Host traits, lifestyle and environment are associated with human skin bacteria

L. Moitinho-Silva [iD],[1,2] N. Boraczynski,[1] H. Emmert [iD],[1] H. Baurecht [iD],[3] S. Szymczak,[4] H. Schulz,[5,6] D. Haller,[7] J. Linseisen,[5,7,8,9] C. Gieger,[5] A. Peters,[5] L. Tittmann,[10] W. Lieb,[11] C. Bang,[2] A. Franke,[2] E. Rodriguez [iD][1] and S. Weidinger [iD][1]

[1]Department of Dermatology and Allergy, University Hospital Schleswig-Holstein, Kiel, Germany

[2]Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany

[3]Department for Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany

[4]Institute of Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein, Kiel, Germany

[5]Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

[6]Comprehensive Pneumology Center Munich, Member of the German Center for Lung Research (DZL), Munich, Germany

[7]ZIEL Institute for Food and Health, Technische Universität München, Freising, Germany

[8]Clinical Epidemiology, Helmholtz Zentrum München – German Research Centre for Environmental Health, Neuherberg, Germany

[9]Chair of Epidemiology, Ludwig-Maximilians-Universität München, at UNIKA-T Augsburg, Augsburg, Germany

[10]Biobank PopGen and Institute of Epidemiology, Kiel University, Kiel, Germany

[11]Institute of Epidemiology, Kiel University, Kiel, Germany

Summary

Background The human skin offers diverse ecosystems for microbial symbionts. However, the factors shaping skin–microbiome interactions are still insufficiently characterized. This contrasts with the broader knowledge about factors influencing gut microbiota.

Objectives We aimed to investigate major patterns of association of host traits, lifestyle and environmental factors with skin bacteria in two German populations.

Methods This is a cross-sectional study with 647 participants from two population-based German cohorts, PopGen (n = 294) and KORA FF4 (n = 353), totalling 1794 skin samples. The V1–V2 regions of the 16S ribosomal RNA (rRNA) gene were sequenced. Associations were tested with two bacterial levels, community (beta diversity) and 16S rRNA gene amplicon sequence variants (ASVs).

Results We validated known associations of the skin microbiota with skin microenvironment, age, body mass index and sex. These factors were associated with beta diversity and abundance of ASVs in PopGen, which was largely replicated in KORA FF4. Most intriguingly, dietary macronutrients and total dietary energy were associated with several ASVs. ASVs were also associated with smoking, alcohol consumption, skin pH, skin type, transepidermal water loss, education and several environmental exposures, including hours spent outdoors. Associated ASVs included members of the genera *Propionibacterium*, *Corynebacterium* and *Staphylococcus*.

Conclusions We expand the current understanding of factors associated with the skin bacterial community. We show the association of diet with skin bacteria. Finally, we hypothesize that the skin microenvironment and host physiology would shape the skin bacterial community to a greater extent compared with a single skin physiological feature, lifestyle and environmental exposure.

> **What is already known about this topic?**
>
> - The skin microbiome is essential for maintaining skin health.

- Skin bacteria abundances are associated with skin physiology patterns (microenvironments), host traits, such as age and sex, and domestic environmental factors, such as pets.
- Evaluation and translation of these associations are difficult because most studies have a limited number of candidate factors.

What does this study add?

- We expand the current knowledge of factors associated with skin microbiota by revealing new factors associated with skin bacteria, including diet.
- We provide a comprehensive view of the factors associated with skin microbiota, which suggests that skin microenvironment and host physiology would shape the skin bacterial community to a greater extent compared with a single skin physiological feature, lifestyle and environmental exposure.

What is the translational message?

- Future clinical research involving skin microbiota should acknowledge the associations found as potential confounders.
- Host factors (age, body mass index and sex) and skin microenvironments should be particularly considered because they were associated with skin bacteria at the community level.

The human skin offers diverse ecosystems that harbour distinct microbial communities, with three postulated major microenvironments: dry, moist and sebaceous.[1] Microenvironments are distinct sets of skin physiological parameters, including pH, temperature, moisture, sebum content and topography.[2] Skin microbiota include beneficial bacteria that can keep potential pathogens at bay.[3] Many common skin diseases are associated with distinct microbiota signatures, including atopic dermatitis[4] and psoriasis.[5] Therefore, understanding the dynamics and functional causations of associated microbiota changes will enable the development of better preventive and therapeutic recommendations for skin health.

There are an increasing number of studies (such as Bouslimani *et al.* and Huang *et al.*) investigating the influence of external and host factors on the skin microbiota of the general population.[6,7] Studies focusing on healthy individuals mostly investigate single or few candidate factors,[8,9] which hampers the generalization and integration of their findings, and the evaluation of their robustness. A recent study with 495 participants and 39 factors indicated the potential of population studies to reveal associations with skin microbiota.[10] Nevertheless, there are knowledge gaps about the forces that shape the skin microbiome, in particular compared with what is known about the human gut microbiota.[11]

Here, we aimed to investigate associations of host traits, lifestyle and environmental factors with skin bacteria in two German population-based cohorts. Skin bacteria were studied at two levels, bacterial community (beta diversity) and bacterial marker gene variants [16S ribosomal RNA (rRNA) gene amplicon sequence variants (ASVs)]. In addition to the investigation of as yet unexplored factors such as diet, known factors associated with skin bacteria – age, body mass index (BMI), sex and skin microenvironment – were integrated in our analysis for validation, as references and as confounders.

Materials and methods

Study population and data acquisition

This study is a cross-sectional survey of skin microbiota from participants of two independent population-based adult German cohorts, PopGen and KORA FF4. For PopGen, 1317 participants (aged 19–77 years; 55% men) were randomly recruited between 2005 and 2007, and reinvited in 2016–2017 (second follow-up) via the local population registry and as blood donors from the region of Kiel, Germany.[12] General health information (standardized questionnaire), dietary patterns and physical activity information (validated, self-administered questionnaires in web-based version and, optionally, on paper[13]) were collected. Macronutrient intake details were obtained using the German Food Code and Nutrient Database.[14]

KORA FF4 (2279 participants) is the second follow-up of the KORA S4 Survey (1999–2001, aged 25–74 years) conducted between 2013 and 2014 in the southern German city of Augsburg and its two surrounding counties.[15] Personal data and lifestyle data were collected from a standardized face-to-face interview.[16] Macronutrient intake details were obtained using the German Food Code and Nutrient Database (BSL III.2)[17] on habitual dietary intake (repeated 24-h food lists

and a food frequency questionnaire).[18] Written informed consent was obtained from all study participants. All protocols were approved by the ethics committees of the Medical Faculty of Kiel University (PopGen) and of the Bavarian Medical Association (KORA).

Biological specimen collection

Skin swabs were collected from participants of PopGen second follow-up (n = 295) and from the youngest age group (39–48 years) of KORA FF4 (n = 376). Participants were asked to avoid bathing/showering and application of any topical agents 24 h prior to the sampling visit. A 4-cm^2 area from the antecubital fossa (PopGen and KORA FF4), retroauricular fold (KORA FF4), forehead, volar and dorsal forearm (PopGen) was firmly swabbed for at least 30 s. Immediately prior to collection, swabs [Catch-All Sample Collection Swab; Epicentre Biotechnologies (Illumina Inc., San Diego, CA, USA)] were soaked in specimen collection fluid.[19] Sampling negative controls were swabs exposed to ambient air for 5 s. After sampling, swabs were immediately stored at –80 °C. DNA was isolated from KORA FF4 samples with MO BIO PowerSoil DNA Isolation Kit (QIAGEN GmbH, Hilden, Germany) and from PopGen samples with QIAamp UCP Pathogen Mini Kit on an automated QIAcube system (QIAGEN).

Dermatological examination

For PopGen participants, skin type, skin pH and transepidermal water loss (TEWL) were recorded by trained dermatologists.[19] At sampling sites, skin pH and TEWL were measured with a Skin pH Meter (HI-99181/HI-1414D; Hanna Instruments, Vöhringen, Germany) and Tewameter® TM 300 (Courage + Khazaka Electronic GmbH, Cologne, Germany). The mean of three measurements was used for analysis.

Microbiota profiling

The V1 and V2 variable regions of the 16S rRNA gene were amplified by polymerase chain reaction (PCR) with the universal primer pair 27F and 338R. Sequencing was performed with MiSeq Reagent Kit v3 on the Illumina MiSeq (Illumina Inc.) (Appendix S1; see Supporting Information).[20,21] Sequence reads were processed with DADA2 v.1.10.[22] The resulting ASV table contained the number of times each ASV was observed in each sample.[23] ASV is a finer scale analogue of the operational taxonomic unit. It resolves the sequenced region variant down to a single-nucleotide difference level. Taxonomic classification was performed with the RDP classifier algorithm based on the Ribosomal Database Project v.16 release.[24,25] Low confidence classifications (< 50) were labelled unclassified. Sequences classified as chloroplasts or mitochondria were removed. Cohort ASV tables were separated into skin site-specific tables, four for PopGen and two for KORA FF4. Samples taken from sites with skin abnormalities (e.g. lesions) or where corticoids or antibiotics were used

within the last 7 days before collection were excluded. To remove possible contaminants, ASVs that were low abundance (< 0·1% of total sequence counts) were excluded (Figure S1; see Supporting Information). Samples that yielded < 5000 sequences were excluded.

Bacterial community diversity (beta diversity) was estimated from Bray–Curtis dissimilarities of rarefied ASV tables (5000 sequences per sample) and visualized with principal coordinates analysis (PCoA). The rarefaction did not impact the ASV diversity recovered (Figure S2; see Supporting Information). Bray–Curtis dissimilarities were calculated using the R package vegan v.2.5-5.[26,27] PCoA was performed using ape package v.5.3.[28] No sequencing batch effects were observed on beta diversities (Figure S2).

Prediction of microenvironment by bacterial genus profile

PopGen and KORA FF4 rarefied microbiota profiles were combined. Sequence counts were grouped by genus. Only the most abundant bacterial genera (> 1%) were used. The machine learning routine consisted of 100 rounds (Appendix S1). Briefly, in each round, one sample per participant was randomly chosen and the dataset was randomly split into training and prediction sets (8 : 2 ratio). The random forest algorithm was trained using the R packages randomForest v.4.6-14[29] and caret v.6.0-84.[30] Predictions of samples' skin microenvironment were evaluated with a multiclass receiver operating characteristic area under the curve (AUC)[31] and with the Matthews correlation coefficient (MCC).[32]

Association of individual traits with skin bacteria

Age, BMI and sex were confounders in all tests, based on the literature[8–10] and on their effects on unadjusted beta diversity association tests (Appendix S1). Total energy intake was included as a confounder for tests with macronutrient intake as recommended.[33] Room temperature and air humidity were included as confounders for tests with TEWL due to their potential impact on its measurements. Therefore, each variable was tested separately in its own confounder-adjusted model (Figure S3; see Supporting Information). This approach was chosen to avoid overfitting, control for main confounders and ensure the interpretability of our results. To address the confounding effect of microenvironment, tests were carried out for each skin site independently. For reference, we tested variables in unadjusted models, which yielded a greater number of significant results (null hypothesis rejections) for beta diversity and fewer for ASV associations (Figure S4; see Supporting Information).

Associations with beta diversity (i.e. Bray–Curtis dissimilarities) were tested with permutational multivariate analysis of variance with adonis2 function in R (vegan package, version 2.4-2) using 999 permutations and sequential effects. False discovery rate (FDR) correction was applied per site and test datasets (as stated in Table S1; see Supporting Information and described in the following section). The adjusted *P*-value

significance cut-off was 0·05. Results are presented as percentage of community variation, i.e. $R^2 \times 100$.

Unlike other tests, assessment of the effect of skin microenvironment on beta diversity includes samples from all skin sites collected within a cohort. To ensure samples were independent of each other, this test was performed in 100 rounds. In each round, one sample per participant was randomly selected. The proportions between microenvironments were kept. R^2 values were summarized with mean $\pm$ SD. Results are presented as a percentage of community variation.

Associations with nonrarefied ASV abundances were tested with the package DESeq2 v.1.24.0,[34] with negative binomial generalized linear models and likelihood ratio tests (Appendix S1). Log fold changes were estimated using the zero-centred normal prior distribution.[34] FDR correction was applied across all ASVs for each variable separately. The adjusted P-value significance cut-off was 0·05.

Study-specific analyses and reproducibility

Age, BMI, sex and lifestyle factors, including dietary macronutrients, were available in both cohorts (Table 1, Figure S5; see

Table 1 Age, body mass index (BMI), sex and lifestyle factors of PopGen and KORA FF4 participants

	PopGen (n = 254)	KORA FF4 (n = 225)
Age/BMI/Sex		
Age (years)	66 (59, 75)	45 (42, 46)
Sex:female	112 (44)	133 (59)
BMI (kg m^{-2})	27 (24, 30)	25 (23, 28)
Diet		
Carbohydrate (g per day)	213 (164, 268)	208 (173, 245)
Energy (kcal per day)	2135 (1752, 2726)	1881 (1672, 2130)
Fat (g per day)	98 (74, 128)	78 (68, 87)
Fibre (g per day)	20 (17, 25)	16 (14, 19)
Protein (g per day)	77 (62, 96)	71 (63, 82)
Systemic antibiotics		
Antibiotics (< 6–8 w)[1]:yes	15 (6)	20 (9)
Smoking		
Smoking:ex-smoker	122 (48)	76 (34)
Smoking:nonsmoker	109 (43)	114 (51)
Smoking:smoker	23 (9)	35 (16)
Alcohol consumption		
Alcohol (g per day)	9 (2, 19)	5 (3, 13)
Physical activity		
Regular sports[2]	228 (90)	–
Sport in summer (h)[2]	2·00 (0·00, 4·75)	–

Values represent n (%) or median (interquartile range; separated by comma). Variables encoding related factors are indicated by a colon, ':'. [1]PopGen collected information about 6 weeks prior sampling. KORA FF4 collected information about 8 weeks prior sampling. [2]Information not available for KORA FF4 participants.

Supporting Information). Only participants with complete information regarding these traits were kept for association analysis. Association analysis with these variables were conducted primarily in PopGen because it includes samples from four different skin sites and three microenvironments, while two skin sites from two microenvironments were sampled in

Table 2 Age, body mass index (BMI), sex, education and environmental factors of KORA FF4 participants

	KORA FF4 (n = 349)
Age/BMI/Sex	
Age (years)	45 (42, 47)
Sex:female	185 (53)
BMI (kg m^{-2})	26 (23, 29)
Education/dwelling	
HVQ:no degree/vocational school/teaching	204 (58)
HVQ:technical school/technical/master	73 (21)
HVQ:engineering school/ polytechnic/college/university	72 (21)
Cohabiting couple:no (single, living alone/married, separated/ divorced/widowed)	75 (21)
Cohabiting couple:yes (living together with partner/married, living together)	274 (79)
Location:city (city of Augsburg)	110 (32)
Location:rural (administrative district of Augsburg/Aichach–Friedberg)	239 (68)
Hours outdoors	
In summer (per week)	3·00 (2·00, 4·00)
In summer (per weekend)	5·00 (3·00, 7·00)
In winter (per week)	1·00 (1·00, 2·00)
In winter (per weekend)	2·00 (1·00, 3·00)
Free time outdoors	
Beach/lake	138 (40)
Green spaces	313 (90)
Mountain region	100 (29)
Urban spaces	103 (30)
Holidays outdoors	
Beach/lake	264 (76)
Green spaces	195 (56)
Mountain region	171 (49)
Urban spaces	121 (35)
Pets	
Animal contact:no (not pet owner/never/ occasionally)	153 (44)
Animal contact:yes (regular or pet owner)	196 (56)
Cat	98 (28)
Dog	52 (15)
Rodent	34 (10)
Other	33 (9)
Ultraviolet protection	
Sun protection in summer:always	50 (14)
Sun protection in summer:mostly	173 (50)
Sun protection in summer:sometimes	72 (21)
Sun protection in summer:rarely/never	54 (15)

Values represent n (%) or median (interquartile range; separated by comma). Variables encoding related factors are indicated by a colon, ':'. HVQ, highest vocational qualification.

KORA FF4. Analysis with these variables in KORA FF4 was used to verify the reproducibility of the results from PopGen. Study-specific analyses with variables available in either cohort were performed. Accordingly, the analyses with variables regarding physical activity (Table 1) and skin physiological parameters (Table S2; see Supporting Information) were performed in Pop-Gen. In addition, the analysis with variables about dwelling, education and environmental factors was performed in KORA FF4 (Table 2). Associations found in at least two sites were regarded as robust, and therefore further inspected in our study. However, associations with environmental variables (Table 2) were also inspected, because little overlap was found between results from the antecubital fossa and retroauricular fold. Data availability is provided in Appendix S1.
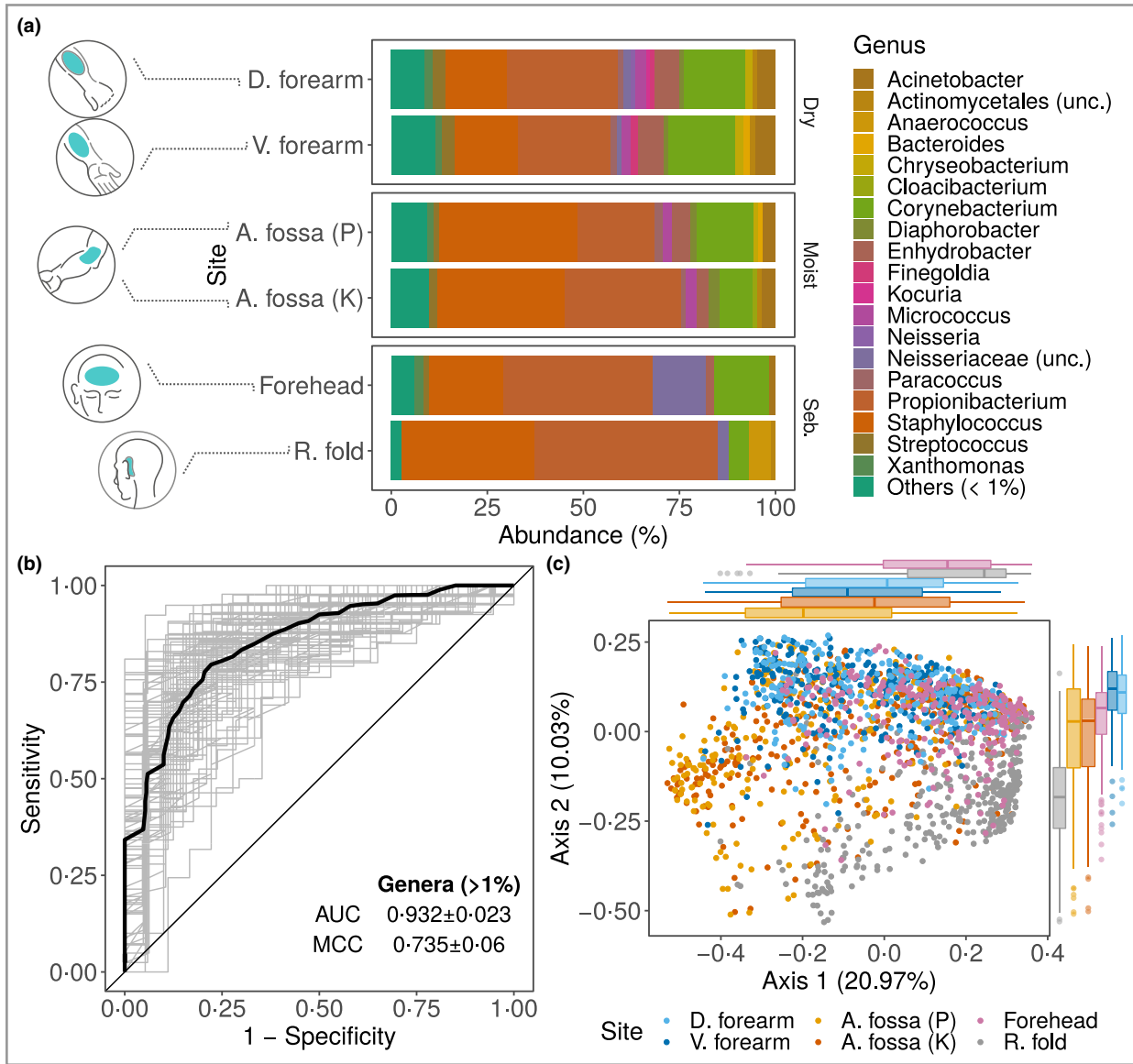

Figure 1 Skin microbiota patterns. (a) Most abundant bacterial genera in skin sites: dorsal (D.) forearm, volar (V.) forearm, antecubital (A.) fossa, forehead and retroauricular (R.) fold. Samples from PopGen (P) and KORA FF4 (K) cohorts are shown. All genera with < 1% abundance were combined in the category 'Others'. Skin sites are grouped by microenvironment, i.e. dry, moist and sebaceous (Seb.). Skin collection sites are depicted. Bacteria from the family *Neisseriaceae* and order *Actinomycetales* that were left unclassified (unc.) at the genus level were also included. (b) Prediction of sample microenvironments by genus profile (> 1% abundance). The receiver operating characteristic area under the curve (AUC) of 100 iterations (light grey lines) and their mean (black bold line) are shown. Median and median absolute deviation (±) of the AUC and Matthews correlation coefficient (MCC) are shown, indicating very high classification performance of random forest predictions. (c) Bacterial community variation (beta diversity). Points represent individual samples. Bray–Curtis dissimilarities were visualized using principal coordinates analysis (PCoA). Marginal boxplots are shown to visualize sample distributions along axes. Percentage of variation explained by each axis is shown. Samples are coloured by microenvironment: dry (light blue and blue), moist (yellow, orange) and sebaceous (pink and grey).
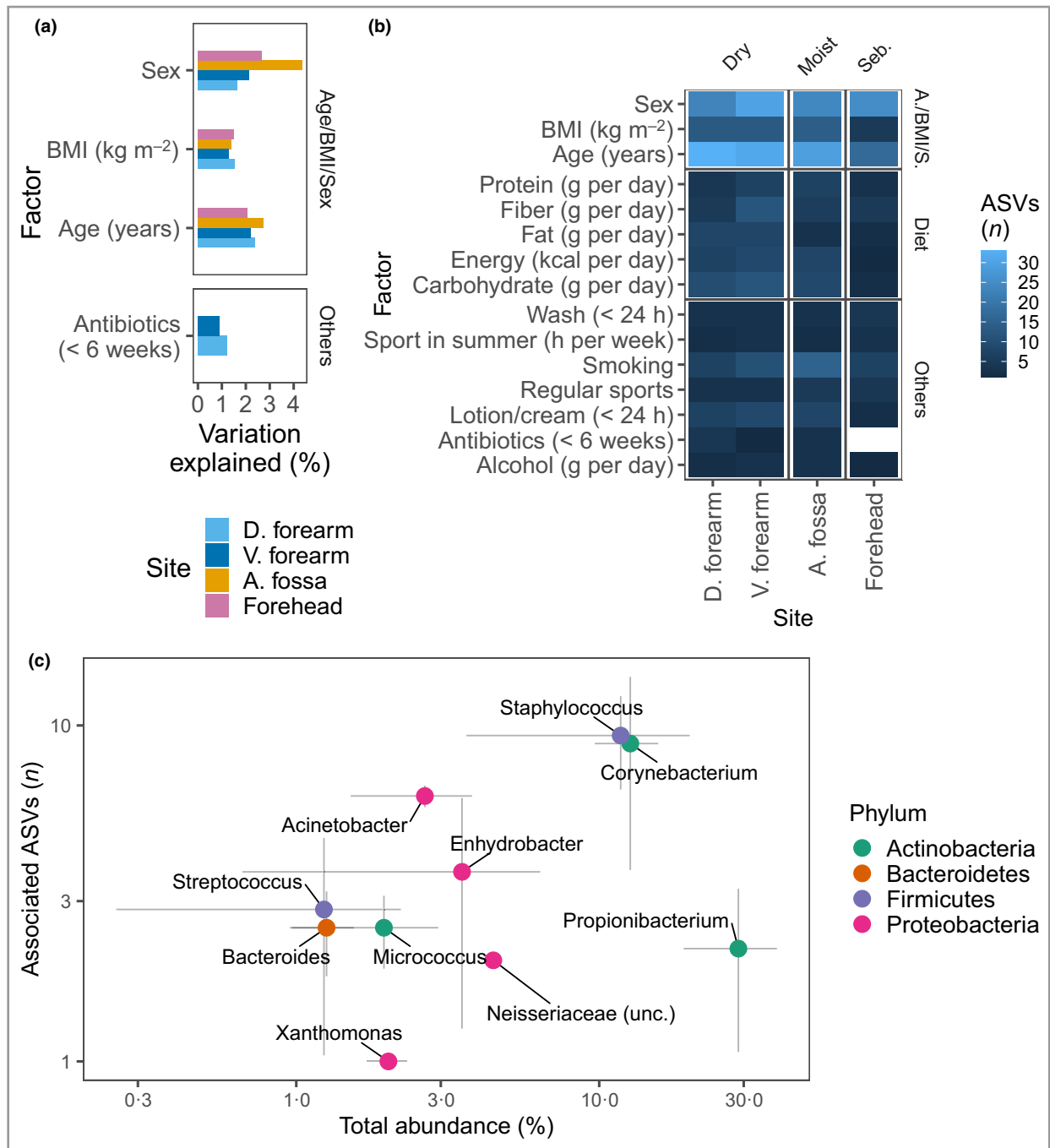
Figure 2 Age, body mass index (BMI), sex and lifestyle were associated with skin bacteria. (a) Factors associated with bacterial community variation in skin sites: dorsal (D.) forearm, volar (V.) forearm, antecubital (A.) fossa and forehead. Only factors that were significantly associated with beta diversity are shown (adjusted $P < 0.05$). (b) Associated bacterial marker gene variants. Number of 16S ribosomal RNA gene amplicon sequence variants (ASVs) significantly associated with each factor, i.e. age (A.), BMI, sex (S.), diet and 'Others', is shown (adjusted P value < 0·05). Skin sites are grouped by microenvironment, i.e. dry, moist and sebaceous (Seb.). (c) Top associated bacterial genera. The *x*-axis shows the sum of all ASV abundances of a given genus averaged by microenvironment, where volar forearm and dorsal forearm together weighted one. The numbers of associated ASVs found in each genus were averaged by microenvironment. Grey lines represent weighted standard deviations. Bacteria from the family *Neisseriaceae* that were left unclassified (unc.) at genus level were also included. Data shown are from analysis of the PopGen cohort.

Results

Skin microbiota patterns reflect skin microenvironments

Microbiota profiles from 647 participants (294 from PopGen and 353 from KORA FF4) were recovered (1794 skin samples). The microbial composition varied among skin sites (Figure 1a). Random forest algorithms trained with the abundant bacterial genera (> 1% of 16S rRNA gene sequences) performed highly on predicting samples' microenvironment (median AUC of 0·932 and median MCC 0·735) (Figure 1b). Bacterial communities were similar within each microenvironment. There was a large overlap between samples collected from the antecubital fossa from PopGen and KORA FF4 participants (Figure 1c). Each sampling site presented distinct, albeit not completely separated, bacterial profiles as observed from density distributions along PCoA axes. The suggested gradients of skin microbiota's structures are supported by the finding that at least 61·65% of the sequences recovered in each site were shared with at least another site (Table S3; see Supporting Information).

Associations of skin microenvironment, host traits and lifestyle with skin bacterial community and marker gene variants

For 86% of PopGen participants (n = 254), individual and lifestyle information (Table 1) was collected. Although participants were asked to avoid washing or using cosmetic cream, lotion or ointment (hereafter referred to as cream/lotion) in the skin collection area 24 h prior to sampling, up to 78% of participants (Table S4; see Supporting Information) did not follow this instruction. Therefore, these behaviours were included as lifestyle variables to be tested. Skin microenvironment was the major trait associated with skin microbiota beta diversity, explaining about 9·2 ± 0·9% of the bacterial community variation. Age, BMI and sex were associated with community variation at all four sampling sites (Figure 2a). The effects of sex and age were the highest, with maximum effect sizes observed in the antecubital fossa (4·3% for sex; 2·7% for age). In addition, the use of systemic antibiotics less than 6 weeks prior to sampling was associated with dorsal and volar forearm bacterial communities (maximum 1·2%)
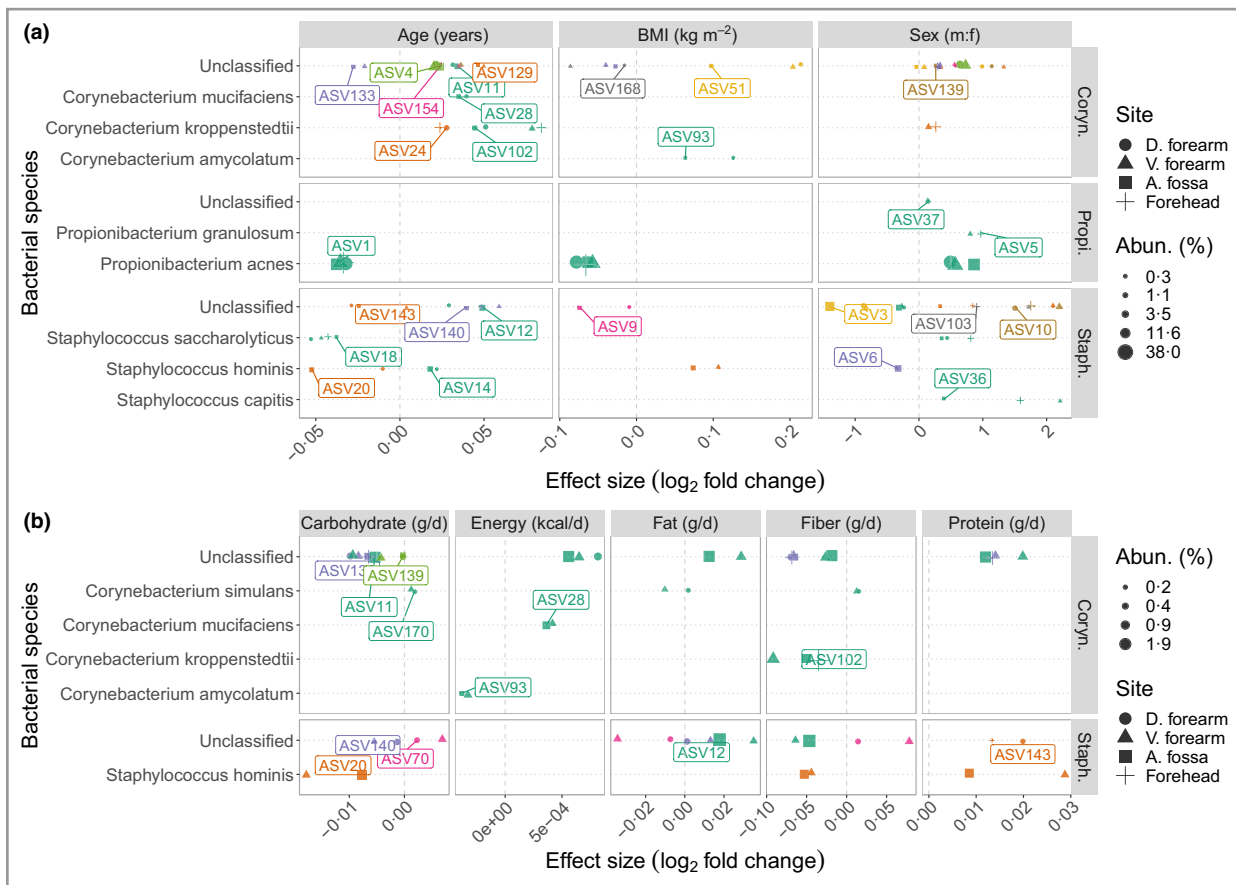

Figure 3 Age, body mass index (BMI), sex and dietary macronutrients were associated with skin bacteria of the genera *Corynebacterium* (Coryn.), *Propionibacterium* (Propi.) and *Staphylococcus* (Staph.). (a) Bacteria associated with age, BMI and sex. Significant effects on bacterial marker gene variants, i.e. 16S ribosomal RNA gene amplicon sequence variants (ASVs), are shown (adjusted P-value < 0·05). Effect size is shown on the x-axis. Species level classification of ASVs is shown when available. ASVs were coloured within each species classification (rows) to facilitate visualization. (b) Bacteria associated with diet macronutrients. Data shown are from analysis of the PopGen cohort. m, male; f, female; D., dorsal; V., volar; A., antecubital; Abun., abundance; g/d, grams per day; kcal/d, kilocalories per day.

(Figure 2a). None of the other factors was associated with beta diversity (Table S1). Associations with microenvironment, sex and BMI were replicated in KORA FF4 (n = 225) [Table 1; Figure S6 (see Supporting Information) and Table S1]. KORA FF4 participants were of a limited age range (median 45 years, interquartile range 42–46), which could explain the absence of significant associations of age with beta diversity in this cohort.

A total of 647 unique significant associations of age, BMI, sex and lifestyle factors with bacterial marker gene variants, i.e. ASVs, were found (Table S5; see Supporting Information). Of these, 209 associations were found in at least two sampling sites, and therefore further inspected. Accordingly, age (maximum of 33 associated ASVs), sex (maximum 30), smoking (maximum 15) and BMI (14) were associated with the highest numbers of variants per site, followed by dietary macronutrients [maximum 12 for carbohydrate (g per day)] (Figure 2b). We then inspected the genus classification of the associated bacteria (Figure 2c). The genus *Propionibacterium* contained few (mean ± SD ASVs per microenvironment of 2·2 ± 1·1, where dorsal and volar forearm together weighted one) but abundant associated variants (mean total sequences of 28·8 ± 9·8%). Other prominent bacterial genera were *Corynebacterium* [n = 8·8 ± 5·1 (12·7 ± 3%)], followed by *Staphylococcus* [n = 9·3 ± 2·9; (11·8 ± 8·1%)]. Similar association patterns were observed in KORA FF4 participants (Figure S6 and Table S5).

The associations with bacteria from the *Propionibacterium*, *Corynebacterium* and *Staphylococcus* genera were further inspected.

With rare exceptions, age, BMI, sex and lifestyle factors were associated with ASVs from these genera with same direction and similar effect sizes across sampling sites (Figure 3; Figure S7; see Supporting Information). Age was positively associated with bacteria from the genus *Corynebacterium*, with exception of one unclassified ASV, including *Corynebacterium mucifaciens*, *C. kroppenstedtii* and *C. amycolatum* (Figure 3a). Age was negatively associated with *Propionibacterium acnes* (recently renamed *Cutibacterium acnes*[35]) and associated with *Staphylococcus* variants both positively and negatively. Similarly, BMI was positively and negatively associated with various *Corynebacterium* and *Staphylococcus* variants. In addition, BMI was negatively associated with *P. acnes*. Men were positively associated with *Corynebacterium*, including *C. kroppenstedtii*, and *Propionibacterium*, including *P. acnes* and *P. granulosum* (renamed *Cutibacterium granulosum*[35]). Men were positively associated with some *Staphylococcus* variants, including *S. saccharolyticus* and *S. capitis*, but negatively with others, such as *S. hominis*.

Dietary intake of macronutrients was associated with 12 ASVs from *Corynebacterium* and *Staphylococcus* genera (Figure 3b). Carbohydrate was positively associated with *C. simulans*, but negatively associated with unclassified *Corynebacterium* variants. Carbohydrate was negatively associated with *S. hominis*, but positively and negatively associated with unclassified *Staphylococcus* variants. Dietary energy, fat and fibre were also found associated in both directions with *Corynebacterium* and *Staphylococcus* variants. Protein was positively associated with
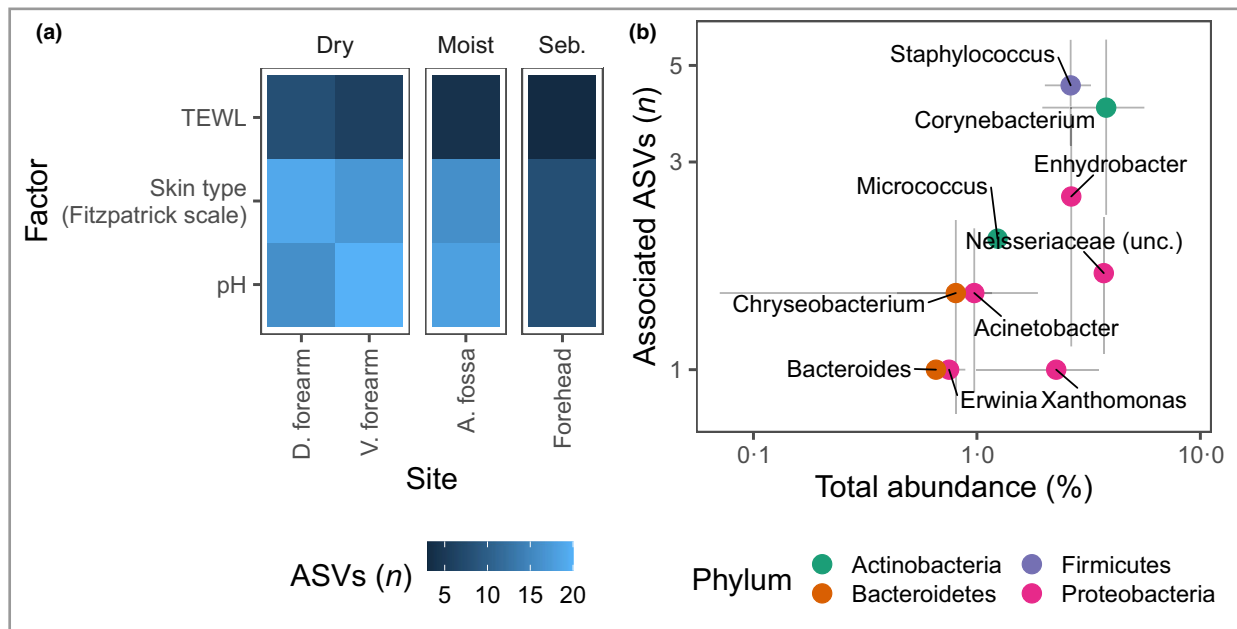

Figure 4 Skin physiology was associated with skin bacteria. (a) Associated bacterial marker gene variants. Number of 16S ribosomal RNA gene amplicon sequence variants (ASVs) significantly associated with each factor is shown (adjusted P-value < 0·05). Skin sites are grouped by microenvironment, i.e. dry, moist and sebaceous (Seb.). Skin sites shown are dorsal (D.) forearm, volar (V.) forearm, antecubital (A.) fossa and forehead. (b) Top associated bacterial genera. The x-axis shows the sum of all ASV abundances of a given genus averaged by microenvironment, where volar forearm and dorsal forearm together weighted one. The number of associated ASVs found in each genus were averaged by microenvironment. Grey lines represent weighted standard deviations. Bacteria from the family *Neisseriaceae* that were left unclassified (unc.) at genus level were also included. Data shown are from analysis of the PopGen cohort. TEWL, transepidermal water loss.
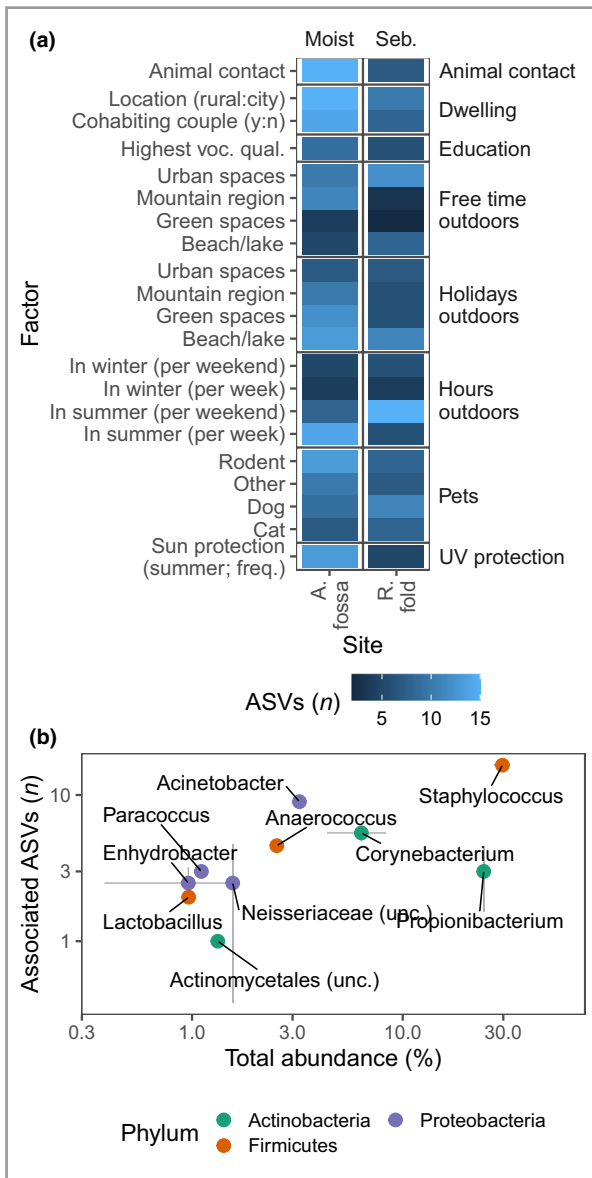
Figure 5 Environmental factors were associated with skin bacteria. (a) Associated bacterial marker gene variants. Number of 16S ribosomal RNA gene amplicon sequence variants (ASVs) significantly associated with each factor is shown (adjusted P-value < 0·05). Skin sites are grouped by microenvironment, i.e. moist and sebaceous (Seb.). Skin sites shown are antecubital (A.) fossa and retroauricular (R.) fold. (b) Top associated bacterial genera. The *x*-axis shows the sum of all ASV abundances of a given genus averaged by microenvironment. The number of associated ASVs found in each genus were averaged by microenvironment. Grey lines represent weighted standard deviations. Absence of deviation bars indicate that the associations were found in a single skin site. Bacteria from the family *Neisseriaceae* and order *Actinomycetales* that were left unclassified (unc.) at genus level were also included. Data shown are from analysis of the KORA FF4 cohort. voc. qual., vocational qualification; UV, ultraviolet.

two *Corynebacterium* and two *Staphylococcus* variants, including *S. hominis*. Remaining lifestyle factors were associated with up to five ASVs from these three genera, including smoking,

alcohol consumption, systemic use of antibiotics, washing, use of lotion/cream and sporting activity (Figure S7). Remarkably, 47 associations of individual traits with ASVs in the antecubital fossa from PopGen participants were replicated in the same skin site of KORA FF4 participants, including the positive association of men with *P. acnes* (ASV1) (Figure S8; see Supporting Information). Thirty of the observed associations had the same direction in both cohorts and 17 were of opposite directions.

Associations of skin physiology, skin type and environmental factors with skin bacterial marker gene variants

After FDR correction, there was no significant association of skin pH, TEWL or skin type (Fitzpatrick scale) with bacterial community variation in PopGen participants that went through dermatological characterization (75 participants, 282 samples) (Tables S1 and S2). Of 239 samples, 61 associations of these factors with ASVs were found in at least two sites and were further inspected. Skin pH was associated with the highest numbers of variants per site (maximum 20 associated ASVs) (Figure 4a), followed by skin type (maximum 19) and TEWL (maximum 8). The ASVs significantly associated with these factors represented a small proportion of the microbiota, where the most prominent genera were *Corynebacterium* (mean total sequences per microenvironment of 3·8 ± 1·8%), unclassified genus/genera of the family *Neisseriaceae* (3·7 ± 5·1%) and *Enhydrobacter* (2·6 ± 3·2%) (Figure 4b).

After FDR correction, no environmental factor (Table 2) was significantly associated with bacterial community variation in KORA FF4 participants (349 participants, 685 samples from two skin sites; Table S1). Although skin sites shared more than 73% of the bacterial 16S rRNA gene sequences recovered (Table S3), 336 of the 350 unique associations of environmental factors with ASVs were found in a single given sampling site. Considering all of the associations found, most of them were recovered in the antecubital fossa (n = 208) in comparison with the retroauricular fold (n = 156) (Figure 5a). Interestingly, hours spent outdoors in summer during weekends were associated with more variants from the retroauricular fold than from the antecubital fossa (n = 15 and n = 8, respectively). In addition, hours spent outdoors in summer were associated with more ASVs in the antecubital fossa (week, n = 14 vs. weekend, n = 8) than in winter (n = 4 vs. n = 5). Other environmental factors with most associations found in a single site, i.e. antecubital fossa, included animal contact (15 ASVs; i.e. ownership or regular contact), housing location (15 ASVs; city vs. rural areas) and belonging to a cohabiting couple (14 ASVs). The most prominent genera associated with environmental factors included *Staphylococcus* (mean total sequences per microenvironment of 29·7 ± 2·6%), *Propionibacterium* (24·2 ± 32·9%) and *Corynebacterium* (6·3 ± 2%) (Figure 5b).
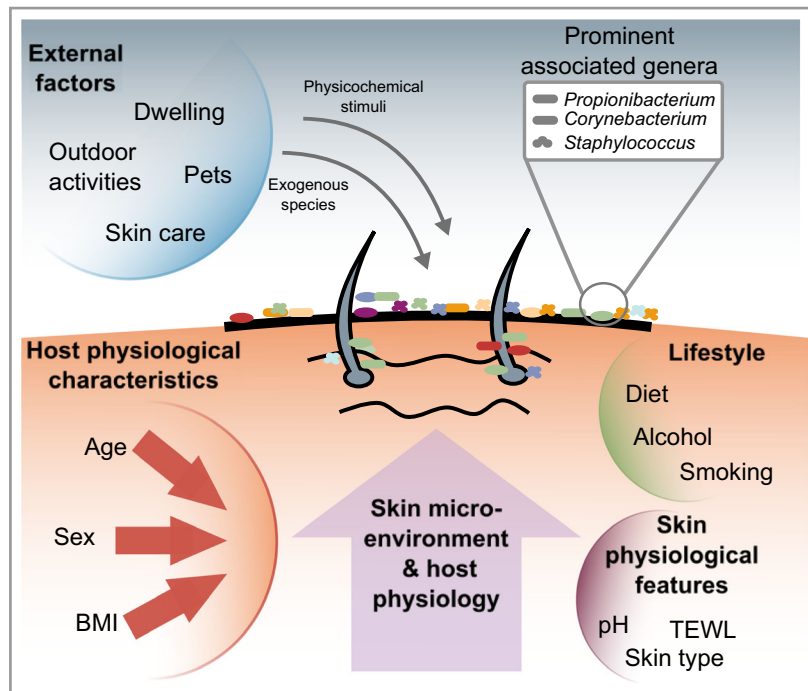
Figure 6 Summary of factors associated with skin microbiota composition in our study panels. Given our identified associations, we hypothesize that skin microenvironment and host physiology may have a larger influence on the skin microbiota in comparison with a single skin physiological feature and environmental factors. TEWL, transepidermal water loss.

Discussion

We show known and unknown associations in two population-based German cohorts. The skin microenvironment was the most prominent trait associated with the skin bacterial community variation (beta diversity). In addition, age, BMI and sex were also associated with beta diversity and with the highest numbers of ASVs. Their community-wide effects are supported by previous association reports.[1,8–10] However, washing and use of emollients (< 24-h period) were not associated with beta diversity but rather with few ASVs, agreeing with their previously shown limited effects.[36,37] Altogether, our results confirm that host factors and skin microenvironments are potential confounders for skin bacterial communities and should be acknowledged in future clinical studies.

We provide the first direct evidence of the association of diet with skin bacteria. Although effect sizes were low, the amount of total energy and macronutrient intake was significantly associated with ASV abundances. Because our study is cross-sectional, further work is required to establish the causal nature of the diet–skin microbiome relationship. Hypothetically, diet could influence skin bacteria by changing skin biochemical composition. For instance, a high fat diet in mice was reported to lead to a change in skin lipid composition and be associated with an increase in *Corynebacterium*.[38] However, such a mechanism has not been studied in humans. Additionally, dietary components may act on skin bacteria by modulating the immune response[39] and through production of gut microbiota metabolites that may reach skin tissue.[40]

The abundance of many bacterial variants was significantly associated with several aspects of human lifestyle, including the surrounding environment. In addition to the likely influence of the domestic environment on skin bacteria, suggested by associations with pets and cohabiting couples also found previously,[41,42] our results indicate the influence of nondomestic environments. For instance, we observed more variants associated with hours spent outdoors in summer compared with winter at the antecubital fossa, which could be explained by seasonal variation in exposure of this skin site. However, no associations of environmental factors with beta diversity were found, suggesting lack of their impact at the community level.

Although no causal effects were investigated, our results allow us to hypothesize that the skin microenvironment and host physiology would shape the skin microbiota to a greater extent in comparison with a single skin physiological feature, lifestyle and environmental exposure (Figure 6). Alterations of the former, as in the case of disease onset[43] or due to ageing,[10] would lead to alterations in skin microbiota. Effects on bacteria variants may occur indirectly through modulation of the skin microenvironment and host physiology, for instance by diet[44] or smoking.[45] Bacterial species abundance may also be modulated by direct exposure to external forces, which are a source of potential colonizer species[46] and physicochemical stimuli, such as ultraviolet. Here, bacterial variants with a broad abundance range (1–30% of skin sites total sequences) were associated with internal and external factors, suggesting that these may influence bacteria from diverse community structural roles.

Although this study is a comprehensive survey of the factors associated with human skin microbiota, it has limitations. Our investigations were based on PCR amplification of the V1 and V2 variable regions of the 16S rRNA gene, which is subject to primer bias[47,48] and has limited taxonomic classification resolution, particularly at the species level.[49] This issue could be addressed by shotgun metagenomics.[50] We observed that a rather small proportion of skin microbiota variation was explained by the factors analysed here, which agrees with previous observations.[10] These findings suggest that additional underexplored factors, such as human genetics, may act on the skin microbiome.

In conclusion, our results expand the current understanding of factors associated with the skin bacterial community. The association of diet with skin bacteria was shown for the first time. The observation of similar association patterns across skin sites, as well as two independent populations, indicates that these results could be generalized to other populations of similar characteristics. Furthermore, we hypothesize that a large proportion of the known putative forces shaping the skin microbiota act through stable structures, such as skin microenvironment and host physiology, rather than by direct influence on the microbial community. Finally, we suggest that these factors should be acknowledged when conducting future clinical research.

Acknowledgments

References

1 Grice EA, Kong HH, Conlan S *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* 2009; **324**:1190–2.

2 Oh J, Byrd AL, Deming C *et al.* Biogeography and individuality shape function in the human skin metagenome. *Nature* 2014; **514**:59–64.

3 Christensen GJM, Brüggemann H. Bacterial skin commensals and their role as host guardians. *Benef Microbes* 2014; **5**:201–15.

4 Paller AS, Kong HH, Seed P *et al.* The microbiome in patients with atopic dermatitis. *J Allergy Clin Immunol* 2019; **143**:26–35.

5 Chang HW, Yan D, Singh R *et al.* Alteration of the cutaneous microbiome in psoriasis and potential role in Th17 polarization. *Microbiome* 2018; **6**:154.

6 Bouslimani A, da Silva R, Kosciolek T *et al.* The impact of skin care products on skin chemistry and microbiome dynamics. *BMC Biol* 2019; **17**:47.

7 Huang S, Haiminen N, Carrieri AP *et al.* Human skin, oral, and gut microbiomes predict chronological age. *mSystems* 2020; **5**: e00630-19.

8 Brandwein M, Katz I, Katz A, Kohen R. Beyond the gut: skin microbiome compositional changes are associated with BMI. *Hum Microbiome J* 2019; **13**:100063.

9 Fierer N, Hamady M, Lauber CL, Knight R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci U S A* 2008; **105**:17994–9.

10 Dimitriu PA, Iker B, Malik K *et al.* New insights into the intrinsic and extrinsic factors that shape the human skin microbiome. *MBio* 2019; **10**:e00839–19.

11 Schmidt TSB, Raes J, Bork P. The human gut microbiome: from association to modulation. *Cell* 2018; **172**:1198–215.

12 Nöthlings U, Krawczak M. [PopGen. A population-based biobank with prospective follow-up of a control group]. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 2012; **55**:831–5 (in German).

13 Nöthlings U, Hoffmann K, Bergmann MM, Boeing H. Fitting portion sizes in a self-administered food frequency questionnaire. *J Nutr* 2007; **137**:2781–6.

14 Dehne LI, Klemm C, Henseler G, Hermann-Kunz E. The German Food Code and Nutrient Data Base (BLS II.2). *Eur J Epidemiol* 1999; **15**:355–9.

15 Holle R, Happich M, Löwel H *et al.* KORA – a research platform for population based health research. *Gesundheitswesen* 2005; **67** (Suppl. 1):S19–25.

16 Rathmann W, Haastert B, Icks A *et al.* High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. *Diabetologia* 2003; **46**:182–9.

17 Hartmann BM, Vasquez-Caicedo AL, Bell S *et al.* The German nutrient database: basis for analysis of the nutritional status of the German population. *J Food Compos Anal* 2008; **21** (Suppl. 1):S115–18.

18 Mitry P, Wawro N, Six-Merker J *et al.* Usual dietary intake estimation based on a combination of repeated 24-H food lists and a food frequency questionnaire in the KORA FF4 cross-sectional study. *Front Nutr* 2019; **6**:145.

19 Baurecht H, Rühlemann MC, Rodríguez E *et al.* Epidermal lipid composition, barrier integrity, and eczematous inflammation are associated with skin microbiome configuration. *J Allergy Clin Immunol* 2018; **141**:1668–76.e16.

20 Wang J, Thingholm LB, Skiecevičienė J *et al.* Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat Genet* 2016; **48**:1396–406.

21 Kozich JJ, Westcott SL, Baxter NT *et al.* Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; **79**:5112–20.

22 Callahan BJ, McMurdie PJ, Rosen MJ *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13**:581–3.

23 Callahan B. DADA2 Pipeline Tutorial (1.16). Available at: https://benjjneb.github.io/dada2/tutorial.html (last accessed 5 April 2021).

24 Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007; **73**:5261–7.

25 Cole JR, Wang Q, Fish JA *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014; **42**:D633–42.

26 Oksanen J, Blanchet FG, Friendly M *et al.* vegan: Community Ecology Package. Available at: https://cran.r-project.org/web/packages/vegan/index.html (last accessed 5 April 2021).

27 R Core Team (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Available at: https://www.R-project.org (last accessed 6 April 2021).

28 Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019; **35**:526–8.

29 Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002; **2/3**:18–22.

30 Kuhn M, Wing J, Weston S *et al.* caret: Classification and Regression Training. R package version 6.0–84. 2020. Available at: https://rdrr.io/cran/caret (last accessed 3 May 2021).

31 Robin X, Turck N, Hainard A *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; **12**:77.

32 Gorman B. mltools: Machine Learning Tools. R package version 0.3.5. 2018. Available at: https://cran.r-project.org/web/packages/mltools/index.html (last accessed 6 April 2021).

33 Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* 1997; **65** (Suppl. 4):1220S–1228S; discussion 9S–31S.

34 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**:550.

35 Scholz CFP, Kilian M. The natural history of cutaneous propionibacteria, and reclassification of selected species within the genus *Propionibacterium* to the proposed novel genera *Acidipropionibacterium* gen. nov., *Cutibacterium* gen. nov. and *Pseudopropionibacterium* gen. nov. *Int J Syst Evol Microbiol* 2016; **66**:4422–32.

36 Two AM, Nakatsuji T, Kotol PF *et al.* The cutaneous microbiome and aspects of skin antimicrobial defense system resist acute treatment with topical skin cleansers. *J Invest Dermatol* 2016; **136**:1950–4.

37 Perez Perez GI, Gao Z, Jourdain R *et al.* Body site is a more determinant factor than human population diversity in the healthy skin microbiome. *PLoS One* 2016; **11**:e0151990.

38 Moestrup KS, Chen Y, Schepeler T *et al.* Dietary control of skin lipid composition and microbiome. *J Invest Dermatol* 2018; **138**:1225–8.

39 Wichers H. Immunomodulation by food: promising concept for mitigating allergic disease? *Anal Bioanal Chem* 2009; **395**:37–45.

40 Visconti A, Le Roy CI, Rosa F *et al.* Interplay between the human gut microbiome and host metabolism. *Nat Commun* 2019; **10**:4505.

41 Ross AA, Doxey AC, Neufeld JD. The skin microbiome of cohabiting couples. *mSystems* 2017; **2**:e00043-17.

42 Song SJ, Lauber C, Costello EK *et al.* Cohabiting family members share microbiota with one another and with their dogs. *eLife* 2013; **2**:e00458.

43 Baurecht H, Hotze M, Brand S *et al.* Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms. *Am J Hum Genet* 2015; **96**:104–20.

44 Cao C, Xiao Z, Wu Y, Ge C. Diet and skin aging – from the perspective of food nutrition. *Nutrients* 2020; **12**:870.

45 Morita A. Tobacco smoke causes premature skin aging. *J Dermatol Sci* 2007; **48**:169–75.

46 Manus MB, Yu JJ, Park LP *et al.* Environmental influences on the skin microbiome of humans and cattle in rural Madagascar. *Evol Med Public Health* 2017; **2017**:144–53.

47 Meisel JS, Hannigan GD, Tyldsley AS *et al.* Skin microbiome surveys are strongly influenced by experimental design. *J Invest Dermatol* 2016; **136**:947–56.

48 Stehlikova Z, Kostovcik M, Kostovcikova K *et al.* Dysbiosis of skin microbiota in psoriatic patients: co-occurrence of fungal and bacterial communities. *Front Microbiol* 2019; **10**:438.

49 Johnson JS, Spakowicz DJ, Hong BY *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 2019; **10**:5029.

50 Segata N. On the road to strain-resolved comparative metagenomics. *mSystems* 2018; **3**:e00190-17.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1 Supplementary materials and methods.

Table S2 Sample details and physiological parameters from PopGen participants.

Table S3 16S ribosomal RNA amplicon sequences shared between sampled sites.

Table S4 Skin sample details of study participants.

Figure S1 Filtering of putative bacterial contaminants from skin samples.

Figure S2 Sequencing runs and rarefaction sample size.

Figure S3 Models used in association analyses.

Figure S4 Effect of confounder adjustment on null hypothesis rejections and effect sizes.

Figure S5 Flowchart of participants and analysis data sets.

Figure S6 Age, body mass index, sex and lifestyle were associated with the skin bacteria in the KORA FF4 cohort.

Figure S7 Lifestyle factors were associated with skin bacteria of the genera *Corynebacterium*, *Propionibacterium* and *Staphylococcus*.

Figure S8 Associations found across PopGen and KORA FF4 cohorts.

Table S1 Results of beta diversity association analyses.

Table S5 Results of amplicon sequence variant association analyses.