Molecular Tracking and Cultivation Reveal Ammonia Oxidizing Archaea as Integral Members of the Human Skin Microbiome

Alexander Mahnert*[,1], Maximilian Dreer*[,2], Ülkü Perier[2], Michael Melcher[2], Stefanie Duller[1], Adina Lehnen[1], Theodora Goessler[1], Daniela Brunner[1], Thomas Graier[3], Peter Wolf[3], Rafael I. Ponce-Toledo[2,#], Logan Hodgskiss[2], Melina Kerou[+,2], Christine Moissl-Eichinger[+,1,4], and Christa Schleper[+,2]

* Shared authorship
[+] Shared corresponding authorship

[1] Medical University of Graz, D&R Institute for Hygiene, Microbiology, and Environmental Medicine, Graz, Austria
[2] University of Vienna, Archaea Biology and Ecogenomics Unit, Vienna, Austria
[3] Medical University of Graz, Department of Dermatology and Venereology, Graz, Austria
[4] BioTechMed Graz, Graz, Austria

[#] Present address: Institut de Systématique, Evolution et Biodiversité (ISYEB), Sorbonne Université, CNRS, Muséum National d'Histoire Naturelle, Paris, France

Abstract

Ammonia oxidizing archaea (AOA) have been repeatedly detected in the human skin microbiome through molecular and biochemical analyses, yet their persistence, physiology, and adaptations remain poorly understood. Here, we describe two cultivated strains, *Candidatus* Nitrosocosmicus epidermidis and *Ca.* Nitrosocosmicus unguis, enriched from human skin samples. These autotrophic strains grow on ammonia and urea as sole energy sources. Genomic islands and expanded gene families in their genomes testify to the capacity of colonizing skin, including specific interactions with host proteins and signaling cascades that distinguish these AOA from their close relatives from soil.

Molecular signatures of *Nitrosocosmicus* ssp. were consistently identified on individuals in cross-sectional and longitudinal cohorts (n=47) with a higher prevalence in sebaceous areas. Co-occurrence network patterns with specific bacteria reinforce the observation that AOA of the genus *Nitrosocosmicus* form a stable component of the healthy skin and represent emerging commensals in the human microbiome.

Introduction

The human body is an ecosystem populated by spatially organized microbial communities composed of bacteria, archaea, viruses/ phages and small eukaryotes, with diverse metabolic activities and interactions: the microbiome [1]. In recent years, numerous studies have led to a better understanding of the mechanisms involved in the complex interplay between the host and its commensals and have revealed the fundamental role of the human microbiome in health and disease [2–4]. However, most studies on human microbial communities have focused exclusively on the bacterial microbiome, while other integral components, such as eukaryotes/fungi, archaea and viruses have remained largely overlooked [5–9]. A systematic metagenomic study has recently revealed the extent of diversity and abundance of archaea in the human gut microbiome [10], where the increased presence of methanogenic archaea affects short-chain fatty acid production and the overall host physiology [11].

The skin microbiome is equally complex and harbors various bacterial species, including *Staphylococcus, Streptococcus, Corynebacterium, and Propionibacterium* (now *Cutibacterium*), and fungi such as Malassezia [12,13]. The composition and diversity of the microbiome varies across different body sites (e.g. oily versus dry skin) and between individuals [14]. As in the gut, microorganisms play essential roles in maintaining skin health as they provide protection against pathogens by competing for nutrients and space, producing antimicrobial compounds and modulating the skin's immune response [15].

Human skin is also colonized by archaea [16,17], in particular by representatives of the ammonia oxidizing archaea (AOA), formerly Thaumarchaeota, now class Nitrososphaeria and phylum Thermoproteota [18–20]. Although rather rare in abundance, archaeal signals were found in human subjects analyzed by 16S rRNA gene sequencing and detection of archaeal *amoA*, the gene for subunit A of the key enzyme ammonia monooxygenase used for ammonia oxidation[16]. AOA-associated Nitrososphaerales signatures were detected in all human subjects and were closely related to those found earlier in cleanrooms and intensive care units [21]. Their presence was confirmed by fluorescence *in situ* hybridization [16], next generation sequencing, and infrared hyperspectral imaging [17]. Although the physiological relevance of AOA to the skin is still subject to speculation, its potential role is intriguing considering that urea and ammonia are components of human sweat and are the main energy sources of ammonia oxidizers. Also, nitric oxide (NO), an important signaling molecule involved in diverse processes in humans, is a byproduct of the ammonia oxidation process [22].

Detection of archaea in human microbiomes is generally hampered by numerous methodological obstacles, such as challenging cultivation [23,24], and molecular detection issues (summarized in [25]), because of the bacteria-centric orientation in sampling, cell lysis, DNA extraction protocols, PCR primers, and classification databases. Consequently, archaeal

signals often remain close to or below the detection limit, making it difficult to understand the role of archaea on human skin [24]. Such obstacles and the lack of cultivated strains of skin AOA have left many questions unanswered, with respect to i) abundance, ii) longitudinal stability, iii) association with human health, iv) physiology and metabolism, v) interaction with the bacterial microbiome, and vi) specific adaptations to the human skin.

In this manuscript, we present the results of extensive cultivation efforts and molecular-based analyses targeting ammonia oxidizing archaea of healthy and diseased individuals. This endeavor resulted in the stable and high enrichment of strains affiliated with the genus *Candidatus* Nitrosocosmicus from human skin. Their genomes reveal genomic features that could facilitate colonization of the skin environment, but also highlight the genomic repertoire shared among all members of the genus *Ca*. Nitrosocosmicus, a widespread environmental group of AOA. The signatures of Nitrosocosmicus could be tracked in different cohorts of healthy volunteers and diseased patients, and correlations with physical parameters and with bacterial occurrence patterns were identified.

Results

Cultivation of AOA from human skin

Different types of inocula were used to initiate enrichment cultures under aerobic conditions in the presence of ammonia as sole energy source. After several months, amplification and sequencing of archaeal 16S rRNA and *amoA* gene fragments identified the presence of one AOA strain per enrichment. The usage of pooled samples of the face, neck, upper body, back, arms, and legs of individuals resulted in three enrichment cultures: Nitrosocosmicus “T1S”, “X2B”, and “Z3A”. A fourth enrichment culture was obtained from cut fingernail edges as initial inoculum (“R2S”).

After repeated transfers over the course of more than a year in medium supplemented with 0.6 mM ammonium, 2 mM of either sodium carbonate (R2S) or sodium bicarbonate (T1S) and an antibiotic mix, two stable cultures displaying near stoichiometric conversion of ammonia to nitrite were obtained within approximately 3-4 weeks (Fig. 1, A-B), with a relative enrichment level of 88-98% based on molecular and microscopic assessments (Extended Data Fig. 1). No bacterial contaminants were detected via amplification of the bacterial 16S rRNA gene (Extended Data Fig. 2), or by light and scanning electron microscopy (Extended Data Fig. 1). Based on the sequenced internal transcribed spacer (ITS) region the remaining contaminant in both enrichments was a single fungal species with 99% sequence identity to *Exophiala* sp. DAOM 216391, a genus of fungi frequently found in manmade environments and known to be an opportunistic pathogen of humans [26,27].

Growth of the two strains, R2S and T1S, correlated with nitrite production in both enrichments, as has been shown previously for other AOA (Fig. 1, A-B) [28–30]. The shortest generation times based on nitrite production were approximately five days at 28°C. After the oxidation of ~500 µM ammonium enrichments R2S and T1S reached cell densities of $2.16*10^5$ (± $1.68*10^4$) and $1.18*10^6$ (± $1.05*10^5$) cells ml^{-1}, respectively. This is similar to the cell densities of *Nitrosocosmicus franklandianus* C13 ($3.8*10^6$ cells ml^{-1}) [31] and *Nitrosocosmicus oleophilus* MY3 ($1.6*10^6$ cells ml^{-1}) [32] after oxidation of ~500 µM ammonium.

In light microscopy, putative AOA were observed as cocci with diameters between 1-2 µm which frequently formed aggregates (Extended Data Fig. 1) and were partly enclosed in putative extracellular polymeric substances (EPS) (Extended Data Fig. 1). These morphological features are comparable with those of previously published members of the genus Nitrosocosmicus [30–33] and were confirmed by scanning electron microscopy (Fig. 1, C-F).

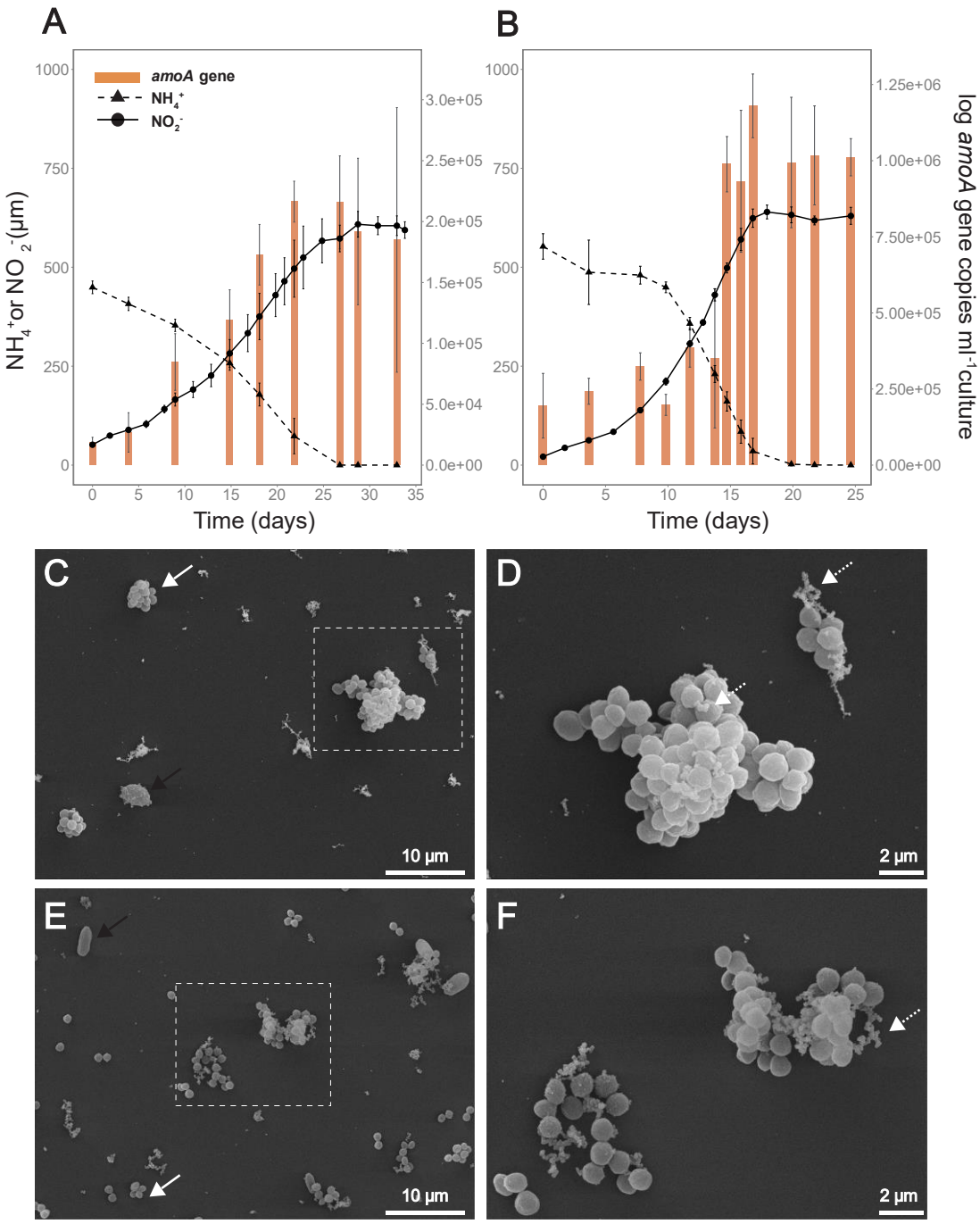


Fig. 1. | Growth and morphology of skin enrichments R2S (A) and T1S (B) displaying near stoichiometric conversion of ammonia (triangles, dashed line) to nitrite (circles, solid line) paralleled by cell growth, estimated by qPCR of archaeal *amoA* gene copies (orange bars). Ammonia and nitrite measurements are averages of biological triplicates. Gene copies of *amo*A are averages of technical triplicates of each biological triplicate. Error bars depict the standard deviation of the mean. Scanning electron micrographs of skin enrichments R2S (C, D) and T1S (E, F) with white arrows indicating AOA cells based on the morphologies seen in Supplementary Figure AOA2 and already published members of the Nitrosocosmicus [30–33]. Dashed white arrows indicate putative extra polymeric substances (D, F). Black arrows indicate the remaining fungal contaminant.

Taxonomic placement and genomic features of the AOA representatives from human skin

Sequencing and assembly of the genomes of the two skin enrichments resulted in one circular contiguous sequence affiliated with AOA for each strain. The genome sizes, G+C content and predicted protein-coding genes of the strains were well within range of other representatives of the genus *Nitrosocosmicus* (Table 1).

Like other representatives of this genus, both strains encoded three identical copies of the 16S and 23S rRNA gene operons, typically indicative of a capacity to respond rapidly to favorable environmental changes [34,35]. The average nucleotide identity (ANI) and average amino acid identity (AAI) values between strains R2S and T1S (87.4% ANI, 86.2% AAI), and between them and other *Nitrosocosmicus* genomes and proteomes were well below the threshold of 95% ANI/AAI proposed for species delineation [36] (Supplementary Tables S1, S2). We therefore propose that they represent two distinct novel species, preliminarily named *Ca.* Nitrosocosmicus unguis R2S (unguis = nail) and *Ca.* Nitrosocosmicus epidermidis T1S ("of the epidermis"), within the candidate genus Nitrosocosmicus.

Table 1: Genomic and growth features of *Ca.* Nitrosocosmicus epidermidis, *Ca.* Nitrosocosmicus unguis compared to other cultivated representatives of the genus *Nitrosocosmicus*

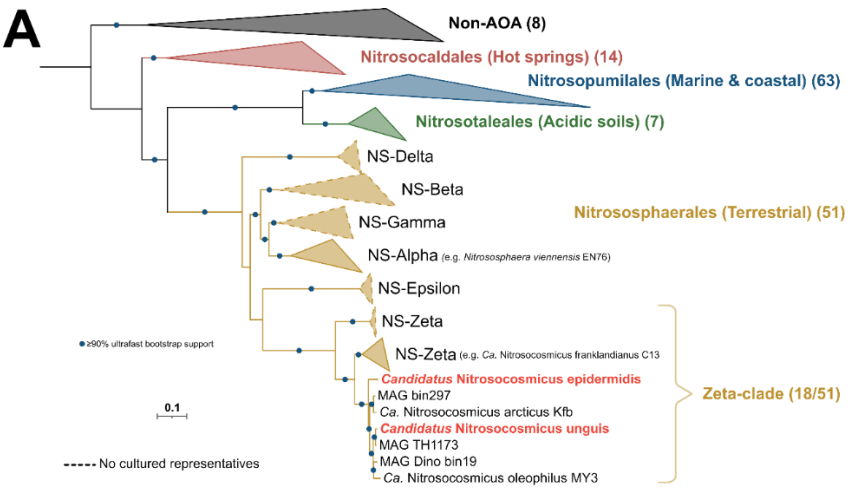
Organism[#]	Status	Growth temp (°C)	pH	Genome size (Mb)	No. of contigs	DNA GC %	protein coding genes	16S/23S/5S rRNA gene copies	reference
Nitrosocosmicus unguis R2S	Enriched	28	7	3.21	1	33.77	3597	3/3/1	this study
Nitrosocosmicus epidermidis T1S	Enriched	28	7	3.07	1	33.55	3401	3/3/1	this study
***N.* oleophilus MY3**	Pure	30	6.5-7	3.43	1	34.14	3722	3/3/1	Jung et al. 2016
***N.* franklandianus C13**	Pure	40	7	2.84	1	34.07	3025	2/2/1	Lehtovirta-Morley et al. 2016
***N.* arcticus Kfb**	Enriched	28	6	2.65*	23	34	2970	3/3/1	Alves et al. 2019
***N.* hydrocola G61**	Enriched	33	8	2.99	1	33.94	3162	2/2/1	Sauder et al. 2017
***N.* agrestis SS**	Enriched	37	6.5-7	3.22**	43	33.42	3524	2/2/1	Liu et al. 2021

[#] all organisms have Candidatus status

* completeness and contamination values: 98.95%, 1.94%

** completeness and contamination values: 96.1%, 2.91%

A maximum likelihood analysis based on 36 concatenated ribosomal protein marker genes present in one copy in at least 100 of the 143 genomes of a dataset comprising representative AOA and non-AOA of Nitrososphaeria confirmed the affiliation of the two novel strains with the NS-zeta lineage as defined by Alves et al. 2018 [37]. Specifically, the two strains formed a highly supported cluster with N. arcticus (obtained from arctic soil), N. oleophilus (from coal-tar contaminated sediment), and other uncharacterized MAGs (Fig 2).



B

Domain desctiption	Domain ID	NS-ζ							NS-α		NT-α
		N.arc	N.frank	N.oleo	N.hydr	N.agr	N.ung	N.epid	N.vie	N.ever	N.deva
Sialidases											
Sialidase superfamily	IPR036278	2	7	13	9	9	4	5	7	11	4
Putative Lipases											
alpha/beta hydrolase fold	PF00561	5	8	14	7	9	13	12	5	6	1
alpha/beta hydrolase fold 1	PF12697	2	4	3	3	2	1	5	1	3	0
Alpha/beta hydrolase fold-3	PF07859	2	2	2	4	2	1	2	0	1	0
Patatin-like phospholipase domain	PF01734 arCOG09020	3	7	15	11	5	14	4	2	6	1
PLipase/COase/thioEstase	PF02230	1	1	1	1	1	1	2	0	1	0
Palmitoyl protein thioesterase	PF02089	1	1	1	0	1	3	2	0	0	0
Adhesion-related and surface recognition-related domains											
Tol-Pal_TolA	IPR014161	1	1	1	1	0	1	1	0	0	0
WD40/YVTN repeat-like-containing domain superfamily, Structurally, both repeated motifs form seven-bladed propellers	IPR015943	12	14	23	20	19	15	13	6	9	9
NHL repeat, Six-bladed beta-propeller (also found in serine/threonine protein kinases in diverse pathogenic bacteria)	PF01436	5	5	5	5	6	6	2	0	0	4
Kelch-type beta propeller	PF01344	2	3	2	4	3	3	2	1	1	0
AdhesinB	IPR006129	2	2	2	2	2	2	2	2	1	1
Hemopexin-like repeats	PF00045	0	0	1	1	0	1	0	0	0	0
HYR domain	PF02494	1	1	1	1	1	1	1	4	2	3
Immunoglobulin-like fold	IPR013783	2	8	4	5	7	4	3	18	20	21
Coagulation factor 5/8 C-terminal domain	IPR000421	1	1	1	4	1	1	1	3	3	0
hyalin	PF02494	1	1	1	1	1	1	1	4	2	3
fibronectin	SSF49265, PS50853	0	0	0	0	0	1	0	1	0	12
Virulence-associated and interaction-associated domains											
Hemolysin-type calcium-binding conserved site	IPR018511	0	3	0	0	2	3	4	1	1	0
1.C.11 The Pore-forming RTX Toxin (RTX-toxin) Family // Serralysin-like metalloprotease, C-terminal // RTX calcium-binding nonapeptide repeat	IPR011049/ PF00353/ arCOG11990	0	4	2	2	4	4	6	2	2	1
Cysteine-rich secretory protein -related	IPR001283	0	1	4	5	0	7	1	0	0	0
Serine rich repeat protein Dentin sialophosphoprotein-like // 1.B.40 The Autotransporter-2 (AT-2) Family	no identifiable domain // OG0001706	1	1	1	1	1	1	1	0	0	0
Peptidases											
Calcineurin-like phosphoesterase domain, ApaH type	PF00149	4	3	4	3	4	4	5	2	2	4
Subtilisin-like serine protease	IPR015500	2	4	3	4	3	4	5	1	1	2
Peptidase M10, metallopeptidase, matrixin	PF00413	1	1	1	1	1	1	1	2	1	4
Peptidase S1C	PF13180	4	4	4	5	3	4	4	4	4	3
Peptidase C1A, papain C-terminal	PF00112	0	1	0	1	0	1	1	0	0	0
Peptidase S8/S53 domain, subtilisin-related	PF00082	2	5	4	4	3	4	5	1	2	2
Peptidase M48	PF01435	1	1	1	2	1	1	1	1	1	0
Peptidase M32, carboxypeptidase Taq	PF02074	1	1	1	1	1	1	1	0	1	0
Serine aminopeptidase, S33	PF12146	2	5	5	5	2	5	2	1	2	1
Peptidase M23	PF01551, arCOG11616	2	1	2	1	1	5	3	0	0	0
Beta-lactamase/transpeptidase-like	SSF56601	1	4	3	6	5	3	0	0	0	0
Peptidase M3A/M3B catalytic domain	PF01432	1	1	2	2	1	1	1	1	1	1
Peptidase_M50	PF02163	1	2	2	1	2	2	3	3	3	2
Peptidase M10, metallopeptidase	PF00413	1	1	1	1	1	1	1	2	1	4

Fig. 2. | Maximum-likelihood phylogenomic tree of AOA and non-AOA reconstructed based on the concatenated alignment of 36 ribosomal proteins comprising a total of 4,707 amino acid sites. Numbers in parentheses represent the number of complete genomes and MAGs in each lineage. NS = Nitrososphaerales (A). Heatmap displaying the detection frequency of specific functional domains discussed in the text in all analyzed *Ca.* Nitrosocosmicus genomes. Numbers refer to the number of occurrences of the respective domain in the genome, but do not always correspond do distinct CDS (i.e. multiple domains on one CDS). Domain IDs refer to Interpro, Pfam. TIGRfam, Superfamily Database entries (B).

To investigate the specific genomic repertoire of *Ca.* N. unguis and *Ca.* N. epidermidis, we reconstructed 3,301 orthologous protein families from a dataset including all available genomes from cultivated strains of the genus *Nitrosocosmicus*: *Ca*. N. franklandianus C13, *Ca.* N. oleophilus MY3, N. hydrocola G61, N. agrestis SS, and *Ca.* N. arcticus Kfb (we refrain from the "Ca." prefix for better readability in the following) (Extended Data: Orthologs). From these orthologous protein families, 1,592 represented the core genome of the genus, comprising about 50% of the predicted CDS of the genomes. In line with their phylogenomic placement, the N. epidermidis shared less protein families with N. oleophilus and N. arcticus (60), while 125 families were shared between N. unguis and N. oleophilus. The content of arCOG functional categories did not differ among the Nitrosocosmicus genomes, indicating that presence and survival on human skin did not require a particular expansion in a functional category (Extended Data Fig. 3). N. unguis and N. epidermidis shared only 20 protein families to the exclusion of all others, while both of them shared more protein families with N. oleophilus (125 and 22 respectively) than any other Nitrosocosmicus genome. The shell (=strain-specific) proteome of each skin isolate consisted of 17 and 16 families for N. unguis and N. epidermidis respectively. Below we discuss specific features of the two skin strains and of the genus Nitrosocosmicus in general, that indicate adaptation to host surfaces and interactions.

A genomic island and expanded gene families implicate factors involved in host interactions

The genome of N. epidermidis contains genomic islands flanked by integrases, with a range of proteins potentially involved in host interaction (Extended Data: Annotations). One of these islands (NET5MPC655_v1_2202- NET5MPC655_v1_2261), encodes a putative autotransporter-1 (AT-1) family (1.B.12) protein (NET5MPC655_v1_2213) with predicted extracellular proteins of unknown function. The C-terminus of autotransporter proteins has the capacity to form β-barrel structures into the outer membrane in order to allow for the export of diverse N-terminal virulence factors [38]. The N-terminal part of the N. epidermidis autotransporter protein consists of leucine-rich repeats involved in protein–protein recognition processes [39]. The island also encodes a putative α-arrestin -like protein (NET5MPC655_v1_2245), matching both arrestin N-terminal (IPR011021) and arrestin C-terminal (IPR014752) -like domains [40]. Arrestins are involved in binding and regulating transmembrane receptors of the 7TM and kinase classes in eukaryotes (including GPCRs in higher eukaryotes), with crucial physiological functions [40]. Although prokaryotic arrestin-like proteins have not been characterized, there are examples of pathogens hijacking β-arrestin signaling pathways of the host to cross epithelial/endothelial barriers [41]. The genomic island also encodes eight copies of membrane attached serine/threonine protein kinases and four

putative serine-rich adhesins, the importance of which in facilitating binding to various tissue types and extracellular matrix proteins in streptococci is well documented [42].

The second strain, N. unguis, encodes expanded gene families whose products are likely to be involved in host interaction (Fig. 2B, Extended Data: Annotations). Among these are seven cysteine-rich secretory protein (CAP) family proteins, which are only found among few members of AOA of Nitrosocosmicus, with only one copy in N. epidermidis. Although present in several bacterial lineages [43], this protein family has only been studied in eukaryotes so far, with essential functions ranging from adhesion to signaling and regulation of ion channels, and can be involved either in pathogenicity or in antimicrobial defense [44].

The genome of N. unguis is also more enriched in beta propeller fold- containing proteins compared to N. epidermidis (Fig. 2B, Extended Data: Annotations) implying different strategies to interact with the surrounding environment and communities. These proteins can act as ligand-binding and signaling proteins or catalyze various hydrolytic or oxidoreductase reactions in the periplasmic space, thereby mediating interactions with the environment, such as the influenza virus neuraminidase [45]. They are generally found in higher numbers in Nitrosocosmicus genomes compared to other AOA lineages (Fig. 2B).

A comprehensive description of the genomic repertoire of both strains, including more putative adaptations to the skin environment can be found in the Supplementary Information and in Extended Data: Annotations.

Members of the genus *Nitrosocosmicus* are equipped for the colonization of host surfaces

Different cultivated Nitrosocosmicus strains have been shown to produce biofilms and exhibit superior adhesion capacities compared to other AOA lineages [32,33], a trait also observed for many bacterial species residing on the human skin [46]. This is mirrored in the proteomes of Nitrosocosmicus ssp. which are enriched in acetamidosugar biosynthesis and adhesion-related proteins, as well as being the only AOA clade encoding LPSE family lipopolysaccharide exporters (Fig. 2B, Extended Data: Annotations) [32,33]. For example, Nitrosocosmicus genomes specifically encode one family of extracellular serine-rich repeat proteins with roles including adhesion and biofilm formation, with 35% sequence identity to dentin sialoprotein-like proteins from eukaryotes [47].

Furthermore, Nitrosocosmicus genomes are characterized by the presence of a suite of gene families whose products can directly interact with and modulate other organisms (Fig. 3). Among these are (i) a putative Zn-dependent endopeptidases related to the RTX toxin family (1.C.11, e.g. 2 clusters in N. epidermidis) which in Gram-negative bacteria are implicated in virulence [48,49],

(ii) enrichment of further families of secreted peptidases (e.g. peptidase family M23, Subtilisin-like serine protease), that can have diverse roles, both in tissue invasion and establishment but also in modulation of other virulence factors [50], (iii) extracellular or membrane-attached putative triacylglycerol esterases/lipases (arCOG06923, PF02089) of which the skin strains encode 2-3 orthologs, that in bacterial colonization/infection of host epithelia enable catabolism of host fatty acids in order to alter host immune response or serve anabolic purposes [51,52]. (iv) patatin-like phospholipase domain proteins (arCOG09020) in N. unguis and N. oleophilus as found in known epithelial colonizing strains [51,52].

Archaeal ammonia oxidizers form a rare, yet prevalent population on human skin

To understand the abundance, prevalence, distribution and stability of archaeal, and particularly Nitrosocosmicus representatives on human skin, 47 healthy individuals were investigated on eight defined body locations (cohorts A1 (age 20-40 y), A2 (60-80+)). According to their microbial profiles (high AOA, low AOA, see methods), 12 of these individuals were then selected and longitudinally sampled on four body sites for an entire year (cohort B; study overview: Extended Data Fig. 4 STORM chart). Samples were quantitatively (qPCR) analyzed using taxonomic (16S rRNA) and functional (*amoA*) marker genes, targeted towards the detection of archaea and archaeal ammonia oxidizers (AOA). As metagenomic sequencing failed due to the low biomass and high background in human DNA, diversity assessments of the skin archaeome were performed via amplicon sequencing to profile the bacterial, archaeal and AOA-specific signatures. A total of 1,770 amplicons (including controls) were processed (details see Materials and Methods, Extended Data Fig. 4 STORM chart).

Throughout the study (cohorts A, B), archaeal signatures were found at least once in all study participants, but were overall low, according to qPCR measurements (approx. 0.2% relative abundance, median across all body sites). AOA-specific qPCR signals accounted for 1.54×10^3 gene copies per 300 cm^2 (Fig. 3A). The highest amount of *amoA* gene copies were found in samples from the forehead (Fig. 3B); there, AOA signatures were found in samples from every tested person at each time point (prevalence: 100%).

In amplicon sequencing, 9.45% of all samples contained archaea when the universal PCR primer combination was applied, whereas 69.98% were positive with the archaea-targeted combination of 16S rRNA gene primers, emphasizing once more the necessity to use archaea-targeted methodology for optimized detection. Also with this approach, all individuals were found to be positive for AOA signatures at least once throughout the study period.

Retrieved 16S rRNA gene sequences of the most abundant ASVs grouped well with the cultivated archaea from human skin (Fig. 3C), because the most prevalent archaeal family

within 326 samples (cohort A and B) was Nitrososphaeraceae (74% positive samples) with *Nitrosocosmicus* being the most prominent genus in this family (24%) (Supplementary Table S3). The joint classification of 16S rRNA to the *amoA* marker gene (see Methods for details) resulted in a prevalence of 38% for the NS-Zeta clade, across all samples.

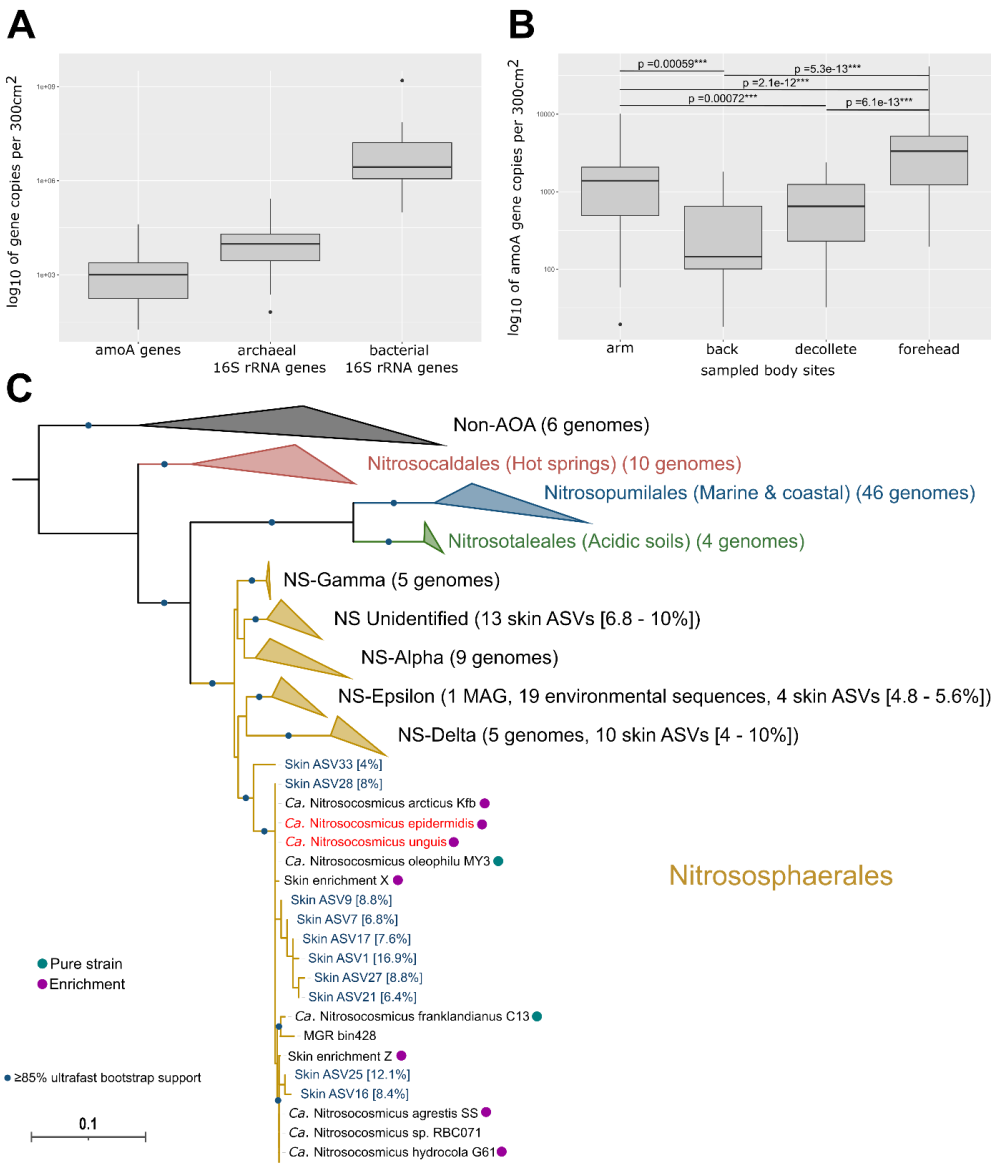


Fig. 3. | Quantitative analysis (qPCR) of AOA, archaeal and bacterial signatures on human skin. The median of 16S rRNA gene copies across all sampling sites is shown; data are corrected for different 16S rRNA gene copy numbers in archaea (median: 1) and bacteria (6); https://rrndb.umms.med.umich.edu/). B: Gene copies per 300 cm^2 of AOA, Archaea and Bacteria, logarithmic scale. C: *AmoA* gene copies per 300 cm^2, across sampling sites (A and B). Maximum-likelihood phylogenetic tree of 16S rRNA gene sequences of AOA and non-AOA, including most prevalent and abundant ASVs of cohort A and B. Blue sequences represent ASVs retrieved from amplicon sequencing of skin samples (cohorts A, B). Values in brackets show the percentage of prevalence of the ASV in our samples (C).

***Nitrosocosmicus* signatures show longitudinal coherence and unique biogeographic patterns across body sites**

Across all eight body sites sampled (cohort A), Nitrososphaeraceae and Nitrosocosmicus signatures were mostly detected on forehead, décolleté, back and arms, indicating a skin physiology-dependent pattern of ammonia-oxidizing archaea in general (Fig. 4, 16S rRNA gene amplicons). High proportions of AOA from the head region could be confirmed quantitatively (qPCR), as highest copy numbers of the *amoA* gene (up to $4.11*10^4$ per 300 cm^2) were retrieved from the head samples (part and forehead; Fig. 3B).

Nitrosocosmicus signatures were stable longitudinally (cohort B, 1 year sampling period), and even individuals that were not Archaea positive at the initial sampling time point became positive over the course of the study period. In addition, none of the participants carrying archaea at the initial time point lost their archaeome during the year of monitoring. Seasons did not have a significant impact on the abundance of AOA (MaAsLin2, q-value = 0.95), indicating that contact with the outer environment, e.g. through contact to soil by gardening work (MaAsLin2, q-value = 0.49), did not influence the skin archaeome. Signatures assigned to Nitrosocosmicus (global mean = 0.17; global variance = 0.09; importance = 0.17) and other genera of the family *Nitrososphaeraceae* (global mean = 0.33; global variance = 0.15; importance = 0.12) achieved high importance in linear regression analysis over time supporting our assumption that AOA represent a coherent longitudinal trait on human skin (Supplementary Table S4). Further information on bacterial ammonia oxidizers can be found in the Supplementary Information.

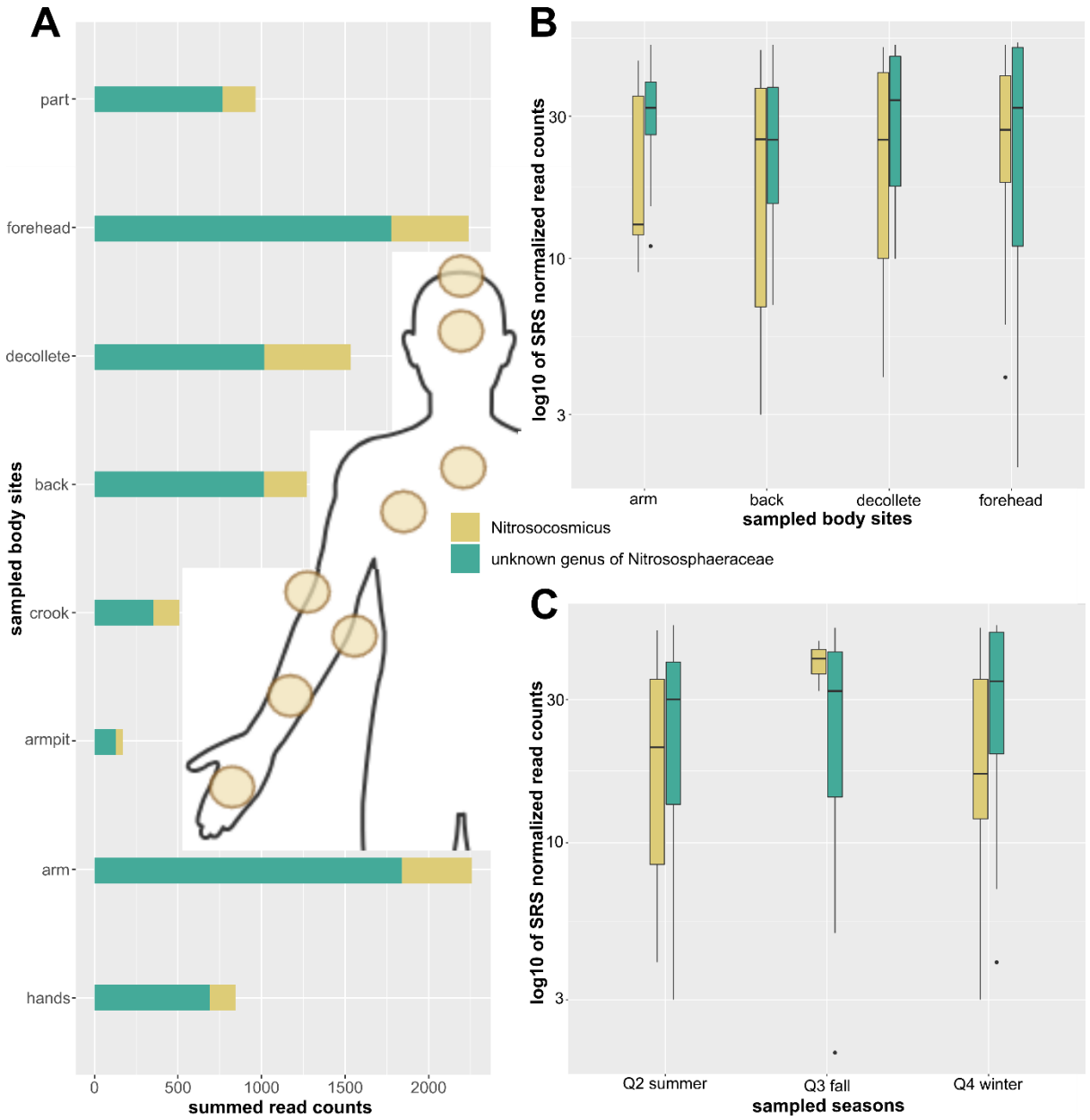


Fig. 4. | AOA prevalence across all samples from different body sites based on 16S rRNA gene sequences, shown by summed read counts (n=326 samples). The reads obtained were assigned to Nitrosocosmicus and other non-higher resolved genera belonging to the family Nitrososphaeraceae (A). Boxplot showing the abundance of AOA (normalized read counts) per body site (B). Boxplot showing AOA dynamics (normalized read counts) across the three covered seasons. Both B) and C) refer to the data of the longitudinal cohort (1 year, 6 time points, 3 seasons, 4 body sites, 283 samples) (C).

AOA are an integral component of the healthy human skin microbiome

Network analyses were performed to understand the integration of AOA into the skin microbiome. These showed a significant positive correlation of Nitrosocosmicus with integral bacterial members of the human skin microbiome (in particular *Lawsonella* and *Finegoldia* SCNIC sparCC p-adjust 0.26; see also initial in silico metabolic modeling in Extended Data Fig. 5). In contrast to other microbiome components, this interaction was relatively stable across all sampling time points (Fig. 5A, as visualized by the dark color of the node, see legend). A specific integration of Nitrosocosmicus into the skin microbiome network was underscored by the lack of positive correlations of other archaeal signatures that are not common on human skin (e.g., *Methanobrevibacter*).

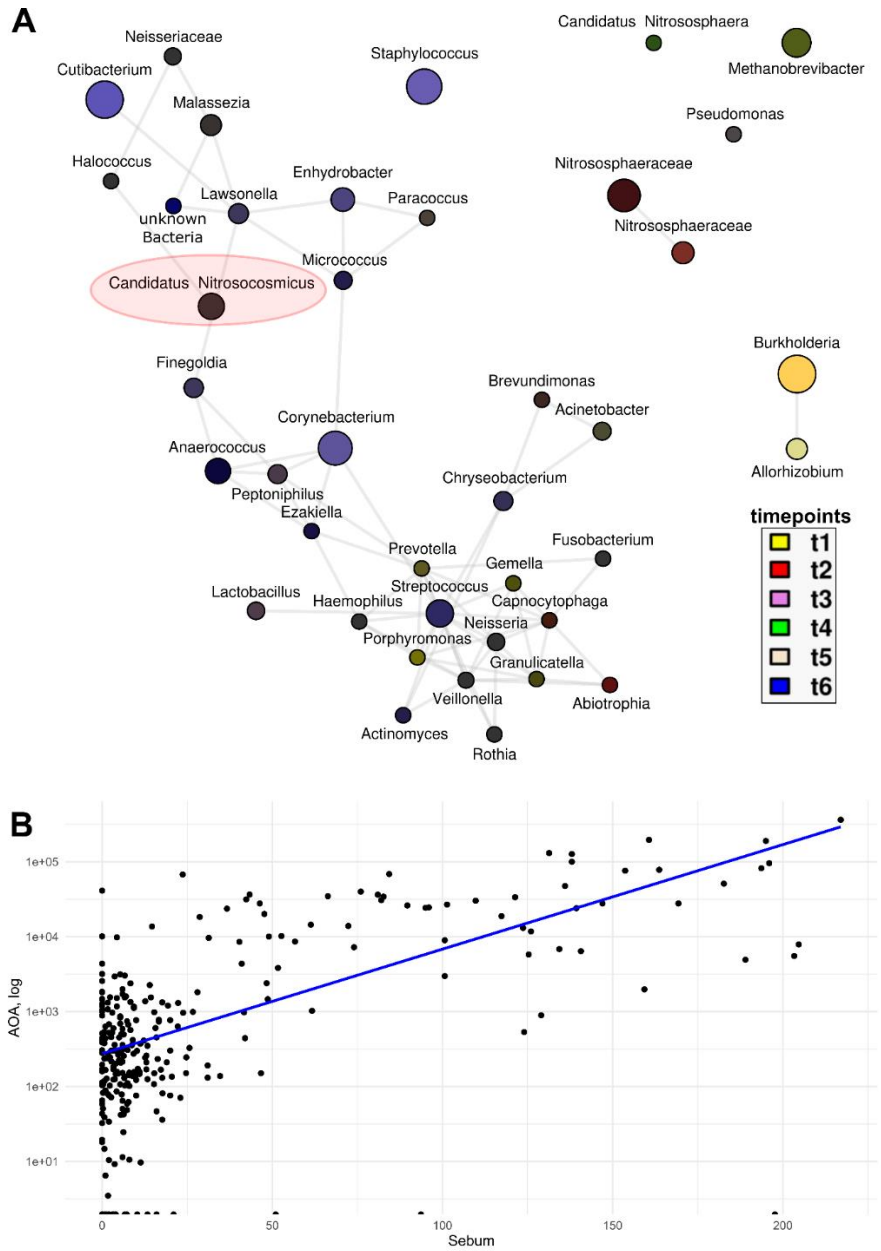


Fig. 5 | Microbial co-occurrence network between the 40 most abundant archaeal and bacterial taxa (SRS normalized Cmin = 100, TSS and log transformed). Significant positive associations from Spearman correlations are highlighted by connecting edges. Taxa (nodes) were colored according to each sampling event (t1 yellow, t2 red, t3 pink, t4 green, t5 beige, t6 blue), resulting in dark gray colors for longitudinally consistent taxa in the dataset. The location of Nitrosocosmicus within the network is highlighted by a red ellipse (A). Spearman correlation plot of AOA signatures and sebum content of the skin (B).

Ordinary least squares (OLS) regression models were used to find associations between archaeal taxa and acquired metadata (skin physiology, demographics of the cohort and their lifestyle as well as health parameters). Considering the compositional nature of our amplicon datasets we used CLR (center log ratio) transformed values and regressed them with all numerical and categorical metadata categories while taking care of several potential confounders (which partly had significant impact on the microbial communities) like a subject's sex, age, BMI, different sampled body sites and the overall microbial diversity (Shannon diversity). According to this analysis, signatures of Nitrososphaeraceae (n = 156) were significantly positively associated (q = 0.02) with qPCR-retrieved quantity for AOA (Supplementary Table S5). This finding supports the congruence of amplicon-based qualitative and qPCR-based quantitative molecular analyses.

The abundance of *amoA* genes as detected by qPCR (tp1, n=48) was not significantly influenced by sex (female vs. male) and age group (group A1 (20-30y) and A2 (>60y)). However, the abundance of amoA signatures correlated with lower pH (rho= -0.081, p=0.174), and significantly with age (rho=-0.16, $p=7.7*10^{-3}$), trans-epidermal water-loss ("Tewameter", rho=0.22, $p=4.8*10^{-4}$), and particularly sebum concentration (rho=0.47, $p=1.0*10^{-16}$) according to Spearman rank correlations.

To assess the possibility whether AOA are involved in psoriatic disease, as hypothesized earlier [16], 20 subjects were recruited and n=72 samples from lesioned and healthy skin samples were taken. However, according to differential abundance analysis based on ANCOM-BC (p adjust = 0.07) and MaAsLin2 (q-value = 0.54) samples from psoriatic skin did not reveal significantly higher or lower AOA counts.

Based on our observations, we can conclude, that Nitrosocosmicus is an integral and stable component of the microbial epidermal skin microbiome that is associated with relevant bacteria via the exchange of ammonia and is characteristic of healthy skin with higher sebum content, less moisture and lower pH.

Discussion

In this study we demonstrate the consistent occurrence of ammonia oxidizing archaea of the genus Nitrosocosmicus by molecular and cultivation techniques on human skin indicating that organisms of this taxon form an integral component of the skin microbiome. Our evidences include:

(i) high prevalence and specific biogeographic patterns in molecular analyses

(ii) correlation with specific bacterial taxa in correlation networks

(iii) independent enrichments of Nitrosocosmicus strains from human samples, and

(iv) adaptations reflected in their genomes.

Based on their central energy metabolism, the ammonia oxidizing archaea on skin should be involved in removing nitrogenous compounds such as urea and NH_3 that are excreted with human sweat [53], thereby lowering pH [17]. Considering that one of their metabolic byproducts is NO and that the final product of ammonia oxidation is nitrite (NO_2^-), they could also be key players in maintaining NOx homeostasis in the human skin. Since NO is one of the most important signaling molecules in mammalian physiology, and is involved in a range of physiological as well as pathological processes on the human skin, reviewed in [54] and [55] , AOA could play a role in skin health or physiology.

Although the overall abundance was relatively low (0.2% of the total microbial signatures in our studies) and our conditions for initial enrichment of AOA were broad enough to include other AOA taxa (such as *Nitrososphaera, Nitrosotalea, Nitrosopumilus*), representatives of the genus Nitrosocosmicus were consistently detected by molecular methods and were the only ones obtained in laboratory cultures. We hypothesize that the actual relative abundance of Nitrosocosmicus could be even higher, as isolation of DNA from Nitrosocosmicus requires harsh cell lysis. However, AOA might also be notoriously low in abundance on the human body due to hygiene practices or clothing preferences since we found sensitivity of AOA to various plastic compounds.

The overall physiology of the genus shows a high congruence with life on the human skin, as Nitrosocosmicus strains exhibit the highest ammonia tolerance among AOA, tolerating between 20 to 200mM of NH_4^+ [33]. This coincides with the elevated concentrations of ammonia (0.6 - 5.8 mM) and urea (22 ± 8 mM) present in human sweat ([56,57] ; Table 2.6). Further, among characterized AOA clades, members of Nitrosocosmicus encode the largest repertoire of putative biofilm-forming genes [32,58], and exhibit strong potential for surface attachment [30–33,59], a general characteristic of the skin microbiota, and a desired trait for skin probiotics [60].

It is also intriguing that the genus Nitrosocosmicus is the most abundant AOA taxon recovered from various plant tissues and was shown to trigger systemic resistance against pathogens [61,62]. Based on these observations, we can hypothesize that the Nitrosocosmicus genus among the AOA is particularly suited to colonize a wide range of habitats, including host surfaces and thus could represent a group of emerging commensals. Parallels with representatives of the archaeal order Methanomassiliicoccales are obvious, whose representatives can be found in diverse ecosystems, with specific adaptations that allow them to thrive in host microbiomes of the gastrointestinal tract [63,64].

It should be noted, that oscillatory behavior of microorganisms, i.e. thriving in natural environments but also host systems is not rare, however, has been mostly studied in (opportunistic) pathogens, such as *Vibrio sp.* ([65]; aquatic environments), *Stenotrophomonas maltophilia* ([66], soil, plant environments), or *Bukholderia pseudomallei* [67].

Further studies of skin archaea will be needed to reveal their evolutionary paths towards commensalism and to elucidate their roles in the skin microbiome and human health. The cultures obtained in this study are crucial prerequisites for advancing our knowledge in this area.

Methods

Ethics and recruitment of cohorts

Subjects for microbiome analyses were recruited as part of the project "Detection of Archaea on Healthy and Diseased Skin" with the ethical vote EK: 27-289 ex 14/15 (Medical University of Graz) and all ethical requirements were followed and fulfilled. The following exclusion criteria were taken into account: Antibiotics therapy within the last three months, pregnancy, ongoing breast-feeding and non-compliance. The participants were asked to not shower or apply skin cream the day before sampling.

Initial cultivation

Prior to enrichment setups, subjects not working in an experimental laboratory were pre-screened for the presence of *Nitrososphaerales* by targeted 16S rRNA gene amplification (see study cohorts below). Inoccula for enrichment cultures were taken from 3 individuals that showed highest presence of AOA. Sterile polyurethane single sponge swabs (BD BBL™ CultureSwab™ EZ) premoistened with sterile freshwater medium (FWM 8,9) were used to sequentially rub the face, neck, upper body, back, arms, and legs. The polyurethane swab was then removed from the plastic shaft and directly added to FWM (Enrichments T1S, X2B and Z3A). For enrichment R2S, free edges of a subjects' fingernails (two to three clippings each) were directly added to FWM (Enrichment R). Earlier studies had shown the suitability to use fingernail microbiomes to assess the skin microbiome, due to scratching [68].

All initial enrichments were set up in sterile 30 ml polystyrene screw cap tubes (Greiner) containing 20 ml FWM supplemented with 0.5 mM NH4Cl, 20 µl non-chelated trace element solution, 7.5 µM FeNaEDTA, 20 µl vitamin solution as well as 25 µg/ml carbenicillin and 25 µg/ml kanamycin (for detailed medium composition see table 1 in [28,69].

Variations between initial enrichments included the carbon source (2 mM Na_2CO_3 or $NaHCO_3$), pH (7 or 8.5), temperature (28°C or 32°C) supplementation with organic substrates (0.5 mM acetate and 0.5 mM pyruvate or none) and filter sterilized spent growth medium of *Nitrososphaera viennensis* EN76 [28] and Ca. Nitrosocosmicus franklandianus C13 [31] (2.5% (v/v) (Supplementary Table S6). Cultures were incubated aerobically without shaking in the dark and routinely transferred (transfer volume 20% v/v) upon reaching nitrite concentrations of approximately 450 µM. After transferring enrichments R2S and T1S into a medium containing 10 µg/ml chloramphenicol, 10 µg/ml novobiocin and 10 µg/ml streptomycin twice, no bacterial contaminants were detected.

Routine culture maintenance

Enrichments were routinely cultivated in sterile 30 ml polystyrene screw cap tubes containing 18 ml FWM supplemented with 0.6 mM NH_4Cl, 2 mM Na_2CO_3 (R) or $NaHCO_3$ (T1S, X2B, or Z3A), 20 µl non-chelated trace element solution without molybdenum ($Na_2MoO_4 \cdot 2H_2O$), 7.5 µM FeNaEDTA, 20 µl vitamin solution as well as 25 µg/ml carbenicillin and kanamycin and are transferred every two to four weeks upon reaching nitrite concentrations of approximately 450 µM. The inoculation volumes were decreased to 10% (v/v).

The medium of enrichment R2S additionally contained 10 mM MOPS ((3-(N-morpholino) propanesulfonic acid), titrated to pH 7 with HCl) and had a final pH of 7. Ammonium and nitrite concentrations were measured colorimetrically by the indophenol method and Griess reagent respectively, as described previously [28,69]. Aliquots used for ammonium measurements were frozen at -20°C and measured jointly at a later time point. Cultures were incubated aerobically at 28°C without shaking in the dark. Solutions were autoclaved (FWM, NH_4Cl) or filter sterilized with a 0.2 µm filter (trace element solution, vitamin solution, FeNaEDTA, antibiotics) prior to use (Extended Data Fig. 6: Enrichment Scheme).

Microscopy

Cultures were 10x concentrated before being imaged with phase contrast light microscopy using 60x and 100x oil immersion objectives. F_{420} autofluorescence was imaged using an F420 filter cube. For scanning electron microscopy six ml of each enrichment at late exponential phase were concentrated by centrifugation (21000 g, 4°C, 30 min). Pellets were fixed in 2.5% glutaraldehyde for 2 h, washed three times with 1x PBS and resuspended in 100 µl 1x PBS. 50 µl of the fixed cell suspension were spotted onto 0.01% poly-L-Lysine coated glass slides (5 mm) and sedimented for 2 h, before being dehydrated in an ethanol series (30-100%, 5 min each). The dehydrated cells were critical point dried in liquid carbon dioxide and subsequently coated with gold in a sputter coater. Samples were imaged in a JEOL IT 300 scanning electron microscope at 20 kV (CIUS imaging facility, Univ. of Vienna).

DNA extraction of cultures

High molecular weight (HMW) DNA of enrichment cultures was extracted by a combination of chemical lysis based on a modified lysis buffer for *Sulfolobus acidocaldarius* [70] and freeze-thawing, followed by a phenol chloroform extraction.
Between 0.3 to 0.5 L of each skin enrichment culture were concentrated by filtration (0.22 µm MF-Millipore™ mixed cellulose ester (MCE) membrane filter), washed off the filter with FWM and pelleted by centrifugation (21000 g, 4°C, 30 min). Pellets were resuspended in a modified TENST buffer (20 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl, 5% N-Lauroylsarcosine, 0.12% Triton X-100) and incubated at 50°C for 1-2 hours. Samples were subsequently subjected to three rounds of freeze-thawing (5 min at -70°C followed by 5 min at 70°C). Proteinase K was added to a final concentration of 100 µg/ml and samples were incubated at 65°C for 1-2 hours. After lysis HMW DNA was extracted by a standard phenol chloroform extraction (2x Phenol:Chloroform:Isoamylalcohol 25:24:1, 2x Chloroform:Isoamylalcohol 24:1) with low-speed centrifugation (4500 g, 4°C, 3 min) between steps. DNA was precipitated overnight at 4°C by adding 1µl glycogen as a carrier and 2x volume of a PEG6000 solution (30% PEG6000 (v/v), 1.6 M NaCl). The precipitated DNA was washed twice with 4°C cold 70% EtOH, air-dried at room temperature for 15-30 min, resuspended in 50-60°C nuclease free water followed by an incubation at 55°C for up to two hours. The extracted HMW DNA was quantified fluorometrically with a Qubit™ 2.0 Fluorometer (Invitrogen) and stored at -20°C.

Quantitative PCR

Archaeal *amo*A genes were quantified in triplicate using primers amoA19F and amoA629R as previously [29,30] with some modifications: 20 µl reactions contained 10 µl Luna® Universal qPCR Master Mix, 3 µl nuclease free water, 0.5 µm of each primer and 5 µl DNA template. DNA was amplified in a CFX Real Time PCR System (Bio-Rad) with following cycling

parameters: 95°C for 1 min, 40x cycles of 95°C 15 s denaturation, 60°C 30 s joint annealing-extension and 60°C 30 s fluorescence measurement. Standard dilutions prepared from DNA of each enrichment respectively using primers amoA19F and amoA629R ranged between 10^2 and 10^8 copies µl $^{-1}$. The efficiency of qPCR assays was between 86.06 and 93.04% with an R^2 of ≥ 0.996.

Sanger and NGS sequencing of marker genes and genomes

In order to confirm purity and identity of cultures, 16S rRNA genes were amplified by PCR with the archaea specific primer A109f [71], the universal primer Univ1492R [72], the crenarchaea specific primer pair 771F/957R [73] and the fungal ITS region-specific primer pair ITS1F and TW13 [74]. Amplification of the *amoA* genes was performed by PCR using the primers amoA19F[75] and amoA629R[76]. PCR products were purified using the Monarch® PCR & DNA Cleanup Kit from New England Biolabs before being sequenced.

Genome sequencing, assembly and comparative genomics

DNA of our enrichments was shotgun sequenced using the NovaSeq 6000 (paired-end, 150 bp) platform at Novogene. Reads were trimmed and Illumina adapters were removed using Trimmomatic (SLIDINGWINDOW:5:20 LEADING:5 TRAILING:5 MINLEN:50 HEADCROP:6) [77]. To obtain the full genomic sequences of the enriched archaeal communities from the selected cultures, Nanopore sequencing on MinION Mk1C (Oxford Nanopore Technologies plc., UK) was performed. For library preparation, extracted genomic DNA was repaired with the NEBNext Companion Module (New England Biolabs GmbH, GER), and then prepared for sequencing on a chemistry version 14 flow cell (R10.4.1, FLO-MIN114) following the Ligation sequencing gDNA – Native Barcoding Kit 24 V14 (SQK-NBD114.24) and flow cell (FLO-MIN114) protocols.

Nanopore sequencing was performed according to the following configurations: run-time of the MinION Mk1C device was set for 72 hours, with a pore scan frequency of 1.5 hours, a minimum read length of 200 bp, high-accuracy base calling, and active channel selection, reserved pores, read splitting, trimming barcodes and mid-read barcode filtering enabled. The following software versions were used: MinKNOW v22.10.7, Bream v7.3.5, Configuration v5.3.8, Guppy v6.3.9 and MinKNOW Core v5.3.1. Resulting sequencing data was analyzed with the following specifications: First, simplex base calling was carried out on a GPU node at the Life Science Compute Cluster (LiSC) University of Vienna (Austria), using guppy-gpu v6.4.2 and the dna_r10.4.1_e8.2_260bps_sup.cfg model for superior base calling. Secondly, duplex base calling was achieved at a GPU Node at the LiSC University of Vienna (Austria), following the recommended tutorial on: https://github.com/nanoporetech/duplex-tools using dorado v0.1.1 (https://github.com/nanoporetech/dorado) with the dna_r10.4.1_e8.2_260bps_fast@v4.0.0 fast model , followed by finding duplex pairs with duplextools v0.2.20-3.10 (https://github.com/nanoporetech/duplex-tools) and dorado stereo/duplex base calling with the dna_r10.4.1_e8.2_260bps_sup@v4.0.0 superior model. Merged simplex and duplex reads were then investigated by performing quality control using NanoPlot nanocomp v1.20.0 [78] followed by filtering with filtlong v0.2.1 (https://github.com/rrwick/Filtlong) (--min-length 200; --keep_percent 90, and --target_bases 20 million). Filtered reads were then assembled using flye v2.9.1 in --nano_hq and --meta mode with a target genome size of 3 Mbpm [79]. Reads were mapped with minimap2 v2.24 [80], and then polished with racon v1.5.0 (https://github.com/isovic/racon) using the following

settings (-m -8 -x -6 -g -8 -w -500). Finally, consensus contigs were obtained with medaka v1.7.2 according to the r1041_e82_260bps_sup_g632 model (https://github.com/nanoporetech/medaka). Basic genome statistics were calculated with genometools by calling gt seqstat [81], and classified the genomes using gtdbtk v2.1.1 with the --full_tree option enabled [82]. Genome completeness and contamination were estimated with checkm2 v1.0.1 including the --allmodels option [83]. After genome comparison using drep v3.4.2 [84], genomes were finally annotated with eggnog-mapper v2.1.9 [85–90] including prodigal for gene identification. Marker genes were identified with Metaxa2 2.2 [91] and the remaining contaminants of enrichments were estimated with kraken2 2.1.2 and bracken 2.8 according to the PlusPFP database [92,93] (https://benlangmead.github.io/aws-indexes/k2).

Gene prediction and annotation of the MAGs assembled in this study was performed automatically using the Microscope annotation platform from Genoscope[94], followed by extensive manual curation. Annotations were supplemented with arCOG assignments from the archaeal Clusters of Orthologous Genes database (2018 release) [95] , Carbohydrate-active enzymes assignments (CAZymes) from dbCAN2 (v7.0) [96] and TCDB family assignments from the Transporter Classification Database [97], using scripts from [98].

Orthologous protein families from a dataset including *Ca*. N. franklandianus C13, *Ca.* N. oleophilus MY3, *Ca.* N. hydrocola G61, *Ca.* N. agrestis SS, and *Ca.* N. arcticus Kfb and the two newly sequenced Ca. Nitrosocosmicus isolates R2S and T1S were constructed using Orthofinder (v.2.5.4) with standard settings [99].

Study-cohorts for molecular-based analyses

Forty-seven healthy individuals (group A1: female=11, male=10, age=20-40y; group A2: f=15, m=11, age=60-85) were recruited for an in-depth analysis of their skin microbiome, and a subgroup of high vs. low skin archaeal carriers was selected based on quantitative (amoA gene and 16S rRNA gene qPCR universal and archaea specific primers) and qualitative (16S rRNA gene amplicons universal and archaea specific primers) observations for longitudinal investigations (group B1 (from A1): f=2, m=4; group B2 (from A2): f=5, m=1) over the period of almost two years (6 time points). The sampling sites for A1/A2 were hand, outer arm, crook, armpit, forehead, part, décolleté and back (n=7), and for B1/B2: forehead, décolleté, arms and back (n=4). Supplementary Table S7 gives an overview on the demographic characteristics of all recruited subjects.
Each subject completed a questionnaire with different categories, e.g., height and weight, skin type, and detailed questions about their respective lifestyle (e.g., eating habits, smoking, alcohol consumption, skin care, medications, animal contact, and hobbies).

Skin measurements and sampling

The forehead and the part, as well as both hands, forearms, crooks and armpits were swabbed with pre-moistened (0.9 % (v/v) NaCl) swabs (BD BBL ™ Culture Swab ™ EZ; Becton Dickinson, USA). NaCl was baked at 260°C for 24 h before solving in PCR-grade water to remove DNA traces. The remaining body sites, including décolleté and back (approximately 300 cm^2), were sampled with a pre-moistened surface-sampling cellulose-sponge kit (VWR, 300-0229P). For the longitudinal sampling (cohort B1/B2, tp2-6), only the forehead was sampled with the swab. After skin sampling, samples were immediately frozen at -80°C until further processing. All skin sampling events were accomplished in a lab that never handled AOA.

Measurements of the skin physiology was performed using the Cutometer(R) MPA580 (Courage + Khazaka, Germany), using the Tewameter (to determine the barrier function by measuring the water loss; 30s, 3 times), the pH-meter, the Sebumeter (to determine the sebum-content, 30s), and the Corneometer (to determine skin moisture). Each skin site was measured three times and the mean of the measurements is reported in Supplementary Tables S8 and S9. Further, the surface temperature of each region was determined using an infrared thermometer (Future Founder, China).

The interpretation of skin parameters was done based on the recommended specifications by the manufacturer, which are provided in Supplementary Table S10.

DNA extraction, controls, amplicon sequencing, and qPCR of skin samples

The DNA was extracted from swabs and sponges using the Purelink ™ Microbiome DNA Purification Kit (Rectal or Environmental Swab Samples; Thermo Fisher Scientific, USA). For the extraction from sponges (liquid sample extract from décolleté and back) one ml of the sample was concentrated by centrifugation (20817 x g, 4°C; 10 min). The largest proportion of the supernatant was removed, and the remaining 70 µl was used for bead beating with 800 µl of the S1 lysis buffer. Material and process controls were included in each step to control for the DNA contamination of the used reagents. The DNA was stored at -20°C until further downstream analyses.
Various targets were assessed using amplicon-based analyses: i) the archaeal 16S rRNA gene region (nested PCR approach as described in [100] and [16]), the ii) amoA targeted approach (described in [28,101,102]), and the iii) “universal” approach, targeting the microbial 16S rRNA gene region [103]. All primer sequences and PCR protocols are provided in Supplementary Table S11. Library construction and next-generation sequencing (Illumina MiSeq®) was performed at the Core Facility Molecular Biology at the Center for Medical Research (Medical University of Graz, Austria). In a first step, the DNA concentrations of the generated amplicons were normalized using a SequalPrepTM normalization plate (Invitrogen) and subsequently each sample was indexed with a unique barcode sequence by 8 cycles of indexing PCR. These indexed Samples were pooled and purified by gel cuts before the library was run on an Illumina MiSeq® instrument and the MiSeq® Reagent Kit v3 with 602 cycles (2 x 301 cycles).

Bioinformatics and data analysis

Resulting demultiplexed fastq reads were processed with QIIME2 [104] versions 2018.11 through 2023.2. After quality control, denoising of reads was achieved with DADA2 [105] and truncation of forward reads at a length of 230bp and 120bp for reverse reads. Potential contaminants were identified with decontam v.1.12 [106] using the prevalence mode with a threshold of 0.5 and removed from downstream analysis. Representative sequences were classified by a pre-trained Naïve-Bayes classifier based on the curated and trimmed SILVA138 database [107] using RESCRIPt [108] and the feature-classifier classify-sklearn plugin [109]. For phylogenetic metrics like UniFrac [110], a rooted phylogenetic tree was generated with FastTree2 [111] based on a masked MAFFT alignment. Data normalization was achieved by scaling with ranked subsampling and the q2-srs QIIME2 plugin [112] at Cmin 50, 100 and 1000 followed by calculating core metrics for alpha and beta diversity. Feature differentials were determined with the QIIME2 plugins aldex2 [113], ANCOM [114], ANCOM-BC [114], and the R packages ANCOM2 (https://github.com/FrederickHuangLin/ANCOM-Code-Archive) and MaAsLin2 [115] including time as a fixed effect and different subjects as a random effect where applicable,

while feature loadings were determined with the QIIME2 plugin DEICODE [116]. Longitudinal analysis covering feature volatility over time as well as other numerical metadata categories like age, BMI or skin physiology measurements including i.e. pH, skin fat or skin wetness were realized with the q2-longitudinal plugin in QIIME2 [109]. The same plugin was used to calculate linear-mixed-effect models of Shannon diversity and weighted UniFrac distances along principal coordinate axis 1. Amplicons of the amoA marker gene were processed in a similar way without read truncations for DADA2 and read classifications by a pre-trained Naïve-Bayes classifier based on the amoA reference database [37,117]. Connection of 16S rRNA to amoA gene information was achieved by following the tutorial at https://github.com/alex-bagnoud/Linking-16S-to-amoA-taxonomy/. Prediction of metadata categories were realized with the q2-sample-classifier QIIME2 plugin [109].

Custom python code for regressing all metadata categories:
To identify potential associations between taxa and recorded numerical and categorical metadata an ordinary least squares (OLS) regression fit method was used with python's statsmodels smf formula (statsmodels.regression.linear_model.OLS - statsmodels 0.15.0 (+8)). Microbial counts of the different investigated cohorts were first clr transformed and filtered for less than 25% missing values. These values were then used as the response variable in dependence of each individual, their sex, age, BMI, body site, the overall microbial diversity indicated by the Shannon index (H'), and where applicable the time of sampling. The model was then evaluated for each genus and recorded metadata according to their q-values (FDR-corrected p-values) and estimated regression coefficients (python code available at GitHub: https://github.com/CME-lab-research/SkinArchaeome).

Network analysis:
Network analyses were conducted to identify mutually exclusive and co-occurring clusters of archaea and bacteria in our merged (archaea-specific and universal) 16S rRNA gene amplicon dataset. After SRS normalization to Cmin = 100 and TSS and log transformations the 40 most abundant taxa were represented as nodes within the network where node size reflects taxa abundance, and edges indicate only positive associations of Spearman correlations between pairs of taxa. Then, we used PCoA to ordinate the nodes in a two-dimensional plot, such that correlating nodes were close together and anti-correlating nodes were far apart. In addition, correlations of the network were further supported by the Sparse Co-Occurrence Network Investigation for Compositional data (SCNIC) package (https://github.com/lozuponelab/SCNIC) using the sparCC metric with default settings for filterings and the R value, and bootstrap adjusted p-values.

Phylogenomic tree reconstruction:
A genome database was created comprising the two AOA genomes presented in this study together with 141 MAGs and completely sequenced genomes (133 AOA and 8 non-AOA genomes) from IMG, DDBJ, or NCBI databases followed by protein prediction using Prodigal v2.6.3 [90]. The workflow proposed by Graham et al. (2018) [101] based on the archaeal single-copy gene collection [118] was followed to identify phylogenetic markers (e-value 10^{-10}) to reconstruct the phylogenomic tree. A total of 36 ribosomal protein families encoded in at least 100 out of the 143 genomes present in our genome database were selected. Each protein family was aligned independently using MAFFT 7.520 [119] (L-INS-i algorithm) followed by the trimming of the alignments using BMGE [120] with default parameters. Trimmed protein families were concatenated using a custom python script and the concatenated alignment was used

to reconstruct a maximum likelihood (ML) phylogenomic tree in IQ-TREE v2.2.2.7 [121] under the LG+C20+F+G model with 1000 ultrafast bootstrap replicates.

Phylogenetic analysis of 16S rRNA sequences:
The identification of 16S rRNA sequences in our AOA genome database (see above) was performed using Barrnap v0.9 (https://github.com/tseemann/barrnap). The 16S rRNA sequence encoded in the MAG GCA_025930575.1 (NS-Epsilon) was used as a query in a BLASTN search against the NCBI database to improve the sequence representation of this sub lineage with environmental sequences. Sequences were aligned with the SINA v1.7.2 aligner [122] using the SILVA database SSU Ref NR 99 release 138.1 [107] with default parameters followed by the alignment of the skin ASVs using MAFFT 7.520 [119] (--addfragments option, L-INS-i algorithm). The alignment was trimmed manually using AliView [123]. The maximum likelihood phylogenetic tree was reconstructed using IQ-TREE v2.2.2.7 [121] with the best-fit model TIM2e+I+R3 chosen by ModelFinder ("-m MFP") [124] according to BIC (Bayesian Information criterion), SH-like approximate likelihood ratio test ("-alrt 1000") [125] and 1,000 ultrafast bootstrap replicates ("-bb 1000").

qPCR specific methods

A SYBR green-based quantitative real-time PCR (qPCR) approach targeting the archaeal amoA gene (19F-629R) and the 16S rRNA gene of archaea (A806F-A958R) and bacteria (Bac_331F-Bac_797R) was used to determine the abundance of AOA, all archaea and all bacteria on selected body sites. Primers, reaction mixtures and thermocycler programs are listed in Supplementary Table S11. All qPCRs followed the MIQE Guidelines.

qPCRs were performed with the Bio-Rad SsoAdvanced™ Universal SYBR® Green Supermix. Each reaction contained 1µl of DNA template and 9µl of mastermix. The mastermix for the archaeal amoA approach consisted of 5µl SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad), 1.25µl of each primer (1.25µM) and 1.5µl of PCR grade water. The mastermix for the archaeal and bacterial 16S rRNA gene approach consisted of 5µl SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad), 0.3µl of each primer (10µM) and 3.4µl of PCR grade water. Amplifications were carried out in a Bio-Rad CFX96 Touch™ Real-Time PCR Detection System. The qPCR cycling conditions were as follows: initial denaturation at 95°C for 1.5min, followed by 45 cycles of denaturation for 15sec at 95°C, annealing for 30sec at 60°C, elongation for 40sec at 72°C, a final denaturation for 15sec at 95°C, 5sec at 60°C and a storage temperature of 8°C. A plasmid containing the amoA gene of *Nitrososphaera viennensis* with a starting concentration of 104.1ng/µl (pscA_Ampkan_AmoA, clone 3) was used as a quantification standard in a 1:10 dilution series from 10^{-2} to 10^{-10}. Product specificity was confirmed by melting curve analysis. All qPCR experiments were performed in triplicates and the results were averaged for further calculations. Data from a previous experiment on the total archaeal community (based on qPCR of the 16S rRNA gene) was also included in the analyses. To account for the different area sizes of the sampled body sites, copy numbers were adjusted to 300 cm^2 (amoA gene copies per 300 cm^2; see Supplementary Table S12). The qPCR files were analyzed using the CFX Manager™ Software, version 3.1 and processed with the R software, version 3.6.1 and Microsoft Excel.

Data availability

All raw sequencing data were submitted to NCBI's Sequence Read Archive (SRA) https://www.ncbi.nlm.nih.gov/sra and are accessible under the following projects: PRJNA1105887 (archaea specific 16S rRNA gene amplicons of cohort A), PRJNA1107153 (universal 16S rRNA gene amplicons of cohort A), PRJNA1107462 (AOA specific amoA gene amplicons of cohort A), PRJNA1107502 (archaea specific 16S rRNA gene amplicons of cohort B), PRJNA1107512 (universal 16S rRNA gene amplicons of cohort B), PRJNA1105887 (archaea specific 16S rRNA gene amplicons of cohort C), PRJNA1108073 (universal 16S rRNA gene amplicons of cohort C), PRJNA1108908 (pooled shotgun metagenomics data of cohort A), PRJNA1108960 (Illumina sequencing data of enrichments), PRJNA1109334 (Nanopore sequencing data of enrichments). Assemblies are publicly available under NCBI BioProjects PRJNA1103140 (Ca. Nitrosocosmicus epidermidis T1S) and PRJNA1103141 (Ca. Nitrosocosmicus unguis). Annotated assemblies are also available at the MicroScope Microbial Genome Annotation and Analysis Platform, developed by the Laboratory of Bioinformatics Analyses for Genomics and Metabolism, Genoscope, Evry, France.

Code availability

The present study did not generate new code, and mentioned tools used for the data analysis were applied with default parameters unless specified otherwise in Methods and our Github repo.

Contributions

AM supervised and performed microbiome sampling and analysis. MD supervised and performed cultivation and microscopy. MM and LH supervised and supported cultivation work. UY performed cultivation. SD, AL, TG, DB performed sampling, sequencing, and data analysis of qPCR and amplicon data. RP performed phylogenetic analyses and genome assembly. TG and PW recruited patients and performed sampling. MK performed genome annotations and contributed to study concept. CME and CS developed the study concept, wrote initial proposal, and wrote, together with MK and AM the manuscript. All authors contributed to the finalization of the manuscript.

Acknowledgments

We acknowledge the technical support by Ignaz Herzeg, Lisa Wink, Polona Mertelj, Sarah Harrer and all collaborators at the Department of Dermatology and Venereology, Medical University of Graz (Andrea Alejandra Catalan Griffiths, Patra Vijaykumar, Isabella Perchthaler, Alexandra Gruber-Wackernagel, Wolfgang Weger, and univ. Prof. Dr. Wolfgang Salmhofer).
The work was funded by the Austrian Science Fund FWF Austrian Science Fund (FWF) grant no. P 30796 (given to CME and CS), and supported by FWF grants P 32697, P 30796, SFB F-83, COE 7, given to CME and FWF Z437 to CS.
We thank Daniela Gruber from the Core Facility Cell Imaging & Ultrastructure Research (LeWi).
We greatly value the computational resources of the MedBioNode provided by the Medical University of Graz and the Life Science Compute Cluster (LiSC) operated by the Computational Systems Biology group at the University of Vienna as well as the Medical University of Graz ZMF Galaxy Team: Core Facility Computational Bioanalytics, Medical

University of Graz, funded by the Austrian Federal Ministry of Education, Science and Research, Hochschulraum-Strukturmittel 2016 grant as part of BioTechMed Graz.
SD was supported by the local PhD program MolMed.

Conflict of interest

The authors declare no conflict of interest.

References

1. Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K. & Corral, G. H. Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8,** 1–22 (2020).
2. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology 2020 19:1* **19,** 55–71 (2020). doi: 10.1038/s41579-020-0433-9
3. Flowers, L. & Grice, E. A. The Skin Microbiota: Balancing Risk and Reward. (2020). doi:10.1016/j.chom.2020.06.017 doi: 10.1016/j.chom.2020.06.017
4. Gomaa, E. Z. Human gut microbiota/microbiome in health and diseases: a review. *Antonie Van Leeuwenhoek* **113,** 2019–2040 (2020). doi: 10.1007/s10482-020-01474-7
5. Borrel, G., Brugère, J. F., Gribaldo, S., Schmitz, R. A. & Moissl-Eichinger, C. The host-associated archaeome. *Nat Rev Microbiol* **18,** 622–636 (2020). doi: 10.1038/s41579-020-0407-y
6. Bang, C. & Schmitz, R. A. Archaea associated with human surfaces: not to be underestimated. *FEMS Microbiol Rev* **39,** 631–648 (2015). doi: 10.1093/femsre/fuv010
7. Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W. & Brugère, J.-F. Archaea and the human gut: New beginning of an old story. *World J Gastroenterol* **20,** 16062 (2014). doi: 10.3748/wjg.v20.i43.16062
8. Nayfach, S., Páez-Espino, D., Call, L., Low, S. J., Sberro, H., Ivanova, N. N., Proal, A. D., Fischbach, M. A., Bhatt, A. S., Hugenholtz, P. & Kyrpides, N. C. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nature Microbiology 2021 6:7* **6,** 960–970 (2021). doi: 10.1038/s41564-021-00928-6
9. Fan, Y., Wu, L. & Zhai, B. The mycobiome: interactions with host and implications in diseases. *Curr Opin Microbiol* **75,** 102361 (2023). doi: 10.1016/J.MIB.2023.102361
10. Chibani, C. M., Mahnert, A., Borrel, G., Almeida, A., Werner, A., Brugère, J.-F., Gribaldo, S., Finn, R. D., Schmitz, R. A. & Moissl-Eichinger, C. A catalogue of 1,167 genomes from the human gut archaeome. *Nat Microbiol* **7,** 48–61 (2022). doi: 10.1038/s41564-021-01020-9
11. Kumpitsch, C., Fischmeister, F. P. S., Mahnert, A., Lackner, S., Wilding, M., Sturm, C., Springer, A., Madl, T., Holasek, S. & Högenauer, C. Reduced B12 uptake and increased gastrointestinal formate are associated with archaeome-mediated breath methane emission in humans. *Microbiome* **9,** 1–18 (2021).
12. Grice, E. A. & Segre, J. A. The skin microbiome. *Nat Rev Microbiol* **9,** 244–253 (2011).
13. Kong, H. H. Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends Mol Med* **17,** 320 (2011). doi: 10.1016/J.MOLMED.2011.01.013
14. Oh, J., Byrd, A. L., Park, M., Kong, H. H. & Segre, J. A. Temporal Stability of the Human Skin Microbiome. *Cell* **165,** 854–866 doi: 10.1016/j.cell.2016.04.008
15. Byrd, A. L., Belkaid, Y. & Segre, J. A. The human skin microbiome. *Nat Rev Microbiol* **16,** 143–155 (2018). doi: 10.1038/nrmicro.2017.157
16. Probst, A. J., Auerbach, A. K. & Moissl-Eichinger, C. Archaea on Human Skin. *PLoS One* **8,** (2013). doi: 10.1371/journal.pone.0065388

17. Moissl-Eichinger, C., Probst, A. J., Birarda, G., Auerbach, A., Koskinen, K., Wolf, P. & Holman, H.-Y. N. Human age and skin physiology shape diversity and abundance of Archaea on skin. *Sci Rep* **7,** (2017). doi: 10.1038/s41598-017-04197-4
18. Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P. A. & Hugenholtz, P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **50,** D785–D794 (2022). doi: 10.1093/NAR/GKAB776
19. Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Davín, A. A., Waite, D. W., Whitman, W. B., Parks, D. H. & Hugenholtz, P. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat Microbiol* 1–14 (2021).
20. Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J. & Hugenholtz, P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **38,** 1079–1086 (2020). doi: 10.1038/s41587-020-0501-8
21. Moissl-Eichinger, C. Archaea in artificial environments: their presence in global spacecraft clean rooms and impact on planetary protection. *ISME J* **5,** 209–19 (2011).
22. Kozlowski, J. A., Kits, K. D. & Stein, L. Y. Comparison of Nitrogen Oxide Metabolism among Diverse Ammonia-Oxidizing Bacteria. *Front Microbiol* **7,** (2016). doi: 10.3389/fmicb.2016.01090
23. Sun, Y., Liu, Y., Pan, J., Wang, F. & Li, M. Perspectives on Cultivation Strategies of Archaea. *Microb Ecol* (2019). doi:10.1007/s00248-019-01422-7 doi: 10.1007/s00248-019-01422-7
24. Umbach, A. K., Stegelmeier, A. A. & Neufeld, J. D. Archaea Are Rare and Uncommon Members of the Mammalian Skin Microbiome. *mSystems* **6,** (2021). doi: 10.1128/MSYSTEMS.00642-21/ASSET/1CFB65AC-B6A0-4252-8E75-629C6EA38AC1/ASSETS/IMAGES/LARGE/MSYSTEMS.00642-21-F004.JPG
25. Mahnert, A., Blohs, M., Pausan, M. R. & Moissl-Eichinger, C. The human archaeome: methodological pitfalls and knowledge gaps. *Emerg Top Life Sci* **2.4,** 469–482 (2018).
26. Usuda, D., Higashikawa, T., Hotchi, Y., Usami, K., Shimozawa, S., Tokunaga, S., Osugi, I., Katou, R., Ito, S., Yoshizawa, T., Asako, S., Mishima, K., Kondo, A., Mizuno, K., Takami, H., Komatsu, T., Oba, J., Nomura, T. & Sugita, M. Exophiala dermatitidis. *World J Clin Cases* **9,** 7963–7972 (2021). doi: 10.12998/wjcc.v9.i27.7963
27. Untereiner, W. A. & Naveau, F. A. Molecular systematics of the Herpotrichiellaceae with an assessment of the phylogenetic positions of Exophiala dermatitidis and Phialophora americana. *Mycologia* **91,** 67–83 (1999). doi: 10.1080/00275514.1999.12060994
28. Tourna, M., Stieglmeier, M., Spang, A., Könneke, M., Schintlmeister, A., Urich, T., Engel, M., Schloter, M., Wagner, M. & Richter, A. Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences* **108,** 8420–8425 (2011).
29. Abby, S. S., Melcher, M., Kerou, M., Krupovic, M., Stieglmeier, M., Rossel, C., Pfeifer, K. & Schleper, C. Candidatus Nitrosocaldus cavascurensis, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome. *Front Microbiol* **9,** 346006 (2018). doi: 10.3389/FMICB.2018.00028/BIBTEX
30. Alves, R. J. E., Kerou, M., Zappe, A., Bittner, R., Abby, S. S., Schmidt, H. A., Pfeifer, K. & Schleper, C. Ammonia oxidation by the arctic terrestrial thaumarchaeote candidatus nitrosocosmicus arcticus is stimulated by increasing temperatures. *Front Microbiol* **10,** 1571 (2019). doi: 10.3389/FMICB.2019.01571/FULL

31. Lehtovirta-Morley, L. E., Ross, J., Hink, L., Weber, E. B., Gubry-Rangin, C., Thion, C., Prosser, J. I. & Nicol, G. W. Isolation of 'Candidatus Nitrosocosmicus franklandus', a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. *FEMS Microbiol Ecol* **92,** 57 (2016). doi: 10.1093/FEMSEC/FIW057
32. Jung, M. Y., Kim, J. G., Sinninghe Damsté, J. S., Rijpstra, W. I. C., Madsen, E. L., Kim, S. J., Hong, H., Si, O. J., Kerou, M., Schleper, C. & Rhee, S. K. A hydrophobic ammonia-oxidizing archaeon of the Nitrosocosmicus clade isolated from coal tar-contaminated sediment. *Environ Microbiol Rep* **8,** 983–992 (2016). doi: 10.1111/1758-2229.12477
33. Liu, L., Liu, M., Jiang, Y., Lin, W. & Luo, J. Production and Excretion of Polyamines To Tolerate High Ammonia, a Case Study on Soil Ammonia-Oxidizing Archaeon ' Candidatus Nitrosocosmicus agrestis'. *mSystems* **6,** (2021). doi: 10.1128/MSYSTEMS.01003-20
34. Roller, B. R. K., Stoddard, S. F. & Schmidt, T. M. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat Microbiol* **1,** 16160 (2016). doi: 10.1038/nmicrobiol.2016.160
35. Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. *Appl Environ Microbiol* **66,** 1328–1333 (2000). doi: 10.1128/AEM.66.4.1328-1333.2000
36. Konstantinidis, K. T., Rosselló-Móra, R. & Amann, R. Uncultivated microbes in need of their own taxonomy. *ISME J* **11,** 2399–2406 (2017). doi: 10.1038/ismej.2017.113
37. Alves, R. J. E., Minh, B. Q., Urich, T., Haeseler, A. & Schleper, C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat Commun* **9,** 1517 (2018).
38. van Ulsen, P., Zinner, K. M., Jong, W. S. P. & Luirink, J. On display: autotransporter secretion and application. *FEMS Microbiol Lett* **365,** (2018). doi: 10.1093/femsle/fny165
39. Kobe, B. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* **11,** 725–732 (2001). doi: 10.1016/S0959-440X(01)00266-4
40. Alvarez, C. E. On the origins of arrestin and rhodopsin. *BMC Evol Biol* **8,** 222 (2008). doi: 10.1186/1471-2148-8-222
41. Marullo, S. & Coureuil, M. Arrestins in host-pathogen interactions. *Handb Exp Pharmacol* **219,** 361–74 (2014). doi: 10.1007/978-3-642-41199-1_18
42. Chan, J. M., Gori, A., Nobbs, A. H. & Heyderman, R. S. Streptococcal Serine-Rich Repeat Proteins in Colonization and Disease. *Front Microbiol* **11,** (2020). doi: 10.3389/fmicb.2020.593356
43. Yeats, C., Bentley, S. & Bateman, A. New Knowledge from Old: In silico discovery of novel protein domains in Streptomyces coelicolor. *BMC Microbiol* **3,** 3 (2003). doi: 10.1186/1471-2180-3-3
44. Gibbs, G. M., Roelants, K. & O'Bryan, M. K. The CAP Superfamily: Cysteine-Rich Secretory Proteins, Antigen 5, and Pathogenesis-Related 1 Proteins—Roles in Reproduction, Cancer, and Immune Defense. *Endocr Rev* **29,** 865–897 (2008). doi: 10.1210/er.2008-0032
45. Chen, C. K.-M., Chan, N.-L. & Wang, A. H.-J. The many blades of the β-propeller proteins: conserved but versatile. *Trends Biochem Sci* **36,** 553–561 (2011). doi: 10.1016/j.tibs.2011.07.004

46. Brandwein, M., Steinberg, D. & Meshner, S. Microbial biofilms and the human skin microbiome. *npj Biofilms and Microbiomes 2016 2:1* **2,** 1–6 (2016). doi: 10.1038/s41522-016-0004-z
47. Cinar, M. S., Niyas, A. & Avci, F. Y. Serine-rich repeat proteins: well-known yet little-understood bacterial adhesins. *J Bacteriol* **206,** (2024). doi: 10.1128/jb.00241-23
48. Mondal, A. K., Lata, K., Singh, M., Chatterjee, S., Chauhan, A., Puravankara, S. & Chattopadhyay, K. Cryo-EM elucidates mechanism of action of bacterial pore-forming toxins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1864,** 184013 (2022). doi: 10.1016/j.bbamem.2022.184013
49. Garcia, C. J., Pericleous, A., Elsayed, M., Tran, M., Gupta, S., Callaghan, J. D., Stella, N. A., Franks, J. M., Thibodeau, P. H., Shanks, R. M. Q. & Kadouri, D. E. Serralysin family metalloproteases protects Serratia marcescens from predation by the predatory bacteria Micavibrio aeruginosavorus. *Sci Rep* **8,** 14025 (2018). doi: 10.1038/s41598-018-32330-4
50. Chua, W., Poh, S. E. & Li, H. Secretory Proteases of the Human Skin Microbiome. *Infect Immun* **90,** (2022). doi: 10.1128/IAI.00397-21
51. Kumar, N. G., Contaifer, D., Wijesinghe, D. S. & Jefferson, K. K. Staphylococcus aureus Lipase 3 (SAL3) is a surface-associated lipase that hydrolyzes short chain fatty acids. *PLoS One* **16,** e0258106 (2021). doi: 10.1371/journal.pone.0258106
52. Chen, X. & Alonzo, F. Bacterial lipolysis of immune-activating ligands promotes evasion of innate defenses. *Proceedings of the National Academy of Sciences* **116,** 3764–3773 (2019). doi: 10.1073/pnas.1817248116
53. Chen, Y.-L., Kuan, W.-H. & Liu, C.-L. Comparative Study of the Composition of Sweat from Eccrine and Apocrine Sweat Glands during Exercise and in Heat. *Int J Environ Res Public Health* **17,** 3377 (2020). doi: 10.3390/ijerph17103377
54. Cals-Grierson, M.-M. & Ormerod, A. D. Nitric oxide function in the skin. *Nitric Oxide* **10,** 179–193 (2004). doi: 10.1016/j.niox.2004.04.005
55. Suschek, C. V., Opländer, C. & van Faassen, E. E. Non-enzymatic NO production in human skin: Effect of UVA on cutaneous NO stores. *Nitric Oxide* **22,** 120–135 (2010). doi: 10.1016/j.niox.2009.10.006
56. Wilson, M. *Microbial Inhabitants of Humans: Their Ecology and Role in Health and Disease.* (Cambridge University Press, 2005).
57. Huang, C.-T., Chen, M.-L., Huang, L.-L. & Mao, I.-F. Uric acid and urea in human sweat. *Chin J Physiol* **45,** 109–15 (2002).
58. Abby, S. S., Kerou, M. & Schleper, C. Ancestral reconstructions decipher major adaptations of ammonia-oxidizing archaea upon radiation into moderate terrestrial and marine environments. *mBio* **11,** 1–20 (2020). doi: 10.1128/MBIO.02371-20/SUPPL_FILE/MBIO.02371-20-SF004.PDF
59. Sauder, L. A., Albertsen, M., Engel, K., Schwarz, J., Nielsen, P. H., Wagner, M. & Neufeld, J. D. Cultivation and characterization of Candidatus Nitrosocosmicus exaquare, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J* **11,** 1142 (2017).
60. Lopes, E. G., Moreira, D. A., Gullón, P., Gullón, B., Cardelle-Cobas, A. & Tavaria, F. K. Topical application of probiotics in skin: adhesion, antimicrobial and antibiofilm *in vitro* assays. *J Appl Microbiol* **122,** 450–461 (2017). doi: 10.1111/jam.13349
61. Song, G. C., Im, H., Jung, J., Lee, S., Jung, M., Rhee, S. & Ryu, C. Plant growth-promoting archaea trigger induced systemic resistance in Arabidopsis thaliana against

Pectobacterium carotovorum and Pseudomonas syringae. *Environ Microbiol* **21,** 940–948 (2019).

62. Taffner, J., Cernava, T., Erlacher, A. & Berg, G. Novel insights into plant-associated archaea and their functioning in arugula (Eruca sativa Mill.). *J Adv Res* (2019).

63. De La Cuesta-Zuluaga, J., Spector, T. D., Youngblut, N. D. & Ley, R. E. Genomic Insights into Adaptations of Trimethylamine-Utilizing Methanogens to Diverse Habitats, Including the Human Gut. *mSystems* **6,** e00939-20 (2021).

64. Cozannet, M., Borrel, G., Roussel, E., Moalic, Y., Allioux, M., Sanvoisin, A., Toffin, L. & Alain, K. New Insights into the Ecology and Physiology of Methanomassiliicoccales from Terrestrial and Aquatic Environments. *Microorganisms* **9,** 30 (2020). doi: 10.3390/microorganisms9010030

65. Soto, W. & Nishiguchi, M. K. Environmental Stress Selects for Innovations That Drive Vibrio Symbiont Diversity. *Front Ecol Evol* **9,** (2021). doi: 10.3389/fevo.2021.616973

66. Brooke, J. S. Stenotrophomonas maltophilia: an Emerging Global Opportunistic Pathogen. *Clin Microbiol Rev* **25,** 2–41 (2012). doi: 10.1128/CMR.00019-11

67. Ghazali, A.-K., Firdaus-Raih, M., Uthaya Kumar, A., Lee, W.-K., Hoh, C.-C. & Nathan, S. Transitioning from Soil to Host: Comparative Transcriptome Analysis Reveals the Burkholderia pseudomallei Response to Different Niches. *Microbiol Spectr* **11,** (2023). doi: 10.1128/spectrum.03835-22

68. Rayan, G. M. & Flournoy, D. J. Microbiologic flora of human fingernails. *J Hand Surg Am* **12,** 605–607 (1987). doi: 10.1016/S0363-5023(87)80217-4

69. Reyes, C., Hodgskiss, L. H., Baars, O., Kerou, M., Bayer, B., Schleper, C. & Kraemer, S. M. Copper limiting threshold in the terrestrial ammonia oxidizing archaeon Nitrososphaera viennensis. *Res Microbiol* **171,** 134–142 (2020). doi: 10.1016/J.RESMIC.2020.01.003

70. Meulenbroek, E. M., Peron Cane, C., Jala, I., Iwai, S., Moolenaar, G. F., Goosen, N. & Pannu, N. S. UV damage endonuclease employs a novel dual-dinucleotide flipping mechanism to recognize different DNA lesions. *Nucleic Acids Res* **41,** 1363–1371 (2013). doi: 10.1093/nar/gks1127

71. Großkopf, R., Janssen, P. H. & Liesack, W. Diversity and structure of the methanogenic community in anoxic rice paddy soil microcosms as examined by cultivation and direct 16S rRNA gene sequence retrieval. *Appl Environ Microbiol* **64,** 960–969 (1998).

72. Lane, D. in *Nucleic Acid Techniques in Bacterial Systematics* (eds. Stackebrandt, E. & Goodfellow, M.) 115–175 (John Wiley & Sons, 1991).

73. Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L. & Schleper, C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ Microbiol* **5,** 787–797 (2003). doi: 10.1046/j.1462-2920.2003.00476.x

74. Taylor, D. L. & Bruns, T. D. Community structure of ectomycorrhizal fungi in a *Pinus muricata* forest: minimal overlap between the mature forest and resistant propagule communities. *Mol Ecol* **8,** 1837–1850 (1999). doi: 10.1046/j.1365-294x.1999.00773.x

75. Tourna, M., Freitag, T. E., Nicol, G. W. & Prosser, J. I. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* **10,** 1357–1364 (2008). doi: 10.1111/j.1462-2920.2007.01563.x

76. Arce, M. I., von Schiller, D., Bengtsson, M. M., Hinze, C., Jung, H., Alves, R. J. E., Urich, T. & Singer, G. Drying and Rainfall Shape the Structure and Functioning of

Nitrifying Microbial Communities in Riverbed Sediments. *Front Microbiol* **9,** (2018). doi: 10.3389/fmicb.2018.02794

77. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014). doi: 10.1093/bioinformatics/btu170

78. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34,** 2666–2669 (2018). doi: 10.1093/bioinformatics/bty149

79. Kolmogorov, M., Bickhart, D. M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T. P. L. & Pevzner, P. A. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* **17,** 1103–1110 (2020). doi: 10.1038/s41592-020-00971-x

80. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37,** 4572–4574 (2021). doi: 10.1093/bioinformatics/btab705

81. Gremme, G., Steinbiss, S. & Kurtz, S. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans Comput Biol Bioinform* **10,** 645–656 (2013). doi: 10.1109/TCBB.2013.68

82. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36,** 1925–1927 (2020). doi: 10.1093/bioinformatics/btz848

83. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *bioRxiv* 2022.07.11.499243 (2022). doi:10.1101/2022.07.11.499243 doi: 10.1101/2022.07.11.499243

84. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal* **11,** 2864–2868 (2017). doi: 10.1038/ismej.2017.126

85. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38,** 5825–5829 (2021). doi: 10.1093/molbev/msab293

86. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T. & Jensen, L. J. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47,** D309–D314 (2019).

87. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7,** e1002195 (2011). doi: 10.1371/journal.pcbi.1002195

88. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12,** 59 (2015). doi: 10.1038/nmeth.3176

89. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35,** 1026–1028 (2017). doi: 10.1038/nbt.3988

90. Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W. & Hauser, L. J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11,** 119 (2010). doi: 10.1186/1471-2105-11-119

91. Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z. & Godhe, A. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS

sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Meth Ecol Evol* **4,** (2013).

92. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20,** 1–13 (2019).

93. Lu, J. & Salzberg, S. L. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome* **8,** 124 (2020). doi: 10.1186/s40168-020-00900-2

94. Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V. & Médigue, C. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res* **48,** D579–D589 (2020). doi: 10.1093/NAR/GKZ926

95. Makarova, K. S., Galperin, M. Y. & Koonin, E. V. Comparative genomic analysis of evolutionarily conserved but functionally uncharacterized membrane proteins in archaea: Prediction of novel components of secretion, membrane remodeling and glycosylation systems. *Biochimie* **118,** 302–312 (2015). doi: 10.1016/j.biochi.2015.01.004

96. Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y. & Yin, Y. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46,** W95–W101 (2018). doi: 10.1093/nar/gky418

97. Saier, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C. & Moreno-Hagelsieb, G. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* **44,** D372–D379 (2016). doi: 10.1093/nar/gkv1103

98. Dombrowski, N., Williams, T. A., Sun, J., Woodcroft, B. J., Lee, J.-H., Minh, B. Q., Rinke, C. & Spang, A. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat Commun* **11,** 3939 (2020). doi: 10.1038/s41467-020-17408-w

99. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20,** 238 (2019). doi: 10.1186/s13059-019-1832-y

100. Pausan, M. R., Csorba, C., Singer, G., Till, H., Schöpf, V., Santigli, E., Klug, B., Högenauer, C., Blohs, M. & Moissl-Eichinger, C. Exploring the Archaeome: Detection of Archaeal Signatures in the Human Body. *Front Microbiol* **10,** (2019). doi: 10.3389/fmicb.2019.02796

101. Arce, M. I., von Schiller, D., Bengtsson, M. M., Hinze, C., Jung, H., Alves, R. J. E., Urich, T. & Singer, G. Drying and Rainfall Shape the Structure and Functioning of Nitrifying Microbial Communities in Riverbed Sediments. *Front Microbiol* **9,** (2018). doi: 10.3389/fmicb.2018.02794

102. Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J. I., Schuster, S. C. & Schleper, C. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442,** 806–809 (2006).

103. Walters, W. A., Caporaso, J. G., Lauber, C. L., Berg-Lyons, D., Fierer, N. & Knight, R. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* **27,** 1159–1161 (2011).

104. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M. & Asnicar, F. *QIIME 2: Reproducible,*

interactive, scalable, and extensible microbiome data science. (PeerJ Preprints, 2018).

105. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. & Holmes, S. P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13,** 581–583 (2016). doi: doi:10.1038/nmeth.3869

106. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & B.J, C. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6,** 226 (2018). doi: 10.1186/s40168-018-0605-2

107. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glockner, F. O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41,** D590–D596 (2013). doi: 10.1093/nar/gks1219

108. Robeson, M. S., O'Rourke, D. R., Kaehler, B. D., Ziemski, M., Dillon, M. R., Foster, J. T. & Bokulich, N. A. RESCRIPt: Reproducible sequence taxonomy reference database management. *PLoS Comput Biol* **17,** e1009581 (2021). doi: 10.1371/journal.pcbi.1009581

109. Bokulich, N., Dillon, M., Bolyen, E., Kaehler, B., Huttley, G. & Caporaso, J. q2-sample-classifier: machine-learning tools for microbiome classification and regression. *J Open Source Softw* **3,** 934 (2018). doi: 10.21105/joss.00934

110. Lozupone, C. & Knight, R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl Environ Microbiol* **71,** 8228–8235 (2005). doi: 10.1128/AEM.71.12.8228-8235.2005

111. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** (2010). doi: 10.1371/journal.pone.0009490

112. Heidrich, V., Karlovsky, P. & Beule, L. 'SRS' R Package and 'q2-srs' QIIME 2 Plugin: Normalization of Microbiome Data Using Scaling with Ranked Subsampling (SRS). *Applied Sciences* **11,** 11473 (2021). doi: 10.3390/app112311473

113. Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. & Gloor, G. B. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One* **8,** e67019 (2013). doi: 10.1371/journal.pone.0067019

114. Lin, H. & Peddada, S. Das. Analysis of compositions of microbiomes with bias correction. *Nat Commun* **11,** 3514 (2020). doi: 10.1038/s41467-020-17041-7

115. Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B. & Schwager, E. H. Multivariable association discovery in population-scale meta-omics studies. *Biorxiv* (2021).

116. Martino, C., Morton, J. T., Marotz, C. A., Thompson, L. R., Tripathi, A., Knight, R. & Zengler, K. A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems* **4,** (2019). doi: 10.1128/mSystems.00016-19

117. Wang, H., Bagnoud, A., Ponce-Toledo, R. I., Kerou, M., Weil, M., Schleper, C. & Urich, T. Linking 16S rRNA Gene Classification to *amoA* Gene Taxonomy Reveals Environmental Distribution of Ammonia-Oxidizing Archaeal Clades in Peatland Soils. *mSystems* **6,** (2021). doi: 10.1128/mSystems.00546-21

118. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., Rubin, E. M., Hugenholtz, P. & Woyke, T. Insights into the

phylogeny and coding potential of microbial dark matter. *Nature* **499,** 431–437 (2013). doi: 10.1038/nature12352

119. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30,** 772–780 (2013). doi: 10.1093/molbev/mst010

120. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10,** 210 (2010). doi: 10.1186/1471-2148-10-210

121. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A. & Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37,** 1530–1534 (2020). doi: 10.1093/molbev/msaa015

122. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28,** 1823–1829 (2012).

123. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30,** 3276–3278 (2014). doi: 10.1093/bioinformatics/btu531

124. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14,** 587–589 (2017). doi: 10.1038/nmeth.4285

125. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* **59,** 307–321 (2010). doi: 10.1093/sysbio/syq010