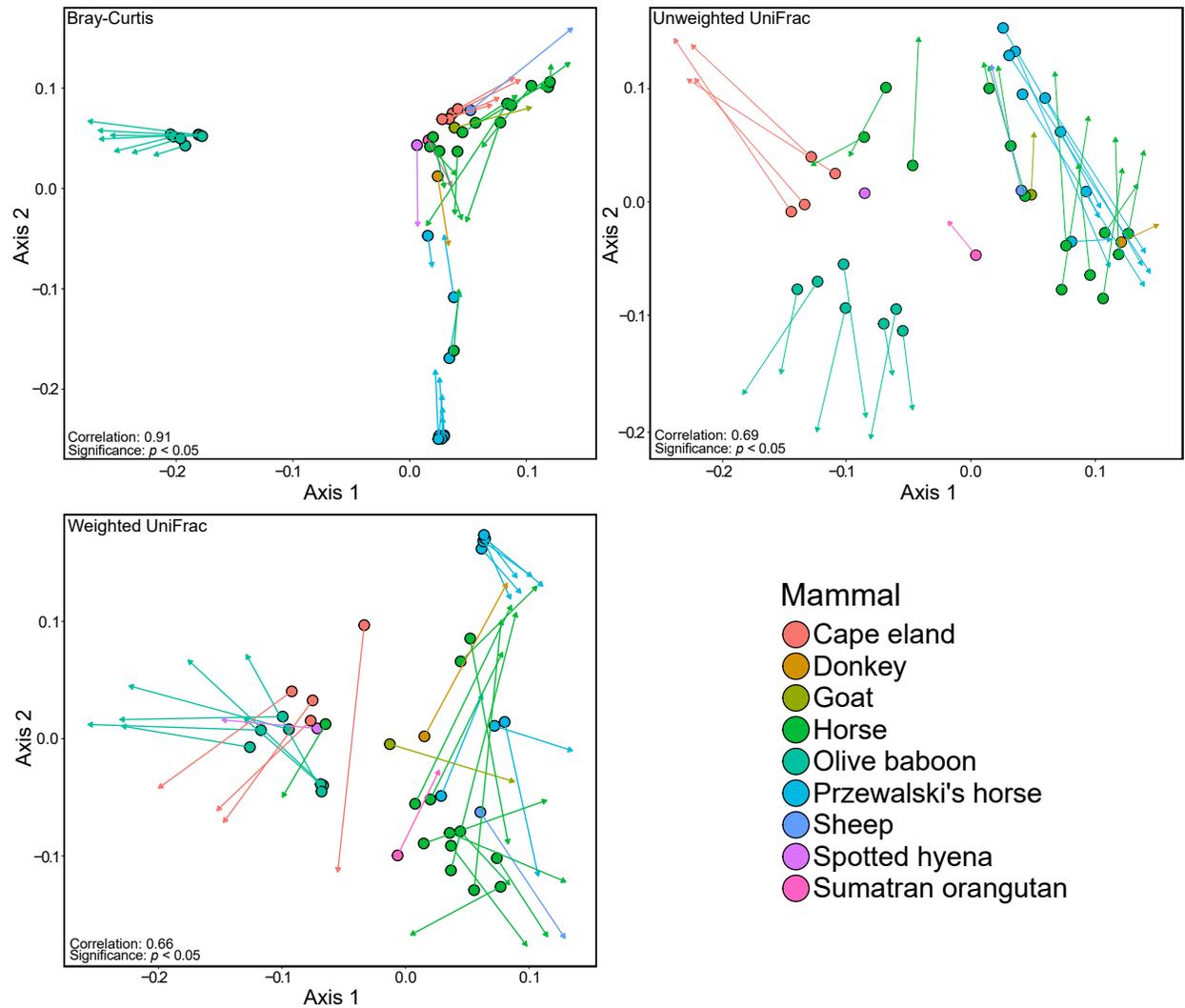


Figure 2-7 Procrustes analysis between ordinations produced by 16S rRNA (primary plot) and *cpn60* (secondary plot) gene profiles. The analysis was completed with 100,000 permutations to calculate significance. Arrow tips indicate the ordination of the secondary plot.

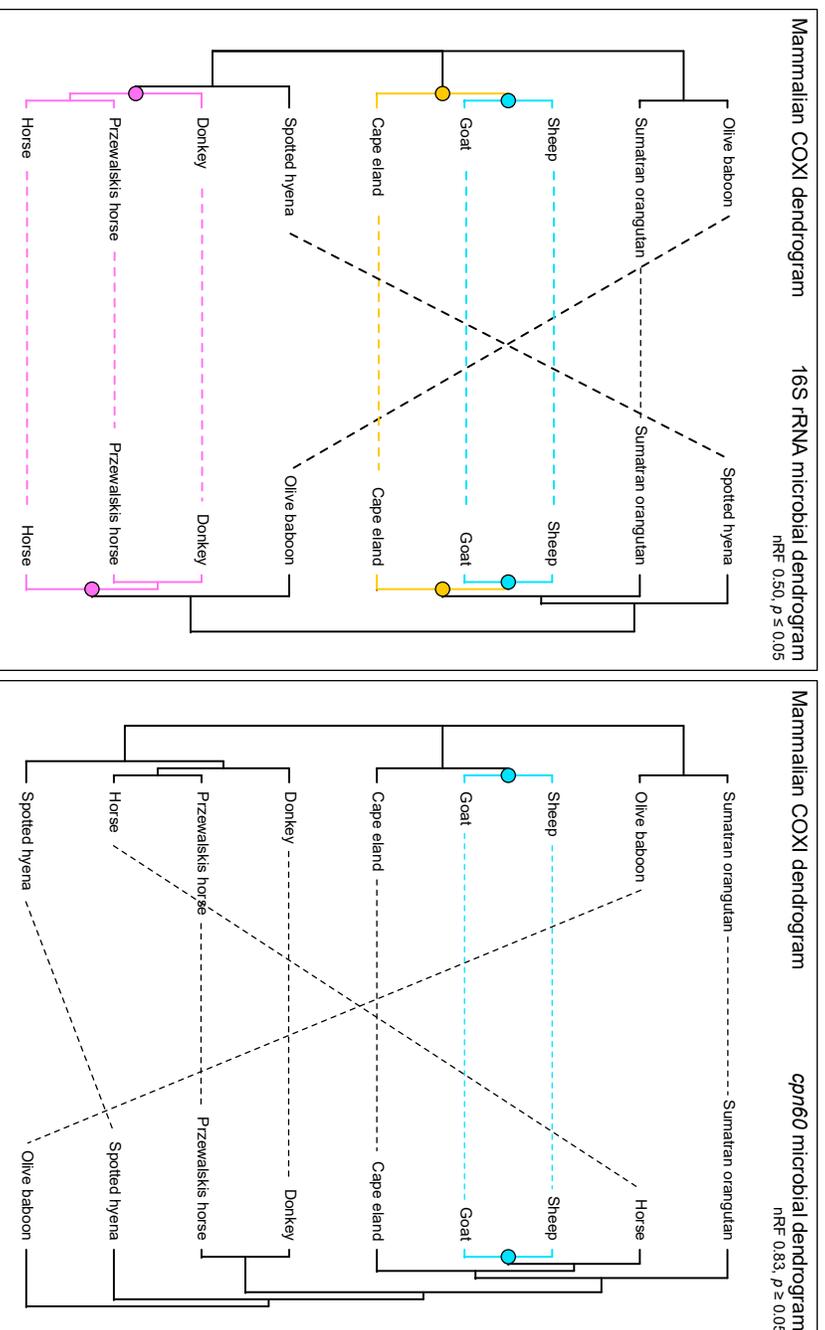


Figure 2-8 Assessment of phyllosymbiosis using Bray-Curtis microbial dendrograms for both the 16S rRNA (left) and *cpn60* (right) genes. Clades that share identical topography between the mammalian COXI and microbial dendrogram, and thus suggest phyllosymbiosis, are indicated with colour. Normalized Robinson-Foulds scores are used to indicate tree similarity, and significance was determined by comparing dendrograms against 100,000 randomly generated trees.

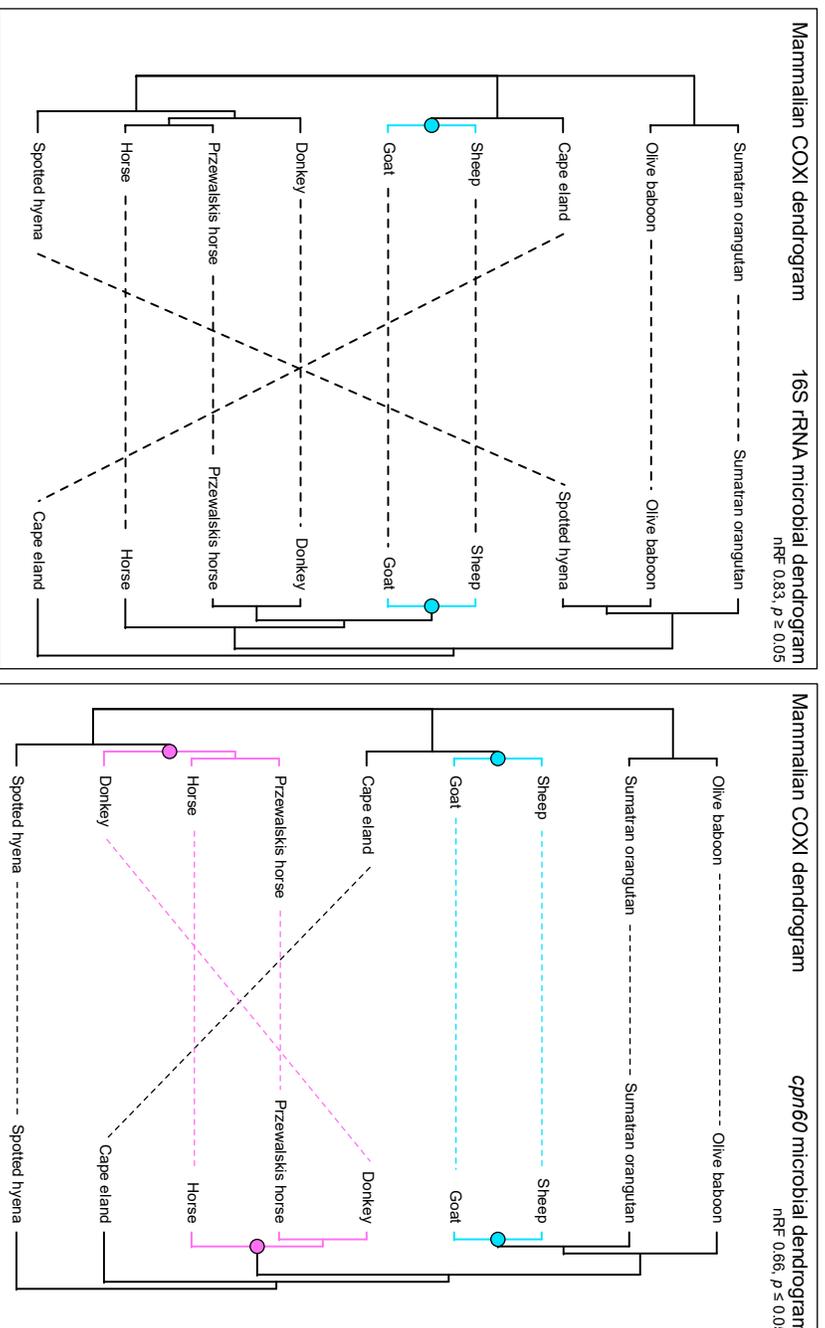


Figure 2-9 Assessment of phylosymbiosis using unweighted UniFrac microbial dendrograms for both the 16S rRNA (left) and *cpn60* (right) genes. Clades that share identical topography between the mammalian COXI and microbial dendrogram, and thus suggest phylosymbiosis, are indicated with colour. Normalized Robinson-Foulds scores are used to indicate tree similarity, and significance was determined by comparing dendrograms against 100,000 randomly generated trees.

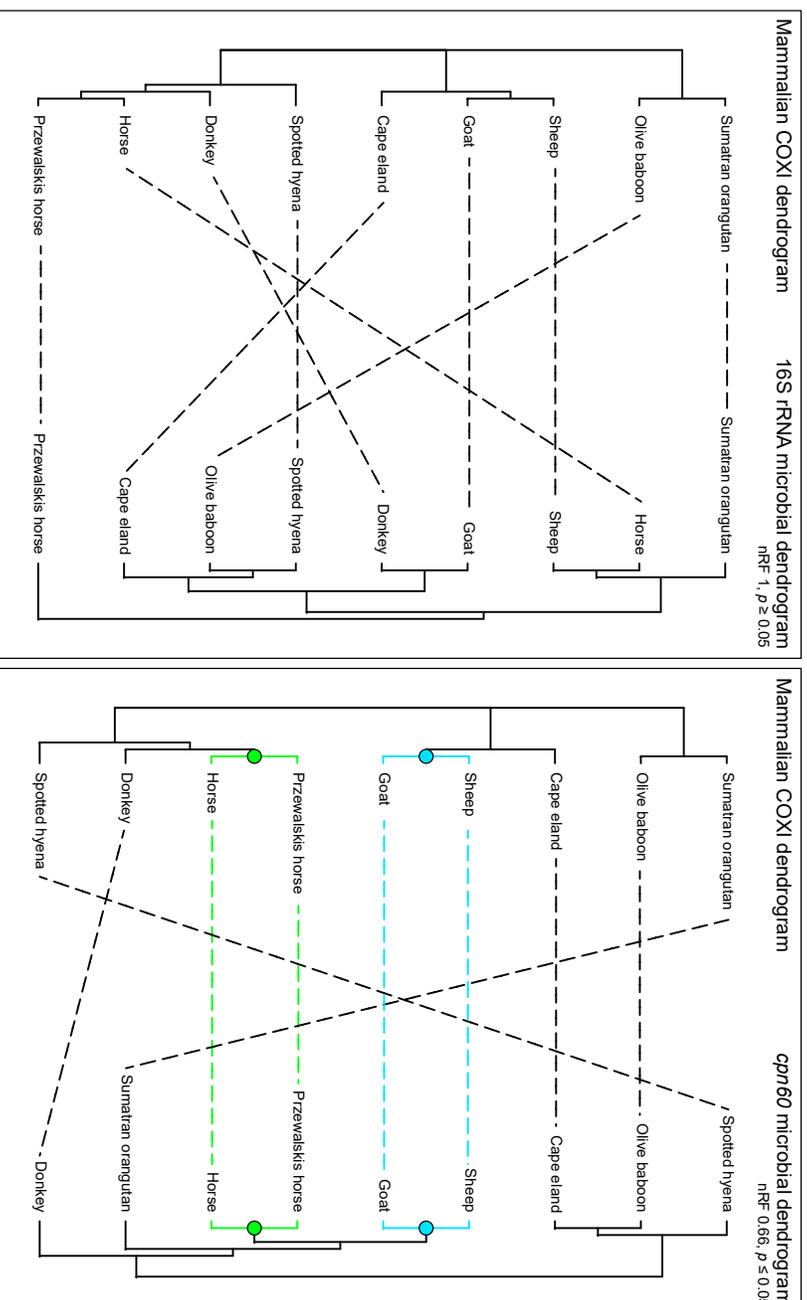


Figure 2-10 Assessment of phylosymbiosis using weighted UniFrac microbial dendrograms for both the 16S rRNA (left) and *cpn60* (right) genes. Clades that share identical topography between the mammalian COXI and microbial dendrogram, and thus suggest phylosymbiosis, are indicated with colour. Normalized Robinson-Foulds scores are used to indicate tree similarity, and significance was determined by comparing dendrograms against 100,000 randomly generated trees.

To demonstrate how the taxonomic resolution provided by the *cpn60* gene might be used to survey specific microbial populations on mammalian skin, reads associated with *Staphylococcaceae* were isolated from both the *cpn60* and 16S rRNA gene datasets and compared. From the 37 samples with greater than 1000 reads, 33 contained *Staphylococcaceae*-associated reads (Figure 2-11). The spotted hyena sample contained no *Staphylococcaceae*-associated reads and was therefore removed from further analysis. The proportion of reads associated with *Staphylococcaceae* varied between mammalian host and ranged from 0.8% (2/2307 reads) to 26.0% (1339/5142 reads), though most samples (23/33, 69.9%) were below 10% relative abundance (Figure 2-12, bottom panel). The olive baboon and Sumatran orangutan samples contained the fewest number of reads associated with *Staphylococcaceae*, with no sample exceeding 1% relative abundance. The distribution of specific *Staphylococcaceae* species differed between mammalian host (Figure 2-12, top panel). Reads associated with *Jeotgalicoccus halophilus* were present among all mammalian hosts excluding the donkey, and an unresolved *Staphylococcaceae* species was similarly present, though absent from the Przewalski's horses and Sumatran orangutan. For the Przewalski's horses, *Macrococcus carouelicus* was the dominant species in most samples though three samples also contained *Salinicoccus* species that were shared with the horse samples. Horse samples were the most variable with overlap between many other mammalian skin samples, though also contained unique sequences associated with *Staphylococcus fleurettii*, *Salinicoccus halodurans*, and *Macrococcus brunensis*. The cape eland samples contained *Staphylococcaceae* populations evenly split between an unresolved *Staphylococcaceae* and *Jeotgalicoccus halophilus*. Comparisons made between the *cpn60* (Figure 2-13, top panel) and 16S rRNA gene (Figure 2-13, bottom panel) *Staphylococcaceae* profiles shows improved definition of *Staphylococcaceae* populations when using the *cpn60* gene. Most of the profiles produced using the 16S rRNA gene have large relative abundances of unresolved *Staphylococcaceae* sequences with only two resolved species (*Macrococcus brunensis* and *Salinicoccus roseus*) present in the dataset. Comparatively, the *cpn60* gene profiles are dominated with species-resolved populations with few exceptions. The disparity in taxonomic resolution between the two phylogenetic markers is most obvious within the Przewalski's horse samples where the large majority of *Staphylococcaceae* associated

sequences become resolved to *Macrocooccus carouselicus* when switching from the 16S rRNA gene to the *cpn60* gene marker.

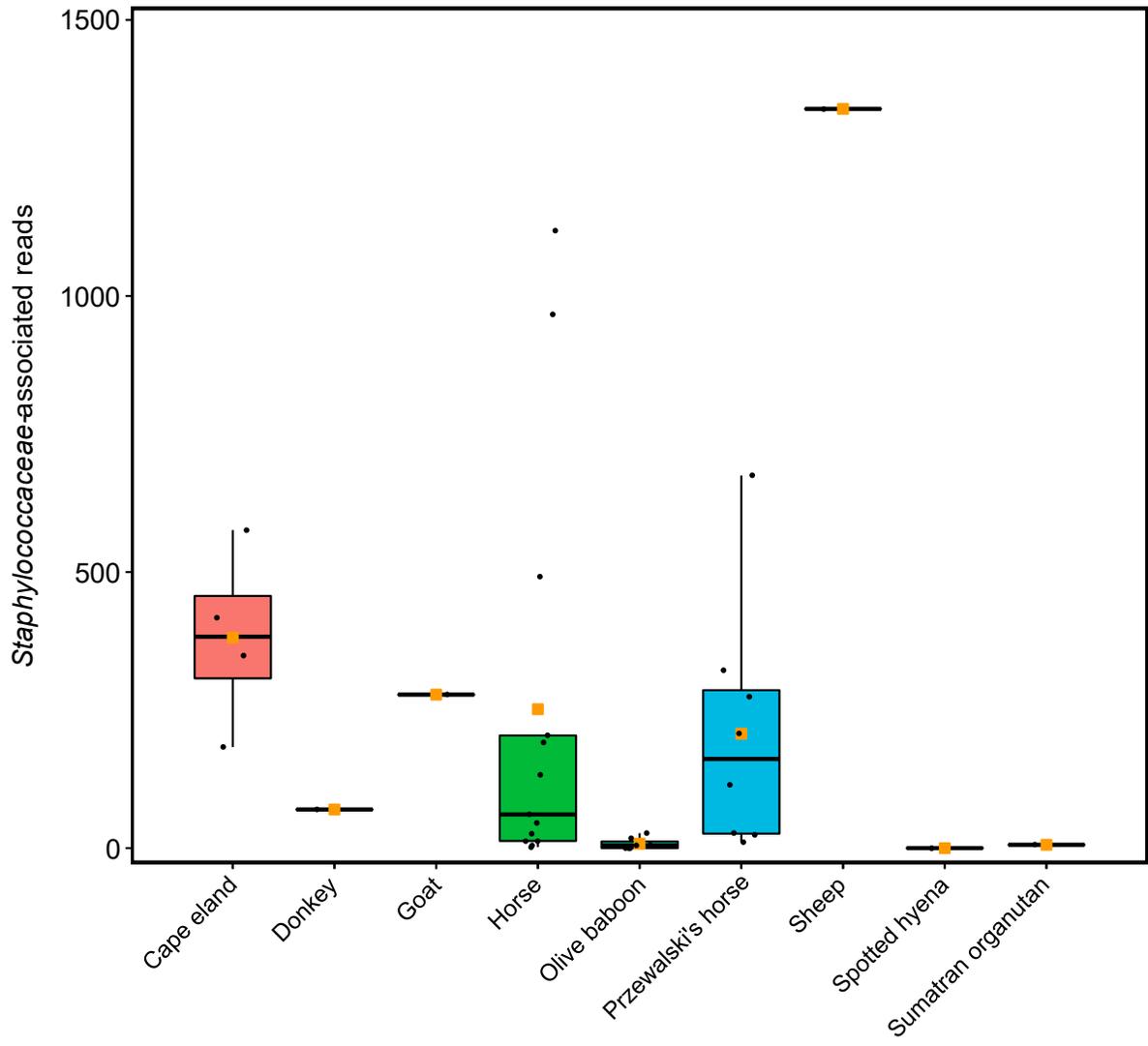


Figure 2-11 Abundance and distribution of *Staphylococcaceae*-associated *cpn60* reads between mammalian hosts. The average number of reads for each host is indicated with an orange square.

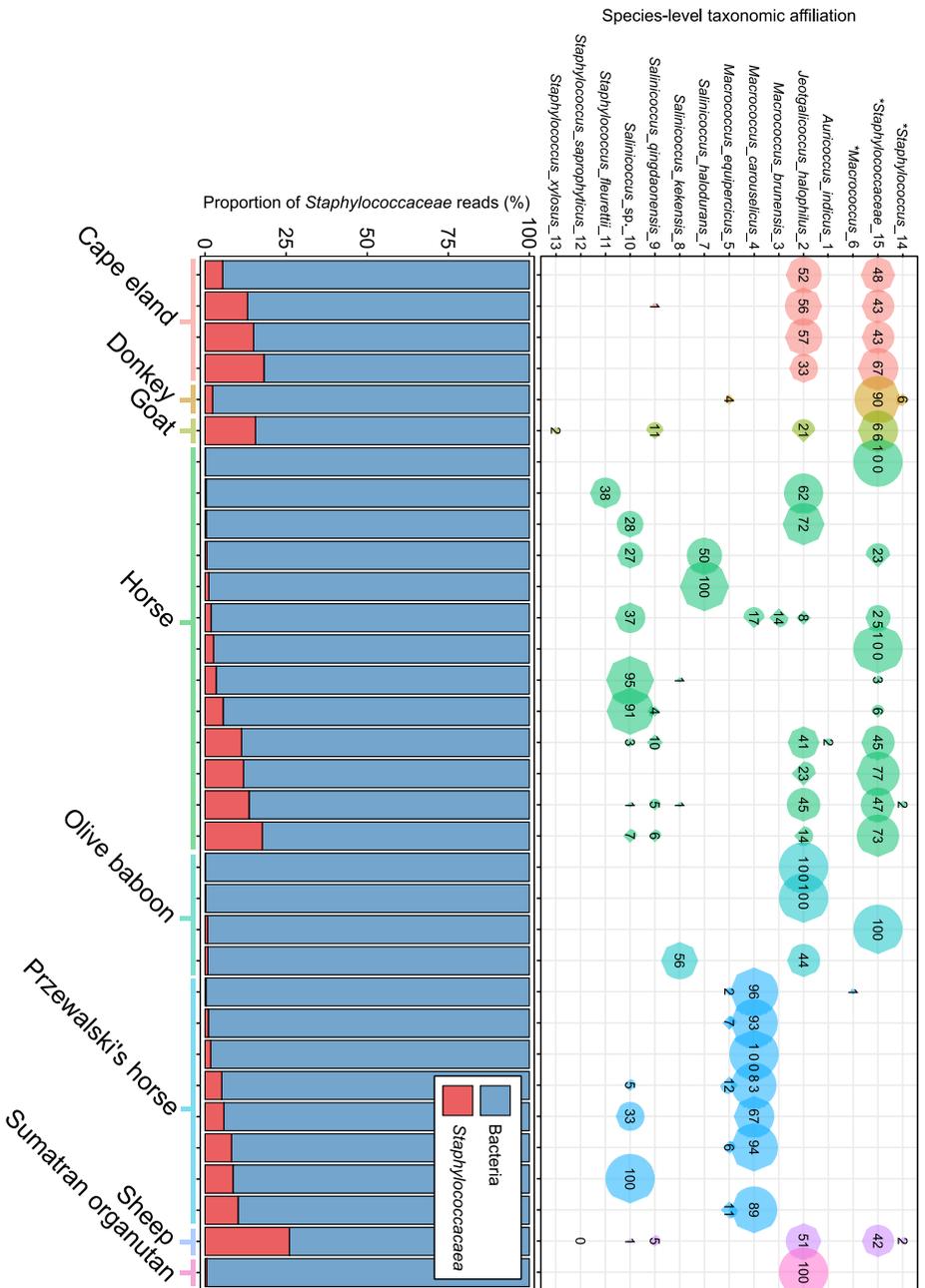


Figure 2-12 The relative abundance of all *Staphylococcaceae* species detected (top panel) and the proportion of *Staphylococcaceae* reads for each mammalian hosts (bottom panel). Taxa unresolved to a species level are indicated with an *. Bubble sizes and associated numeric values represent the relative abundances of taxa for a given sample.

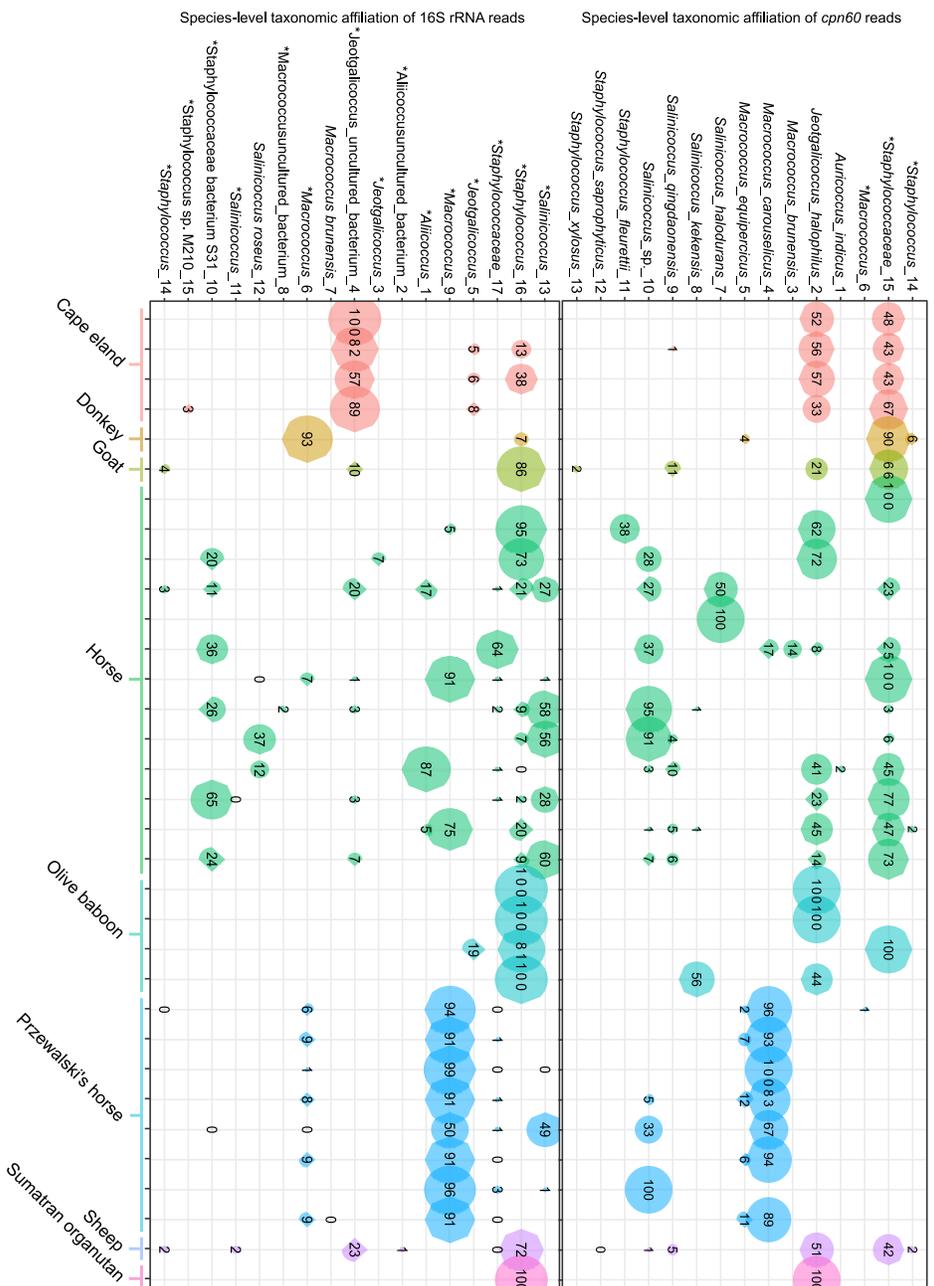


Figure 2-13 Relative abundance and distribution of *Staphylococcaceae* species between mammalian hosts generated by amplification of the *cpn60* (A) and 16S rRNA (B) genes. Taxa unresolved to a species level are indicated with an *. Bubble sizes and associated numeric values represent the relative abundances of taxa for a given sample.

2.4 Discussion

2.4.1 Validating the *cpn60* gene for microbial profiling of mammalian skin and assessing phylosymbiosis

An important goal of this study was to compare microbial profiles generated by the *cpn60* and 16S rRNA genes with respect to community composition and host differentiation. To do this, The Procrustes analysis [130] was used to compare ordinations prepared with these genes and different distance metrics. When applied to PCoA ordinations, the Procrustes analysis scales, translates, and rotates two plots, one atop the other, to produce a final plot that minimizes the sum of squares between all points [130]. This process can be completed repeatedly to assess significance of the correlation between ordinations. With respect to the *cpn60* and 16S rRNA genes, the Procrustes analysis allows for the comparison between microbial community composition without the requirement of identical taxonomic nomenclature. Each of the metrics tested (Bray-Curtis, unweighted/weighted UniFrac) were all strongly correlated and significant, indicating that community compositions generated with the *cpn60* and 16S rRNA gene were similar (Figure 2-7). However, the microbial profiles generated by these genes could be very different but produce the same pattern in multidimensional ordination space. For example, a dataset containing three samples, A, B, and C with unique ASVs 1, 2, and 3 would separate in ordinate space within a PCoA plot. A second dataset containing samples A, B, and C with unique ASVs 4, 5, and 6 would separate similarly, and when assessed with a Procrustes analysis would produce a strong correlation with the first. This is a particular issue for the Bray-Curtis metric, where differences in taxonomic profiles are not accounted for, because the distances among samples are based on ASV presence/absence and relative abundances independent of phylogeny. The inclusion of phylogeny in the UniFrac metric compensates for the potential differences in taxonomy between the *cpn60* and 16S rRNA gene datasets. Referring to the previous example, if ASVs are instead collapsed to a species level and assigned taxonomic identifiers, dataset A, B, and C could now contain sp. 1, sp. 1, and sp. 2. Because two ASVs have been collapsed into identical taxonomic identifiers, the distances between A and B will be reduced and the correlation of the Procrustes analysis between plots will also decrease. This is likely the

reason that the UniFrac metrics tested showed a lower correlation compared to the Bray-Curtis metric.

Comparing microbial profiles generated by two separate universal prokaryotic phylogenetic markers should ideally produce profiles with a similar taxa and relative abundances, which should reflect the true microbial communities of the samples. Although both the *cpn60* and 16S rRNA gene profiles were compositionally similar, the taxonomic profiles had large differences (Figure 2-5). Different gene markers used for amplicon-based studies are subject to different biases and may not always be directly comparable. For example, the 16S rRNA gene is influenced by gene copy number [76, 131], nucleotide GC content [132, 133], and certain bacterial proportions [134], all of which have been shown to impact microbial profiles. Profiles generated with the *cpn60* gene could contain similar biases with exception to the gene copy number, although recent developments have sought to eliminate biases based on primer degeneracy and annealing temperature [106]. The largest influence on the microbial profiles and the cause for discrepancy between sample pairs within this dataset is the separate sequence reference database used for each dataset. The availability of curated *cpn60* gene reference databases is limited. The only currently maintained *cpn60* database is the cpnDB [104], which contains approximately 7,000 sequences in its non-redundant database. Although this was increased to 17,713 sequences by combining it with nucleotide BLAST results, the cpnDB is surpassed by the SILVA 138.1 database, which contains over 510,000 non-redundant 16S rRNA gene reference sequences [93]. A large proportion of the *cpn60* gene reads were unclassified bacteria (i.e., unresolved_1582) or remained unresolved to the genus level. An increase in database coverage for the *cpn60* gene would result in increased classification of sequences and assist in comparing the taxonomic overlap between the *cpn60* and 16S rRNA gene datasets.

In addition to the lack of database coverage, a more immediate factor affecting sample pair congruity within this study is the use of two separate taxonomy databases. In combination with the reference sequence databases, taxonomy databases provide the full taxonomic lineage for affiliated sequences. The direct comparisons of taxonomy profiles

between two phylogenetic markers requires that ASVs be collapsed into taxonomic levels (e.g., genus or species), which is entirely dependent on the taxonomy reference database provided. Much like sequences with a single nucleotide difference may be interpreted as unique ASVs, taxa with names that differ by as little as one character in their lineage will be classified as separate taxa. In this study, the 16S rRNA gene database used the SILVA database [93, 94] that obtains its taxonomic information from a combination of sources, including the NCBI and GTDB [92], and undergoes further manual annotations. In contrast, the cpnDB does not maintain a taxonomy reference database. Instead, the *cpn60* gene dataset uses a taxonomy reference database based on NCBI taxonomy that was specifically generated for this thesis (Chapter 2.2.2). Although manual curation of the taxonomy database was completed at a broad level, fine-level curation was beyond the scope of this study. As such, differences between the *cpn60* and 16S rRNA gene profiles may have been affected by incongruous taxonomies between databases. For example, two *Corynebacterium* genera (i.e., *Corynebacterium* 1_94 and *Corynebacterium*_93) were present in the data (Figure 2-5). A successful collapse to the genus level should have placed the associated ASVs into a single common *Corynebacterium* genus. In this case, the difference is the inclusion of a “_1” at the end the *Corynebacterium* lineage within the 16S rRNA gene taxonomy file (or inversely, a missing “_1” in the *cpn60* gene taxonomy file). Similarly, *Massilia* is separated into *Massilia*_1203 and *Massilia*_1347. Here, the differences are a result of higher taxonomic reclassification: *Massilia* has been reclassified under class *Gammaproteobacteria* within the 16S rRNA taxonomy based on the GTDB [92], whereas the NCBI taxonomy and SILVA database currently maintains the original class *Betaproteobacteria* lineage. As such, ASVs associated with the same genus are listed as two separate taxa instead of one, and the difference is “invisible” at the genus level. Should the *cpn60* gene continue to be used as an alternative phylogenetic marker to the 16S rRNA gene, it would be important that a reference taxonomy database, containing compatible taxonomic lineages with the SILVA database, be maintained alongside the cpnDB so that direct comparisons can be made. Alternatively, a combination of BLAST and Smith-Waterman alignments (watered-BLAST) has been used

previously to assign taxonomy to *cpn60* gene datasets [120], though this method also uses the NCBI taxonomy database and would be subject to the same incompatibility issues.

Despite difficulties with direct taxonomic comparisons, compositional analysis of the *cpn60* gene microbial profiles observed phylosymbiosis within Perissodactyla (Donkey, horse, and Przewalski's horse) and Artiodactyla (sheep and goat) [27] with moderate congruency, using the unweighted and weighted UniFrac metrics (Figures 2-8, Figure 2-9). These observations have been made previously using the 16S rRNA gene, although required Artiodactyla and Perissodactyla dendrograms to be isolated from other mammalian classes [27]. Because the mammalian hosts included in the study varied in location and age, and are more impacted by potentially transient environmental microorganisms, phylosymbiosis patterns could be masked when nonhuman mammals from multiple mammalian classes are included together for analysis [27]. Because this current study relies on samples originally obtained from the previous publication [27], the same confounding factors are present and could be impacting the analysis. Notwithstanding, *cpn60*-based microbial dendrograms produced significant and congruent phylosymbiosis results for Artiodactyla and Perissodactyla, without requiring isolating them from other mammalian classes. These observations were not present when using the 16S rRNA gene profiles and the same collection of mammalian hosts, suggesting that the taxonomic resolution provided by the *cpn60* gene has allowed for more delineation among genus- and species-level taxa. However, mammalian host representation is limited within this current study, and the inclusion of additional samples might result in masking of the phylosymbiosis pattern, as seen previously [27].

One of the largest influences on the *cpn60*-based UniFrac diversity metrics is the abundance of unclassified and unresolved high-level taxa (i.e., phylum, class, order). These taxa can represent a large proportion of sample reads, in some cases more than 50% of the reads within a sample and are present in almost every sample. The presence of abundant and prevalent unclassified taxa would be expected to reduce distances between samples, masking more subtle differences in the microbiota and additional phylosymbiosis patterns when using both the weighted and unweighted UniFrac measures. Although the 16S rRNA gene

reference database is considerably more extensive, the V4-V5 fragment of the 16S rRNA gene lacks sufficient nucleotide diversity to confidently resolve species [95]. Thus, ASVs are classified into less-informative genus-level groups that are shared generally among mammalian hosts. Because it is more likely for common genera to be dispersed among mammalian skin samples, mammalian host microbial profiles become less phylogenetically distinct in the absence of species-level classification. However, the capability of the *cpn60* gene to differentiate species [101] allows for common genera to be further resolved. Specifically for the unweighted UniFrac metric, the species-level phylogenetic resolution provided by the *cpn60* gene would have a considerable impact because these shallow-branch taxa contribute nearly 90% of the sample distance [135]. For the weighted UniFrac, deep-branch taxa are largely responsible for sample distances [135] and therefore should be less influenced by an increase in phylogenetic resolution unless it results in changes to deep-branch topology. That phylogenetic-based UniFrac metrics produced significant phyllosymbiosis results only with the *cpn60* microbial profiles could suggest that the phylogenetic resolution provided by the *cpn60* gene reveals subtle compositional differences not observable when using the 16S rRNA gene.

2.4.2 Using *cpn60* to define species within *Staphylococcaceae*

A proposed benefit of the *cpn60* gene is its ability to universally resolve specific communities of mammalian skin to the species level and reveal host-microorganism associations and potential co-evolutionary relationships. To that end, *Staphylococcaceae* was chosen given the ubiquity of affiliated taxa among mammalian hosts [27] and relevance of certain members to skin health and disease. When using the *cpn60* gene, *Staphylococcaceae* communities had more species-resolved taxa compared to profiles generated using the 16S rRNA gene. The species *Macrococcus carouselicus*, first observed on horses and thus named after a carousel [136], is predominant on Przewalski's horses although absent on the other horses within the study. Similarly, *Macrococcus equipercicus* was resolved alongside *M. carouselicus*, indicating that complex communities exist within more general 16S rRNA gene genus classifications. The distribution of these taxa suggests either a strong host-specificity for Przewalski's horses, potential co-evolutionary histories, or an environmental influence.

The Przewalski's horse is considered a true "wild horse" having remained undomesticated and relatively isolated to the Asian steppes [137]. The skin of Przewalski's horses could be different to that of the common domesticated horses included in this study, and thus harbour distinct microbial communities. However, although the horses included in this analysis were regularly brushed (daily to weekly), the Przewalski's horses were not groomed in any capacity. As such, the *Staphylococcaceae* community of the Przewalski's horse could represent a more "natural" state common to all horses, in contrast to the domesticated horses where their microbiomes are continually disturbed by human interaction. Regardless, differences in, or disruption to, *Staphylococcaceae* populations could impact horse species and their susceptibility to disease given that *Staphylococcaceae* has a major role in equine pastern dermatitis [138].

The most prominent *Staphylococcaceae* member affiliated with the *cpn60* dataset is *Jeotgalicoccus halophilus*. The comparable 16S rRNA gene dataset [27] also contains the genus *Jeotgalicoccus* with similar proportions for the cape eland and several horse samples, although most *Staphylococcaceae* affiliated sequences belong to the *Macrococcus* genus (83.7%). None of the *Jeotgalicoccus* ASVs within the 16S rRNA gene dataset were resolved to *J. halophilus*. The genus *Jeotgalicoccus* was originally isolated from a traditional Korean fermented seafood [139], with other members of the genus captured as aerosols from pig [140] and turkey [141] farms. Specifically, *J. halophilus* was first isolated from a salt lake [142] and has since been detected as an airborne bacterium in hatcheries [143] and in association with marine corals [144]. There is no additional literature that mentions *J. halophilus* in association with mammalian skin, although this does not exclude their previous detection. Querying the NCBI database for "Jeotgalicoccus halophilus" returned sequences obtained from the environmental studies previously mentioned, though also included a bovine mastitis study in which *J. halophilus* was detected [145].

Although the genus *Jeotgalicoccus* was detected previously on mammalian skin using the 16S rRNA gene [27], the *cpn60* gene has allowed for taxonomic delineation of the *J. halophilus* from the other species therein. Within the Primates, *J. halophilus* represents the totality of the *Staphylococcaceae* associated reads (Figure 2-12, top panel), although this is

likely due to limited sequencing depth and low read count. However, the Cape eland samples have a considerably higher *Staphylococcaceae* read depth, with *J. halophilus* representing at least half the total reads, and are the predominant species-resolved taxa. Similarly, the sheep sample, which contained the highest proportion of *Staphylococcaceae*-associated reads, also had a high proportion of *J. halophilus*, suggesting that the relative abundance and prevalence of this taxon is not an artifact of limited read depth. Additionally, given its absence in several samples and two mammalian hosts (i.e., donkey and Przewalski's horse), it is unlikely to be a contaminant introduced from within the lab environment during sample extraction and processing.

The importance or role of *J. halophilus* in the context of mammalian skin is unknown. As a facultative anaerobe with basic metabolic requirements, a growth range of 4 to 40 °C, and salt tolerability of 0.1 to 16% w/v [142], it is well suited for survival on the skin. As well, it is coagulase and oxidase positive and resistant to several natural antibiotics [142], which could indicate its ability to act as an opportunistic pathogen. Similarly, for all the *Staphylococcaceae* observed in association with mammalian hosts, it is difficult to make conclusions about their involvement in skin function, health, and disease. For example, *Staphylococcus fleuretti*, found in a single horse sample, is a coagulase-negative organism associated with various animal diseases and has been indicated as a potential contributor to methicillin resistance within the environment [146]. However, the horse from which the sample was taken had no reported skin health issues, though did have a mild respiratory infection [27]. Even well-established pathogens, like *S. aureus* or *S. epidermidis*, can exist in non-disease states within the skin microbiome only causing issues when the skin barrier is broken or when the community is disrupted [68, 69]. Thus, the detection of disease-associated genera or species on mammalian skin can provide only a limited insight into host-microbe dynamics. Ultimately, more work is necessary to further define the interactions these bacterial species might have with their mammalian hosts. Regardless, that they have been detected at all on mammalian skin demonstrates the ability for *cpn60* to resolve species and its application towards profiling the mammalian skin microbiome.

Despite the taxonomic resolution provided by the *cpn60* gene and the observations of specific *Staphylococcaceae* communities of the mammalian skin microbiome, further analysis of the *cpn60* gene dataset is limited by shallow sequencing depth and mammalian host representation. Many of the samples originally sent for sequencing (58%) did not produce a sufficient read depth (i.e., > 1000 reads) and had to be removed to avoid bias during analysis. The microbial biomass on mammalian skin is variable and can be comparably low, with dry-swab sampling of the skin producing the least amount of biomass compared to other methods [147]. The samples used in this current study were collected via the dry-swabbing method [27] and therefore are likely to have relatively low biomass and associated DNA yields. Additionally, the samples are several years old (>3 at time of sequencing), have been stored at -20°C instead of the recommended -80°C for skin samples [147]. Combined, these factors may have impacted the integrity of the extracted DNA, resulting in reduced amplification yield and subsequent sequencing depth. The *cpn60* gene itself has a median copy number of one copy per genome [101] and so would also limit the amount of template available in PCR and the number of reads available for sequencing. Future exploration of specific communities of the mammalian skin microbiome using the *cpn60* gene would benefit greatly from using genomic DNA that has been recently extracted from mammalian skin samples, ideally collected using wet-swabbing or tape-stripping to increase biomass [147].

2.5 Conclusion

The increased taxonomic resolution provided by the *cpn60* marker gene provided greater phylogenetic context for the mammalian skin microbiome profiles generated here. Using the *cpn60* gene, phylosymbiosis was observed in mammals belonging to class Perissodactyla and Artiodactyla when using the phylogeny-based UniFrac distance metric, which was not possible for the 16S rRNA gene. Resolving species from within genus classifications provide important insight into the distribution and presence of specific taxa among mammalian hosts and their potential impacts on host skin health and disease. The *cpn60* gene also revealed previously unobserved associations between mammalian hosts and specific taxa, such as the case of *Jeotgalicoccus halophilus*, that might have otherwise

remained unnoticed. Importantly, amplification of the *cpn60* gene does not exclude specific bacterial communities over others (i.e., archaea-specific or species-specific primers) and can thus be used for both whole microbial community and species-specific profiling. Although the 16S rRNA gene is likely to remain as the dominantly used marker for amplicon-based studies, the *cpn60* gene is complementary to microbiome studies where universal low-level taxonomic resolution is desired. However, if *cpn60* amplicon studies are to be compared with those produced using the 16S rRNA gene, it is imperative that a standardized taxonomy database be maintained alongside the cpnDB sequence reference database. In lieu of a separate taxonomy database, progress towards integrating the cpnDB with existing taxonomy databases, such as SILVA [93, 94] or the ribosomal database project [148] will be valuable for future *cpn60*-based research.

Chapter 3¹

Archaea are rare and infrequent members of mammalian skin microbiome

3.1 Archaea on mammalian skin

Mammalian skin, which includes both humans and non-human mammals, hosts spatially and temporally diverse microbial communities due to extensive chemical and physical variability. Skin topography and epithelial cell type [149], underlying vasculature and endocrine system physiology [150], moisture and oil content [151], and pheromones [11] can all influence microbial colonization and establishment, and these vary according to mammalian host and body site. For instance, sebaceous and apocrine gland secretions create local anoxic areas that provide metabolic substrates for microbial growth [151, 152]. Human apocrine glands are located primarily in the armpits (i.e., axillae), whereas rhesus monkeys and baboons have a more diffuse apocrine system [5]. The microbial communities that develop according to these physiological differences, primarily represented by *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* [21], have direct and measurable impacts on host health. For humans, shifts away from a “normal” microbial community composition are associated with eczema [153, 154] and psoriasis [155]. Similar links between mammalian dysbiosis and disease have been reported for dogs [23, 156], bovines [157], and camels [158]. Therefore, the interconnectivity between mammalian skin physiology, host health, and skin microbiota underscores the importance of elucidating factors that control the diversity and composition of skin-associated microbiomes.

The microbiota of mammals are dominated by bacteria [22, 44, 151, 159, 160], although high-throughput sequencing approaches have captured archaeal signatures as well. Early research exploring archaea in mammals focused primarily on the gastrointestinal tract, where methanogenic *Methanobacteriota* (formerly *Euryarchaeota* [91]) members were first detected [72, 161–163]. In this context, archaeal communities have been characterized sufficiently to predict mutualistic contributions to host metabolism [42, 73] as well being

¹ A version of this chapter has been accepted for publication as: Umbach AK, Stegelmeier AA, Neufeld JD. Archaea are rare and uncommon members of the mammalian skin microbiome. *mSystems* 2021. DOI:10.1128/mSystems.00642-21

implicated in disease etiology [164, 165]. Archaea are now widely accepted as members of the gut and mucosal microbiota of mammals and more recently have been reported within human breast milk [166]. In contrast to many gut-associated studies, the mammalian skin “archaeome” is poorly characterized, to the extent that archaea have often been excluded from comprehensive skin microbiome reviews due to insufficient data [151, 167, 168]. However, in the last half decade archaea have been reported as skin microbiota members of multiple individuals and body locations [39–42]. Limited existing research on the skin archaeome demonstrates the need for additional study of mammalian skin and its associated archaeal populations in this emerging field of study.

To our knowledge, Probst and colleagues were the first to demonstrate that archaea can be detected as representatives of human skin microbiota [43]. By sampling torsos of 13 individuals, their study using archaea-targeting methods estimated that human skin harbours an average archaeal relative abundance of 0.60% and as high as 4.2% of the total microbial community [43]. In considering bacterial and archaeal 16S rRNA gene copy number, Probst et al. suggests that these values could increase to an average of 1.40% and maximum of 9.86% [43]. A subsequent study also using an archaea-targeted approach detected an average human skin archaeal community of 1.1% for ages 1 to 11 years, 0.2% for ages 12 to 60, and 4.7% for ages 61 to 75, with a maximum of up to 10.4%, and were dominated by putative ammonia-oxidizing archaea (AOA) of the class *Nitrososphaeria* (previously phyla *Crenarchaeota/Thaumarchaeota*, now *Thermoproteota*) [40]. In contrast, studies using universal prokaryotic (i.e., *Bacteria* and *Archaea*) detection methods suggest a limited skin-associated archaea community. Extensive skin sampling of cohabitating couples revealed average archaeal sequence relative abundances of less than 0.5%, with archaea detected only for a few samples [22]. A human skin study evaluating the impact of polycyclic aromatic hydrocarbons pollutants observed similarly low archaeal relative abundances of less than 0.01% [169]. Furthermore, a large-scale human metagenome survey revealed that archaea accounted for approximately 1% of all human metagenome sequences, with the majority of these sequences affiliated with the *Methanobacteriota* phylum from gut or mucosal membranes [44]. A large study consisting of 589 mammalian skin swab samples concluded

that less than 0.1% of all sequences were associated with archaea, with most sequences affiliated with the *Methanobacteriota* and *Halobacteriota* phyla [27]. More recently, a shotgun metagenome study investigating udder cleft dermatitis on dairy cows observed archaea associated reads nearing 7% relative abundance, represented primarily by *Methanobacteriota* [170].

Skin-associated built environments, such as keyboards and door handles, can potentially act as an extension of the skin environment through high frequency contact and deposition of skin microorganisms. Both keyboard and phone microbiomes are influenced by the corresponding finger microbiomes of their users [24, 171] and keyboard microbiomes reflect the skin of their users such that they can be used to identify their specific user [24]. From a survey of campus door handle microbiomes, the results revealed that door handles produced microbial profiles that were more similar to skin than to soil or other external environments [118]. Archaeal taxa detected on these handles were primarily affiliated with *Methanobacteriota* and *Halobacteriota* phyla, although archaea were detected at a low relative abundance of less than 0.01% of all amplicon sequences [118]. Thus, profiling the archaeome of skin-associated surfaces will enable a better understanding of the detected skin-associated archaea and their allochthonous or autochthonous origins.

In order to help further address archaeal diversity and relative abundance on mammalian skin, this work explored archaeal sequences associated with skin and skin-associated environments using previously published data from human skin [22], non-human mammalian skin [27], and door handles [118], as well as newly generated fingertip and keyboard sample data. Using multiple primers sets for a subset of samples, this study evaluated 16S rRNA gene amplicon sequence profiles from 1,058 skin samples (i.e., 458 human, 600 non-human mammalian) and 630 skin-associated samples (i.e., 240 keyboard, 390 door handles), for a total of 1,688 sample profiles. The results demonstrate infrequently detected presence and low archaeal relative abundances on skin and skin-associated surfaces, with only few exceptions. When detected, mammalian skin sequences that affiliated with archaea were primarily assigned to the *Methanobacteriota* and *Halobacteriota* phyla; putative AOA from the *Nitrososphaeria* were largely undetectable.

3.2 Skin archaea methods

3.2.1 Sample collection, selection, and processing

Keyboard and finger swab samples from individuals between the ages of 18 and 70 were collected for this study in accordance with the University of Waterloo Office of Research Ethics (ORE) project 40212. After participants washed their hands, the index and middle fingers on the left and right hands were swabbed for 30 seconds in a circular motion with a sterile foam swab (Puritan Medical Products, ME, USA). Participants were then asked to perform a typing exercise ten times (“The quick brown fox jumped over the lazy dog 1234567890”). The same fingers were swabbed again using new sterile swabs. Twenty-four keyboard keys per keyboard were selected to capture different usage frequencies based on an analysis with WhatPulse version 2.8.0. Each individual key was swabbed for 30 seconds with a new sterile swab. All swabs were then stored in applicator tubes in a -20°C freezer. Genomic DNA was extracted from swab samples using the PowerSoil DNA Isolation Kit (Qiagen, Canada) using the manufacturer’s protocol with minor protocol modifications. Swab tips were removed with a flame-sterilized sterile scalpel and deposited into a PowerSoil bead-beating tube and incubated at 70°C for 10 minutes on a rotating holder. The tubes were then subjected to mechanical lysis using a FastPrep24 agitator (MP Biomedicals, OH, USA) at 5.5 m/s for 45 seconds and extracted following manufacturer protocols. The samples were eluted in 10 mM Tris and stored at -20°C prior to PCR amplification and sequencing.

Previously published data from mammalian [27] and human [22] skin studies, and a campus door handle survey [118], were compiled here for comparison. The Pro341F/Pro805R V3-V4 primers used in these studies were selected based on their original design for increased archaeal detection [172]. In order to test whether archaeal distributions were influenced by universal prokaryotic primers used in generating results for these earlier studies (Pro341F/Pro805R; V3-V4 regions; [172]), this research obtained extracted DNA from 38 human and 54 mammalian skin samples (92 samples total) to generate new 16S rRNA gene profiles using additional universal prokaryotic primers as detailed below. Samples included in the primer set comparison were selected based on previously observed

archaeal sequences in those samples such that samples with greater archaeal relative abundances were selected to increase the likelihood of obtaining sequences for direct comparison.

3.2.2 PCR and sequencing

The V4-V5 regions of the 16S rRNA gene were amplified from all fingertip swabs, keyboard swabs, and 90-sample subset of human/mammalian skin (see above) using universal prokaryotic primers 515F-Y (5'- GTGYCAGCMGCCGCGGTAA - 3') [87] and 926R (5' - CCGYCAATTYMTTTRAGTTT - 3') [89]. Both primers were modified to include a 6-base barcode sequence used for identification of amplicons, an adaptor sequence for flow cell binding, and an Illumina primer binding site [173]. The PCR was performed in a sterile ISO 5 HEPA PCR hood, which was cleaned with 70% ethanol before being treated with UV for 15 minutes prior to use. A PCR master mix was created using UV-treated PCR-grade water, 1x ThermoPol buffer, 0.2 μ M forward primer, 0.2 μ M reverse primer, 200 μ M dNTPs, 15 μ g BSA, 0.625 units of Hot Start *Taq* DNA polymerase (New England Biolabs, MA, USA), and 1 μ L of DNA template in each 25- μ L reaction. Positive and negative PCR controls were included, as were extraction kit controls. Amplification was performed using a T100 thermal cycler (Bio-Rad Laboratories, Canada) using the following reaction conditions: 95°C initial denaturation for 3 minutes, and 40 cycles of 95°C denaturation for 30 seconds, 55°C annealing for 30 seconds, 68°C extension for 1 minutes, with a final extension at 68°C for 7 minutes. All PCR amplifications were performed in triplicate then pooled in equimolar quantities before purifying on a 1% ethidium bromide gel. Amplicons were extracted from the gel and purified using a Wizard SV Gel and PCR Clean-Up System (Promega, WI, USA). The library was diluted to 8 pM and 15% PhiX control v3 (Illumina, Canada) was added prior to sequencing. The 515F-Y/926R samples were sequenced on a 2x250 cycle TG MiSeq Reagent Nano Kit v2 (Illumina Canada, MS-103-1003) on a MiSeq (Illumina). The keyboard and fingers samples were sequenced using a 2x250 cycle MiSeq Reagent Kit v2 (Illumina Canada, MS-102-2003).

3.2.3 Processing of sequencing reads

Sequence reads were demultiplexed using the MiSeq Reporter software version 2.5.0.5 (Illumina). Demultiplexed sequences were processed to generate amplicon sequence variants (ASVs) using QIIME2 [174], managed by Automation, eXtension, and Integration of Microbial Ecology (AXIOME) v3.0 [175]. The impact that truncation length and primer sequence removal methods might have on downstream archaeal and bacterial read processing archaeal proportions was assessed. Forward and reverse reads were trimmed using cut-adapt [176] version 2019.7.10 to a shared 515 to 805 nucleotide V4 region across all datasets using primers 515F-Y and Pro805R (3' - GACTACNVGGGTATCTAATCC - 5'). Trimmed reads were denoised, merged, and chimeras removed using DADA2 version 2019.10.0 [123] while maintaining a minimum of 12-base overlap for the forward and reverse reads. The ASVs were classified using a naïve Bayes classifier trained with the SILVA 132 SSURef NR99 database [94], with additional taxonomic annotation and reassignment using the GTDB in order to ensure compliance with SILVA 138 classifications [91]. Additional 454 pyrosequencing data from a previous human skin microbiome study (“Roche 454”) [40] was imported as single-end reads through QIIME2 using the *qiime dada2 denoise-pyro* command and trimmed using an identical workflow to the paired-end reads, excluding the merge-denoising step. Negative controls were manually inspected to ensure that they were distinct from sample profiles and were treated appropriately. Only two ASVs had overlap among three controls and four non-human mammalian skin samples. Because these ASVs were observed in only a small number of controls (3/81) and had low read proportions for both the controls and samples, these ASVs were left in to avoid removal of potentially valid skin-associated archaeal reads. Because the Roche 454 dataset was generated with archaea-specific primers, these data are limited to analysis of the identities and distributions of archaeal genera rather than their relative abundances in relation to bacteria.

All 515F-Y/926R sequences generated for the current study were deposited in the European Nucleotide Archive (ENA) under project accession number PRJEB42587 (human/mammalian skin 92-sample subset) and PRJEB42589 (finger and computer keyboard swabs). Sequence data from project IDs PRJNA385010 (mammalian skin [27]),

PRJNA345497 (human skin [22]), and PRJNA313528 (human skin [40]) were retrieved from the Sequence Read Archive. Sequence data from project PRJEB10962 (campus door handles [118]) was retrieved from the European Nucleotide Archive.

3.2.4 Tree generation

The 16S rRNA gene multiple sequence alignment of all archaeal ASVs was performed using ClustalW [126], with a gap opening penalty of 15.0 and a gap extension penalty of 6.66, in MEGA X version 10.1.8 [127]. Sequence alignments were trimmed for gaps as appropriate. A maximum likelihood tree was generated in MEGA X using a GTR +G +I nucleotide substitution model; confidence was assessed with 1000 bootstraps.

3.2.5 Assessment of primer coverage

Universal prokaryotic and archaea-specific primers used in this current and previous studies were analyzed *in silico* for their coverage of the domain *Archaea*. Forward and reverse primer sequences were entered into SILVA TestPrime version 1.0 using the SSU 138 database [177]. Primers were tested without allowing mismatches (“Zero mismatch”) and by allowing a single mismatch provided that it did not occur within five bases proximal to the 3’ end (“One mismatch”). Database coverage was parsed from the “taxonomy browser” into three categories: “Bacteria”, “Archaea” and “Nitrososphaeria”.

3.3 Results

This study evaluated archaeal distributions within eight 16S rRNA gene amplicon datasets (Table 3-1; Supplemental dataset S1 [https://figshare.com/articles/dataset/Supplemental_dataset_S1_xlsx/14248580]). Four were collected from previously published data (“mammalian skin”, “human skin”, and “campus door handles”, “Roche 454”), two were from newly obtained samples (“keyboard” and “fingers”), one represented a subset of the mammalian and human skin data with 92 samples prioritized by archaeal presence (“Pro341F/Pro805R”), and another was composed of the same sample subset that was processed again for sequencing by using an alternate universal prokaryotic primer set (“515F-Y/926R”). All primers used to generate amplicon data included in this study were tested *in silico* for primer coverage. The Pro341F/Pro805R and 515F-Y/926R comparison datasets were generated to test for possible primer bias against archaeal 16S rRNA gene sequences in previous data. Overall, coverage of the *Archaea* domain for the universal prokaryotic primers (i.e., Pro341F/Pro805R, 515F-Y/926R) was extensive (Table 3-2). Both primers had greater than 65% coverage of *Archaea* at zero mismatches, although *Nitrososphaeria* coverage was limited depending on the primer set (515F-Y/926R, 83%; Pro341F/Pro805R, 16%). However, with only one mismatch permitted (not proximal to the 3' end), both primers had greater than 85% coverage for both *Archaea* and *Nitrososphaeria*. Conversely, archaea-specific primers (i.e., 344af/517ur, 344af/915ar) that were used to generate Roche 454 data had below 50% coverage for *Archaea* at zero mismatch but increased to 65% to 71% for one non-3'-end mismatch. Coverage of *Nitrososphaeria* with these archaea-specific primers remained below 40% for both zero and one mismatch.

Table 3-1 Summary of skin and skin-associated datasets. The total reads represent the number of reads output following the merge-denoise and chimera removal steps in the workflow. The data presented for Pro341F/Pro805R is a subset of data from the human and mammalian skin datasets.

Study	Samples	Sample type	Method	Primers	Total reads	Average reads per		PCR cycles
						sample	sample	
Non-human								
Mammalian Skin [27]	546	Skin	Swab	Pro341F/Pro805R	5,238,782	9594 ± 7840		*40
Human Skin [22]	340	Skin	Swab	Pro341F/Pro805R	7,631,068	22444 ± 17611		*40
		Human and						
Pro341F/Pro805R	92	mammalian skin	Swab	Pro341F/Pro805R	1,000,284	10872 ± 9490		*40
		Human and						
515F-Y/926R	92	mammalian skin	Swab	515F-Y/926R	573,461	6233 ± 4803		40
Fingers	80	Fingers	Swab	515F-Y/926R	1,722,646	21533 ± 12498		45
Keyboard	240	Keyboards	Swab	515F-Y/926R	5,972,034	24883 ± 15353		45
Door Handles [118]	390	Door handles	Swab	Pro341F/Pro805R	4,859,381	12252 ± 5359		35
Roche 454 [40]	22	Human skin	Swab	344af/917ar	54,554	2,479 ± 5466		35

* Nested PCR was performed by first amplifying the V3-V4 hypervariable region of the 16S rRNA gene for 25 cycles, followed by 15 cycles for Illumina adapter ligation.

Table 3-2 Coverage comparison of the bacterial and archaeal coverage of universal prokaryotic and archaea-specific primers. Each primer set was tested *in silico* using SILVA TestPrime version 1.0 [177] using the SSU 138 database. Primers were tested without allowing mismatches (“zero mismatches”) or allowing a single mismatch provided that it did not occur within five bases proximal to the 3’ end (“one mismatch”).

	Coverage with zero mismatches			Coverage with one mismatch		
	<i>Bacteria</i> (%)	<i>Archaea</i> (%)	<i>Nitrososphaeria</i> (%)	<i>Bacteria</i> (%)	<i>Archaea</i> (%)	<i>Nitrososphaeria</i> (%)
515F-Y/926R	84.5	81.0	82.5	90.7	88.8	86.8
Pro341F/Pro805R	82.6	68.2	16.2	88.4	86.3	92.2
344af/517ur	0	30.7	15.1	0	71.2	36.8
344af/915ar	0	44.6	12.9	0	65.4	32.0

To assess the impact that truncating the 16S rRNA gene amplicons from different studies to a common 515 to 805 region might have on archaeal taxonomic proportions and relative abundance, several truncation strategies were evaluated. Archaeal taxonomic proportions remained largely unchanged regardless of method, except for the quality-trimmed “515_Tr” approach (Figure 3-1). The proportion of archaea and bacteria-associated reads was also unaffected for most of the datasets with the chosen method, although the nonhuman mammalian skin and door handle datasets showed an increased in bacteria-associated reads and a decrease in archaea-associated reads (Figure 3-2).

Archaeal relative abundances (Figure 3-3A) and ASV profiles (Figure 3-3B) were similar based on profiles generated with the 515F-Y/926 and Pro341F/Pro805R primer pairs. Archaea were detected in 26 and 18 of the 92 samples (28% and 20% of samples) for the 515F-Y/926R and Pro341F/Pro805R primer pairs, respectively (Table 3-3), with an average archaeon-associated read count of 25 ± 49 and 14 ± 19 (Supplemental dataset S1 [https://figshare.com/articles/dataset/Supplemental_dataset_S1_xlsx/14248580]). Of the 20 samples with detectable archaea generated with only one of the two primer pairs, archaeal relative abundances did not exceed 0.3%, except for a single mammalian skin sample with a relative abundance of 0.65%. Archaea on human skin were largely undetected, regardless of the primer pair used. For samples with archaea detected using both primers sets, archaeon-associated ASV profiles were similar (Figure 3-3B). Samples with a low number of reads varied the most from their paired samples, whereas pairs with higher read counts were highly similar (Figure 3-3B). The 16S rRNA gene profiles generated using the 515F-Y/926R primer set typically contained additional archaeon-associated ASVs (Figure 3-3B), albeit in lower abundance compared to the dominant ASVs common to paired samples analyzed with both primer sets.

Of the eight datasets included in our analysis (Table 3-3), the subset mammalian skin samples that were amplified with 515F-Y/926R universal prokaryotic primers had the highest archaeal relative abundances, representing $0.12 \pm 0.22\%$ of all reads overall (Figure 3-4A). The non-human mammalian skin sample set had the second highest archaeal relative abundance ($0.08 \pm 0.04\%$) and accounted for the majority of archaea-positive samples (194;

65.5%) within the study, although most samples (352; 64.5%) did not contain any detected archaea. Human skin had a very low archaeal relative abundance ($2.0 \times 10^{-3} \pm 0.02\%$) with only a fraction of samples (5.9%) containing archaeon-associated sequences. The archaeal relative abundance of the finger dataset was similarly low ($0.01 \pm 0.06\%$) and confined to two samples (2.5% of dataset), and keyboard and door handle datasets were also low ($0.02 \pm 0.09\%$ and $7.0 \times 10^{-3} \pm 0.04\%$). Despite these average relative abundances, most samples from all eight datasets (excluding the previously published Roche 454 dataset, 1,392; 82.5%) had no detected archaeal reads, despite thousands of reads per sample. Although most samples had no detected archaea, several samples (296; 17.5%) contained archaeon-associated sequences, though none exceeded 2% relative abundance. For example, sample “16SOB” was sourced from an olive baboon and had the highest archaeal relative abundance of any mammalian skin sample (1.51% with primers Pro341F/Pro805R; 1.82% with primers 515F-Y/926R). The only other sample with an archaeal relative abundance greater than 1% belonged to the keyboard dataset, at 1.12%, sourced from a female participant between 20 and 29 years of age. Together, the data generated with both universal primer sets yielded only a small proportion of archaea-positive skin or skin-associated surface samples.

The taxonomic distribution of archaea (Figure 3-4B) and the number of archaeal ASVs varied among datasets and were dominated primarily by only a few phyla. Non-human mammalian skin was dominated by *Methanobacteriota* sequences, which constituted 82.2% of all archaeal reads. *Halobacteriota* and *Nitrososphaeria* were the next most prevalent phyla, corresponding to 14.0% and 1.99% of archaeal reads, respectively. Conversely, the human skin archaeome was comprised of *Halobacteriota* (43.8% of archaeal reads), *Methanobacteriota* (24.4% of archaeal reads), and *Nanoarchaeota* (22.7% of archaeal reads). Similar to the mammalian skin archaeome, *Nitrososphaeria* represented a small proportion of archaea sequences (4.6% of archaeal reads). The finger swab dataset contained archaea belonging to only the *Methanobacteriota* and *Halobacteriota* phyla, at 74.6% and 25.4% of archaeal reads, respectively. The Pro341F/Pro805R and 515F-Y/926R datasets, representing a subset of all human and mammalian skin samples, both revealed high proportions of *Methanobacteriota* and *Halobacteriota*. Overall, within all archaeal sequences from the

subset of all mammals and humans, the 515F-Y/926R dataset showed an increase in *Nitrososphaeria* proportions (18.7% of archaeal reads) compared to Pro341F/Pro805R (7.3% of archaeal reads). Campus door handles were dominated by *Methanobacteriota* (46.9% of archaeal reads) and *Halobacteriota* (42.4% of archaeal reads). In contrast, sampled computer keyboards were dominated by *Nitrososphaeria* (54.3% of archaeal reads) and *Nanoarchaeota* (25.7% of archaeal reads).

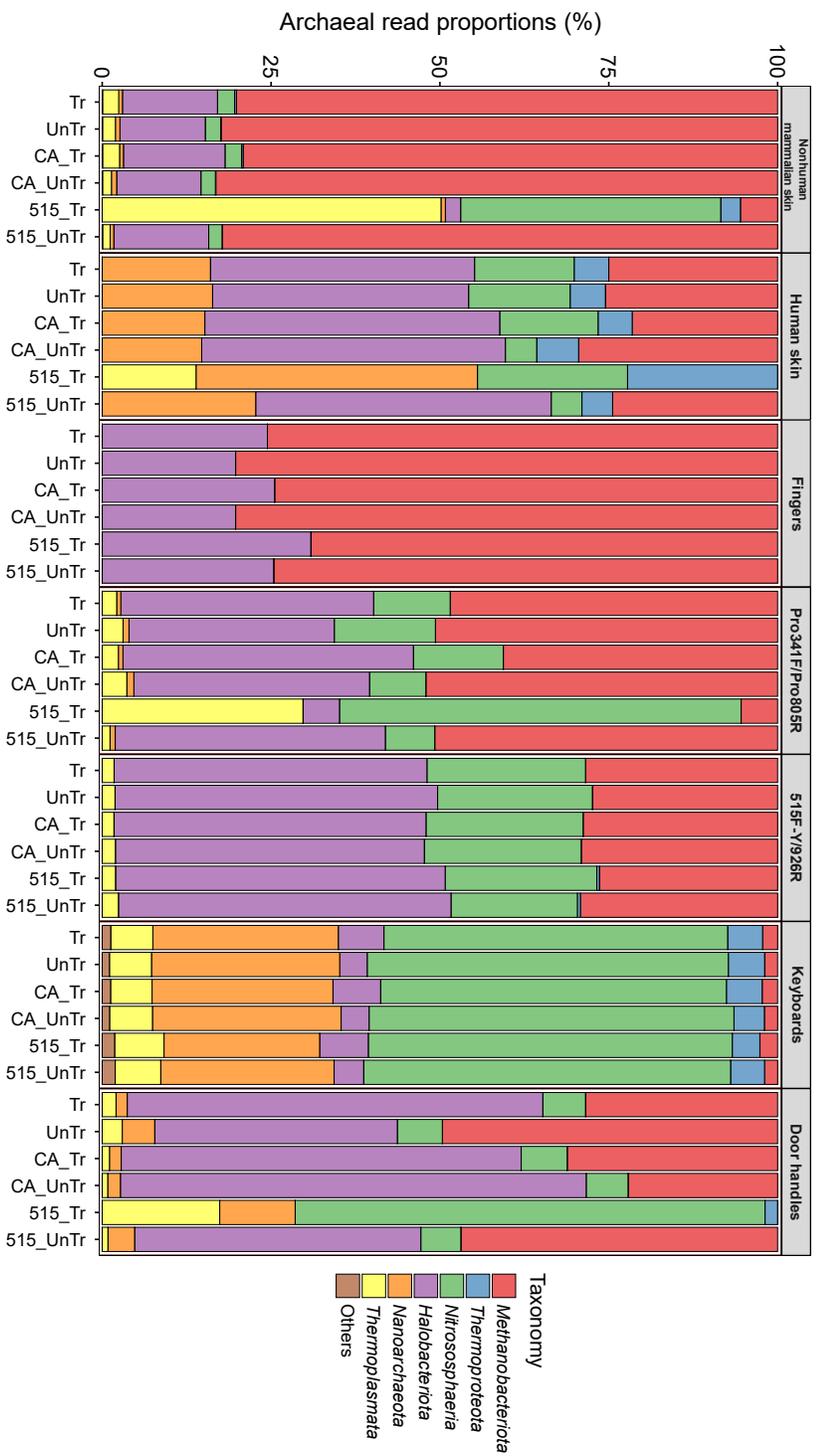


Figure 3-1 The effect of truncation and primer removal methods on archaeal taxonomic proportions following processing through DADA2. The “Tr” and “UnTr” samples refer to the truncated and untruncated datasets respectively, with primer sequences removed from the reads using basic trimming. The “CA” samples refers to cut-adapt [176] which was used for primer sequence removal. Samples with the “515” prefix indicate those in which the 16S rRNA gene reads were truncated to the common 515 to 805 region of the 16S rRNA gene used in the study.

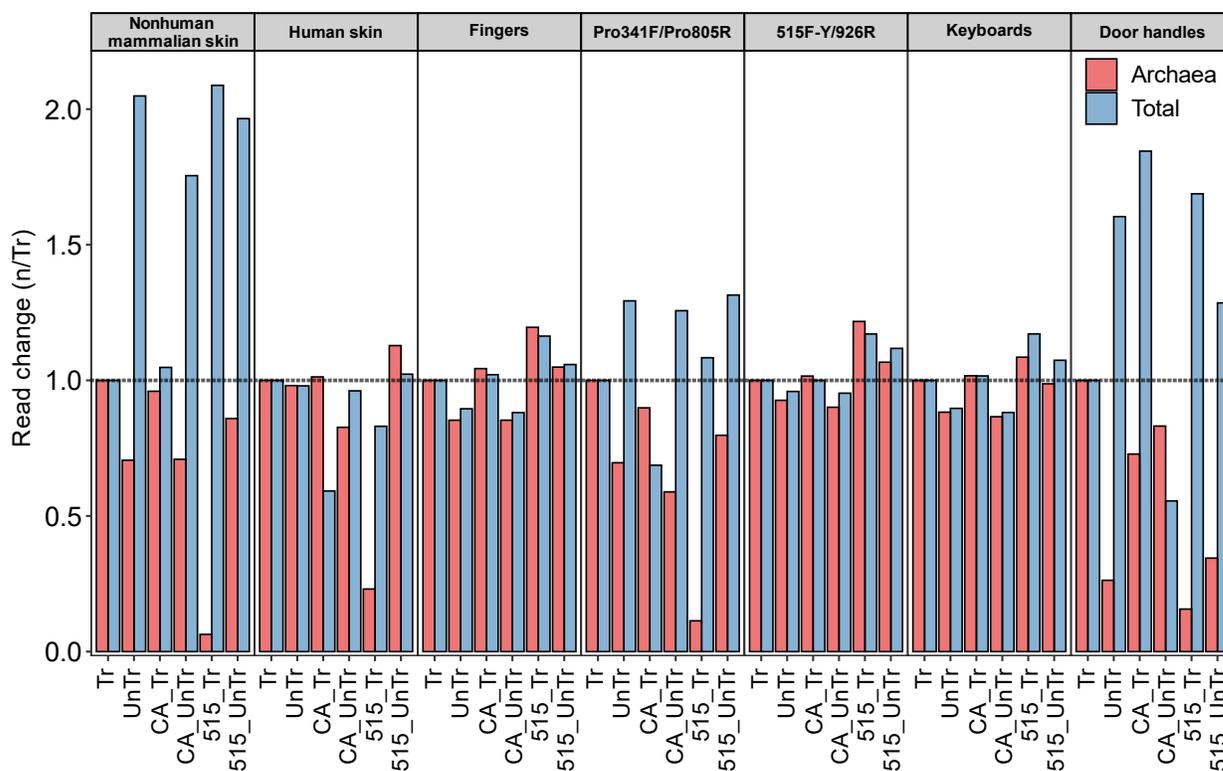
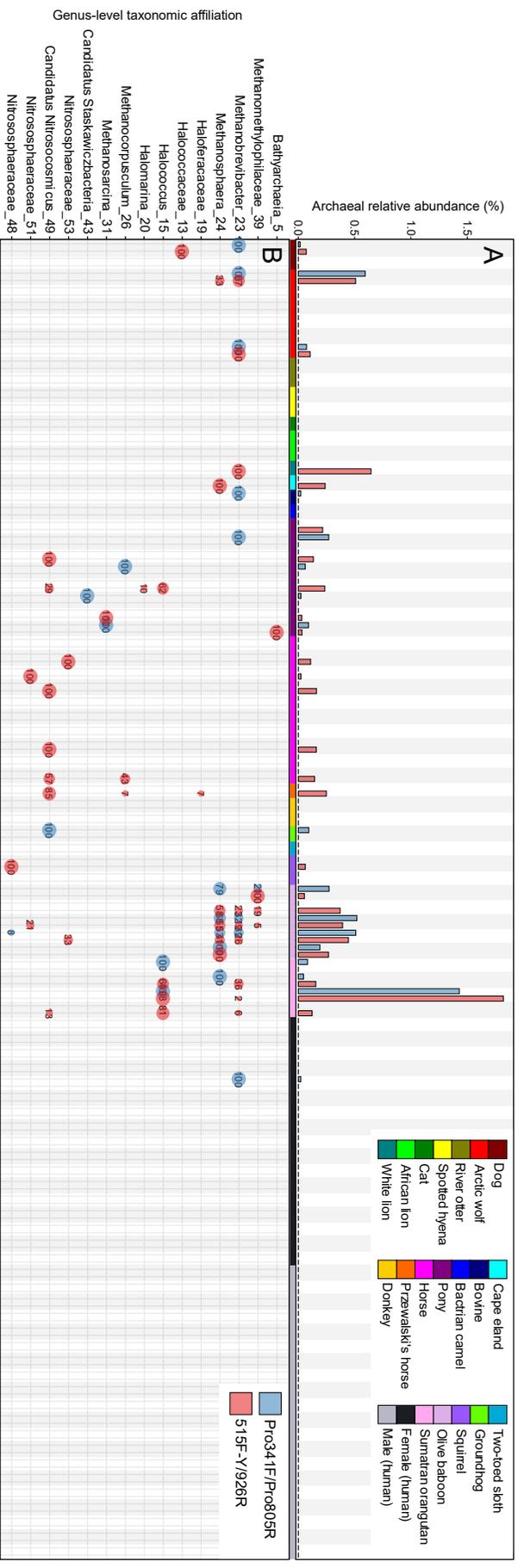


Figure 3-2 The effect of truncation and primer removal methods on archaeal and bacterial read numbers following processing through DADA2. The “Tr” and “UnTr” samples refer to quality-trimmed and untrimmed datasets respectively, with primer sequences removed from the reads by manually trimming the 5’ end. The “CA” samples refers to cut-adapt [176] which was used for primer sequence removal. Samples with the “515” prefix indicate those in which the 16S rRNA gene reads were excised to the common 515 to 805 region used in the study. All changes within the sample datasets were compared to their original quality-truncated dataset (Tr), indicated by the dashed line at 1.0.

Table 3-3 Summary of archaeal reads and ASVs of the skin and skin-associated environment. The relative abundance of archaea was calculated using all reads for each dataset. Minimum (Min) and maximum (Max) archaeal relative abundances were calculated for each dataset excluding zero-archaea samples. The data presented for Pro341F/Pro805R is a subset of data from the mammalian and human skin datasets.

Dataset	Total samples	Archaea-detected samples	Archaeal ASVs	Total ASVs	Archaeal reads	Bacterial reads	Archaeal		
							relative abundance (%)	Min (%)	Max (%)
Non-human mammalian									
skin	546	194	121	24306	4179	5,238,782	0.079	0.013	1.510
Human skin	340	20	25	9613	176	7,631,068	0.002	0.005	0.264
Pro341F/Pro805R	92	18	20	5399	260	1,000,284	0.026	0.016	1.429
515F-Y/926R	92	26	31	5827	668	573,461	0.116	0.012	1.820
Fingers	80	2	2	1604	193	1,722,646	0.011	0.397	0.418
Keyboard	240	15	33	6623	935	5,972,034	0.015	0.015	1.116
Door handles	392	39	28	13546	354	4,859,381	0.007	0.009	0.522



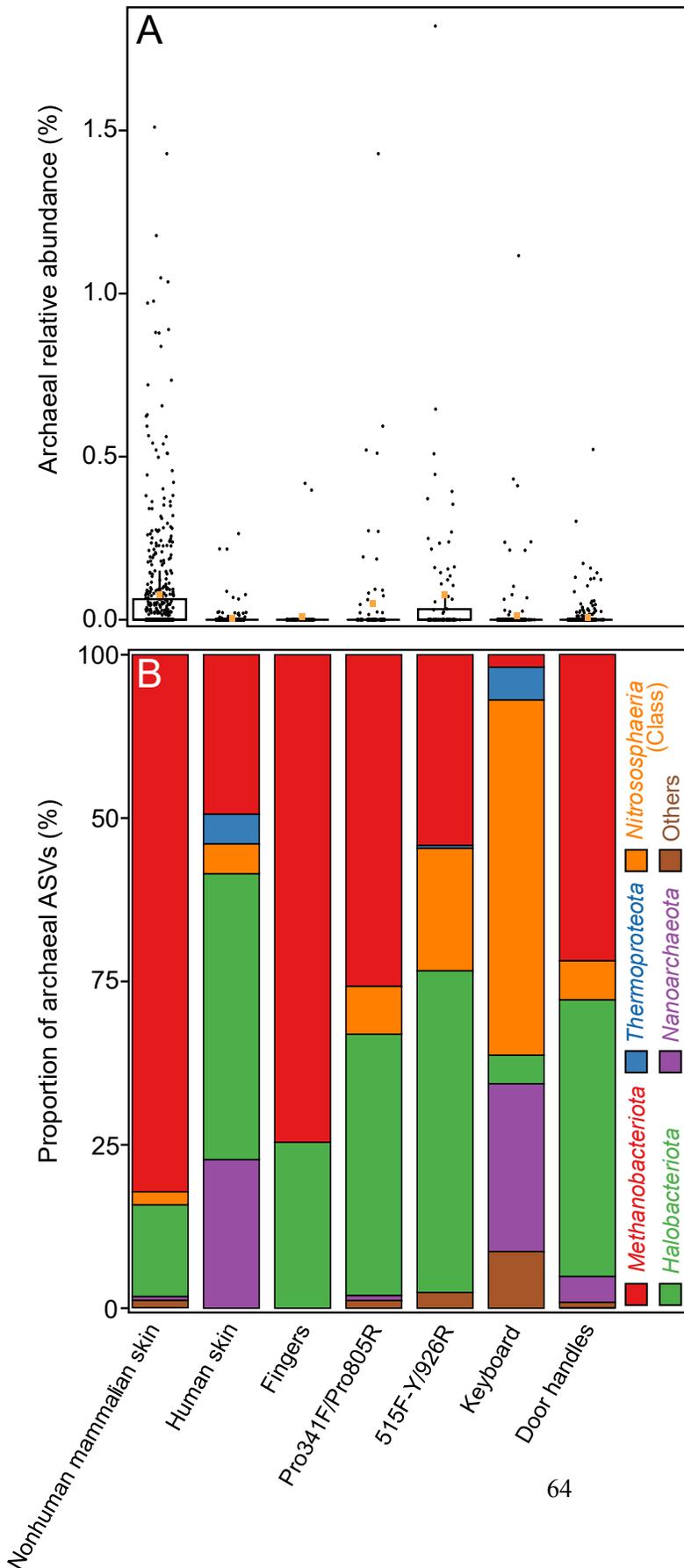


Figure 3-4 Archaeal 16S rRNA gene relative abundances (A). The relative abundance of archaeal sequences was calculated by dividing the number of sequences affiliated with archaeal ASVs by the total number of sequences for each sample. Relative abundance averages for all samples in each dataset is indicated by orange squares within the boxplot. The taxonomic proportion of the archaeome of the skin and skin-associated surfaces is separated by phyla or class (B). Archaeal taxonomic proportions include archaeal 16S rRNA gene reads only and represent the proportion of archaeal reads belonging to each phylum or class. Class Nitrososphaeria was separated from the phylum Thermoproteota to highlight putative AOA-associated archaea specifically. The Thermoproteota category thus does not contain any Nitrososphaeria-associated reads.

Analysis of all archaeal genera present within each dataset (including the Roche 454) showed some overlap, with 31 of 54 genera observed in at least two datasets (Figure 3-5). The remaining 23 genera were unique to a single dataset. These unique genera span the archaeal DPANN and TACK superphyla and include ASVs present at various relative abundances. Within the *Methanobacteriota*, *Methanobrevibacter* was the most common genus, found in nearly all samples, and was also present in the previously published Roche 454 human skin amplicon dataset [40]. Another genus affiliated with the *Methanobacteriota* is *Methanosphaera*, which was found associated with several samples, either alone or in combination with *Methanobrevibacter* sequences. Various *Halobacteriota* were present, with *Halococcus/Halococcaceae* and *Methanocorpusculum* found within several mammalian skin samples, including the Roche 454 dataset. Although the *Halobacteriota* (excluding *Methanocorpusculum*) were not prevalent across samples within a dataset, they accounted for a considerable proportion of all archaeal reads (Figure 3-4B). The AOA-associated *Nitrososphaeria* were present but were rare and detected in only 79 of 1,710 samples (4.6%) included in this analysis (Roche 454 included). Of those 79, 49 (62.0%) belonged to the Roche 454 dataset. The non-human mammalian skin and Roche 454 datasets shared the same *Nitrososphaeria*-classified genera, except for one ASV that could not be resolved to the genus level (i.e., “Nitrosotaleaceae_54”), which was associated exclusively with the Roche 454 dataset (Figure 3-5). On non-human mammalian skin, AOA-associated genera were found on the skin of several mammals (i.e., olive baboon, squirrel, donkey, dog, groundhog, cheetah, and asian elephant). The genus *Nitrosocosmicus* was absent from human skin datasets, although ASVs associated with this genus were present in all other datasets (i.e., non-human mammalian skin, keyboards, fingers, Roche 454, door handles) and was one of the more common archaeal genera detected.

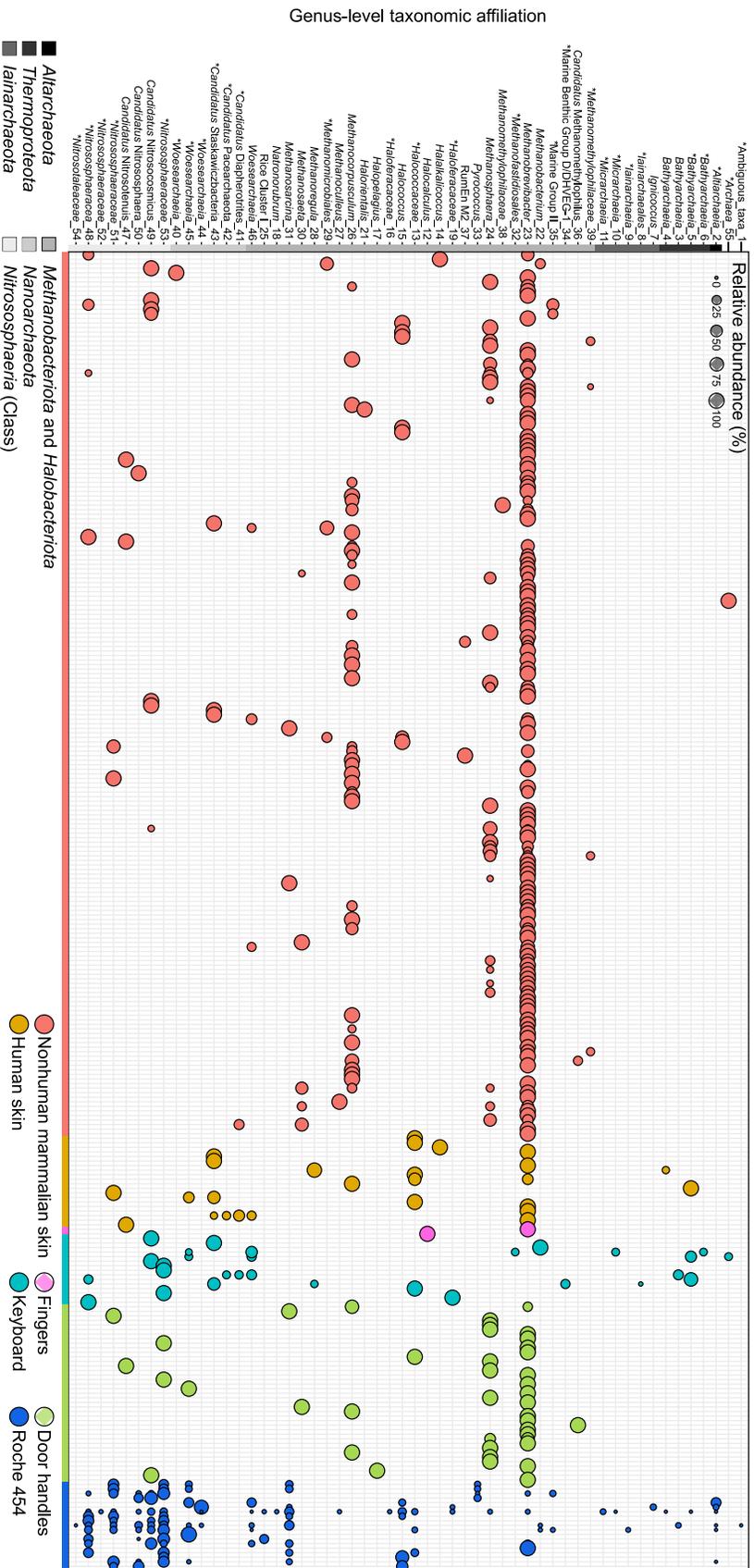


Figure 3-5 Distribution of all archaeal genera across skin and skin-associated surfaces. The ASV table was collapsed to the genus level and then filtered for archaeal taxa and contains any sample with a non-zero number of archaeal 16S rRNA gene reads. The size of the bubbles represents the relative abundance of the genera with respect to the total number of archaeal 16S rRNA gene reads within a sample. Archaeal ASVs not resolved to a genus level were collapsed to their most resolved taxonomic level and are indicated with an “*” and are referred to as “genera” within the results and discussion. A human skin dataset [32] was included for comparison.

An analysis of 193 archaeal ASVs associated with new and previously published datasets analyzed in this study (Roche 454 excluded) revealed that detected archaeal ASVs were differentially distributed between datasets, and that the overlap between datasets was limited (Figure 3-6). The non-human mammalian dataset contained 57 of the 69 (82.6%) *Methanobacteriota* ASVs detected; 49 (71.0%) of those were unique to the dataset. The remaining eight ASVs were observed in two or more datasets, including three *Methanobacteriota* ASVs that were present in three datasets (44329_*Methanobacteriaceae*, 12677_*Methanobacteriaceae*, 8955_*Methanobacteriaceae*). The *Halobacteriota* were similarly dominant, with 35 of 52 (67.3%) total ASVs were unique to non-human mammalian skin. Only a single *Halobacteriota* ASV, “43418_*Methanosaetaceae*”, was shared between more than one dataset and belonged to both non-human mammalian skin and human skin samples. The 21 *Nitrososphaeria* ASVs originated solely from non-human mammalian skin (n=9; 42.9%), keyboard (n=4; 19.0%), door handles (n=4; 19.0%), and human skin (n=1; 4.8%) datasets, with *Nitrososphaeria* absent from the separate finger dataset. The remaining three (14.2%) *Nitrososphaeria* ASVs were shared between datasets, with two of the ASVs were shared between two datasets, and a single ASV, “26019_*Nitrososphaeraceae*” shared among three datasets: non-human mammalian skin, keyboard, and campus door handles. The ASVs associated with *Thermoproteota*, *Iainarchaeota*, and *Nanoarchaeota* represented the remaining 51 (26.4%) ASVs and were observed more broadly within multiple datasets but had no overlap among datasets.

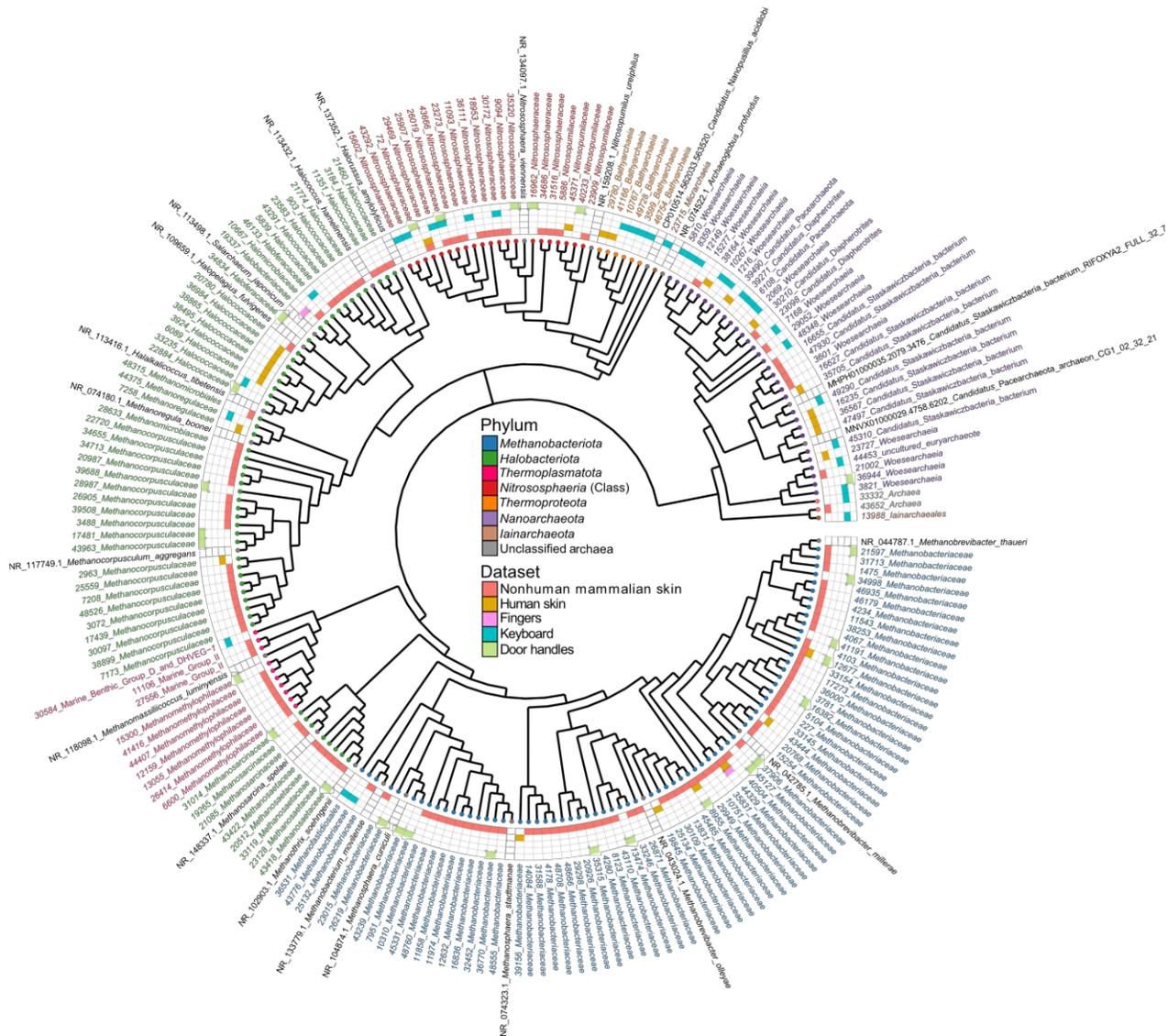


Figure 3-6 The prevalence and overlap of archaeal ASVs on skin and skin-associated surfaces. The tree contains all archaeal ASVs from each dataset. ASVs in black are 16S rRNA gene reference sequences retrieved from the NCBI and SILVA databases, whereas remaining ASVs are coloured by their respective class or phylum. Because not all ASVs were resolved to a species or genus level, all ASVs were renamed to a family level for consistency. ASV overlap between datasets is indicated through the heatmap squares. The maximum likelihood tree was constructed using a GTR +G +I model with 1000 bootstrap replicates.

3.4 Discussion

3.4.1 Archaea on mammalian skin are rare and uncommon

This amplicon-based assessment of skin and skin-associated microbiota revealed an infrequent and low-abundance distribution of archaea among all 1,688 samples from skin and skin-associated amplicon datasets. The limited detection of archaea on mammalian skin suggests that archaeal communities are relatively rare, of low relative abundance, and below common detection limits when using PCR with universal prokaryotic primers and high-throughput sequencing. The disparity between human skin and non-human mammalian skin is expected considering the large differences in their local environment and hygiene practices. Most non-human mammalian skin samples contained no detectable archaeal sequences and no sample exceeded 2%. Given differences in skin physiology, living conditions, and geographic origins of the host, it was not unexpected to observe large variations in archaeal distributions among mammalian skin samples, particularly within samples where archaea were detected. For example, although human and non-human mammalian skin both contained sequences associated with *Methanobrevibacter*, non-human mammalian skin had a much greater comparable abundance and sample prevalence. Although *Methanobrevibacter* taxa are typically host-associated within the gut [178, 179], they are suggested to be distributed in soil [180] and water [181] and previous results suggest that a transient environmental skin layer is common for mammalian hosts [27]. It is perhaps more likely that gut-associated *Methanobrevibacter* are instead being deposited into the environment, via animal feces, which are then taken up onto the skin. The AOA are abundant members of soil (and other environments) where they contribute to the biogeochemical cycling of nitrogen [182] and were also found sporadically and at low relative abundance among mammalian skin samples. *Nitrososphaeria*-associated reads comprised only a small portion of the total archaeal reads in non-human mammalian skin (1.9%) and human skin (4.5%) and are likely derived from environmental sources [182, 183]. Terrestrial and semi-aquatic non-human mammals in close contact with their respective environments and debris (e.g., feces, urine) without the routine hygiene practices of humans, are perhaps more likely to have skin communities containing allochthonous archaea from environmental reservoirs.

The 12 shared archaea-associated ASVs in all datasets were associated with the *Methanobacteriota*, *Nitrososphaeria*, and *Halobacteriota*. These taxa have been observed in environmental [180, 181] or commercial (e.g., food, cosmetics) [184] samples which might explain their ubiquity across datasets.

The skin-associated built environments were equally absent of detected archaea, with the majority of keyboard (93.8%) and campus door handle (90.0%) samples containing no detected archaea. The keyboard and door handle built environments have different chemical and physical environments from skin, but via frequent human contact can develop skin-like microenvironments and microbial communities [24, 171]. The few archaeal sequences present on campus door handles were associated with archaea in a similar pattern as observed in human and non-human mammalian skin: primarily *Methanobacteriota* and *Halobacteriota*, and likely derived from environmental reservoirs as suggested elsewhere [118]. Additionally, food and cosmetic reservoirs for *Halobacteriota*-associated genera have been identified [184] and might serve as an explanation for the observed archaeal sequences. However, although only representing less than 0.02% of the total reads, the archaeal community proportions observed on keyboards were markedly different from any other analyzed dataset within this study. Instead of a dominance of *Methanobacteriota*, as seen in other datasets analyzed here, the AOA-associated class *Nitrososphaeria* collectively contributed more than 50% of archaeal reads. Most of these reads (>90%) can be attributed to two individuals: a female between the age of 20 and 29, and a female of greater than 60 years of age. The detected *Nitrososphaeria* may not be entirely representative of the typical keyboard microbiota given that *Nitrososphaeria* associated reads were absent from most other keyboard sample data. Instead, these sequences might be attributed to individual differences in the keyboard users themselves or their activity habits (e.g., eating multiple meals at the keyboard, local environment).

Metagenomic studies focusing on skin-associated archaea are rare, making assessment of the skin archaeome difficult. Only a small number of studies address skin archaea directly with archaea-targeted approaches [39–41, 43]. All other skin-associated archaea data is produced from holistic skin microbiome studies utilizing universal

prokaryotic primers and 16S rRNA gene amplicon sequencing [22, 27, 169] or metagenome-assembled genomes [19, 44, 170], where skin-associated archaea attention is minimized due to their low abundance. The data presented on the rarity of skin-associated archaea in this current work, along with the comparisons made to existing skin archaea literature, represents the near-entirety of skin archaeome research. However, skin microbiome profiling can be impacted by several factors which could affect the profiling of the skin archaeome. For example, our current study samples only the outermost portion of the epidermis (i.e., stratum corneum) although it is suggested that the lower epidermis and subepidermal areas of the skin might harbour microbial communities [4]. As well, our study encompasses a human population specific to a region in North America, and regionality and race has been shown to impact skin communities [185]. Furthermore, individual hygiene habits, such as deodorant use [186] and showering [147] alter microbial community composition. Within this current study, there was the potential that shortening the 16S rRNA gene amplicons to the common 515 to 805 region could impact the classification of archaeal sequences and taxonomic proportions. Taxonomic proportions were shown not to be impacted by quality trimming and truncation method (Figure 3-1), although some datasets (i.e., nonhuman mammalian skin and door handles) experienced changes in proportions. In the case of the nonhuman mammalian skin and door handles datasets, proportional changes in archaeal relative abundance are a result of up to a two-fold increase in the number of bacteria-associated reads output from DADA2 after removing manual quality-based read truncation steps (i.e., “Tr” vs “UnTr”). Quality trimming can have a large impact on read assignment and read merging within high-throughput amplicon sequencing datasets, particularly with lower quality sequencing data [187, 188]. Both the nonhuman mammalian skin and door handle datasets contained lower-quality sequence data that might have been impacted by excess subjective quality-based trimming. In absence of manual quality trimming, DADA2 still applies an automatic quality filtering step that removes spurious and chimeric sequences [123]. Therefore, the increased number of bacteria-associated reads in the untruncated “UnTr” datasets is likely an accurate representation of the expected read numbers, and by extension, archaeal proportions. Despite these areas of concern, our results provide a meaningful contribution to the knowledge of the

skin archaeome but highlights an area of research that is lacking and that would benefit from additional contributions and comparisons.

3.4.2 Implications for previous studies of human skin archaea

Although the same genera were shared between this study and those from a Roche 454 dataset (e.g., Figure 3-5), our data suggest much lower archaeal relative abundances for human skin samples than what has been previously estimated using qPCR archaeal and bacterial 16S rRNA proportions [40, 43]. Primary causes for these differences may be primer- and method-specific. Successful detection of archaeal sequences is dependent on using primers that include adequate coverage of the domain *Archaea*. This current study leveraged two universal prokaryotic primer pairs that, at least *in silico*, have extensive coverage of the archaeal and bacterial domains (Table 3-2). The same DNA extraction method and 515F-Y/926R prokaryotic primers were recently used to detect archaeon-associated ASVs of the AOA *Nitrosocosmicus* genus in wastewater treatment plants [189], with very high agreement between archaeal abundances (~1%) observed among amplicon-based, qPCR, metagenomic, and fluorescent in situ hybridization (FISH) methods [189, 190]. Additionally, an evaluation of several common universal prokaryotic primers concluded that 515F-Y/926R was the optimal primer pair for detecting archaea in marine ecosystems [191]. This work also demonstrates with direct comparisons that archaeal detection between both primer pairs is similar. Universal prokaryotic primers allow archaea to be co-amplified alongside bacteria, providing a sample-by-sample characterization of archaeal relative abundances. Although this provides community context, it has been suggested that co-amplification of bacteria and archaea could introduce bias against archaea because of low template exclusion [192, 193]. Alternatively, studies by Probst *et al.* [43] and Moissl-Eichinger *et al.* [40] estimated absolute abundances based on qPCR using archaea-specific 344af/517ur primers. Taxonomic information was generated using the 344af/915ar primers separately for 16S rRNA gene sequencing [40]. Although community context is lost, because of the exclusion of bacterial 16S rRNA genes during amplification, the archaea-specific primers could amplify rare archaeal 16S rRNA gene sequences that otherwise might be outcompeted by bacterial 16S rRNA template.

Extraction bias and uncertainty associated with qPCR amplicon specificity and the capacity for the 344af/517ur primers to unintentionally amplify bacterial template is a concern. Using biased extraction methods that do not account for the physiological differences between bacteria and archaea cell structures (e.g., S-layer, methanochondroitin) [194] has been suggested as a reason for low archaeal abundances observed on skin [40]. The PowerSoil kits used in this current study do not impact archaeal diversity [195] and the additional pre-treatment steps included in this work should minimize bias. However, post-extraction non-specific amplification of bacterial sequences could result in an over-estimation of archaeal abundances during qPCR. Justification for this concern is found in the 344af/915ar sequencing data of Moissl-Eichinger *et al.* [40]. Although this primer pair was used only to generate archaeal taxonomic data, several samples contained bacterial sequence contamination ranging from 10% to 50% of all reads. The primers used for qPCR share the same archaea-specific forward primer that was used in sequencing (344af), but also a universal prokaryotic 517ur reverse primer. Although the primer pair shows very low bacterial coverage *in silico* [41], the less specific 344af/517ur pair might have amplified bacterial 16S rRNA gene sequences. If so, the co-amplification of bacterial sequences during archaeal qPCR quantification could be undetected, and both the bacterial and archaeal products would be unintentionally combined and used to quantify archaea. The resulting abundances would be overestimated and could explain the high reported archaeal abundances. If not attributed to regional differences (e.g., cultural or regional differences in skin communities, surrounding environment, lifestyles), relatively high skin archaeal abundances previously reported could be attributed to methodological bias.

Although PCR and primer biases that confound archaeal amplification may exist, this work maintains that if archaea are routinely abundant on mammalian skin, the large sample size presented in our study would have revealed this to be the case. In the least, it would be expected that the number of samples with detected archaeal sequences would be higher than currently presented, regardless of abundance. Notwithstanding, the differences in quantitation method do not explain the taxonomic profile differences observed in literature. Whereas this work shows considerable *Methanobacteriota* and *Halobacteriota* dominance, some prior

literature suggests that *Nitrososphaeria* were the most abundant taxa [40, 43]. If *Nitrososphaeria* were a common skin community member, it would be expected that the increased archaeal coverage and large sample size of this current study would reveal more *Nitrososphaeria* prevalence or relative abundance among samples. Indeed, this is the case for the keyboard microbiome, where more than 50% of the archaeal sequences belong to *Nitrososphaeria*.

Other than potential methodological bias, the differences between our amplicon data and previous qPCR or Roche 454 data might be explained by geographical origins of human subjects (such as personal hygiene, activity type, or geographic region; Canada versus Northern Europe) or some other yet unidentified factor. For example, the gut microbiome of individuals in areas of Europe contain more *Methanobacteriota* sequences compared to individuals in North America [196]. Although there is extensive literature that attempts to define a core skin pan-microbiome [21, 160, 197], archaea are often absent from these studies and reviews. If archaea are established members of the skin microbiome, then the potential regional (e.g., forest vs grassland) [198] and population differences that drive the observed variation in diversity and abundance could have a considerable and currently poorly quantified impact on archaeal abundance on human skin.

Previous research of human skin archaea also used fluorescence *in situ* hybridization (FISH) [43] and Fourier transform infrared focal plane array hyperspectral imaging (FTIR-FPA) [40]. Both methods suggested the presence of archaea on human skin, although they do not offer quantitative information and are not linked to any other sample data [40]. Although FISH was intended to verify the presence of *archaea* on human skin [43], the ARC915 probe has extensive coverage for several archaeal phyla including *Methanobacteriota* and *Halobacteriota*. Thus, cells observed with this method should be interpreted cautiously and should be confirmed in the future with multiple probes and samples to help verify the identity of FISH-positive cells.

3.4.3 Skin as a potential habitat for archaea

Although the presented data suggests that archaeal occurrences on human skin are infrequent, and that occurrences on mammalian skin are comparatively more common though

still rare, the presence of archaeon-associated sequences in some samples may nonetheless be relevant to host habitat and health. Archaeon-associated genera and ASVs shared amongst datasets could provide insight into common archaeal detection, colonization, or contamination of the skin and skin-associated surfaces. For example, many of the prominent archaeal genera observed in our datasets were observed in the Roche 454 human skin dataset produced in a separate lab, with different methods, and in a distant geographic location [40]. The detection of these shared archaea, despite differences in detection techniques, might suggest the existence of a more “core” archaeome comprised of ubiquitous soil-associated archaea, albeit at very low abundance. The one *Nitrososphaeria* and three *Methanobacteriota* associated ASVs which overlapped concomitantly with human and non-human mammalian skin could represent archaeal skin ecotypes that might be more adapted to the physical and chemical environment of mammalian skin.

With such low archaeal abundances detected in our study, it is difficult to make conclusions about whether the skin archaea represent autochthonous populations or allochthonous environmental contaminants. However, there are several features of mammalian skin that might allow archaeal populations to establish. For example, mutualistic relationships between acetogenic bacteria and methanogenic *Methanobacteriota* archaea have been documented [199] and it could be that the same community interactions are occurring on the skin, although it would require localized zones of anoxia. These anoxic zones could exist on the skin as a natural result of sebum and apocrine secretion onto the skin surface [151, 152]. Although sequences associated with acetogenic bacteria and methanogenic archaea were found within the same samples [27], it is uncertain whether such syntrophic interactions occur.

Several halophilic genera were detected on mammalian skin in addition to other skin-associated environments. Although mammalian skin might be a suitable environment for *Halobacteriota* archaea, due to its varying salinity compared to other surface environments, these microorganisms are more likely to be contamination from food and skin care products due to their frequent consumption and use by humans [184]. However, halophilic archaea

have been previously reported in the human oral and intestinal tracts [72, 200], which could provide a route through which they might colonize skin or renew their populations.

Putative AOA associated genera were detected within a relatively small number of mammalian skin samples. Non-human mammalian skin could be exposed to ammonia-containing substrates from their environment (e.g., animal waste) and provide an energy source to AOA established on skin surfaces. Mammalian skin might also passively diffuse ammonia [201, 202] or excrete it in sweat [203], and there is the potential for localized skin diseases to increase ammonia production [204]. Furthermore, common shampoos and hand soaps often contain ammonia derivatives that could facilitate AOA growth. Notwithstanding, *Nitrososphaeria* associated reads made up only a small portion of the total archaeal reads (non-human mammalian skin 1.9%; human skin 4.5%). If non-human mammalian skin can maintain a suitable environment for AOA metabolism, these archaea appear to represent rare and relatively low abundance populations.

3.5 Conclusion

The exploration and microbial profiling of the skin and of skin-associated environments is an integral component to understanding the interactions of the skin microbiome and host health. This work provides evidence that archaea are rare and infrequent members of human and non-human mammalian skin, and skin-associated surfaces. For samples with detectable archaea, individual host and environmental variation might explain archaeal distributions. Shared ASVs and genera among datasets provide insight into this rare biosphere skin archaeome, regardless of whether they are autochthonous or allochthonous community members. Overall, this work challenges recent literature suggesting an unexpected abundance of archaea on human skin specifically and suggest that future research using shared samples and validated and standardized methods could help confirm our downward estimates of the numerical relevance of the skin archaeome. For example, future skin archaeome characterization using universal prokaryotic primers could benefit from an internal sequencing workflow standard made of both archaeal and bacterial 16S rRNA gene sequence templates [205]. Overall, our amplicon-based data agree with

comprehensive skin microbiome metagenomic assessment, demonstrating that “archaea were nearly absent on skin” [19].

Chapter 4

Conclusion and recommendations

4.1 Summary

The mammalian skin microbiome is complex and an integral component of mammalian skin health and function. Although the microbial communities of human skin have been thoroughly characterized, non-human mammalian skin research has received less attention. Identifying the microbiota in association with mammalian skin is a step towards understanding mammalian skin itself and the host-microorganism relationships that helped define it [1]. Importantly, by establishing baseline knowledge of what constitutes a “normal” mammalian skin microbiome, researchers interested in skin disease might better understand how microbial communities impact, or are impacted by, various skin pathologies. Additionally, co-evolutionary histories shared among mammalian hosts and their skin microbiota cannot be fully explored without first establishing which microbiota are present.

Chapter 2 of this thesis demonstrates how the alternative phylogenetic marker, *cpn60*, can be used to generate high-resolution microbial profiles of mammalian skin that produce confident species-level taxonomic classifications. Through these high-resolution profiles, hypotheses regarding phylosymbiosis and co-evolutionary relationships between the mammalian host and skin microbiota can be better tested. This work highlights the importance of species-level resolution in amplicon-based studies and the emergence of phylosymbiosis patterns and host-microorganism associations that remain undetected with traditional 16S rRNA gene sequencing. In doing so, this chapter contributes to an existing framework for using the *cpn60* gene to profile microbial communities [110, 117], validates its use for profiling the mammalian skin microbiome, and presents observations of host-microorganism associations that merit additional *cpn60*-based research.

Chapter 3 of this thesis bridges a considerable knowledge gap of the archaeal communities that reside on mammalian skin. In surveying the mammalian skin archaeome, I expand on existing skin archaea literature [40, 43] and provide an overview of the archaeal communities present on human and nonhuman mammalian skin, as well as skin-associated

built environments of the keyboard and door handles. My research shows that archaea are infrequently detected on both skin and skin-associated surfaces when using universal prokaryotic primers, exist at relatively low abundance, and are dominated by *Methanobacteriota* and *Halobacteriota* when detected. Moreover, this research provides a summary of archaeal sequences detected in association with mammalian hosts and will serve as starting point for exploring the potential co-evolutionary relationships that archaea might share with skin-associated bacteria or the host skin environment itself.

4.2 Validation of *cpn60*, observations of phylosymbiosis, and high-resolution profiling of the mammalian skin microbiota

Although the 16S rRNA gene has extensive support for resolving taxonomy in amplicon-based studies [92, 93, 174], it has limited taxonomic resolution when used with short-read sequences that are typically produced using high-throughput sequencing [95]. Thus, use of the 16S rRNA gene for profiling the mammalian skin microbiome provides a thorough survey of the skin microbiota to the genus level, but prevents confident species-levels analysis. Although phylosymbiosis of mammalian skin has been observed previously using the 16S rRNA gene [27], additional co-evolutionary patterns could be present but masked by the limited resolving power of the marker gene. Previous co-evolutionary studies exploring patterns of phylosymbiosis [56] have benefitted from non-16S rRNA gene, species-specific markers, because their increased taxonomic resolution and species-level analyses have enabled the observation of more subtle co-evolutionary relationships. However, use of bacterium-specific markers requires an *a priori* knowledge of the microbial community that might not be available and excludes other microbial populations of interest. Alternatively, the *cpn60* gene marker is both universal in coverage and capable of species-level resolution, though has not yet been used for profiling mammalian skin. My work presented in Chapter 2 validates *cpn60* for use in profiling microbial communities of mammalian skin and provides a high-resolution analysis of the microbiota within the skin microbiome. In doing so, this work provides a framework for future *cpn60* use on mammalian skin and has highlighted host-microorganism associations of specific *Staphylococcaceae* species that could have a co-evolutionary influence.

My work using the *cpn60* gene provided a high-detail microbial profile of mammalian skin that included skin swab samples from Artiodactyla, Perissodactyla, Carnivora, and Primate hosts. This work also produced direct paired-sample validation of *cpn60* and 16S rRNA gene microbial profiles using both taxonomic and community composition comparisons. Direct taxonomic comparisons of *cpn60* and 16S rRNA gene microbial profiles have been made in the past using small well-annotated databases and have shown strong similarity [120, 206]. The taxonomy-based comparisons of the *cpn60* and 16S rRNA gene datasets in this thesis showed incongruities (Figure 2-5), although this work used a considerably larger database that could not be as meticulously annotated to match the taxonomies of the 16S rRNA gene database. My work reinforces the need for a *cpn60* gene taxonomy database that is compatible with the current 16S rRNA gene taxonomy classifiers if extensive taxonomic comparisons between the *cpn60* and 16S rRNA gene markers is desired. However, the *cpn60* and 16S rRNA gene profiles were compositionally similar, producing significant correlations using Procrustes analysis for Bray-Curtis and weighted and unweighted UniFrac ordinations (Figure 2-7). Therefore, direct taxonomic compatibility between the *cpn60* and 16S rRNA gene databases are not required for analysis of phyllosymbiosis or additional co-evolutionary relationships but would nonetheless be useful for future *cpn60*-based research.

The *cpn60* gene microbial profiles were examined for evidence of phyllosymbiosis identified previously in Artiodactyla and Perissodactyla using the 16S rRNA gene [27]. This work was able to confirm evidence of phyllosymbiosis when using the UniFrac distance metric (Figure 2-8, Figure 2-9) for Artiodactyla and Perissodactyla, suggesting that the *cpn60* gene microbial profiles contained phylogenetic differences in microbial community composition that went previously unseen when using the 16S rRNA gene. The observation of phyllosymbiosis could suggest co-evolutionary relationships between mammalian hosts and their microbiota. To further explore these relationships, this work leveraged the taxonomic resolution provided by the *cpn60* gene microbial profiles to survey *Staphylococcaceae* communities common among mammalian hosts (Figure 2-12). Several *Staphylococcaceae* species were found in association with mammalian hosts (i.e., *Macrococcus carouselicus* and

Przewalski's horse; *Salinicoccus* sp. with horses) and *Jeotgalicoccus halophilus* was observed in proportionally high relative abundance and prevalence among mammalian hosts. The species *J. halophilus* has limited association with the mammalian skin microbiome in literature and thus highlights a potentially important host-microorganism interaction, and the detection of additional resolved skin-associated species suggests that the future use of the *cpn60* gene could reveal subtle differences in skin microbial communities between mammalian hosts. Although this work attempted to further assess these *Staphylococcaceae* populations for evidence of phylosymbiosis and co-evolution, such analyses were not possible due to shallow sequencing depth and limited mammalian host representation.

Future microbiome studies using the *cpn60* gene would benefit from increased sequencing depth to help reduce bias and provide more information for read and proportion-based analyses. Minimal time between sample collection, DNA extraction, and sequencing, along with skin sampling practices that maximizes the collected biomass, would help ensure a sufficient sequencing depth is achieved. Additionally, taxonomy for the cpnDB reference sequences must be maintained separately or standardized with other commonly used 16S rRNA gene databases [93, 94, 148] for direct comparisons. Although *Staphylococcaceae* was selected because it represents a core skin community member and has importance to skin disease, alternative microbiota might be equally important to survey using the *cpn60* gene. For example, several samples within this study had proportionally very few *Staphylococcaceae*-associated reads and thus could not be used for analysis. Future work might benefit from identifying proportionally more abundant core genera observed previously [27], or focus on specific populations on a sample-by-sample basis for further species-level analysis, all of which can be accomplished with a single dataset generated with the *cpn60* gene. As such, this work demonstrates the versatility of the *cpn60* gene for both broad and species-level profiling of microbial communities and as a complement to traditional 16S rRNA gene microbial profiling.

4.3 Skin-associated archaea are infrequent and in low relative abundance on mammalian skin

The 16S rRNA gene is widely used in amplicon-based studies [76] for profiling bacterial and archaeal communities. Although bacterial skin-associated communities have been thoroughly profiled for human skin [149, 167, 197], archaeal communities of human and non-human mammalian skin have been comparatively neglected. Only a small number of studies focused on the human skin archaeome exist [40, 43] and non-human mammalian skin associated archaea are often excluded from discussion or included in microbial profiling studies as an aside [27, 170]. As such, there exists a substantial gap in knowledge of the mammalian skin archaeome and the potential host-microorganism interactions occurring therein.

Chapter 3 of my thesis work contributes a necessary review of mammalian skin-associated archaea research and reports the relative abundances, distribution, and taxonomic proportions of archaea across mammalian skin and skin-associated surfaces. The limited detection of archaea on skin and skin-associated surfaces using the 16S rRNA gene suggests that archaea are infrequently detected and in relatively low abundance (i.e., <0.1% of the total community, for all environments) (Figure 3-4A). When detected, skin-associated archaea were primarily represented by *Methanobacteriota* and *Halobacteriota* (Figure 3-4B), in contrast to the *Nitrososphaeria* dominance previously observed [40, 43]. These observations suggest either a methodological bias resultant from detection differences in universal prokaryotic and archaea-specific primers, or a geographic influence on archaeal distribution and abundance on mammalian skin. My thesis work also revealed that archaeal sequences associated with *Methanobrevibacter*, *Methanocorpusculum*, and various halophilic archaea were prevalent across multiple datasets (Figure 3-5, Figure 3-6). The prevalence of these archaeon-associated sequences suggests either frequent uptake from the environment (e.g., gut-associated archaea deposited via feces and transferred to the skin) or a currently undefined host-archaeon association on mammalian skin. Additional work is necessary to identify whether skin-associated archaea are transient features of the skin microbiome or if the skin can support environments that would allow archaeal metabolism and persistent

archaeal populations. Methanogenic archaea could exist in anoxic areas of the skin and survive using acetoclastic methanogenesis in combination with acetogenic bacteria [199], and sweat [203] and breath [204] could provide *Nitrososphaeria* with the required ammonia. However, the development of co-evolutionary relationships between a host and microorganism requires relatively stable ecological conditions and a moderate degree of microbial prevalence [64], which does not exist for archaea on mammalian skin. Thus, co-evolution might play a small or non-existent role in host-archaeon associations on mammalian skin.

The work presented in Chapter 3 highlights several areas of improvement for future skin-related archaeal research. My work has demonstrated that the primer pairs Pro341F/Pro805 and 515F-Y/926R are similarly capable of detecting archaea on mammalian skin (Figure 3-3). However, future work should further evaluate the potential methodological biases between universal prokaryotic primers and their archaeon-specific complements. Although these comparisons do exist [39, 41], they are limited in the number and variety of skin samples. Comparisons between universal prokaryotic and archaeon-specific primers should be made using the same samples and should include a variety of hosts and sampling sites. As well, the use of internal sequencing standards [205] could help quantify potential bias. Additionally, surveying the archaeome of geographically distant human populations could provide additional insight into the diversity of the human skin archaeome.

Bibliography

1. Maderson PFA. Mammalian skin evolution: a reevaluation. *Exp Dermatol* 2003; **12**: 233–236.
2. Bereiter-Hahn J, Matoltsy AG, Richards KS (eds). *Biology of the integument: 2 vertebrates*. 1986. Springer, Berlin, Heidelberg.
3. Kolarsick PAJ, Kolarsick MA, Goodwin C. Anatomy and physiology of the skin. *J Dermatol Nurses Assoc*; **3**: 203–213.
4. Nakatsuji T, Chiang H-I, Jiang SB, Nagarajan H, Zengler K, Gallo RL. The microbiome extends to subepidermal compartments of normal skin. *Nat Commun* 2013; **4**: 1–8.
5. Folk GE, Semken A. The evolution of sweat glands. *Int J Biometeorol* 1991; **35**: 180–186.
6. Boer M, Duchnik E, Maleszka R, Marchlewicz M. Structural and biophysical characteristics of human skin in maintaining proper epidermal barrier function. *Postepy Dermatol Alergol* 2016; **33**: 1–5.
7. Ramasastry P, Downing DT, Pochi PE, Strauss JS. Chemical composition of human skin surface lipids from birth to puberty. *J Invest Dermatol* 1970; **54**: 139–144.
8. Stewart ME, Downing DT. Chemistry and function of mammalian sebaceous lipids. In: Elias PM (ed). *Advances in Lipid Research*. 1991. Elsevier, pp 263–301.
9. Elias PM, Goerke Jon, Friend DS. Mammalian epidermal barrier layer lipids: composition and influence on structure. *J Invest Dermatol* 1977; **69**: 535–546.

10. Nielsen BL, Rampin O, Meunier N, Bombail V. Behavioral responses to odors from other species: introducing a complementary model of allelochemics involving vertebrates. *Front Hum Neurosci* 2015; **9**: 226.
11. Wyatt TD. Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2010; **196**: 685–700.
12. Bouslimani A, Porto C, Rath CM, Wang M, Guo Y, Gonzalez A, et al. Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci USA* 2015; **112**: E2120–E2129.
13. Bouslimani A, da Silva R, Kosciolk T, Janssen S, Callewaert C, Amir A, et al. The impact of skin care products on skin chemistry and microbiome dynamics. *BMC Biol* 2019; **17**: 47.
14. Clarke SA. An essay on diseases of the skin: containing practical observations on sulphureous fumigations in the cure of cutaneous complaints, with cases. 1821. Colburn.
15. Burgess TH. Practical observations on diseases of the skin. *Prov Med J Retrosop Med Sci* 1843; **5**: 307–310.
16. Damman GW. Preliminary note on some micro-organisms of normal skin. *Brit Med J* 1892; **2**: 122–123.
17. Evans CA, Smith WM, Johnston EA, Giblett ER. Bacterial flora of the normal human skin. *J Invest Dermatol* 1950; **15**: 305–324.
18. Schommer NN, Gallo RL. Structure and function of the human skin microbiome. *Trends Microbiol* 2013; **21**: 660–668.

19. Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA. Biogeography and individuality shape function in the human skin metagenome. *Nature* 2014; **514**: 59–64.
20. Somerville DA. The normal flora of the skin in different age groups. *Br J Dermatol* 1969; **81**: 248–258.
21. Grice EA, Kong HH, Renaud G, Young AC, Bouffard GG, Blakesley RW, et al. A diversity profile of the human skin microbiota. *Genome Res* 2008; **18**: 1043–1050.
22. Ross AA, Doxey AC, Neufeld JD. The skin microbiome of cohabiting couples. *mSystems* 2017; **2**: e00043-17.
23. Bradley CW, Morris DO, Rankin SC, Cain CL, Misic AM, Houser T, et al. Longitudinal evaluation of the skin microbiome and association with microenvironment and treatment in canine atopic dermatitis. *J Invest Dermatol* 2016; **136**: 1182–1190.
24. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. *Proc Natl Acad Sci USA* 2010; **107**: 6477–6481.
25. Council S, Skene JHP, Platt ML, Dunn RR, Horvath JE. Diversity and evolution of the primate skin microbiome. *Proc Royal Soc B* 2016; **283**.
26. Ross AA, Rodrigues Hoffmann A, Neufeld JD. The skin microbiome of vertebrates. *Microbiome* 2019; **7**: 79.
27. Ross AA, Müller KM, Weese JS, Neufeld JD. Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class *Mammalia*. *Proc Natl Acad Sci USA* 2018; **115**: E5786–E5795.

28. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. *Science* 2009; **326**: 1694–1697.
29. Oh J, Byrd AL, Park M, Kong HH, Segre JA. Temporal stability of the human skin microbiome. *Cell* 2016; **165**: 854–866.
30. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and temporal diversity of the human skin microbiome. *Science* 2009; **324**: 1190–1192.
31. Zeeuwen PLJM, Kleerebezem M, Timmerman HM, Schalkwijk J. Microbiome and skin diseases. *Curr Opin Allergy Clin Immunol* 2013; **13**: 514–520.
32. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 1977; **74**: 5088–5090.
33. Sato T, Atomi H. Novel metabolic pathways in archaea. *Curr Opin Microbiol* 2011; **14**: 307–314.
34. Valentine DL. Adaptations to energy stress dictate the ecology and evolution of the *Archaea*. *Nat Rev Microbiol* 2007; **5**: 316–323.
35. Koga Y, Morii H. Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects. *Biosci Biotechnol Biochem* 2005; **69**: 2019–2034.
36. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the *Thaumarchaeota*. *Nat Rev Microbiol* 2008; **6**: 245–252.

37. Spang A, Caceres EF, Ettema TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 2017; **357**: eaaf3883.
38. Moissl-Eichinger C, Pausan M, Taffner J, Berg G, Bang C, Schmitz RA. Archaea are interactive components of complex microbiomes. *Trends Microbiol* 2018; **26**: 70–85.
39. Koskinen K, Pausan MR, Perras AK, Beck M, Bang C, Mora M, et al. First insights into the diverse human archaeome: specific detection of archaea in the gastrointestinal tract, lung, and nose and on skin. *mBio* 2017; **8**: e00824-17.
40. Moissl-Eichinger C, Probst AJ, Birarda G, Auerbach A, Koskinen K, Wolf P, et al. Human age and skin physiology shape diversity and abundance of archaea on skin. *Sci Rep* 2017; **7**: 4039.
41. Pausan MR, Csorba C, Singer G, Till H, Schöpf V, Santigli E, et al. Exploring the archaeome: detection of archaeal signatures in the human body. *Front Microbiol* 2019; **10**: 2796.
42. Lurie-Weinberger MN, Gophna U. Archaea in and on the human body: health implications and future directions. *PLoS Pathog* 2015; **11**: e1004833.
43. Probst AJ, Auerbach AK, Moissl-Eichinger C. Archaea on human skin. *PLoS One* 2013; **8**: e65388.
44. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 2019; **176**: 649–662.

45. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci USA* 2013; **110**: 3229–3236.
46. Telford MJ, Littlewood DTJ (eds). *Animal evolution: genomes, fossils, and trees*. 2009. Oxford University Press, Oxford, New York.
47. Sanford JA, Gallo RL. Functions of the skin microbiota in health and disease. *Semin Immunol* 2013; **25**: 370–377.
48. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science* 2012; **336**: 1268–1273.
49. Lavrinienko A, Tukalenko E, Mappes T, Watts PC. Skin and gut microbiomes of a wild mammal respond to different environmental cues. *Microbiome* 2018; **6**: 209.
50. Fargione J, Brown CS, Tilman D. Community assembly and invasion: An experimental test of neutral versus niche processes. *Proc Natl Acad Sci USA* 2003; **100**: 8916–8920.
51. Brooks AW, Kohl KD, Brucker RM, Opstal EJ van, Bordenstein SR. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biol* 2016; **14**: e2000225.
52. Davenport ER, Sanders JG, Song SJ, Amato KR, Clark AG, Knight R. The human microbiome in evolution. *BMC Biol* 2017; **15**: 127.
53. Lim SJ, Bordenstein SR. An introduction to phylosymbiosis. *Proceedings of the Royal Society B: Biological Sciences* 2020; **287**: 20192900.
54. Lavrinienko A, Tukalenko E, Mappes T, Watts PC. Skin and gut microbiomes of a wild mammal respond to different environmental cues. *Microbiome* 2018; **6**.

55. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. *Science* 2008; **320**: 1647–1651.
56. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, et al. Cospeciation of gut microbiota with hominids. *Science* 2016; **353**: 380–382.
57. Moeller AH, Gomes-Neto JC, Mantz S, Kittana H, Segura Munoz RR, Schmaltz RJ, et al. Experimental evidence for adaptation to species-specific gut microbiota in house mice. *mSphere* 2019; **4**: e00387-19.
58. Ochman H, Worobey M, Kuo C-H, Ndjango J-BN, Peeters M, Hahn BH, et al. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol* 2010; **8**: e1000546.
59. Grond K, Bell KC, Demboski JR, Santos M, Sullivan JM, Hird SM. No evidence for phyllosymbiosis in western chipmunk species. *FEMS Microbiol Ecol* 2020; **96**: fiz182.
60. Foster T. Staphylococcus. In: Baron S (ed). *Medical Microbiology*, 4th ed. 1996. University of Texas Medical Branch at Galveston, Galveston, Texas.
61. Rich M. Staphylococci in animals: prevalence, identification and antimicrobial susceptibility, with an emphasis on methicillin-resistant *Staphylococcus aureus*. *Br J Biomed Sci* 2005; **62**: 98–105.
62. Bletz MC, Archer H, Harris RN, McKenzie VJ, Rabemananjara FCE, Rakotoarison A, et al. Host ecology rather than host phylogeny drives amphibian skin microbial community structure in the biodiversity hotspot of Madagascar. *Front Microbiol* 2017; **8**: 1530.

63. Lutz HL, Jackson EW, Webala PW, Babyesiza WS, Kerbis Peterhans JC, Demos TC, et al. Ecology and host identity outweigh evolutionary history in shaping the bat microbiome. *mSystems* 2019; **4**: e00511-19.
64. Sieber M, Traulsen A, Schulenburg H, Douglas AE. On the evolutionary origins of host–microbe associations. *Proc Natl Acad Sci USA* 2021; **118**: e2016487118.
65. Naik S, Bouladoux N, Wilhelm C, Molloy MJ, Salcedo R, Kastenmuller W, et al. Compartmentalized control of skin immunity by resident commensals. *Science* 2012; **337**: 1115–1119.
66. Wang G, Sweren E, Liu H, Wier E, Alphonse MP, Chen R, et al. Bacteria induce skin regeneration via IL-1 β signaling. *Cell Host Microbe* 2021; **29**: 777-791.e6.
67. Scharschmidt TC. Establishing tolerance to commensal skin bacteria: timing is everything. *Dermatol Clin* 2017; **35**: 1–9.
68. Kobayashi T, Glatz M, Horiuchi K, Kawasaki H, Akiyama H, Kaplan DH, et al. Dysbiosis and *Staphylococcus aureus* colonization drives inflammation in atopic dermatitis. *Immunity* 2015; **42**: 756–766.
69. Williams MR, Gallo RL. Evidence that human skin microbiome dysbiosis promotes atopic dermatitis. *J Invest Dermatol* 2017; **137**: 2460–2461.
70. Otto M. *Staphylococcus epidermidis* – the “accidental” pathogen. *Nat Rev Microbiol* 2009; **7**: 555–567.
71. Robinson NP, Bang C, Schmitz RA. Archaea: forgotten players in the microbiome. *Emerg Top Life Sci* 2018; **2**: 459–468.

72. Oxley APA, Lanfranconi MP, Würdemann D, Ott S, Schreiber S, McGenity TJ, et al. Halophilic archaea in the human intestinal mucosa. *Environ Microbiol* 2010; **12**: 2398–2410.
73. Samuel BS, Gordon JI. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc Natl Acad Sci USA* 2006; **103**: 10011–10016.
74. Martin W. Pathogenic archaeobacteria: do they not exist because archaeobacteria use different vitamins? *Bioessays* 2004; **26**: 592–593.
75. Gill EE, Brinkman FSL. The proportional lack of archaeal pathogens: Do viruses/phages hold the key? *Bioessays* 2011; **33**: 248–254.
76. Větrovský T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 2013; **8**: e57923.
77. Fox GE, Woese CR. 5S RNA secondary structure. *Nature* 1975; **256**: 505–507.
78. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**: 4576–4579.
79. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 1990; **345**: 60–63.
80. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 2008; **74**: 2461–2470.
81. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 1991; **173**: 697–703.

82. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 2019; **10**: 5029.
83. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 2007; **69**: 330–339.
84. Bukin YS, Galachyants YP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data* 2019; **6**: 190007.
85. Gilbert JA, Meyer F, Antonopoulos D, Balaji P, Brown CT, Brown CT, et al. Meeting report: the terabase metagenomics workshop and the vision of an earth microbiome project. *Stand Genom Sci* 2010; **3**: 243.
86. Gilbert JA, Jansson JK, Knight R. The earth microbiome project: successes and aspirations. *BMC Biol* 2014; **12**: 69.
87. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016; **18**: 1403–1414.
88. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* 2015; **75**: 129–137.
89. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced amplicons. *BMC Bioinform* 2011; **12**: 38.

90. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2016; **44**: D7–D19.
91. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for *Bacteria* and *Archaea*. *Nat Biotechnol* 2020; **38**: 1079–1086.
92. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarszewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat biotechnol* 2018; **36**: 996–1004.
93. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013; **41**: D590–D596.
94. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 2012; **28**: 1823–1829.
95. Zeigler DR. Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int J Syst Evol Microbiol* 2003; **53**: 1893–1900.
96. Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* 2007; **73**: 278–288.
97. Ogier J-C, Pagès S, Galan M, Barret M, Gaudriault S. *rpoB*, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiol* 2019; **19**: 171.

98. La Duc MT, Satomi M, Agata N, Venkateswaran K. *gyrB* as a phylogenetic discriminator for members of the *Bacillus anthracis-cereus-thuringiensis* group. *J Microbiol Methods* 2004; **56**: 383–394.
99. Peeters K, Willems A. The *gyrB* gene is a useful phylogenetic marker for exploring the diversity of *Flavobacterium* strains isolated from terrestrial and aquatic habitats in Antarctica. *FEMS Microbiol Lett* 2011; **321**: 130–140.
100. Poirier S, Rué O, Peguilhan R, Coeuret G, Zagorec M, Champomier-Vergès M-C, et al. Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using *gyrB* amplicon sequencing: a comparative analysis with 16S rDNA V3-V4 amplicon sequencing. *PLoS One* 2018; **13**: e0204629.
101. Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE. The chaperonin-60 universal target is barcode for bacteria that enables de novo assembly of metagenomic sequence data. *PLoS One* 2012; **7**: e49755.
102. Goh SH, Potter S, Wood JO, Hemmingsen SM, Reynolds RP, Chow AW. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *J Clin Microbiol* 1996; **34**: 818–823.
103. Goh SH, Facklam RR, Chang M, Hill JE, Tyrrell GJ, Burns ECM, et al. Identification of *Enterococcus* species and phenotypically similar *Lactococcus* and *Vagococcus* species by reverse checkerboard hybridization to chaperonin 60 gene sequences. *J Clin Microbiol* 2000; **38**: 3953–3959.
104. Hill JE. cpnDB: a chaperonin sequence database. *Genome Res* 2004; **14**: 1669–1675.

105. Hill JE, Seipp RP, Betts M, Hawkins L, Kessel AGV, Crosby WL, et al. Extensive profiling of a complex microbial community by high-throughput sequencing. *Appl Environ Microbiol* 2002; **68**: 3055–3066.
106. Johnson LA, Chaban B, Harding JCS, Hill JE. Optimizing a PCR protocol for *cpn60*-based microbiome profiling of samples variously contaminated with host genomic DNA. *BMC Res Notes* 2015; **8**: 253.
107. Verbeke TJ, Sparling R, Hill JE, Links MG, Levin D, Dumonceaux TJ. Predicting relatedness of bacterial genomes using the chaperonin-60 universal target (*cpn60* UT): application to *Thermoanaerobacter* species. *Syst Appl Microbiol* 2011; **34**: 171–179.
108. Vermette CJ, Russell AH, Desai AR, Hill JE. Resolution of phenotypically distinct strains of *Enterococcus* spp. in a complex microbial community using *cpn60* universal target sequencing. *Microb Ecol* 2010; **59**: 14–24.
109. Sakamoto M, Ohkuma M. Usefulness of the *hsp60* gene for the identification and classification of Gram-negative anaerobic rods. *J Med Microbiol* 2010; **59**: 1293–1302.
110. Albert AYK, Chaban B, Wagner EC, Schellenberg JJ, Links MG, Schalkwyk J van, et al. A Study of the vaginal microbiome in healthy Canadian women utilizing *cpn60*-based molecular profiling reveals distinct *gardnerella* subgroup community state types. *PLoS One* 2015; **10**: e0135620.
111. Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome* 2014; **2**: 23.

112. Bondici VF, Lawrence JR, Khan NH, Hill JE, Yergeau E, Wolfaardt GM, et al. Microbial communities in low permeability, high pH uranium mine tailings: characterization and potential effects. *J Appl Microbiol* 2013; **114**: 1671–1686.
113. Levy-Rimler G, Bell RE, Ben-Tal N, Azem A. Type I chaperonins: not all are created equal. *FEBS Letters* 2002; **529**: 1–5.
114. Braig K, Joachimiakt A, Horwich AL, Sigler PB. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 1994; **371**: 9.
115. Hayer-Hartl M, Bracher A, Hartl FU. The GroEL–GroES chaperonin machine: a nanocage for protein folding. *Trends Biochem Sci* 2016; **41**: 62–76.
116. Fayet O, Ziegelhoffer T, Georgopoulos C. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J Bacteriol* 1989; **171**: 1379–1385.
117. Vancuren SJ, Santos SJD, Hill JE, The Maternal Microbiome Legacy Project Team. Evaluation of variant calling for *cpn60* barcode sequence-based microbiome profiling. *PLoS One* 2020; **15**: e0235682.
118. Ross AA, Neufeld JD. Microbial biogeography of a university campus. *Microbiome* 2015; **3**: 66.
119. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* 2017; **11**: 2639–2643.
120. Schellenberg J, Links MG, Hill JE, Dumonceaux TJ, Peters GA, Tyler S, et al. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. *Appl Environ Microbiol* 2009; **75**: 2889–2898.

121. Hill JE, Town JR, Hemmingsen SM. Improved template representation in *cpn60* polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. *Environ Microbiol* 2006; **8**: 741–746.
122. Fernando C, Hill JE. *cpn60* metagenomic amplicon library preparation for the Illumina Miseq platform. 2021. Protocol Exchange.
123. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13**: 581–583.
124. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016; **4**: e2584.
125. Kans J. Entrez direct: e-utilities on the UNIX command line. *Entrez Programming Utilities Help [Internet]*. <https://www.ncbi.nlm.nih.gov/sites/books/NBK179288/>.
126. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673–4680.
127. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018; **35**: 1547–1549.
128. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 2012; **9**: 772–772.
129. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003; **52**: 696–704.
130. Gower JC. Generalized Procrustes analysis. *Psychometrika* 1975; **40**: 33–51.

131. Louca S, Doebeli M, Parfrey LW. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* 2018; **6**: 41.
132. Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One* 2012; **7**: e43093.
133. Laursen MF, Dalgaard MD, Bahl MI. Genomic GC-content affects the accuracy of 16S rRNA gene sequencing based microbial profiling due to PCR bias. *Front Microbiol* 2017; **8**.
134. Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, et al. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* 2015; **15**: 66.
135. Fukuyama J. Emphasis on the deep or shallow parts of the tree provides a new characterization of phylogenetic distances. *Genome Biol* 2019; **20**: 131.
136. Kloos WE, Ballard DN, Webstert JA, Hubner RJ, Ludwig W. Delimiting the genus *Staphylococcus* through description of *Macrococcus caseolyticus* gen. nov., comb. nov. and *Macrococcus equiperficus* Spm nov., *Macrococcus bovicus* sp. nov. and *Macrococcus carouselicus* sp. nov. *Int J Sys Bacteriol* ; **48**: 859–877.
137. Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr Biol* 2015; **25**: 2577–2583.
138. Kaiser-Thom S, Hilty M, Axiak S, Gerber V. The skin microbiota in equine pastern dermatitis: a case-control study of horses in Switzerland. *Vet Dermatol*.

139. Yoon J-H, Lee K-C, Weiss N, Kang KH, Park Y-H 2003. *Jeotgalicoccus halotolerans* gen. nov., sp. nov. and *Jeotgalicoccus psychrophilus* sp. nov., isolated from the traditional Korean fermented seafood jeotgal. *Int J Syst Evol Microbiol*; **53**: 595–602.
140. Glaeser SP, Kleinhagauer T, Jäckel U, Klug K, Kämpfer P. *Jeotgalicoccus schoeneichii* sp. nov. isolated from exhaust air of a pig barn. *Int J Syst Evol Microbiol* 2016; **66**: 3503–3508.
141. Kämpfer P, Busse H-J, Glaeser SP, Clermont D, Criscuolo A, Mietke H. *Jeotgalicoccus meleagridis* sp. nov. isolated from bioaerosol from emissions of a turkey fattening plant and reclassification of *Jeotgalicoccus halophilus* Liu et al. 2011 as a later heterotypic synonym of *Jeotgalicoccus aerolatus* Martin et al. 2011. *Int J Syst Evol Microbiol* 2019; **71**: 004745.
142. Liu W-Y, Jiang L-L, Guo C-J, Yang SS. *Jeotgalicoccus halophilus* sp. nov., isolated from salt lakes. *Int J Syst Evol Microbiol* 2011; **61**: 1720–1724.
143. Brauner P, Klug K, Jäckel U. Eggshells as a source for occupational exposure to airborne bacteria in hatcheries. *J Occup Environ Hyg* 2016; **13**: 950–959.
144. Peng J. Diversity and chemical defense role of culturable non-actinobacterial bacteria isolated from the South China Sea gorgonians. *J Microbiol Biotechnol* 2013; **23**: 437–443.
145. Bautista-Trujillo GU, Solorio-Rivera JL, Rentería-Solórzano I, Carranza-Germán SI, Bustos-Martínez JA, Arteaga-Garibay RI, et al. Performance of culture media for the isolation and identification of *Staphylococcus aureus* from bovine mastitis. *J Med Microbiol* 2013; **62**: 369–376.

146. Bhargava K, Zhang Y. Characterization of methicillin-resistant coagulase-negative staphylococci (MRCoNS) in retail meat. *Food Microbiol* 2014; **42**: 56–60.
147. Kong HH, Andersson B, Clavel T, Common JE, Jackson SA, Olson ND, et al. Performing skin microbiome research: a method to the madness. *J Invest Dermatol* 2017; **137**: 561–568.
148. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014; **42**: D633–D642.
149. Gallo RL. Human skin is the largest epithelial surface for interaction with microbes. *J Invest Dermatol* 2017; **137**: 1213–1214.
150. Neuman H, Debelius JW, Knight R, Koren O. Microbial endocrinology: the interplay between the microbiota and the endocrine system. *FEMS Microbiol Rev* 2015; **39**: 509–521.
151. Grice EA, Segre JA. The skin microbiome. *Nat Rev Microbiol* 2011; **9**: 244–253.
152. James AG, Austin CJ, Cox DS, Taylor D, Calvert R. Microbiological and biochemical origins of human axillary odour. *FEMS Microbiol Ecol* 2013; **83**: 527–540.
153. Chng KR, Tay ASL, Li C, Ng AHQ, Wang J, Suri BK, et al. Whole metagenome profiling reveals skin microbiome-dependent susceptibility to atopic dermatitis flare. *Nat Microbiol* 2016; **1**: 16106.
154. Fyhrquist N, Muirhead G, Prast-Nielsen S, Jeanmougin M, Olah P, Skoog T, et al. Microbe-host interplay in atopic dermatitis and psoriasis. *Nat Commun* 2019; **10**: 4703.

155. Alekseyenko AV, Perez-Perez GI, De Souza A, Strober B, Gao Z, Bihan M, et al. Community differentiation of the cutaneous microbiota in psoriasis. *Microbiome* 2013; **1**: 31.
156. Pierezan F, Olivry T, Paps JS, Lawhon SD, Wu J, Steiner JM, et al. The skin microbiome in allergen-induced canine atopic dermatitis. *Vet Dermatol* 2016; **27**: 332-e82.
157. Zinicola M, Lima F, Lima S, Machado V, Gomez M, Döpfer D, et al. Altered microbiomes in bovine digital dermatitis lesions, and the gut as a pathogen reservoir. *PLoS One* 2015; **10**: e0120504.
158. Ayalew Y, Assefa A, Mekonen N, Belete S, Ayisheshim A. A review on camel dermatophilosis. *Adv Biol Res* 2015; **9**: 363–372.
159. Timm CM, Loomis K, Stone W, Mehoke T, Brensinger B, Pellicore M, et al. Isolation and characterization of diverse microbial representatives from the human skin microbiome. *Microbiome* 2020; **8**: 58.
160. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. *Nature* 2012; **486**: 207–214.
161. Gaci N, Borrel G, Tottey W, O’Toole PW, Brugère J-F. Archaea and the human gut: new beginning of an old story. *World J Gastroenterol* 2014; **20**: 16062–16078.
162. Raymann K, Moeller AH, Goodman AL, Ochman H. Unexplored archaeal diversity in the great ape gut microbiome. *mSphere* 2017; **2**: e00026-17.

163. Borrel G, Brugère J-F, Gribaldo S, Schmitz RA, Moissl-Eichinger C. The host-associated archaeome. *Nat Rev Microbiol* 2020; **18**: 622–636.
164. Cavicchioli R, Curmi PMG, Saunders N, Thomas T. Pathogenic archaea: do they exist? *Bioessays* 2003; **25**: 1119–1128.
165. Bang C, Schmitz RA. Archaea associated with human surfaces: not to be underestimated. *FEMS Microbiol Rev* 2015; **39**: 631–648.
166. Togo AH, Grine G, Khelaifia S, des Robert C, Brevaut V, Caputo A, et al. Culture of methanogenic archaea from human colostrum and milk. *Sci Rep* 2019; **9**: 18653.
167. Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. *Nat Rev Microbiol* 2018; **16**: 143–155.
168. Kong HH, Segre JA. Skin microbiome: looking back to move forward. *J Invest Dermatol* 2012; **132**: 933–939.
169. Leung MHY, Tong X, Bastien P, Guinot F, Tenenhaus A, Appenzeller BMR, et al. Changes of the human skin microbiota upon chronic exposure to polycyclic aromatic hydrocarbon pollutants. *Microbiome* 2020; **8**: 100.
170. Ekman L, Bagge E, Nyman A, Waller KP, Pringle M, Segerman B. A shotgun metagenomic investigation of the microbiota of udder cleft dermatitis in comparison to healthy skin in dairy cows. *PLoS One* 2020; **15**: e0242880.
171. Lax S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, et al. Forensic analysis of the microbiome of phones and shoes. *Microbiome* 2015; **3**: 21.

172. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. *PLoS One* 2014; **9**: e105592.
173. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl Environ Microbiol* 2011; **77**: 3846–3852.
174. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019; **37**: 852–857.
175. Min D, Doxey AC, Neufeld JD. AXIOME3: automation, extension, and integration of microbial ecology. *GigaScience* 2021; **10**: giab006.
176. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011; **17**: 10–12.
177. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013; **41**: e1–e1.
178. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, et al. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci USA* 2007; **104**: 10643–10648.
179. Hansen EE, Lozupone CA, Rey FE, Wu M, Guruge JL, Narra A, et al. Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proc Natl Acad Sci USA* 2011; **108**: 4599–4606.

180. Angel R, Claus P, Conrad R. Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. *ISME J* 2012; **6**: 847–862.
181. Bomberg M, Montonen L, Jurgens G, Münster U. Diversity and function of archaea in freshwater habitats. *Curr Trends Microbiol* 2008; **4**: 61–89.
182. Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 2006; **442**: 806–809.
183. Erguder TH, Boon N, Wittebolle L, Marzorati M, Verstraete W. Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol Rev* 2009; **33**: 855–869.
184. Oren A. Industrial and environmental applications of halophilic microorganisms. *Environ Technol* 2010; **31**: 825–834.
185. Leung MHY, Wilkins D, Lee PKH. Insights into the pan-microbiome: skin microbial communities of Chinese individuals differ from other racial groups. *Sci Rep* 2015; **5**: 11845.
186. Urban J, Fergus DJ, Savage AM, Ehlers M, Menninger HL, Dunn RR, et al. The effect of habitual and experimental antiperspirant and deodorant product use on the armpit microbiome. *PeerJ* 2016; **4**: e1605.
187. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One* 2013; **8**: e85024.
188. Mohsen A, Park J, Chen Y-A, Kawashima H, Mizuguchi K. Impact of quality trimming on the efficiency of reads joining and diversity analysis of Illumina paired-end reads in

- the context of QIIME1 and QIIME2 microbiome analysis frameworks. *BMC Bioinform* 2019; **20**: 581.
189. Spasov E, Tsuji JM, Hug LA, Doxey AC, Sauder LA, Parker WJ, et al. High functional diversity among *Nitrospira* populations that dominate rotating biological contactor microbial communities in a municipal wastewater treatment plant. *ISME J* 2020; **14**: 1857–1872.
190. Sauder LA, Albertsen M, Engel K, Schwarz J, Nielsen PH, Wagner M, et al. Cultivation and characterization of *Candidatus Nitrosocosmicus exaquare*, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J* 2017; **11**: 1142–1157.
191. McNichol J, Berube PM, Biller SJ, Fuhrman JA. Evaluating and improving small subunit rRNA PCR primer coverage for bacteria, archaea, and eukaryotes using metagenomes from global ocean surveys. *mSystems* 2021; **6**: 13.
192. Kennedy K, Hall MW, Lynch MDJ, Moreno-Hagelsieb G, Neufeld JD. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol* 2014; **80**: 5717–5722.
193. Chandler DP, Fredrickson JK, Brockman FJ. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. *Mol Ecol* 1997; **6**: 475–482.
194. Albers S-V, Meyer BH. The archaeal cell envelope. *Nat Rev Microbiol* 2011; **9**: 414–426.

195. Roopnarain A, Mukhuba M, Adeleke R, Moeletsi M. Biases during DNA extraction affect bacterial and archaeal community profile of anaerobic digestion samples. *3 Biotech* 2017; **7**: 1–12.
196. Mancabelli L, Milani C, Lugli GA, Turrone F, Ferrario C, Sinderen D van, et al. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ Microbiol* 2017; **19**: 1379–1390.
197. Bay L, Barnes CJ, Fritz BG, Thorsen J, Restrup MEM, Rasmussen L, et al. Universal dermal microbiome in human skin. *mBio* 2020; **11**: e02945-19.
198. Starke R, Siles J, Fernandes M, Schallert K, Benndorf D, Plaza C, et al. The structure and function of soil archaea across biomes. *J Proteomics* 2021; **237**: 104147.
199. Schuchmann K, Müller V. Energetics and application of heterotrophy in acetogenic bacteria. *Appl Environ Microbiol* 2016; **82**: 4056–4069.
200. Kim JY, Whon TW, Lim MY, Kim YB, Kim N, Kwon M-S, et al. The human gut archaeome: identification of diverse haloarchaea in Korean subjects. *Microbiome* 2020; **8**: 114.
201. Nose K, Mizuno T, Yamane N, Kondo T, Ohtani H, Araki S, et al. Identification of ammonia in gas emanated from human skin and its correlation with that in blood. *Anal Sci* 2005; **21**: 1471–1474.
202. Czarnowski D, Górski J, Józwiuk J, Boroń-Kaczmarek A. Plasma ammonia is the principal source of ammonia in sweat. *Eur J Appl Physiol* 1992; **65**: 135–137.
203. Brusilow SW, Gordes EH. Ammonia secretion in sweat. *J Physiol* 1968; **214**: 513–517.

204. Schmidt FM, Vaittinen O, Metsälä M, Lehto M, Forsblom C, Groop P-H, et al. Ammonia in breath and emitted from skin. *J Breath Res* 2013; **7**: 017109.
205. Venkataraman A, Parlov M, Hu P, Schnell D, Wei X, Tiesman JP. Spike-in genomic DNA for validating performance of metagenomics workflows. *Biotechniques* 2018; **65**: 315–321.
206. Hill JE, Fernando WMU, Zello GA, Tyler RT, Dahl WJ, Van Kessel AG. Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the *cpn60* universal primer cocktail. *Appl Environ Microbiol* 2010; **76**: 4550–4552.