

RESEARCH

Open Access



Integrated analysis of facial microbiome and skin physio-optical properties unveils cutotype-dependent aging effects

Chuqing Sun^{1,2,3†}, Guoru Hu^{1†}, Liwen Yi⁴, Wei Ge⁵, Qingyu Yang⁵, Xiangliang Yang^{4,6,7}, Yifan He^{8,9,10*}, Zhi Liu^{4*} and Wei-Hua Chen^{1,11*}

Abstract

Background Our facial skin hosts millions of microorganisms, primarily bacteria, crucial for skin health by maintaining the physical barrier, modulating immune response, and metabolizing bioactive materials. Aging significantly influences the composition and function of the facial microbiome, impacting skin immunity, hydration, and inflammation, highlighting potential avenues for interventions targeting aging-related facial microbes amidst changes in skin physiological properties.

Results We conducted a multi-center and deep sequencing survey to investigate the intricate interplay of aging, skin physio-optical conditions, and facial microbiome. Leveraging a newly-generated dataset of 2737 species-level metagenome-assembled genomes (MAGs), our integrative analysis highlighted aging as the primary driver, influencing both facial microbiome composition and key skin characteristics, including moisture, sebum production, gloss, pH, elasticity, and sensitivity. Further mediation analysis revealed that skin characteristics significantly impacted the microbiome, mostly as a mediator of aging. Utilizing this dataset, we uncovered two consistent cutotypes across sampling cities and identified aging-related microbial MAGs. Additionally, a Facial Aging Index (FAI) was formulated based on the microbiome, uncovering the cutotype-dependent effects of unhealthy lifestyles on skin aging. Finally, we distinguished aging related microbial pathways influenced by lifestyles with cutotype-dependent effect.

Conclusions Together, our findings emphasize aging's central role in facial microbiome dynamics, and support personalized skin microbiome interventions by targeting lifestyle, skin properties, and aging-related microbial factors.

Keywords Facial microbiome, Cutotype, Metagenome-assembled genomes, Aging, Lifestyle

Background

Our facial skin homes millions of microorganisms, predominately bacteria, playing vital roles in skin health. These functions include maintaining the physical barrier, protection against pathogens, modulating immune response, and metabolizing bioactive materials [1–4]. Facial skin has unique characteristics compared with other skin regions [5], such as increased sebaceous glands density, particularly in the T-zone (forehead, nose, and chin), contribute to enhanced sebum production, preventing dryness, and aiding skin protection [6, 7]. Moreover, constantly exposure to environmental factors such

[†]Chuqing Sun and Guoru Hu contributed equally to this study.

*Correspondence:

Yifan He

heyifan@btbu.edu.cn

Zhi Liu

zhiliu@hust.edu.cn

Wei-Hua Chen

weihuachen@hust.edu.cn

Full list of author information is available at the end of the article



as UV radiation, pollutants, and microorganisms shapes the distinct facial microbiome that differs from other skin area [8].

The facial microbiome plays an indispensable role in skin health, with profound implications for both academia and industry. Aging is a recognized factor influencing the composition and function of the facial microbiome [9, 10]. Studies show that elderly individuals exhibit significantly higher microbiome diversity [9, 11], potentially linked to the reduced immune capabilities associated with increased age [12]. Significant distinctions *Actinobacteria* and *Proteobacteria* populations have been observed between the age groups, with specific bacterial enrichment of *Cutibacterium* and *Staphylococcus* among Korean women in their 20–30 s, and *Streptococcus* and *Pseudomonas* in individuals in their 60 s [13]. Other studies [12, 14] have come across similar results. Notably, facial microbiome composition demonstrates a stronger prediction capacity on chronological age than the gut microbiome, indicating a stronger association between aging and changes in facial microbiome structure [15, 16].

Conversely, the facial microbiome impacts the host skin phenome by regulating skin immunity and inflammation [17, 18], inhibiting pathogen invasion [19, 20], and maintaining pH and stratum corneum hydration [21, 22] through various mechanisms [23], opening avenues for facial microbiome interventions to enhance host skin physiological properties [24].

However, targeting aging-related facial microbes can be challenging due to correlations with skin physiological properties, with themselves change with age. Aging-related physiological changes, such as decreased collagen, sebaceous glands, and sweat [25], affect skin surface physiology [26, 27], influencing the facial microbiome [28]. Notably, certain age-related physio-optical features shape common indicator microbes with age. Particularly, studies show abundances of *Cutibacterium* positively while *Streptococcus* negatively correlated with sebaceous gland area [29], and many other factors [10, 25, 30]. These results raise questions about the comprehensive understanding of how age influences the skin microbiota.

To address the role of aging in shaping the facial microbiome while controlling for skin physio-optical properties, we conducted a comprehensive study involving 479 volunteers across three regions in China, spanning ages 18 to 64 with a well-distributed representation across age-intervals. Employing deep metagenome sequencing on the facial microbiome and state-of-art equipment to assess skin characteristics, we established the Facial Microbiome Genome Compendium (FMGC). This compendium better represented our and surprisingly public facial microbiome datasets than existing skin

microbial catalogs such as the Skin Microbial Genome Collection (SMGC) [31], and the Unified Human Skin Genome (UHSG) catalog [32]. We then further characterized our data and focused on four key aspects: identifying distinct cutotypes (a classification or grouping of individuals based on the composition and abundance of microbial species found on their skin) correlated with age and skin properties, revealing aging as the primary driver of the facial microbiome, developing a microbiome-based Facial Age Index (FAI) to assess the impact of lifestyle habits on skin aging, and identifying aging-related microbial species and pathways. Notably, our research underscores that the acceleration of aging induced by unhealthy lifestyles is cutotype-dependent. This nuanced understanding of the relationship between aging and the facial skin microbiome sheds light on potential personalized interventions for skin microbiome management. In summary, our study provides a comprehensive dataset, sheds light on the role of aging in skin microecology, and lays the foundation for tailored intervention approaches targeting lifestyle habits, skin characteristics, and microbial factors.

Results

A facial microbiome genome compendium that better represented our and public facial metagenomic datasets through multi-center sampling and deep-sequencing

To enhance our understanding of the human facial microbiome and its association with aging and the physio-optical conditions of the skin, we conducted a study involving 479 healthy Chinese individuals of both genders and various ages (ranging from 18 to 64) from Beijing, Wuhan, and Guangzhou in China. These regions represent the Northern, Central, and Southern parts of China, respectively (Table S1). Inclusion of these regions is beneficial to the broad representation of constructed metagenome-assembled genomes (MAGs). We collected 498 skin samples from both cheeks of the volunteers, subjected them to deep metagenomic sequencing (Fig. 1A). After eliminating vector sequences and low-quality bases, we obtained an average of 102 million read pairs per sample. We estimated that 71.35% of these reads (interquartile range (IQR): 63.20~85.24%) were attributed to human DNAs, consistent with findings from previous studies [5]. Subsequent removal of host contaminations yielded an average of 29 million read pairs per sample. To mitigate potential impacts induced by physiological characteristics and personal lifestyle habits, we assess the physio-optical features of participants' facial skin, including moisture, sebum, gloss, skin elasticity, and sensitivity (Tivi), using state-of-the-art equipment ("Methods" section), while relevant lifestyle information from the participants were collected (Table S1).

A comprehensive skin reference microbial genome catalogue is essential to characterize our samples. Recent studies have generated several such catalogs, including the SMGC [31] and the UHSG [32] catalog. SMGC was assembled from a total of 2479 skin metagenome samples including 506 facial samples and 1973 from other body sites, while UHSG used a total of 450 facial and 2069 other skin metagenomes. We thus evaluated their representativeness on our dataset by aligning our clean reads to these catalogs and calculated the mapping rates. Surprisingly, only 66.9% and 59.4% of our reads could be mapped to the two catalogs, respectively. Additionally, metaPhlAn4 [34], a widely used taxonomic profiling tool, reported ~45% unaccounted abundances by existing reference genomes (Fig. S1). Therefore, we have made the decision to assemble our own MAG as the basis to enhance the annotation and characterization of our data.

To construct our own set of MAGs, we employed a bioinformatics pipeline similar to SMGC [31] (“Methods” section). Briefly, we filtered out MAGs that were ≤ 200 kbp or deemed low-quality, high-contamination, or both based on CheckM [35] and de-replicated the remaining MAGs at an average nucleotide identity (ANI) of 95% (“Methods” section). We finally generated a comprehensive set of 2737 non-redundant species-level MAGs and referred it as the Facial Microbiome Genome Compendium (FMGC) (Table S2).

As expected, the FMGC genomes better represented our facial microbiome samples by recruiting a median of 80.0% sequencing reads, significantly higher than the public catalogues ($p < 0.001$, Wilcoxon rank sum test, Fig. 1B). Strikingly, FMGC also better represented available public facial microbiome datasets. For example, on the SMGC facial samples ($n = 506$), FMGC recruited 95.8% of the metagenomic sequencing reads, significantly higher than that by UHSG (91.3%, $p < 0.001$, Wilcoxon rank sum test; Fig. 1C). Similarly, FMGC recruited 90.37% of the sequencing reads of the UHSG

facial samples ($n = 120$), higher than SMGC (~86.61%, $p < 0.001$, Wilcoxon rank sum test; Fig. 1C). For fair comparison, we conducted the reads recruitment using the three above mentioned catalogs on an independent public dataset used by the integrated Human Skin Microbial Gene Catalog (iHSMGC) [5]. FMGC recruited 68.5% of the sequencing reads of the iHSMGC samples ($n = 822$), higher than the SMGC (56.3%) and UHSG (50.98%, $p < 0.001$, Wilcoxon rank sum test; Fig. 1C).

As the iHSMGC dataset encompassed facial samples from multiple regions, while FMGC only collected from the cheek, we also examined whether the representativeness of the FMGC varied across different facial locations. The results revealed no significant differences in the representativeness of FMGC among the cheeks, forehead, and nose regions. We further conducted principal coordinate analysis (PCoA) on the iHSMGC samples, employing FMGC for abundance calculations to evaluate the overall microbial structure across these three facial regions. The results demonstrated a lack of significant differences among these areas (Fig. S2, $p = 0.7$), underscoring the consistent representativeness of FMGC across distinct facial regions.

To quantify the novelty in FMGC genomes to facial microbiome, we compared the genomic sequences with those in public reference skin microbial catalogs. At an ANI threshold of 95%, 752 and 807 of the FMGC MAGs could be found in SMGC and UHSG, respectively, totaling 1029 MAGs (Fig. S3A). In addition, 1644 and 1014 could also be found in the reference microbial genomes used by GTDB-tk [33] and MethPhlAn4 [34] at such an ANI threshold. Specifically, using GTDB-tk [33], we assigned all FMGC MAGs to either bacteria ($n = 2734$) or archaea ($n = 3$) at the kingdom level, 94–100% of them to known genus (94.26%) and higher taxonomic levels, and ~60% to known species (Table S1; Fig. S3B). Together, a total of 809 MAGs were considered as novel, representing 30% of the FMGC genomes (Fig. 1D and

(See figure on next page.)

Fig. 1 Construction and analysis of the Facial Microbiome Genome Compendium (FMGC). **A** Overall workflow of this study including the multi-center sample and meta info collection, measurement of skin physio-optical properties using industrial-grade instruments such as Corneometer, pH meter, Glossometer, Sebumeter, and Cutometer from Courage and Khazaka, Visia-CR from Canfield and Tivi700 from WheelsBridge, generation of a representative Facial Microbiome Genome Compendium (FMGC) consisting 2737 species level genomes and integrative analyses of the (meta)genomic features with the host meta-data and skin features. **B–D** Mapping rates comparison: Comparison of mapping rates of facial metagenome sequencing reads to MAGs from FMGC, SMGC, and UHSG. The facial metagenome data included **B** 498 sample from this study; **C** 506 sample from SMGC study and randomly selected 120 sample from UHSG study; and 822 sample including 3 sample sites from the iHSMGC study [5]. **D** Novelty assessment: Evaluation of FMGC MAGs’ novelty based on their overlap with reference microbial genomes in the SMGC and those used by GTDB-tk and metaPhlAn4. **E** Phylogenetic analysis of the non-redundant 3359 MAGs combined from the FMGC and SMGC genomes. The orange color strip for FMGC MAGs, grey for SMGC MAGs, and blue for UHSG MAGs, and light pink for novel MAGs in FMGC. The outer ring is colored by phyla that were annotated using GTDB-tk [33]. **F** Phylogenetic diversity (branch length of selected MAGs) expansion of the FMGC over the SMGC genomes at the phylum level. The top ten 10 phyla with the highest number of MAGs were selected. **G** Expansion of the FMGC MAGs over the SMGC genomes at the genus level. Top 10 genera with the largest number of MAGs were selected. Statistical significance indicated by: ns $p > 0.05$, $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Wilcoxon rank sum test

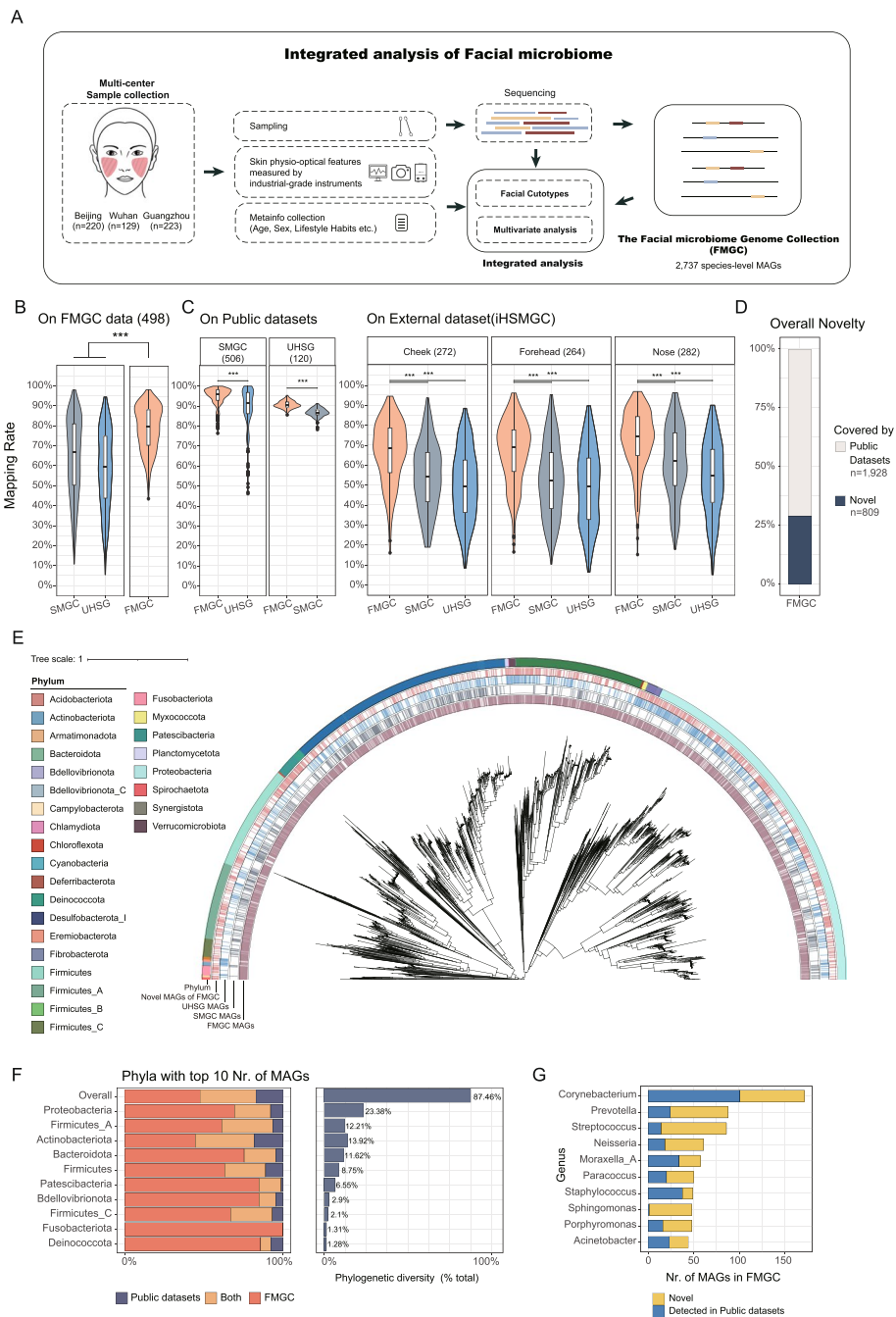


Fig. 1 (See legend on previous page.)

S3B). Importantly, these novel MAGs exhibited even higher qualities in genome size, N50 length and completeness while maintaining the same contamination level (~2.58%) as compared with those overlapping with the aforementioned datasets (Fig. S4; Wilcoxon rank sum test). These findings suggest that the novel MAGs likely represent genuine novel genomes rather than low-quality fragments resulting from mis-assembly.

We also quantified the contribution of FMGC to the increased genome diversity at different taxonomic levels. To achieve this, we merged all species-level MAGs from FMGC, SMGC, and UHSG, and obtained a total of 4167 MAGs. Remarkably, FMGC MAGs exhibited an expansive representation across diverse branches in the phylogenetic tree (Fig. 1E), in contrast to certain branches in public datasets that lacked representation (Fig. S5). In

total, the FMGC contributed to a 47% increase in phylogenetic diversity throughout the entire tree (Fig. S5). Further examination at the phylum level revealed that most phyla displayed similar levels of increase (Fig. 1F). Similar trends were observed at the genus level (Fig. 1G). Notably, the *Sphingomonas* genus of the *Proteobacteria* phylum stands out, with 47 out of 48 MAGs exclusively detected in FMGC and only one MAG found in existing datasets (Fig. 1G).

Together, FMGC demonstrated a higher representativeness compared to existing public catalogues, contributing to a substantial increase in phylogenetic diversity across various taxonomic levels.

FMGC reveals two consistent cutotypes in facial samples across regions

Cutotypes offer a comprehensive view of the inter-individual variation in facial microbial composition [5, 36]. A cutotype is a classification or grouping of facial microbiome samples (or based on the microbial composition into different skin microbiome profiles or community types. Community types in the gut bacteriome and mycobiome are associated with disease risks [37] and aging [38]. Similarly, we would expect that cutotypes can also help researchers to better understand the role of the skin microbiome in maintaining health and preventing skin-related issues. To identify and characterize cutotypes in our samples, we calculated the relative abundances at the genus-level by aligning the sequenced reads to the FMGC MAGs using BWA (“Methods” section). Genus-level abundances were chosen because community types are commonly determined at this level [39, 40], and the GTDB-tk tool we used for taxonomic classification could assign most of our MAGs (94.26%) to known genera (Fig. S6), ensuring the reliability of our results. Applying the Partitioning Around Medoids (PAM) clustering algorithm (“Methods” section), we obtained two cutotypes: the *Cutibacterium*-cutotype (*C*-cutotype) and *Streptococcus*-cutotype (*S*-cutotype; Fig. 2A). Importantly, we consistently observed these cutotypes when analyzing samples from individual regions or when combining them (Fig. 2A). And only a small fraction of samples (~6.2%) exhibited an additional cutotype, namely *Enterobacter*-cutotype.

In comparison, we constructed cutotypes using the genus-level abundances annotated by metaPhlAn4 [34], one of the most widely used taxonomic profiling tools. However, it is important to note that metaPhlAn4 assigned a median of only 54% abundances to known genera, resulting in fewer annotated genera and lower α -diversity compared to FMGC in our samples (Fig. S6). When analyzing all samples combined, metaPhlAn4 generated two cutotypes: the *Prevotella*-cutotype and the

Cutibacterium-cutotype. However, significant inconsistencies emerged when analyzing the samples separately by region, ~51% of samples’ cutotypes were discordant. Specifically, ~42% *Prevotella*-cutotype samples shift to *Streptococcus*-cutotype (~29% to *Corynebacterium*-cutotype; ~18.22% to *Malikia*-cutotype; ~9.81% to *Porphyromonas*-cutotype; ~0.93% to *Cutibacterium*-cutotype), while ~4.2% *Cutibacterium*-cutotype samples shift to *Corynebacterium*-cutotype (~2.9% to *Porphyromonas*-cutotype).

These findings highlight the importance of using FMGC as a better catalogue of reference genomes for further characterizing the facial microbiome, emphasizing the need for more complete and accurate reference genomes to obtain stable and reliable cutotype classifications for further investigations.

Age as a determinant of facial microbiome composition

With the availability of skin physio-optical features and participants’ metadata, we investigated the influence of these factors on facial microbiome. Remarkably, age, moisture, and sebum were identified as key factors that correlated with both cutotypes and driver taxa, exhibiting distinct trends (Fig. 2C). Specifically, we observed a notable decline in the proportion of the *C*-cutotype with increasing age and higher levels of skin moisture, while it exhibited a significant increase with elevated sebum levels (Fig. 2C). Conversely, the *S*-cutotype displayed an opposite pattern, indicating a contrasting response to these skin properties (Fig. 2C, S7). These findings were further supported by the relative abundance of the two driver genera, *Streptococcus* and *Cutibacterium*, which exhibited similar patterns to their respective cutotypes (Fig. 2D, S8). These results are consistent with previous findings [25, 30].

We further quantified the influence of the aforementioned factors on the facial microbiome using an Envfit test (“Methods” section). Consistent with the cutotype analysis, all of these factors demonstrated significant correlations with the metagenome structure (measured with axis 1, and 2 of the principal coordinate analysis (PCoA) of facial microbial profiles), with the exception of R2 (an elasticity measurement) and pH (Fig. 2E). Among them, age, moisture, sebum, and sensitivity emerged as the top four factors (Fig. 2E). Notably, significant inter-correlations were observed among these factors, with age demonstrating correlations with most of the other factors (Fig. 2E). Specifically, age displayed negative correlations with skin elasticity measures (R2, R5, and R7), sebum, and gloss, while demonstrating a positive correlation with Sensitivity (Tivi). To better understand the individual contributions of these factors, we conducted a mediation analysis [41–43], which revealed that age had

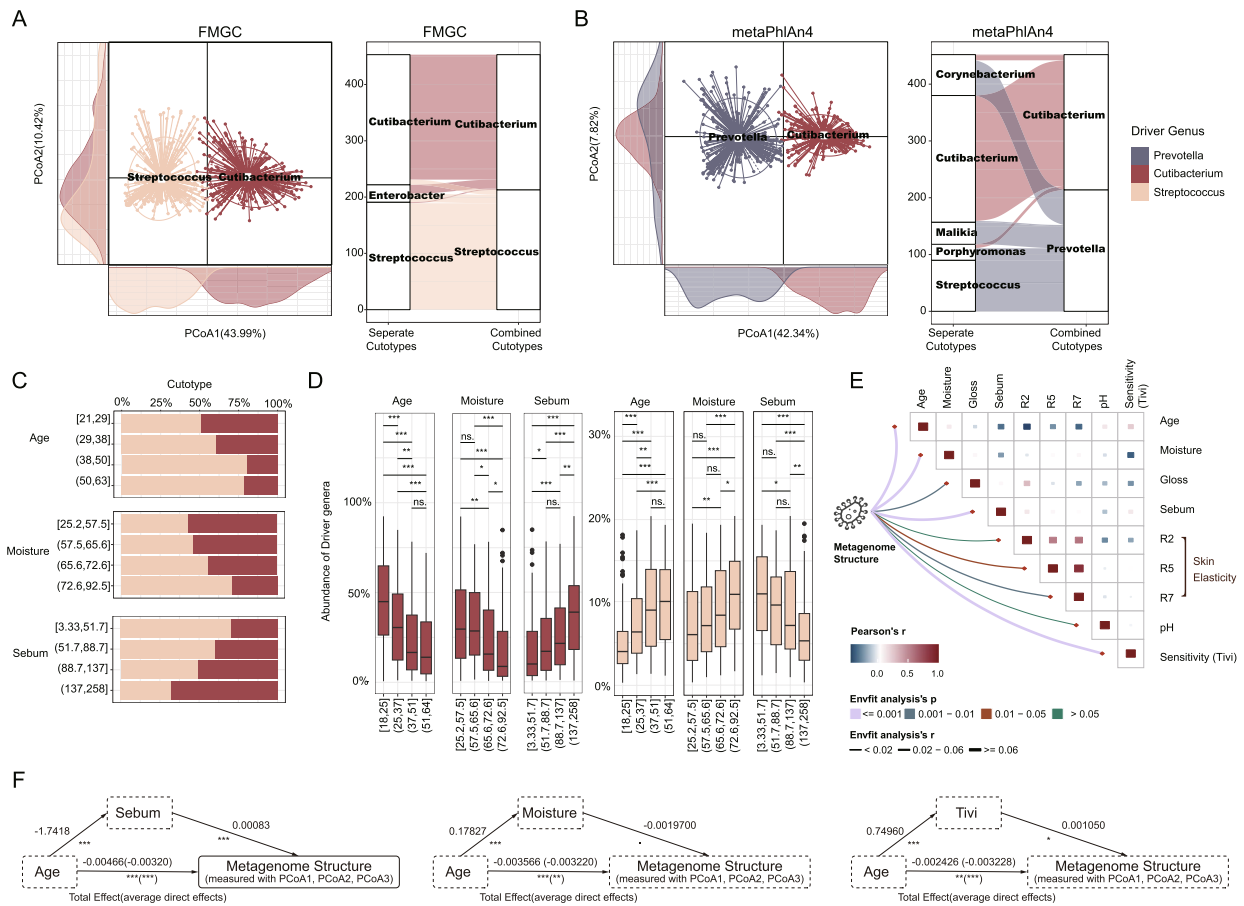


Fig. 2 Two consistent cutotypes built with FMGC and the correlation between microbial community structure and aging and skin physio-optical properties. Cutotypes constructed using genus-level abundances derived from the **A** FMGC MAGs and **B** metaPhiAn4. The PCoA plots display the cutotypes built with all samples together. The Sankey diagram illustrates the analysis of samples separated by region and the combined analysis of all samples, showcasing the variation in cutotypes. **C** The proportion of different cutotypes observed, when age, moisture, and sebum are evenly grouped according to their distributions. **D** The relative abundance of the driver of the cutotypes, when age, moisture, and sebum are evenly grouped according to their distributions. **E** The tile map showing the Pearson correlation between collected meta info and skin physio-optical properties. The width of curves is the explained variance of these factors to metagenome structure calculated with Envfit test, and colors for its *p*-value. **F** Mediation analysis examining relationship between age and microbiome structure, considering significant influencing factors such as sebum, moisture, and sensitivity. The arrow from the independent variable (age) to the mediating variable (sebum, moisture, and Tivi) illustrates the influence of the independent variable on the mediating variable. Likewise, the arrow from the mediating variable to the dependent variable (microbiome structure) demonstrates the impact of the mediating variable on the dependent variable. Finally, the arrow from the independent variable to the dependent variable indicates the overall effect of the independent variable on the dependent variable, encompassing both total effect and average direct effect. Average direct effects (ADE) refers to the direct effects of the independent variable on the dependent variable. It represents the effect of the independent variable on the dependent variable while controlling for the mediating variable(s). Total effect, on the other hand, refers to the overall effect (direct + indirect) of the independent variable on the dependent variable. $p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, Wilcoxon rank sum test

the most substantial and direct impact on the community structure of the facial microbiome (measured with axis 1, 2, and 3 of PCoA, Fig. 2F). For example, age exerts a significant influence on both sebum levels and the structure of the microbiome, where the latter is influenced both directly and indirectly by age; when accounting for the mediator, age continues to exhibit a significant average direct effect, which is stronger than that of the sebum level to the facial microbiome (Fig. 2F).

Together, age stood out as the factor exhibiting the strongest correlation with the abundances of specific bacterial species in our exploration of the correlation between individual bacterial species and the physio-optical properties of the skin, as well as other meta-information (partial Pearson correlation, $p < 0.05$; Fig. S9). Considering that age is an independent characteristic not influenced by any other factor, while other physiological properties are correlated with age (e.g., the older the age,

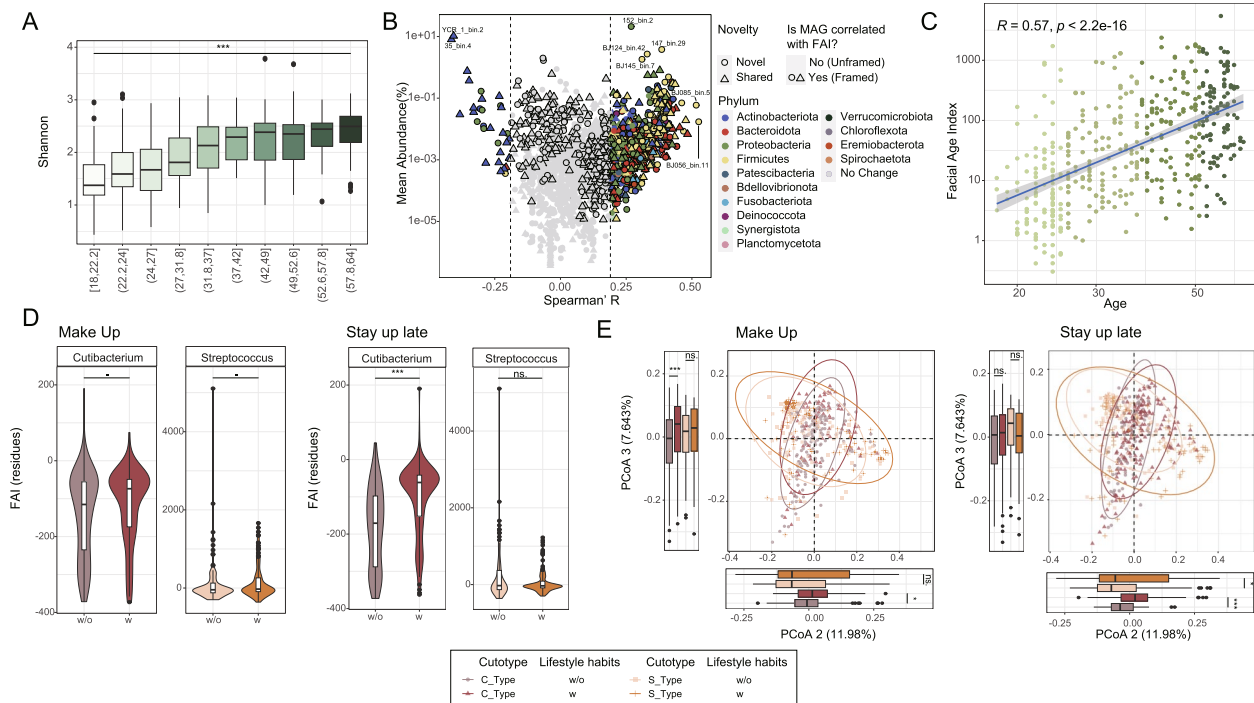


Fig. 3 A microbiome-based facial age index (FAI) reveals effects of suboptimal lifestyle habits on skin aging. **A** Barplot showing the increase in Shannon diversity with advancing age. **B** Scatter plot displaying MAGs correlated with the calculated Facial Age Index (FAI). Labeled dots represent MAGs with mean abundance higher than 1% or relatively higher absolute Spearman's R (≥ 0.5). Framed dots indicate MAGs that also exhibit strong correlations with FAI. **C** Facial age index (FAI) built with marker microbes and its significantly positive correlation with chronological age, confirming the practicability of using facial skin microbes to access skin age. Noted that Wuhan samples were not included in the construction of FAI due to its limited age range and inclusion of acne samples. **D** Boxplot showing the FAI-residues (with the effects of the chronological age removed) in participants with/without certain lifestyle habits and different cutotypes. **E** PCoA plot demonstrating the difference of overall microbiome structure among participants with/without certain lifestyle habits and different cutotypes. $p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, Wilcoxon rank sum test

the less sebum), we used it as a primary variable in our mediation analysis. Following the same logic, we argue that the observation of a stronger age-microbe correlation may suggest a causal relationship from age to microbial composition.

FMGC reveals hundreds of novel age-related MAGs

We next focused on the age-related effects in the facial microbiome. We found a significant increase in the α -diversity of the facial microbiome, as indicated by the Shannon index, with advancing age (Fig. 3A). This observation was supported by an overall Pearson correlation coefficient of 0.47 and a p value < 0.001 (Fig. S10), consistent with previous research [11, 16]. The overall trend was partially driven by the higher α -diversity of the *S*-cutotype comparing to the *C*-cutotype (Fig. S11) and the increased proportion of the *S*-cutotype in the older age groups (Fig. 2C). Furthermore, similar patterns, such as the increased α -diversity with advancing age, were consistently observed within individual cutotypes (Fig. S12). These trends may reflect the declining functionality of the

facial immune system, resulting from continuous exposure to environmental and/or chemical factors [44]. Furthermore, they suggest a potential connection between aging, increased microbiome diversity and susceptibility to certain diseases.

To identify individual microbial genomes (MAGs) significantly correlating with age, we scrutinized the dataset, excluding 269 samples with potential confounding factors such as acne and make-up habits [45, 46]. The refined dataset of 229 samples revealed a total of 685 age-associated MAGs ($p < 0.05$, Pearson's correlation; Fig. 3B, Table S3). Among these, 652 MAGs (338 novel MAGs) displayed positive correlations with age (pMAGs), while only 31 (21 novel MAGs) exhibited negative correlations (nMAGs). Interestingly, a closer inspection revealed that the *Prevotella* genus, known as a keystone genus in human gut while often found associated with human infection in other body sites [47], contained the highest number of pMAGs (47 out of 652; Fig. S13A), followed by *Streptococcus* and *Neisseria* (44 and 30, respectively), two of the known members of the oral microbiota. These

two genera include several pMAGs capable of opportunistic pathogenicity, such as *Streptococcus oralis* [48] (5 MAGs, including 3 novel ones), *Streptococcus mitis* [49] (21 MAGs, 14 novel ones), *Neisseria subflava* [50] (5 MAGs, 1 novel). Surprisingly, there are 8 pMAGs (6 novel ones) belong to the notorious pathogen *Streptococcus pseudopneumoniae* [49]. On the other hand, *Corynebacterium* contains MAGs from different species correlated positively (27 MAGs) or negatively (12 MAGs) with age (Fig. S13A). For example, BJ042_bin.19 is a novel pMAG of *Corynebacterium macginleyi*, reported as a pathogen in case of unilateral conjunctivitis [51], that positively correlates with age showing a Spearman's R of 0.42, while HC31_bin.3 is a *Corynebacterium singular* negatively correlates with age exhibiting a Spearman's R of -0.33 (Table S3), suggesting heterogeneity exists within the *Corynebacterium* genus. Besides, the *Cutibacterium* contains the second highest number of negatively correlated MAGs (7 nMAGs), consistent with our above observations.

Focusing on MAGs with stronger correlations (absolute Spearman's $R \geq 0.5$) or with high mean relative abundance ($\geq 1\%$) across samples (Fig. S13B), we identified a subset of eight MAGs, comprising six pMAGs and two nMAGs. Among which, two nMAGs, 35_bin.4 and YCR_1_bin.2, from *Cutibacterium acnes* stood out, with a mean relative abundance ranging from ~8 to 10% and high Spearman's R, consistent with the known negative associations between *C. acnes* and aging [16]. Surprisingly, all of the six pMAGs are identified as novel MAGs, including four with no species-level assignment from four different genera (*Leuconostoc*, *Enterobacter*, *Anaerococcus*, and *Streptococcus*). Specifically, the two low-abundant yet strongly correlated pMAGs both belong to the *Anaerococcus* genus, including one from *Anaerococcus nagyae*, a species previously reported to be associated with fatal sepsis in patients undergoing transarterial chemoembolization treatment [52].

Experimental validation is imperative to elucidate the roles of these age-related MAGs, particularly the novel ones. Subsequent experiments should shed light on their functional significance.

A microbiome-based Facial Age Index reveals cutotype-dependent effects of lifestyles on skin aging

To quantitatively access the age-related effects to the facial microbiome, we further identified age-related taxa at higher taxonomic levels (Table S3) and used them together with the age-related MAGs to develop a Facial Aging Index (FAI). The FAI takes into account both variations in taxa abundance and the correlations between bacterial composition and age ("Methods" section). A higher FAI score indicates a greater degree of skin aging.

To validate the effectiveness of our approach, we constructed the FAI using data from a specific region (e.g., Guangzhou) and then tested it on samples from different regions (e.g., Beijing). We assessed the correlation between chronological age and FAIs in the test dataset. Regardless of the region used to build the FAI, the test results remained consistent across other regions (Fig. S14). Applying this method to all our samples, we discovered a positive correlation between the FAI and the chronological age of the participants, both when combined ($p < 2.2e-16$, Pearson's $r = 0.55$, Fig. 3C) and within individual cutotypes (C-cutotype: $p < 2.2e-16$, Pearson's $r = 0.54$; S-cutotype: $p = 2.3e-11$, Pearson's $r = 0.47$, Fig. S15). Specifically, 557 out of the 652 pMAGs were found to correlate with FAI, while all 31 of the nMAGs also exhibited correlations with FAI (Fig. 3B, S13C).

Interestingly, when examining the impact of different lifestyles on the facial microbiome, which are often related to skin conditions [46, 53, 54], we observed significantly higher FAI-residuals (controlled with age) in individuals with certain lifestyles, including "Make up" (excessive or heavy makeup application, defined as three or more cosmetic applications per week; see "Methods" section) and "Stay up late" (going to bed after 11 pm on three or more occasions per week, indicating habitual sleep restriction; see "Methods" section). After accounting for the participants' chronological age, we found significantly elevated FAI residuals in the "Make up" group compared to the group without make-up in both the C- and S-cutotypes (Fig. 3E). We observed similar trends in the "Stay up late" group, although the difference was only statistically significant in the C-cutotype, not the S-cutotype. Notably, since the C-cutotype was enriched with young participants (Fig. 2C), our analysis suggests that the "Stay up late" lifestyle has a greater impact on the younger population than the older population. Furthermore, we observed relatively lower sebum-level-residuals ($p < 0.1$, Wilcoxon rank sum test), significantly lower R5-residuals (skin elasticity, the higher the better; $p < 0.05$, Wilcoxon rank sum test), and higher Tivi-residuals (sensitivity, the lower the better; $p < 0.05$, Wilcoxon rank sum test), in S-cutotype samples with unhealthy lifestyles (Fig. S16). Additionally, other host conditions such as constipation and "Mood Swings" (referring to frequently significant emotional fluctuations) were also found to significantly affect skin aging (Fig. S17). It is noteworthy that "Mood Swings" specifically impact participants with the C-cutotype (Fig. S16).

We further confirmed the effects of lifestyles on facial microbiome using PCoA analysis, which taking into account differences in all microorganisms. We again observed significant differences in the overall between group diversities in participants with and without "Make

UP” and “Stay up late” lifestyles. As shown in Fig. 3E, we used the main axis 2 and 3 of the PCoA analysis to analyze the influence of lifestyles on facial microorganisms, because the main axis was mainly driven by the cutotype (Fig. S18). Our results showed that “Stay up late” significantly affected the microbial structure in both cutotypes, while “Make up” only affected the C-cutotype.

Specifically, we observed a higher relative abundance of the pMAGs in samples with makeup habits, while the nMAG were enriched in samples without makeup habits (Fig. S13D). Furthermore, we also assessed the total abundance of the pMAGs and the nMAGs, and the results remained consistent. This suggests a potential association between makeup habits and MAG-level changes in facial microbes.

Together, our findings suggest that age-associated microbial markers can serve as indicators to characterize skin age, and distinct cutotypes exhibit varied responses to external influences.

Cutotype-dependent aging effect in microbial metabolite pathways due to unhealthy lifestyles

We next investigated the microbial functional changes related to facial skin aging. We annotated the microbial functions using HUMAnN3 [55]. These functions were then grouped into seven categories based on their MetaCyc annotations [56], namely Inflammation, Nicotinamide adenine dinucleotide (NAD) consumption, Oxidation/Glycosylation, Anti-aging, Anti-Oxidation, Biosynthesis, and Others. We then conducted a PERMANOVA analysis to determine the role of aging in the abundance structure of the microbial pathways, and discover that age is the only significant factor (Fig. S19A). To investigate age-related pathways, we classified the participants into four equal-sized groups based on their age: Group A (18–25 years), B (26–37 years), C (37–51 years), and D (51–64 years). Using LefSe analysis [57], we compared the youngest group (Group A) with the oldest group (Group D) and identified 61 and 237 pathways that were significantly more abundant in Group A and D, respectively (Table S4).

Notably, pathways associated with the first three categories were known to be associated with aging [58–66]. All pathways in these three groups showed significant enrichment in the Group D (Fig. 4). We collectively refer to these pathways as AG-pathways (Table S4). For instance, the OANTIGEN-PWY pathway, categorized under Inflammation, is responsible for producing O-antigen, a repeating unit of lipopolysaccharides (LPS) known to trigger inflammatory responses [67]. We observed a significantly enrichment of this pathway in oldest group (Fig. S20) with the highest linear discriminant analysis (LDA) score of 3.30. Another pathway, PWY-5136, falls

under NAD consumption and employs NAD in the process of fatty acid β -oxidation. This pathway generates reactive oxygen species (ROS), leading to mitochondrial damage and oxidative stress [68–70]. The increasing prevalence of the NAD consumption pathway underscores its potential significance, explaining the inclusion of NAD in skincare products designed for the advanced age group. Additionally, besides the individual pathways, the cumulative abundance of pathways within each of the broad functional categories demonstrated an upward trend with advancing age (Fig. 4A).

Conversely, pathways associated with Anti-Aging, Biosynthesis, and Anti-Oxidation are found to be enriched in Group A, thus referred to as anti-AG-pathways. For example, the ARGSYN-PWY (L-arginine biosynthesis pathway) of the Anti-Aging group generates L-arginine. Arginine is often used in anti-aging skincare products because of its potential ability to support collagen production [72, 73]. Meanwhile, the PWY-6126 (super-pathway of adenosine nucleotides de novo biosynthesis II) of the Anti-Oxidation group is a metabolic pathway that participates in the de novo biosynthesis of adenosine nucleotides, which are involved in oxidative stress response, glutathione biosynthesis, and DNA synthesis under anaerobic conditions. Overall, the total abundance of the pathways within each group also displayed a decreasing trend with the advancing age (Fig. 4A). It is worth to mention that this grouping is based on sample size, and is merely for better presenting the result. Using an alternative grouping method according to skin wrinkles [74] generated essentially the same results (Fig. S21). In fact, when age was used as numerical values, we observed significant correlations between age and the pathway abundances in manners that are consistent with the grouped ages (Fig. S19B). In particular, the chronological age negatively correlated with Anti-Aging ($R = -0.31$ and $p = 1.7e-11$), Biosynthesis ($R = -0.3$ and $p = 4.9e-11$) and Anti-Oxidation ($R = -0.29$ and $p = 1.6e-10$), and positively correlated with Inflammation ($R = 0.2$ and $p = 2.4e-05$), NAD consumption ($R = 0.2$ and $p = 2.4e-05$), and Oxidation/Glycosylation ($R = 0.21$ and $p = 5.3e-06$). These results indicate that our observations are robust again the age grouping methods.

We further investigated whether the impact of the above mentioned lifestyles on the facial microbiome is reflected in microbial functions. To explore this, we controlled for chronological age differences between groups and compared pathway abundances among groups with different lifestyles. Consistent with the FAI analysis, we observed significantly higher abundances of AG-pathways and lower abundances of anti-AG-pathways in the “Make up” group (Wilcoxon rank sum test; Fig. 4B). These trends were consistently observed in both the

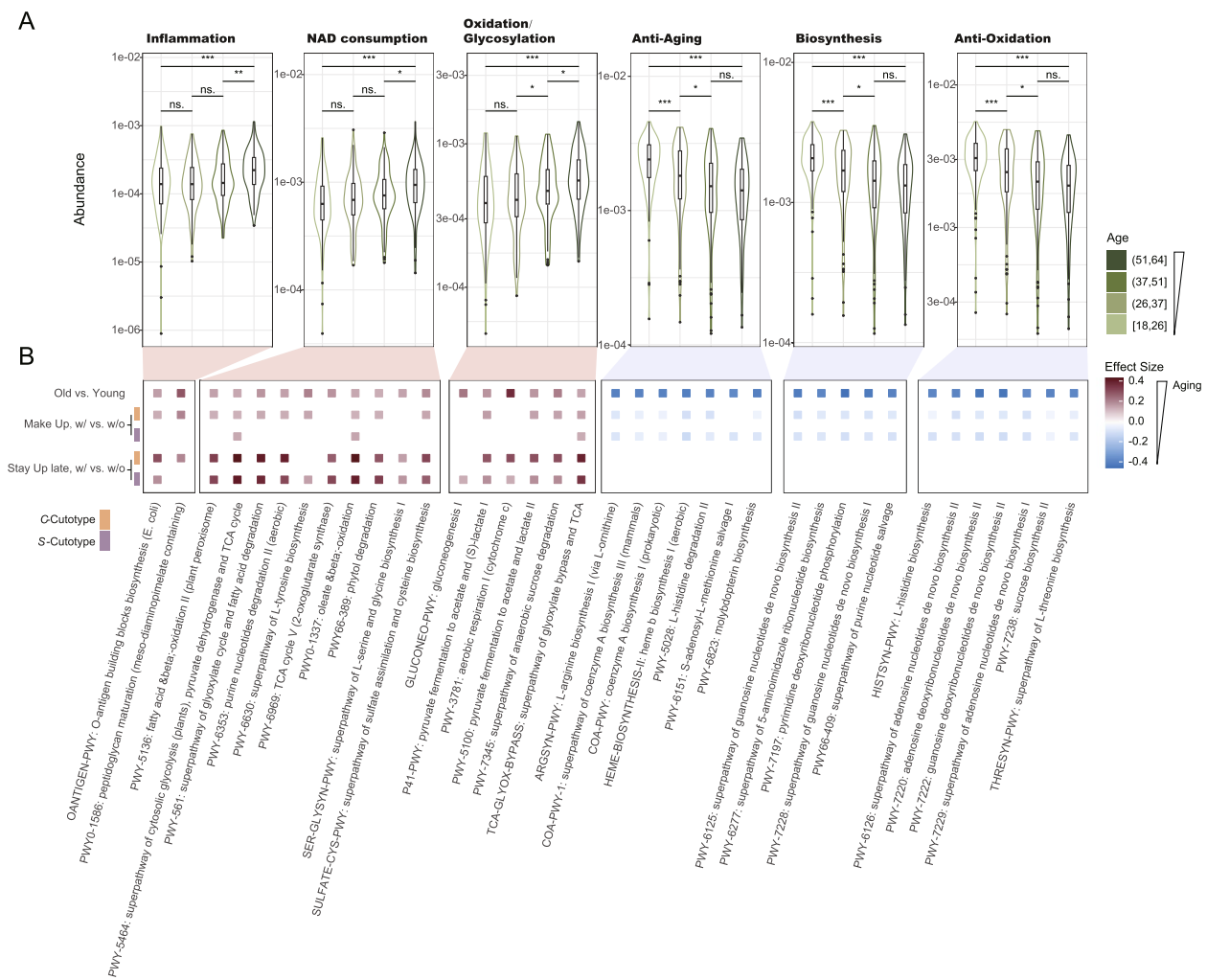


Fig. 4 Functional analysis reveals molecular pathways underlying skin aging and their associations with suboptimal lifestyle habits. **A** Boxplot showing the accumulated abundance of aging-related (Inflammation [58, 59], NAD consumption [60–62, 71], and Oxidation/Glycosylation [63–66]; red shadows) and anti-aging-related (Anti-Aging, Biosynthesis, and Anti-Oxidation; green shadows) pathways across different age groups. **B** Tile map plot displaying the effect size, as determined by the Wilcoxon rank sum test (filtered with adjusted one tailed p value < 0.1), between the youngest and oldest age groups, with or without make up habits and with or without stay-up-late habits. $p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, Wilcoxon rank sum test

C- and S-cutotype groups. Interestingly, while the FAI analysis indicated that “Stay Up Late” primarily affected the C-cutotype group, our findings revealed that metabolite pathways in participants belonging to both cutotypes were influenced by this lifestyle. However, the influence of “Stay Up Late” on microbial functions were only statistically significant in the AG-pathways but not on the anti-AG-pathways (Fig. 4B).

Should these alterations in metabolic pathways be experimentally validated, they could elucidate the potential molecular mechanisms underlying skin aging and provide potential intervention targets to counteract the aging process. Our analysis further emphasizes the

impact of specific lifestyles on aging and highlights the importance of personalized intervention strategies tailored to individual lifestyles. For example, exploring the supplementation of NAD and anti-aging metabolites may be considered as strategies to mitigate skin aging in older age groups or in individuals with habitual sleep restriction.

Discussion

Aging and alterations skin physio-optical conditions could exert similar changes to the facial microbiome, obscuring the aging-microbiome associations. Here, we elucidated the pivotal role of aging in facial microbiome

while effectively controlling for skin physio-optical properties through a multi-center and deep-sequencing survey on the facial microbiome. We sampled 479 volunteers with a broader age distribution (spanning ages 18 to 64) and well-distributed age representations across age-intervals (Fig. S22), and employed state-of-art industrial grade equipment to measure the skin physio-optical properties.

Strikingly, existing skin reference microbial catalogs including SMGC, and UHSG had low representation of our sequencing data (Fig. 1B). To address this, we constructed our own MAG catalog, the Facial Microbial Genome Compendium (FMGC), consisting of 2373 species-level MAGs. FMGC better represented our samples and effectively expanding the diversity of known skin microbial genomes by up to 47% (Fig. 1F, S5). FMGC also better represented the public skin microbiome datasets than these two public MAG catalogs (Fig. 1B, C).

Despite that SMGC ($n=2479$) and UHSG ($n=2519$) used more samples and included skin microbiome data from other body sites, we obtained significantly more MAGs. This may be attributed to two reasons. First, due to high host DNAs in the facial microbiome samples (e.g., median: 76.43%; interquartile range: 63~85%; Fig. S23), we employed deep sequencing and obtained an average of ~4.2 G usable bases after the removal of human DNA, vector sequences, and low-quality bases. This value is significantly higher than the samples used in SMGC (~0.14 G). Second, in order to determine contribution of aging in the composition of skin microbes, we sampled a relatively even distribution of age groups ranging from 18 to 64 compared to that of UHSG, which consisted of mostly young adults (Fig. S22). Considering the pivotal role of aging in skin microbiome, the limited age group might result in the loss of certain MAGs. This emphasizes the critical importance of sequencing depth and age distribution of the samples to enhance the FMGC due to the high human DNA contents and age dependent composition in facial microbiome sequencing data.

We identify two consistent cutotypes in our samples, both within and across different regions. In our study, the high rate of unclassified taxa (~45%) obtained from public taxonomic profiling tools has further contributed to the observed inconsistencies among samples from various regions (Fig. 1A). Cutotypes have been previously reported in facial skin microbiome by Li and colleagues [5] but the results also suffered from high fractions of unknown taxa according to metaPhlan4 (~45%, Fig. S24). Consequently, the authors reported two cutotypes, namely *Cutibacterium*- and *Moraxella*-cutotype [5]. With FMGC, we recapitulated the two cutotypes but also identified an additional *Enterobacter*-cutotype that accounted for 33% ($n=268$) of the total samples

(Fig. S25). Interestingly, in these samples, *Enterobacter* is predominantly undetectable when analyzed using metaPhlan. However, when employing FMGC, the relative abundance of *Enterobacter* averages closes to 40%. These findings underscore the critical role of a highly representative genome catalog in facilitating more reliable and robust subsequent analyses. These findings are also in line with previous results that the microbiome community types can vary significantly in different populations and/or are influenced by sample sizes [40].

We then highlighted the central role of aging in shaping the facial microbiome, while properly controlling the skin physio-optical properties. Our findings demonstrated significant impacts of all examined factors on the skin microbiome compositions including the cutotypes, moisture, sebum, gloss, and pH, with the exception of skin elasticity, while also revealing significant inter-correlations among these factors. Specifically, age exhibited strong correlations with all the factors, including decreased skin elasticity, gloss, and sebum, as well as increased moisture level, pH, and sensitivity with advancing age (Fig. 2E). These observations align with prior research [10, 25, 30, 45, 75]. Notably, our discovery that *Cutibacterium* abundance positively correlates with sebum content and negatively correlates with moisture level echoes previous reports [10, 25, 30]. These conclusions are further confirmed through our multi-center sampling, affirming their robustness. Diverging from existing literature, our study employs multivariate and Mediation analyses to extensively disentangle the contributions of these factors, effectively establishing the pivotal role of aging in shaping the skin microbiome, either directly or indirectly by mediating the physio-optical properties. Consequently, our study offers a more comprehensive and unique perspective on the factors that contribute to the composition of the skin microbiome compared to previous research.

Having established the key role of aging, we further revealed its impacts on the facial microbiome from several aspects. Firstly, we have successfully identified key microbes that are specifically associated with the aging process. Building upon this, we have developed and rigorously validated a microbiome-based facial aging index (FAI). By utilizing the FAI, we have been able to effectively quantify and uncover the effects of lifestyles on the skin aging index. Furthermore, our investigation has revealed aging-related metabolite pathways and shed light on the differential impacts of lifestyles on different cutotypes. These findings imply that behavior can influence skin microbiome, and modifying behavior can enhance skin microbiome health, e.g., by reversing the facial aging index. These results offer potential avenues for targeted interventions that can specifically

target aging-related microbes, pathways, or metabolites. However, further longitudinal studies, tracking individuals over time, would be more effective in elucidating the impact of age on the skin microbiome and establishing a causal relationship. If future experiments validate our observed correlations and demonstrate their impact on skin aging, they could serve as a foundation for potential skincare interventions, and pave the way for the development of tailored skincare products and treatments that cater precisely to the unique and specific requirements of individuals.

In conclusion, our findings highlight the central role of aging on facial microbiome and strongly support personalized skin microbiome interventions by targeting lifestyle, skin properties, and aging-related microbial factors including cutotypes, key microbes, metabolic pathways, and metabolites, laying a foundation for future research and applications.

Conclusions

Our study shed light on the significant role of aging in shaping the facial microbiome, carefully considering skin physio-optical properties through a comprehensive multi-center deep-sequencing survey. These findings emphasize the potential value of personalized skin microbiome interventions aimed at addressing aging-related microbial factors, offering valuable insights for future research and applications.

Methods

Microbial sampling, skin physiology assessment, and lifestyle survey

A total of 479 volunteers were recruited from three cities, namely Beijing, Wuhan, and Guangzhou, representing the northern, central, and southern of China, respectively. A total of 498 samples were collected from these volunteers. This study was approved by the Ethics Committee of Tongji Medical College of Huazhong University of Science and Technology (No, S083).

For the volunteers in Beijing ($n=201$) and Guangzhou ($n=170$), the following requirements must be met:

1. Age 18 to 65
2. In good health and free from serious illnesses such as cancer, cardiovascular, and cerebrovascular diseases
3. No serious skin conditions
4. No antibiotics in the last 2 months
5. No acid skin care products in the last month

A two-day procedure was carried out to assess the physio-optical properties and sample the skin microbiome, as the following. Firstly, on the first day, participants are required to gently wash their faces only with

Table 1 Test indicators and the corresponding instruments used in this study

Test indicators	Instruments used
Skin moisture content	Corneometer (Courage and Khazaka, Germany)
pH value	pH meter (Courage and Khazaka, Germany)
Glossiness	Glossometer (Courage and Khazaka, Germany)
Skin surface sebum content	Sebumeter (Courage and Khazaka, Germany)
Skin elasticity (R2, R5, R7)	Cutometer (Courage and Khazaka, Germany)
Sensitivity (Tivi)	Tivi700 (WheelsBridge, Sweden)
Facial photo	Visia-CR (Canfield, USA)

water in the morning and avoid excessive rubbing, then arrived at the laboratory and rested for 20 min before any tests. During which, they were asked to fill out a questionnaire about personal data and lifestyles (Supplementary Table 1). Then tests were performed using the following instruments to measure the physical, chemical and optical properties of the facial skin, as detailed in Table 1. Secondly, to prepare for the second day, the volunteers were asked to wash their face only with water and apply a minimal amount of skin care products in the evening. They were advised to avoid acid and antibacterial skin care products, as well as to maintain a consistent diet and emotional stability for the duration of the experiment. Lastly, on the second day, the volunteers were asked not to wash their face or use any skin care products and cosmetics, and arrive at the laboratory in the morning. They rested for 20 min before the sampling. To avoid contaminations, the participants were instructed to very gently splash distilled water on their faces in our facilities.

We then used sterilized microbe-free swabs (Cat #: LC34, HYNAUT, China) to sample on both cheeks of each volunteer, wearing latex gloves that were disinfected with alcohol. We used two cotton swabs soaked with sterilized saline water (0.9%) on each cheek, wiping each cotton swab on the face for about 40 s. We used a total of four cotton swabs on both cheeks. After sampling, we broke off the cotton swab head and put it into a tube containing 2 mL preservation solution (Cat #: CY-98000A, Huachenyang, Shenzhen, China).

For volunteers in Wuhan ($n=127$), only the facial microbiome samples were collected, following the same procedure mentioned above. Volunteers may have either healthy skin ($n=71$) or suffer from acne ($n=56$), which were diagnosed by dermatologists. Other requirements remain the same as for volunteers in Beijing and Guangzhou.

In particular, several swabs were exposed to air for the same duration as the sampling process as a negative control. However, further validation indicated their low DNA content, rendering them unsuitable for sequencing.

To prevent DNA degradation, a 2 mL room-temperature stabilizing reagent kit (Cat. No. CY-98000A, Huachenyang, Shenzhen, China) was used to mix with the swab collected from each participant. The collected samples were transferred to freezers at -80°C within 12 h of collection and stored until DNA extraction.

DNA extraction, library preparation, and metagenomic sequencing

DNA extraction was carried out by adding 1000 μL of CTAB lysis buffer and 20 μL of lysozyme to the sample, followed by a 2–3-h incubation at 65°C . The supernatant was collected by centrifugation at 950 rpm, and an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added, mixed thoroughly, and centrifuged at 12,000 rpm for 10 min. The supernatant was transferred, and an equal volume of chloroform:isoamyl alcohol (24:1) was added, mixed well, and centrifuged at 12,000 rpm for 10 min. The supernatant was transferred to a 1.5 mL centrifuge tube, and 3/4 volume of isopropanol was added for precipitation at -20°C . After centrifugation at 12,000 rpm for 10 min, the liquid was discarded, and the DNA pellet was washed twice with 1 mL of 75% ethanol. The DNA sample was dissolved in 51 μL of ddH₂O, and if necessary, incubated at $55\text{--}60^{\circ}\text{C}$ for 10 min for better dissolution. RNase A (1 μL) was added to digest RNA, and the mixture was incubated at 37°C for 15 min.

Subsequently, genomic DNA was randomly fragmented by Covaris LE220 (Covaris, Inc.) according to the manufacturer's instructions. The fragmented DNA was selected by magnetic beads to an average size of 200–500 bp. The selected fragments underwent end-repair, 3' adenylated, adapters-ligation, PCR amplifying, and the products were purified by the magnetic beads. The double-stranded PCR products were heat denatured and circularized using the splint oligo sequence. The single-strand circle DNA (ssCirDNA) was formatted as the final library and qualified by QC. Sequencing was performed using the BGI Genomics DNBSEQ-T7 (PE150) sequencer.

Metagenome data processing and construction of the facial microbiome genome compendium

All the raw metagenomic data were processed using Fastp v0.23.4 [76] to eliminate adapter contamination, low-quality bases, N' base, and duplication. The resulting trimmed reads were then aligned to the human genome reference sequence (hg38) using Bowtie2 (version 2.4.5) with default parameters and filtered to

obtain clean data. In the pursuit of optimum results, we explored the efficacy of two recently introduced human reference genomes: the T2T-CHM13v2.0 genome [77] and the Chinese human pan-genome (CPC) [78]. To achieve this, we employed a sample set consisting of 10 randomly selected samples. To our surprise, our analysis revealed that both the T2T and CPC genomes yielded significantly lower read removal rates compared to the hg38 genome ($p=0.00098$ and 0.004 , respectively, as per paired Mann–Whitney test; depicted in Fig. S26). Consequently, we opted to employ the hg38 genome for the removal of human reads in our dataset, ensuring data integrity and quality.

Subsequently, the clean data were assembled primarily using metaSPAdes [79] (version 3.15.5) due to its superior assembly lengths and quality. MetaSPAdes first tested the k-mer values, and selected the optimal value in 21–141 as the optimal strategy for subsequent assembly. However, MetaSPAdes failed to assemble a small fraction of samples ($n=56$; Supplementary Table 1) due to the high memory demand and/or large number of sequencing reads. The failed samples were then assembled by MEGAHIT [80] (version 1.2.9) with default parameters. The resulting contigs were further refined with metaMIC [81] (<https://github.com/ZhaoXM-Lab/metaMIC>).

Next, MetaWRAP [82] (version 1.3.2) was employed to bin the assembled contigs of each sample. MetaWRAP integrated the binning results from three binning tools, namely CONCOCT [83], MaxBin [84], and metaBAT [84], refined the bins using its “bin_refinement” module, and generated a consolidated, high-quality bin set. This process yielded a genome set (bins) for each sample.

CheckM [35] was performed on the final bin set (default parameters) and generate the final report file for quality assessment.

The bins from all sample were then combined and de-replicated using dRep [85] (version 3.4.1) with parameters “-comp 50 -con 10 -pa 0.9 -sa 0.95 -nc 0.30 -cm larger,” which would generate a set of non-redundant metagenome-assembled genomes (MAGs) with average nucleotide identity (ANI) $>95\%$ (i.e., species-level MAGs) with $>50\%$ completeness and $<10\%$ contamination, and result in 2737 MAGs. Additionally, we also generate a set with ANI $>99\%$ (i.e., strain-level MAGs), and resulted in 5029 MAGs.

Furthermore, Barrnap (version 0.9, <https://github.com/tseemann/barrnap>) and tRNAscan-SE [86] (version 2.0.11) were used to identify rRNA and tRNA genes in these MAGs, respectively. The quality of these MAGs was evaluated based on the criteria defined by the MIMIG [87] which was also used by the SMGC [31], as detailed below:

1. High quality: MAGs with >90% completeness, <5% contamination and having 23SrRNA, 16SrRNA, 5S rRNA genes and at least 18 tRNA genes;
2. Near complete: MAGs with >90% completeness, <5% contamination;
3. Medium quality: MAGs with $\geq 50\%$ completeness and <10% contamination;
Of note, MAGs with <50% completeness or $\geq 10\%$ contamination were considered to be low quality and removed from further consideration.

In the end, we obtained a total of species-level 2737 MAGs and referred them as to the Facial Microbiome Genome Compendium (FMGC). Of which, 62 were high quality, 732 (26.74%) were near complete, while 1943 (70.99%) were medium quality.

Detection of eukaryotic genomes and viral contigs in FMGC

The detection of eukaryotic genomes involved an initial filtration step using EukRep v0.6.7 [88] to selectively retain bins containing substantial amounts of eukaryotic DNA. Bins with more than 1 Mbp of eukaryotic bases were considered as potential eukaryotic bins. Next, Busco v5.5.0 [89] was used to estimate bin completeness and contamination. We obtained a total of 35 putative eukaryotic bins with very low completeness, ranging from 1 to 50%, suggesting no useful eukaryotic genomes could be identified in our dataset. Notably, only 2 bins reached 50% completeness (50.6% and 54.5%).

Additionally, VirSorter2 [90] ($-\text{min-score } 0.7$) was used to identify viral contigs. Contigs with VirSorter2 score over 0.7 were considered as viral contigs. As short contigs may only represent fragments of viral genomes, contigs that were longer than 5 kb were kept for further analysis. The kept dataset was dereplicated using CD-HIT [91] (v4.8.1, parameters: $-c 0.95 -n 8$) using a global identity threshold of 95%. We obtained a total of 21,777 non-redundant viral contigs. The completeness and contamination were estimate with CheckV [92]. In total, 2417 (9.86%) and 165 (0.76%) viral contigs were considered complete or high-quality (>90% completeness).

The viral contigs could be obtained at https://figshare.com/projects/Integrated_analysis_of_facial_microbiome_illustrates_impacts_of_aging_and_skin_physiological_properties/173451.

Of note, our analysis focused on mainly the facial bacteria and did not incorporate the viral contigs or the eukaryotic bins in our further analyses.

Taxonomic annotation and estimation of the novelty of FMGC in comparison to public datasets

The taxonomy annotation of FMGC MAGs was conducted using the `classify_wf` module of GTDB-tk [33]

with default parameters. MAGs lacking taxonomy annotation at the species level were identified as potential novel MAGs. To determine the proportion of novel MAGs in FMGC, the FastANI [93] (v1.1, default parameters) tool was utilized to compare all MAGs against various public datasets, including the reference microbial genomes in the SMGC and those used by metaPhlan4. A MAG is considered novel if it exhibits <95% ANI compared to the sequences in the public datasets and was not annotated by the GTDB-tk tool at the species level.

Phylogenetic analysis of MAGs

The `de_novo_wf` module of GTDB-tk [33] was used to construct a phylogenetic tree using the nucleotide sequences of selected MAGs/genomes as input. The resulting tree was visualized with iTol [94] and Evolvew v3 [95].

Estimation of the relative abundance of the FMGC MAGs at all taxa level

Two methods were used to calculate the relative abundance of FMGC MAGs in each sample. The first method was MetaPhlan4 [96] (version 4.0.3), which uses clade-specific marker genes to provide taxonomic assignments to the sequencing data and estimate the composition of microbial communities with relative abundances from species level to higher taxonomic clades such as genus, order, family, phylum, and kingdom. These profiles were filtered to retain species with a relative abundance of $\geq 1e-4$.

The second method was adopted from the SMGC paper [31], which first involved aligning the clean sequencing reads from each sample to the FMGC MAGs using BWA-MEM (version 0.7.17) [97]. Then the “presence” of a MAG in a sample was estimated if over 50% of its length was covered by the aligned reads from the sample with >4X overall sequencing depth. For each of MAGs, a reads per kilobase million (RPKM) value was calculated. However, the abundance values of the MAGs that were not “present” in the sample were set to zero. Relative abundance at the MAG (species) level was calculated by dividing the RPKM of a specific MAG by the total RPKM of all MAGs that presented in the sample.

The relative abundance of higher taxa levels was calculated as the summation of the abundances of all its members, using taxonomic annotations from GTDB-tk.

Identification of cutotypes of facial microbiome in our and public samples

To define the cutotypes based on the skin microbiome, similar to the construction of enterotype in previous studies [39, 40], samples were clustered with Jensen-Shannon distance (JSD) [98] using genus level relative

abundance estimated by FMGC MAGs and metaPhlAn4. Samples were further clustered using partitioning around medoid (PAM) clustering. The PAM algorithm is an implementation of k-medoids clustering. The core idea of the k-medoids algorithm is similar to that of k-means, but the main difference lies in the way the cluster centers are updated. The optimal number of clusters was estimated using the Calinski-Harabasz (CH) index of the Cluster-Sim package. The CH index measures the compactness within each cluster by calculating the sum of squared distances between each point and the centroid of its respective cluster. It also evaluates the separation between the cluster centroids and the centroid of the entire dataset by calculating the sum of squared distances.

Between-class analysis (BCA) was used for clustering the samples for visualization and identifying the driver taxa using “bca” function of ade4 package [99]. BCA is a special case of principal component analysis (PCA) that utilizes instrumental variables. In this case, the variables are qualitative factors, such as skin types. BCA analysis allows us to identify the main components based on the centroids of each group, highlighting the differences between the groups, and then associating each sample with its corresponding category. It is a supervised projection of the data, where the distances between predefined classes (in this case, cutotypes) are maximized.

Correlation analysis of age, skin properties, and facial microbiome profiles

The correlations between the participant-related data including age and skin physio-optical characterizes the facial microbiome-related data including relative abundances of the taxa and alpha- and beta-diversities using either the “psych” package or the Envfit test implemented in the “vegan” package. R v.4.0.5 (<https://www.r-project.org/>) was used throughout the study unless otherwise stated.

Mediation analysis of age and skin properties to microbial structure

Mediation analysis implemented in the mediation R package (v4.5.0, <https://www.rdocumentation.org/packages/mediation/versions/4.5.0>) to determine whether age mediates the microbial structural changes through the skin physio-optical properties such as moisture, sebum production, and sensitivity. The results were visualized in Fig. 2F. Specifically, as shown in Fig. 2F, the arrow from the independent variable (age) to the mediating variable (sebum, moisture, and Tivi) illustrates the influence of the independent variable on the mediating variable. Likewise, the arrow from the mediating variable to the dependent variable (microbiome structure) demonstrates the impact of the mediating variable on the dependent

variable. Finally, the arrow from the independent variable to the dependent variable indicates the overall effect of the independent variable on the dependent variable, encompassing both total effect and average direct effect. Average direct effects (ADE) refers to the direct effects of the independent variable on the dependent variable. It represents the effect of the independent variable on the dependent variable while controlling for the mediating variable(s). Total effect, on the other hand, refers to the overall effect (direct + indirect) of the independent variable on the dependent variable.

Construction of age index

To eliminate the influence of confounding factors, we exclude individuals with acne ($n=55$), and those who use makeup ($n=191$), resulting in a total of 229 usable samples (Table S1) in the construction of Age Index. A Pearson correlation analysis was performed on age and species abundances for the selected samples to obtain correlations and p values for each species with age. The Facial Age Index (FAI) was constructed using species that significantly correlated with age (i.e., p value < 0.05 in Pearson correlation analysis), with the equations below.

$$mR_i = \frac{R_{ij}}{\text{median}(R_i)}$$

$$I_p = \sum_{i \in P} mR_i \times cor_i \times \log_{10}(p_i)$$

$$I_n = \sum_{i \in N} mR_i \times cor_i \times \log_{10}(p_i)$$

$$\text{Facial Age Index} = \frac{I_p}{I_n}$$

R_{ij} : The relative abundance of species i in sample j ;
 $\text{median}(R_i)$: The median relative abundance of species i in all samples;

cor_i : The absolute correlation coefficient of Spearman’s analysis between species i and age;

p_i : The correlation p value of Spearman’s analysis between species i and age.

P : Species positively correlated with age;

N : Species negatively correlated with age.

The FAI represents the quantification of age through skin microbes.

Impact of lifestyle on Facial Age Index while controlling for chronic age

Lifestyle information were extracted from a self-reported questionnaire, including habitual sleep restriction (whether they went to bed after 11 pm [100] more than 3 times a week, referred to as stay up late), heavy cosmetic

application (whether they excessively or heavily wear cosmetics, more than 3 times a week, referred to as make up), emotional dysregulation (whether they frequently experience significant emotional fluctuations, more than 2 times a day), and constipation (less than two bowel movements per week).

To analyze the effects of these lifestyle on FAI, participants were divided into two groups according to each lifestyle. To control for the chronological age of the participants, the residuals of the FAI as a function of age was calculated using the “resid” function in R using formula of “resid(lm(FAI~age)).”

Estimation of the relative abundance of FMGC pathways

HUMAnN [55] (version 3.6.1) was used to annotate the microbial functions and estimate their relative abundances for each sample. HUMAnN utilizes a comprehensive database to assign functional annotations to the microbial genes present in the samples, allowing for the identification of specific biological pathways and functions associated with the microbial communities.

Identification of microbial pathways associated with aging

To identify pathways associated with age and lifestyle, participants were divided into four equal-sized groups according to their age, namely Group A (18–25 years), B (26–37 years), C (37–51 years), and D (51–64 years). Aging-associated pathways were identified as those they showed significantly different abundance between the youngest (Group A) and oldest (Groups D) groups using a Linear discriminant analysis Effect Size (LEfSe) [57]. A linear discriminant analysis (LDA) score cutoff of 3 (adjusted p value < 0.05) was used as the threshold of statistical difference. Statistical analysis using Wilcoxon rank sum test (adjusted p value < 0.1 to identified pathways with slight difference among age groups) was then carried out to compared the relative abundance across different lifestyles, and the effect sizes were visualized in Fig. 4B.

Statistical analysis

All processed data, if not otherwise stated, were loaded into R (v4.0.5, <https://www.r-project.org/>), analyzed or visualized. Adonis analysis were carried out with Adonis2 using adonis2 function from R package, vegan [101], with default parameters.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-024-01891-0>.

Additional file 1: Table S1. A list of 498 collected samples and related information in detail. And the used questionnaire in sheet 2. Table S2. A list of 2,737 assembled MAGs and related information in details. Table S3. 1,315

age related bacteria and calculated Person's R and p -value. Table S4. 348 age related pathways and calculated LDA score, p -value and adjust p -value. Table S5. Softwares and resources used in this paper.

Additional file 2: Figure S1, Boxplot showing the distribution of unclassified ratio of our dataset using metaPhlan4. Figure S2, PcoA plot showing the beta-diversity difference across the three facial regions (Cheek, Forehead, and Nose) in iHSMGC. Figure S3, A) Venn plot showing the overlap MAGs of FMGC with SMGC and UHSG under 95% ANI. B) Bar chart illustrates the annotation rates of FMGC in different software or libraries, with the portion annotated using GTDB-tk further categorized and presented at each taxonomy level. Out of the 2,737 FMGC MAGs, 1,644 and 1,014 were found in the GTDB-tk and MetaPhlan4 datasets, respectively, while 918 were present in both datasets (Figure 1E and S2). Alongside the 752 FMGC MAGs that overlapped with the SMGC dataset, and 807 overlapped with the UHSG dataset, a total of 809 MAGs (30% out of total) remained classified as novel. Figure S4, Box plot showing detailed information of the novel MAGs. These MAGs exhibited even higher qualities in genome size, N50 length and completeness while maintaining the same contamination level (2.58%) as compared with those overlapping with the public datasets. Figure S5, Phylogenetic diversity of FMGC subtree, subtree of public datasets (SMGC and UHSG) and Unique-FMGC (Novel MAGs of FMGC compared to public datasets' MAGs) subtree. Figure S6, Violin plot showing the observed genera, unclassified rate and Shannon diversity from FMGC calculated relative abundance and metaPhlan4 calculated relative abundance. Figure S7, Bar plot displaying the distribution of the two assigned cutotypes of people with different physio-optical properties. Figure S8, Box plot displaying the relative abundance of the driver genera in the people with different physio-optical properties. Figure S9, Heatmap showing the partial pearson correlation of bacteria and skin physio-optical properties. Figure S10, Scatter plot and fitted curve illustrating a positive correlation between age and Shannon diversity. The color of the data points indicates increasing age, with greener points representing older age groups. Figure S11, Boxplot depicting the differences in Shannon diversity between two cutotypes and the differences in Shannon residuals after removing the age background. The S-type population exhibits significantly higher Shannon diversity. Figure S12, Boxplot demonstrating the noticeable upward trend of Shannon diversity with increasing age in two cutotypes. Figure S13, A) Barplots showing the Nr. of MAGs positively or negatively correlated with Age and their belonging genera. B) Detailed information about the MAGs significantly correlated with age and with mean abundance higher than 1% or with relatively higher absolute Spearman's R (≥ 0.5). C) The scatter plot displaying the MAGs correlated with the calculated FAI. D) The relative abundance of the labeled age-related MAGs controlled by age, along with the total abundance of all age-related MAGs in samples with or without a makeup habit. Figure S14, The methodology validation of Facial Aging Index involved establishing FAI using population data from both Guangzhou and Beijing regions, and subsequently validating it using data from other regions. The findings revealed a significant positive correlation between FAI and actual age, indicating a strong association between FAI and chronological age. Figure S15, The scatter plot and fitted curve demonstrate a significant positive correlation between FAI and age in both cutotypes. Figure S16, Boxplot showing the Skin properties including Sebum level, R5, and Tivi-residues (with the effects of the chronological age removed) in participants with/without certain lifestyle habits and different cutotypes. Figure S17, Other host conditions such as constipation and “Mood Swings” (referring to significant emotional fluctuations) were also found to significantly accelerate skin aging. Figure S18, Principal Coordinate Analysis (PCoA) plot displaying the structural differences between the two cutotypes. Axis 1 represents the primary axis that illustrates the variation in structure between the two cutotypes. Figure S19, Age-related pathways' abundance correlate with advancing age. A) Adonis2 analysis result displaying co-variants' effect to microbiome structure. B) Scatter plot showing the accumulated abundance of age-related pathways and advancing age. Figure S20, Lefse analysis revealed metabolic pathways that exhibited significant differences between the young and old age groups. Figure S21, Accumulated Aging/anti-aging related pathway abundance in different aging stages. Figure S22, Age distribution of FMGC and UHSG samples. Figure S23, The violin plot displays the analysis results of our dataset, indicating that

human DNAs accounted for 76.43% (interquartile range: 63%–85%) of the total sequencing reads. This finding emphasizes the necessity for higher sequencing depth in order to capture a more comprehensive representation of non-human DNA components. Figure S24, Violin plot showing the unclassified rate of FMGC calculated relative abundance and metaPhlan4 calculated relative abundance of iHSMGC samples. Figure S25, Cutotypes of iHSMGC samples. Figure S26, Boxplot showing the human-origin reads removed by T2T-CHM13v2.0, hg38 and Chinese human pan-genome (CPC). The data shows that hg38 removed a significantly higher number of reads, demonstrating a stricter criteria in human reads removal.

Acknowledgements

We thank all members of the Chen, and Liu labs for their help with the facial sample collection, and insightful discussions.

Authors' contributions

WHC and ZL designed and directed the research. CS and GH analyzed the data; LY, WG, QY managed the sampling and helped the experiments. CS wrote the paper with results from all authors; WHC, ZL, YH, and XY polished the manuscript through multiple iterations of discussions with all authors. All authors have read and approved the final manuscript.

Funding

This research is supported by National Key Research and Development Program of China (2019YFA0905600 to W.H.C and Z. L).

Availability of data and materials

The clean sequencing data (i.e., sequencing reads with vectors, low-quality bases and human DNA removed) used in this study are available in the CNCB GSA database under accession PRJCA018420 (<https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA018420>). Sequences of the MAGs, Eukaryotic bins and viral contigs generated by this study and related meta-data have been uploaded to a Figshare repository at https://figshare.com/projects/Integrated_analysis_of_facial_microbiome_illustrates_impacts_of_aging_and_skin_physiological_properties/173451; please consult Table S5 for the detailed URL for each dataset.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Tongji Medical College of Huazhong University of Science and Technology (No, S083).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular Imaging, Center for Artificial Intelligence Biology, Huazhong University of Science and Technology, Wuhan 430074, China. ²Center for Research and Development, Xiamen Treatgut Biotechnology Co., Ltd., Xiamen, China. ³School of Life Sciences, Xiamen University, Xiamen, China. ⁴Department of Biotechnology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China. ⁵Department of Dermatology, Huazhong University of Science and Technology Hospital, Wuhan 430074, China. ⁶College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China. ⁷National Engineering Research Center for Nanomedicine, Wuhan 430074, China. ⁸The GBA National Institute for Nanotechnology Innovation, Guangzhou 510799, China. ⁹School of Biomedical Science and Engineering, South China University of Technology, Guangzhou 510641, China. ¹⁰College of Chemistry and Materials Engineering and Institute of Cosmetic Regulatory Science, Beijing Technology and Business

University, Beijing 100048, People's Republic of China. ¹¹Institution of Medical Artificial Intelligence, Binzhou Medical University, Yantai 264003, China.

Received: 13 March 2024 Accepted: 29 July 2024

Published online: 05 September 2024

References

- Grice EA. The intersection of microbiome and host at the skin interface: genomic- and metagenomic-based insights. *Genome Res.* 2015;25(10):1514–20.
- Belkaid Y, Segre JA. Dialogue between skin microbiota and immunity. *Science.* 2014;346(6212):954–9.
- Scharschmidt TC, Fischbach MA. What lives on our skin: ecology, genomics and therapeutic opportunities of the skin microbiome. *Drug Discovery Today: Disease Mechanisms.* 2013;10(3):e83–9.
- Flowers L, Grice EA. The skin microbiota: balancing risk and reward. *Cell Host Microbe.* 2020;28(2):190–200.
- Li Z, Xia J, Jiang L, Tan Y, An Y, Zhu X, Ruan J, Chen Z, Zhen H, Ma Y, et al. Characterization of the human skin resistome and identification of two microbiota cutotypes. *Microbiome.* 2021;9(1):47.
- Swaney MH, Nelsen A, Sandstrom S, Kalan LR. Sweat and sebum preferences of the human skin microbiota. *Microbiol Spectr.* 2023;11(1):e0418022.
- Murphrey MB, Miao JH, Zito PM. Histology, Stratum Corneum. In: StatPearls [Internet]. Treasure Island: StatPearls Publishing; 2022.
- Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. *Nat Rev Microbiol.* 2018;16(3):143–55.
- Kim JH, Son SM, Park H, Kim BK, Choi IS, Kim H, Huh CS. Taxonomic profiling of skin microbiome and correlation with clinical skin parameters in healthy Koreans. *Sci Rep.* 2021;11(1):16269.
- Ying S, Zeng DN, Chi L, Tan Y, Galzote C, Cardona C, Lax S, Gilbert J, Quan ZX. The influence of age and gender on skin-associated microbial communities in urban and rural human populations. *PLoS ONE.* 2015;10(10):e0141842.
- Howard B, Bascom CC, Hu P, Binder RL, Fadayeel G, Huggins TG, Jarrold BB, Osborne R, Rocchetta HL, Swift D et al. Aging-associated changes in the adult human skin microbiome and the host factors that affect skin microbiome composition. *J Invest Dermatol.* 2022;142(7):1934–1946 e1921
- Kim G, Kim M, Kim M, Park C, Yoon Y, Lim DH, Yeo H, Kang S, Lee YG, Beak NI, et al. Spermidine-induced recovery of human dermal structure and barrier function by skin microbiome. *Commun Biol.* 2021;4(1):231.
- Kim H-J, Oh HN, Park T, Kim H, Lee HG, An S, Sul WJ. Aged related human skin microbiome and mycobiome in Korean women. *Sci Rep.* 2022;12(1):2351.
- Leung MH, Wilkins D, Lee PK. Insights into the pan-microbiome: skin microbial communities of Chinese individuals differ from other racial groups. *Sci Rep.* 2015;5:11845.
- Huang S, Haiminen N, Carrieri AP, Hu R, Jiang L, Parida L, Russell B, Allaband C, Zarrinpar A, Vázquez-Baeza Y, et al. Human skin oral and gut microbiomes predict chronological age. *mSystems.* 2020;5(1):e00630–19.
- Larson PJ, Zhou W, Santiago A, Driscoll S, Fleming E, Voigt AY, Chun OK, Grady JJ, Kuchel GA, Robison JT, et al. Associations of the skin, oral and gut microbiome with aging, frailty and infection risk reservoirs in older adults. *Nat Aging.* 2022;2(10):941–55.
- Krejner A, Bruhs A, Mrowietz U, Wehkamp U, Schwarz T, Schwarz A. Decreased expression of G-protein-coupled receptors GPR43 and GPR109a in psoriatic skin can be restored by topical application of sodium butyrate. *Arch Dermatol Res.* 2018;310(9):751–8.
- Sanford JA, O'Neill AM, Zouboulis CC, Gallo RL. Short-chain fatty acids from *Cutibacterium acnes* activate both a canonical and epigenetic inflammatory response in human sebocytes. *J Immunol.* 2019;202(6):1767–76.
- Midorikawa K, Ouhara K, Komatsuzawa H, Kawai T, Yamada S, Fujiwara T, Yamazaki K, Sayama K, Taubman MA, Kurihara H, et al. *Staphylococcus aureus* susceptibility to innate antimicrobial

- peptides, & defensins and CAP18, expressed by human keratinocytes. *Infect Immun*. 2003;71(7):3730–9.
20. Wanke I, Steffen H, Christ C, Krismer B, Götz F, Peschel A, Schaller M, Schittek B. Skin Commensals amplify the innate immune response to pathogens by activation of distinct signaling pathways. *J Investig Dermatol*. 2011;131(2):382–90.
 21. Pinto D, Ciardiello T, Franzoni M, Pasini F, Giuliani G, Rinaldi F. Effect of commonly used cosmetic preservatives on skin resident microflora dynamics. *Sci Rep*. 2021;11(1):8695.
 22. Zheng Y, Hunt RL, Villaruz AE, Fisher EL, Liu R, Liu Q, Cheung GYC, Li M, Otto M. Commensal *Staphylococcus epidermidis* contributes to skin barrier homeostasis by generating protective ceramides. *Cell Host Microbe*. 2022;30(3):301–313.e309.
 23. Chen H, Zhao Q, Zhong Q, Duan C, Krutmann J, Wang J, Xia J. Skin microbiome, metabolome and skin phenome, from the perspectives of skin as an ecosystem. *Phenomics*. 2022;2(6):363–82.
 24. Zhou H, Shi L, Ren Y, Tan X, Liu W, Liu Z. Applications of human skin microbiota in the cutaneous disorders for ecology-based therapy. *Front Cell Infect Microbiol*. 2020;10:570261.
 25. Boxberger M, Cenizo V, Cassir N, La Scola B. Challenges in exploring and manipulating the human skin microbiome. *Microbiome*. 2021;9(1):125.
 26. Farage MA, Miller KW, Elsner P, Maibach HI. Functional and physiological characteristics of the aging skin. *Aging Clin Exp Res*. 2008;20(3):195–200.
 27. Cotterill JA, Cunliffe WJ, Williamson B, Bulusu L. Age and sex variation in skin surface lipid composition and sebum excretion rate. *Br J Dermatol*. 1972;87(4):333–40.
 28. Lehtimäki J, Karkman A, Laatikainen T, Paalanen L, von Hertzen L, Haahntela T, Hanski I, Ruokolainen L. Patterns in the skin microbiota differ in children and teenagers between rural and urban environments. *Sci Rep*. 2017;7(1):45651.
 29. Howard B, Bascom CC, Hu P, Binder RL, Fadayeel G, Huggins TG, Jarrold BB, Osborne R, Rocchetta HL, Swift D, et al. Aging-associated changes in the adult human skin microbiome and the host factors that affect skin microbiome composition. *J Investig Dermatol*. 2022;142(7):1934–1946.e1921.
 30. Rozas M, Hart de Ruijter A, Fabrega MJ, Zorngani A, Guell M, Paetzold B, Brillet F: **From dysbiosis to healthy skin: major contributions of *Cutibacterium acnes* to skin homeostasis**. *Microorganisms*. 2021;9(3):628.
 31. Saheb Kashaf S, Proctor DM, Deming C, Saary P, Holzer M, Program NCS, Taylor ME, Kong HH, Segre JA, Almeida A, et al. Integrating cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and functions. *Nat Microbiol*. 2022;7(1):169–79.
 32. Li Z, Ju Y, Xia J, Zhang Z, Zhen H, Tong X, Sun Y, Lu H, Zong Y, Chen P, et al. Integrated human skin bacteria genome catalog reveals extensive unexplored habitat-specific microbiome diversity and function. *Adv Sci (Weinh)*. 2023;10(28):e2300050.
 33. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*. 2019;36(6):1925–7.
 34. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, Manghi P, Dubois L, Huang KD, Thomas AM, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlan 4. *Nat Biotechnol*. 2023;41(11):1633–44.
 35. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25(7):1043–55.
 36. Leung MHY, Tong X, Shen Z, Du S, Bastien P, Appenzeller BMR, Betts RJ, Mezzache S, Bouroukba N, Cavusoglu N, et al. Skin microbiome differentiates into distinct clotypes with unique metabolic functions upon exposure to polycyclic aromatic hydrocarbons. *Microbiome*. 2023;11(1):124.
 37. Vandeputte D, Commer LD, Tito RY, Kathagen G, Sabino J, Vermeire S, Faust K, Raes J. Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nat Commun*. 2021;12(1):6740.
 38. Lai S, Yan Y, Pu Y, Lin S, Qiu J-G, Jiang B-H, Keller MI, Wang M, Bork P, Chen W-H, et al. Enterotypes of the human gut mycobiome. *Microbiome*. 2023;11(1):179.
 39. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, et al. Enterotypes of the human gut microbiome. *Nature*. 2011;473(7346):174–80.
 40. Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol*. 2018;3(1):8–16.
 41. Imai K, Keele L, Tingley D, Yamamoto T: Causal mediation analysis using R. In: 2010; New York, NY: Springer New York: 129–154.
 42. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15(4):309–34.
 43. Park S, Steiner PM, Kaplan D. Identification and sensitivity analysis for average causal mediation effects with time-varying treatments and mediators: investigating the underlying mechanisms of kindergarten retention policy. *Psychometrika*. 2018;83(2):298–320.
 44. Jensen JM, Proksch E. The skin's barrier. *G Ital Dermatol Venereol*. 2009;144(6):689–700.
 45. Dréno B, Dagnelie MA, Khammari A, Corvec S. The skin microbiome: a new actor in inflammatory acne. *Am J Clin Dermatol*. 2020;21(1):18–24.
 46. Bouslimani A, da Silva R, Kosciulek T, Janssen S, Callewaert C, Amir A, Dorrestein K, Melnik AV, Zaramela LS, Kim J-N, et al. The impact of skin care products on skin chemistry and microbiome dynamics. *BMC Biol*. 2019;17(1):47.
 47. Tett A, Pasolli E, Masetti G, Ercolini D, Segata N. Prevotella diversity, niches and interactions with the human host. *Nat Rev Microbiol*. 2021;19(9):585–99.
 48. Do T, Jolley KA, Maiden MCJ, Gilbert SC, Clark D, Wade WG, Beighton D. Population structure of *Streptococcus oralis*. *Microbiology (Reading)*. 2009;155(Pt 8):2593–602.
 49. Sadowy E, Hryniewicz W. Identification of *Streptococcus pneumoniae* and other Mitis streptococci: importance of molecular methods. *Eur J Clin Microbiol Infect Dis*. 2020;39(12):2247–56.
 50. Pollack S, Mogtader A, Lange M. *Neisseria subflava* endocarditis Case report and review of the literature. *Am J Med*. 1984;76(4):752–8.
 51. Jousseaume AM, Funke G, Jousseaume F, Herberich G. *Corynebacterium macginleyi*: a conjunctiva specific pathogen. *Br J Ophthalmol*. 2000;84(12):1420–2.
 52. Toprak NU, Sayin E, Akilli FM, Gundogdu A. Sepsis caused by *Anaerococcus nageyae* after transarterial-chemoembolization for hepatocellular carcinoma: Case report and literature review. *Anaerobe*. 2021;72:102464.
 53. Skowron K, Bauza-Kaszewska J, Kraszewska Z, Wiktorczyk-Kapischke N, Grudlewska-Buda K, Kwieceńska-Piróg J, Walecka-Zacharska E, Radtke L, Gospodarek-Komkowska E. Human skin microbiome: impact of intrinsic and extrinsic factors on skin microbiota. *Microorganisms*. 2021;9(3):543.
 54. Hillebrand GG, Dimitriu P, Malik K, Park Y, Qu D, Mohn WW, Kong R. Temporal variation of the facial skin microbiome: a 2-year longitudinal study in healthy adults. *Plast Reconstr Surg*. 2021;147(15–2):50s–61s.
 55. Beghini F, McIver LJ, Blanco-Miguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife*. 2021;10:e65088.
 56. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*. 2014;42(Database issue):459–71.
 57. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60.
 58. Chen B, Yang J, Song Y, Zhang D, Hao F. Skin immunosenescence and type 2 inflammation: a mini-review with an inflammaging perspective. *Front Cell Dev Biol*. 2022;10:835675.
 59. Pilkington SM, Bulfone-Paus S, Griffiths CEM, Watson REB. Inflammaging and the Skin. *J Invest Dermatol*. 2021;141(4S):1087–95.
 60. Aman Y, Qiu Y, Tao J, Fang EF. Therapeutic potential of boosting NAD+ in aging and age-related diseases. *Translational Medicine of Aging*. 2018;2:30–7.
 61. Covarrubias AJ, Perrone R, Grozio A, Verdin E. NAD(+) metabolism and its roles in cellular processes during ageing. *Nat Rev Mol Cell Biol*. 2021;22(2):119–41.

62. Gardell SJ, Coen PM. NAD⁺ and human muscle health. *Nature Aging*. 2022;2(3):195–6.
63. Kim CS, Park S, Kim J. The role of glycation in the pathogenesis of aging and its prevention through herbal products and physical exercise. *J Exerc Nutrition Biochem*. 2017;21(3):55–61.
64. Pageon H, Zucchi H, Rousset F, Monnier VM, Asselineau D. Skin aging by glycation: lessons from the reconstructed skin model. *Clin Chem Lab Med*. 2014;52(1):169–74.
65. Pageon H, Zucchi H, Pennacchi PC, Asselineau D: Glycation and skin aging. In: *Textbook of Aging Skin*. Edited by Farage MA, Miller KW, Maibach HI. Berlin, Heidelberg: Springer Berlin Heidelberg; 2017: 1247–1270.
66. Zheng W, Li H, Go Y, Chan XHF, Huang Q, Wu J. Research advances on the damage mechanism of skin glycation and related inhibitors. *Nutrients*. 2022;14(21):4588.
67. Yücel G, Zhao Z, El-Batrawy I, Lan H, Lang S, Li X, Buljubasic F, Zimmermann W-H, Cyganek L, Utikal J, et al. Lipopolysaccharides induced inflammatory responses and electrophysiological dysfunctions in human-induced pluripotent stem cell derived cardiomyocytes. *Sci Rep*. 2017;7(1):2935.
68. Silva S, Michniak-Kohn B, Leonardi GR. An overview about oxidation in clinical practice of skin aging. *An Bras Dermatol*. 2017;92(3):367–74.
69. Nguyen D, Samson SL, Reddy VT, Gonzalez EV, Sekhar RV. Impaired mitochondrial fatty acid oxidation and insulin resistance in aging: novel protective role of glutathione. *Aging Cell*. 2013;12(3):415–25.
70. Rinnerthaler M, Bischof J, Streubel MK, Trost A, Richter K. Oxidative stress in aging human skin. *Biomolecules*. 2015;5(2):545–89.
71. Nadeeshani H, Li J, Ying T, Zhang B, Lu J. Nicotinamide mononucleotide (NMN) as an anti-aging health product—promises and safety concerns. *J Adv Res*. 2022;37:267–78.
72. Ahmed IA, Mikail MA, Zamakshari N, Abdullah AH. Natural anti-aging skincare: role and potential. *Biogerontology*. 2020;21(3):293–310.
73. Burnett CL, Heldreth B, Bergfeld WF, Belsito DV, Hill RA, Klaassen CD, Liebler DC, Marks JG Jr, Shank RC, Slaga TJ, et al. Safety assessment of α -amino acids as used in cosmetics. *Int J Toxicol*. 2013;32(6 Suppl):41s–64s.
74. Yang XX, Zhao MM, He YF, Meng H, Meng QY, Shi QY, Yi F. Facial skin aging stages in Chinese females. *Front Med (Lausanne)*. 2022;9:870926.
75. Lam M, Hu A, Fleming P, Lynde CW. The impact of acne treatment on skin bacterial microbiota: a systematic review. *J Cutan Med Surg*. 2022;26(1):93–7.
76. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics*. 2018;34(17):i884–90.
77. Nurk S, Koren S, Rhie A, Rautiainen M, Bizkadez AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. The complete sequence of a human genome. *Science*. 2022;376(6588):44–53.
78. Gao Y, Yang X, Chen H, Tan X, Yang Z, Deng L, Wang B, Kong S, Li S, Cui Y, et al. A pangenome reference of 36 Chinese populations. *Nature*. 2023;619(7968):112–21.
79. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27(5):824–34.
80. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674–6.
81. Lai S, Pan S, Sun C, Coelho LP, Chen WH, Zhao XM. metaMIC: reference-free misassembly identification and correction of de novo metagenomic assemblies. *Genome Biol*. 2022;23(1):242.
82. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*. 2018;6(1):158.
83. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. Binning metagenomic contigs by coverage and composition. *Nat Methods*. 2014;11(11):1144–6.
84. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome*. 2014;2(1):26.
85. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 2017;11(12):2864–8.
86. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49(16):9077–96.
87. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloie-Fadrosch EA, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35(8):725–31.
88. West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res*. 2018;28(4):569–80.
89. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647–54.
90. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*. 2021;9(1):37.
91. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28(23):3150–2.
92. Nayfach S, Camargo AP, Schulz F, Eloie-Fadrosch E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*. 2021;39(5):578–85.
93. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114.
94. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*. 2021;49(W1):gkab301.
95. Subramanian B, Gao S, Lercher MJ, Hu S, Chen WH. Evolvview v3: a web-server for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res*. 2019;47(W1):W270–5.
96. Aitor B-M, Francesco B, Fabio C, Lauren JM, Kelsey NT, Moreno Z, Paolo M, Leonard D, Kun DH, Andrew Maltez T et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlan 4. *bioRxiv* 2022:2022.2008.22.504593.
97. Li H: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In.; 2013. [arXiv:1303.3997](https://arxiv.org/abs/1303.3997).
98. Fuglede B, Topsoe F. Jensen-Shannon divergence and Hilbert space embedding. In: *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, Chicago, IL, USA, 2004, pp. 31. <https://doi.org/10.1109/ISIT.2004.1365067>.
99. Dray S, Dufour A-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J Stat Softw*. 2007;22(4):1–20.
100. Nikbakhtian S, Reed AB, Obika BD, Morelli D, Cunningham AC, Aral M, Plans D. Accelerometer-derived sleep onset timing and cardiovascular disease incidence: a UK Biobank cohort study. *Eur Heart J Digit Health*. 2021;2(4):658–66.
101. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;14(6):927–30.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.