IFSCC 2025 full paper IFSCC2025-1653

"Predicting Facial Microbial Distribution: A Machine Learning Pilot Study"

Martin Stalder [1], Sergi Raurich [1], Ana Finzel[2], and Martin Pagac[1,*]

[1] dsm-firmenich; [2] Bio-me

1. Introduction

The facial microbiome is a complex and dynamic ecosystem composed of various microbial species living on the skin. This environment is continuously influenced by both external factors, such as temperature and humidity, and internal factors, such as age, gender and hormone-driven sebaceous gland activity (i.e. sebum levels). Proper management of the skin's microbiome can support skin regeneration, and changes in microbial distribution have been linked to skin aging, potentially exacerbating wrinkle formation.[1], [2]

Although many key microbial species on the skin have been identified, their distribution varies across different areas of the skin, with some microbes preferring oily environments and others dry or humid ones. The facial skin, being one of the most exposed areas, is frequently subject to external environmental influences and the application of various treatments aimed at protecting the skin and reducing the signs of aging. Despite this, there is limited understanding of how microbial species are distributed across the face and their impact on skin health.[3]

In this pilot study, we investigated the facial skin microbiome of five healthy Caucasian female subjects. We collected 23 samples from each subject, focusing on a panel of 54 microbial species relevant to human skin. Additionally, we propose multiple approaches to predict the distribution of these species and its visualization using our previously developed 3D facial color mapping technology tool.[4] Modeling the distribution of microbes on facial skin is an essential step towards understanding the preferential localization of microbial species, given that species are not evenly distributed across the face and different locations, e.g. cheek and forehead, show significantly different microbial abundances for the same species. This understanding will be key for uncovering interpersonal differences and developing personalized and effective strategies to modulate the microbiome for healthy aging.

The primary contribution of this research is the identification and dissemination of strategies to minimize the number of microbial measurements required across the facial surface while still obtaining a representative profile of the facial skin microbiome. Additionally, we test a further hypothesis: even though microbe exhibits unique behavior, they tend to distribute predictably on the face. Therefore, although our model is tested on a limited sample of species, it can

generalize to infer the distribution of untested species based on their relative abundance at selected sites on healthy female subjects.
This implementation has been inspired by techniques developed for optimal sensor placement and subset selection for spatial interpolation, particularly relevant in geostatistics.
In contrast to the machine learning approach, we also present a model-free strategy for selecting relevant sites to generate facial maps using greedy backward elimination. This alternative method is efficient and rapid, making it suitable for targeted investigations of specific microbes.
Finally, we identify an important single facial site that provides the best representation of the facial microbiome, the cheek. This site has already been described as an area representing the whole facial microbiome using 16S rRNA gene sequencing.[5] Thanks to our trial design and our targeted approach, we have increased coverage of the face by sampling 23 facial site and can better measure inter- and intrapersonal differences with a larger subject cohort which enables us to unveil further insights.

2. Materials and Methods

Python (>=3.11) scripts have been developed using the following packages: "ipykernel>=6.29.5", "ipython>=9.0.2", "joblib>=1.4.2", "jupyterlab>=4.3.5", "matplotlib>=3.10.1", "numpy>=2.2.3", "pandas>=2.2.3", "pykrige>=1.7.2", "scikit-bio>=0.6.3"[6], "scikit-learn>=1.6.1"[7], "scipy>=1.15.2"[8], "seaborn>=0.13.2", "statsmodels>=0.14.4".

Diversity Metrics

Alpha diversity is presented as Shannon diversity. Shannon diversity objective is to provide a measure on the “evenness” of the microbiome in a specific sample. “Evenness” refers to the similarity in abundances between species in a sample.
The formula is given as follows:

$$H' = -\sum_{i=1}^{R} p_i \ln(p_i)$$

Where p_i is the proportion of a species in a sample.

Beta diversity provides a measure of how microbial community composition varies between different samples. In our study, we assess these differences by calculating Bray-Curtis dissimilarity. It is calculated as follows:

$$BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}$$

Where C_{ij} represent the sum of minimum occurrence for common species between two samples and S_i and S_j represent the total number of microbes found in each sample.

Human facial skin samples collection

A total of 116 human facial skin samples from five healthy female Caucasian volunteers, aged 28–49 years, were collected from 23 facial sites (coordinates of pre-defined sampling sites are listed in Table x). All healthy volunteers provided written consent for participation before

enrolment. Skin microbiome sampling protocols and analytical procedures were approved by the internal Ethics Committee of DSM-Firmenich (Kaiseraugst, Switzerland) and conducted under the guidelines and regulations of the Helsinki Declaration. Facial skin of volunteers was devoid of visual skin conditions at the time of sample collection. Exclusion criteria for volunteers included the use of any topical medication over the previous month. Volunteers were also instructed to avoid application of facial skin care and cosmetic products for 24 hours and to refrain from washing the face for 12 hours prior to sample collection.
Samples from volunteers were collected by applying D-Squame Discs (Clinical and Derm, Dallas, USA) on pre-defined hemifacial sites using a uniform pressure for 5 seconds on each sampling disc with a D-Squame Pressure Instrument. Discs were removed from the skin using sterile tweezers and stored in 2 mL microcentrifuge tubes at -80 °C until DNA extraction.

DNA extraction

Extraction and DNA purification of the human facial skin samples was performed at Bio-Me´s laboratory (Oslo, Norway). Beads and lysis buffer (MagMAX™ Microbiome Ultra Kit, Thermo Fisher Scientific, Waltham, US) were added to a 2 mL microcentrifuge tube containing a D-Squame Disc. The D-Squame Disc was then processed using the VWR Star-Beater for 2 minutes at 30 Hz. Afterward, the tube was centrifuged for 5 minutes at 3700 rpm. Next, 440 µL of the resulting supernatant and 40 µL of proteinase K (MagMAX™ Microbiome Ultra Kit) were transferred to a well of a KingFisher™ 96 Deep-Well Plate (Thermo Fisher Scientific, Waltham, US). The plate was placed into the KingFisher™ Apex Instrument (Thermo Fisher Scientific, Waltham, US), and subsequent steps were carried out according to the manufacturer's instructions for the MagMAX™ Microbiome Ultra Kit. The purified DNA was then transferred to a storage plate and stored at -20°C.

Quantitative PCR analysis

A PCR-based preamplification step was performed before running the samples on the OpenArray™. In this step, 2.5 µL of sample DNA were mixed with 5 µL of TaqMan™ PreAmp Master Mix (Thermo Fisher Scientific, Waltham, US) and 2.5 µL of Bio-Me PMP™ primer mix[1], and amplified for 12 cycles. The resulting preamplified DNA was diluted 1:10.
Next, TaqMan™ Universal DNA Spike-In Control was added to TaqMan™ OpenArray™ Genotyping Master Mix (1:6) (Thermo Fisher Scientific, Waltham, US). 3 µL of the resulting mix and 2 µL of preamplified DNA were transferred to a well of a 384-well plate. The 384-well plate was spun at 220 rpm for two minutes and placed into the OpenArray™ AccuFill™ System (Thermo Fisher Scientific, Waltham, US), to load the samples into the custom-manufactured PMP™ qPCR panels. The next steps were carried out according to the manufacturer´s instructions. Panels were analyzed by the QuantStudio™ 12K Flex Real-Time PCR System following the

[1] The PMP™ Comprehensive Skin Microbiome panel v1.0 (Bio-Me, Oslo, Norway) was used in this study. The Precision Microbiome Profiling (PMP™) panel consists of TaqMan™ qPCR assays targeting 48 bacterial species and subspecies, 5 *Malassezia* species, the genus *Malassezia*, the bacterial 16S rRNA gene, and an internal PCR control

default TaqMan™ OpenArray™ Real-Time PCR Plates Protocol (Thermo Fisher Scientific, Waltham, US). The work was performed at Bio-Me´s laboratory (Oslo, Norway).
Standard curves generated for the in vitro performance evaluation of qPCR assays were used to convert the quantification cycle (Cq) value obtained for each assay target into an absolute quantification of the target taxon (reported as number of genomic copies per µL DNA).

Data Processing

Table 1. List of 32 species targeted by qPCR analysis and retained after data processing.

Species		
Acinetobacter johnsonii	Cutibacterium namnetense	Peptoniphilus coxii
Anaerococcus nagyae	Dermacoccus nishinomiyaensis	Peptoniphilus lacrimalis
Anaerococcus octavius	Finegoldia magna	Staphylococcus aureus
Corynebacterium amycolatum	Kocuria rhizophila	Staphylococcus capitis
Corynebacterium jeikeium	Lactobacillus iners	Staphylococcus caprae
Corynebacterium kroppenstedtii	Malassezia arunalokei	Staphylococcus epidermidis
Corynebacterium matruchotii	Malassezia globosa	Staphylococcus haemolyticus
Corynebacterium tuberculostearicum	Malassezia restricta	Staphylococcus hominis
Cutibacterium acnes	Malassezia sympodialis	Staphylococcus warneri
Cutibacterium avidum	Micrococcus luteus	Streptococcus mitis
Cutibacterium granulosum	Moraxella osloensis	

Out of the panel of 50 species, we filtered out undetected ones, leaving a set of 32 (Table 1). Gene copies were converted to pseudocounts by adding a unit value. We calculated relative abundances for each sample and performed centered log transformation (CLR).
The CLR values were obtained as following:
The geometric mean of the distribution was calculated as:

$$g_m = \left(\prod_{i=1}^{n} a_i \right)^{1/n}$$

And the log ratio was computed as follows:

$$c_i = \log\left(\frac{a_i}{g_m}\right)$$

Where a_i is the pseudocount for a given species i and n is the total number of species.

Scoring metric

Root mean squared error (RMSE) was used to asses model performance on the obtained predicted vector. Based on the current workflow, the application of RMSE to the CLR transformed data is equivalent to the Aitchison distance.[9] Unless otherwise stated we use the raw RMSE for which a lower score is better. In particular, we set an arbitrary threshold of achieving a RMSE < 1 as a definition of acceptable performance. This value can be understood as the 2.7 fold change of the median sample.

Model selection

Table 2. Regression models tested along with the parameters and hyperparameters space ranges used for randomized grid search. Description of each parameter can be found in the scikit-learn documentation.

Model	Parameter	Range
Random Forest	n_estimators	50 to 200
	max_depth	2 to 8
	min_samples_split	2 to 5
	min_samples_leaf	1 to 4
Linear Regression	fit_intercept	True, False
	positive	True, False
Decision Tree	max_depth	2 to 8
	min_samples_split	2 to 6
	min_samples_leaf	1 to 4
PLS	n_components	1 to 3
	scale	True, False
SVR	C	0.1 to 5
	epsilon	0.1 to 0.5
	kernel	linear, rbf
Lasso	alpha	0.01 to 1
	max_iter	100 to 500

Variance Mitigation Strategies

We avoided assumptions about site performance by generating random splits of the sites and testing models with 7 to 12 sites. Multiple random seeds were tested to prevent bias.

Leave-One-Subject-Out (LOSO) Cross Validation

A nested LOSO cross-validation approach was implemented to prevent data leakage between subjects. Each model was tested by iteratively leaving one subject out, optimizing hyperparameters using random grid search on the training set, and selecting the best model for site selection.

Site Selection

Mixed-effects linear models were used to evaluate the effect of interpersonal variation, random initialization, and the number of sampled sites on model performance. The best overall cross-validation score determined the selected model.

Brute Force Algorithm

This approach involved testing 2,097,151 site combinations to solve the multioutput regression problem.

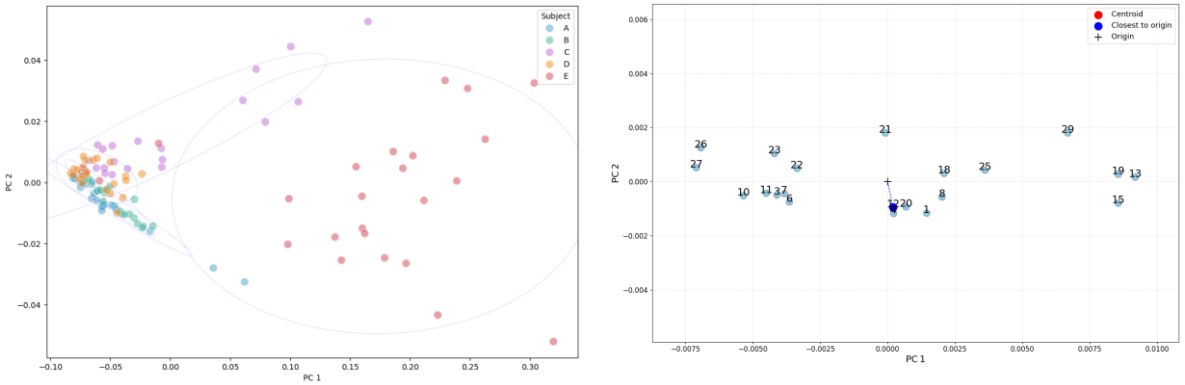
Greedy Breadth First Search Algorithm

This more computationally efficient method explored solution space layer by layer, selecting the site combination that provided the highest immediate improvement in model performance at each step.

Facial Maps Generation

Microbial distributions at experimental and predicted sites were inferred using thin-plate splines (TPS) from the scikit-learn library to generate complete facial maps.

3. Results



(a) (b)

Figure 1. Principal coordinate analysis of β-diversity dissimilarity.(a) All samples colored by the donor subject;(b) Diversity calculated based on average microbial load aggregated by facial site. Each site is represented by a filled light blue dot, annotated by the location on the facial mapping tool. The cross shows the origin of the plot, and the blue line shows the distance between the closest point and the origin.

Prior to modeling, we investigated the qPCR results by analyzing diversity revealing strong clustering across subjects (Figure 1a). This revealed strong interpersonal differences within our cohort and further highlighted some intrapersonal differences on the facial skin of the same subject. We further investigated the dissimilarity between sites by averaging microbial loads across sites (Figure 1b) and identified site 12 (under eye, inner) as the least dissimilar site compared to all others. Site 4 (forehead, middle left, upper) and 20 (cheek, middle, lateral) lie also in close proximity.

Model selection results

Table 3. Statistical evaluation of model performance (RMSE) given the model and the held-out subjects.

Variable	Coefficient	[0.025	0.975]
Intercept	-1.9274	-1.951	-1.904
Subject A	-0.6260	-0.646	-0.606
Subject B	0.1735	0.153	0.194
Subject C	-0.6099	-0.630	-0.589
Subject D	-0.2731	-0.294	-0.253
Subject E	-0.2523	-0.273	-0.232
DecisionTree	0.2744	0.255	0.294
LassoLars	**0.5173**	**0.498**	**0.537**
LinearRegression	0.5235	0.504	0.543
PLS	0.5689	0.549	0.589
RandomForest	0.4426	0.423	0.462
SVR	0.5639	0.544	0.584
Number of sites	0.0124	0.010	0.015

Lasso regression was the top-performing model for a specific combination of 12 sites (coeff = 0.5173). While decision trees and random forests showed lower efficiency, other regression methods like multi linear regression (coeff = 0.5235), Partial Least Squares (PLS, coeff = 0.5689), and Support Vector Regression (SVR, coeff = 0.5639) performed similarly to Lasso.

A subject-specific bias was noted, impacting predictions for Subject A (coeff = -0.6260) and Subject C (coeff = -0.6099). Subjects D, E showed similar results (coeff = -0.2731 and -0.2523, respectively). Subject B presented a positive coefficient (coeff = 0.1735). The number of sites in the model showed only a marginal effect in improving model performance in this setup (coeff = 0.0124).

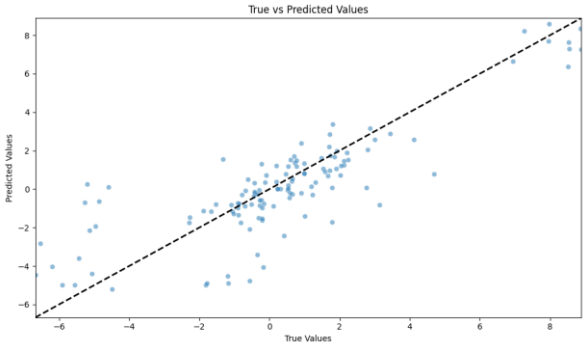


Figure 2. Experimental versus predicted values plot. The diagonal line represents exact predictions. Light blue dots represent the predicted CLR values (RMSE = 1.77). The specific model shown has been trained with 12 sites.

The obtained model results are presented in Figure 2 for a representative test set, based on the cross-validated model. Negative CLR values can be attributed to low abundant species, where the model shows the highest variation in results. Prediction accuracy tends to increase with increasing relative abundance. At the largest CLR values, most often where C. Acnes is located in the distribution, the variation also seems to increase.

Site selection

Brute force vs Greedy Breadth First Search

Compared to brute force, which exhaustively tests all possible combinations, the greedy breadth first search algorithm significantly reduces the number of evaluations required, thus saving time and computational resources (Figure 3). It consistently finds solution in the top percentiles of the distribution for each number of sites and on average improves the RMSE by 0.09.

With the obtained selected sites (lowest RMSE) and model, we predicted the distribution of all investigated species (Figure 3). The sparsest data was held out as an unseen test set. In the training set, *C. Acnes*, *K. Rhizophila*, *M. Restricta* and *S. Epidermidis* are the best fitted species. From the held out species, we observe that *L. Iners*, *S. Aureus* and *S. Caprae* have a comparable RMSE to the best fitted species in the training set. The average RMSE for the training set was 1.36, whereas the one for testing was 1.64.

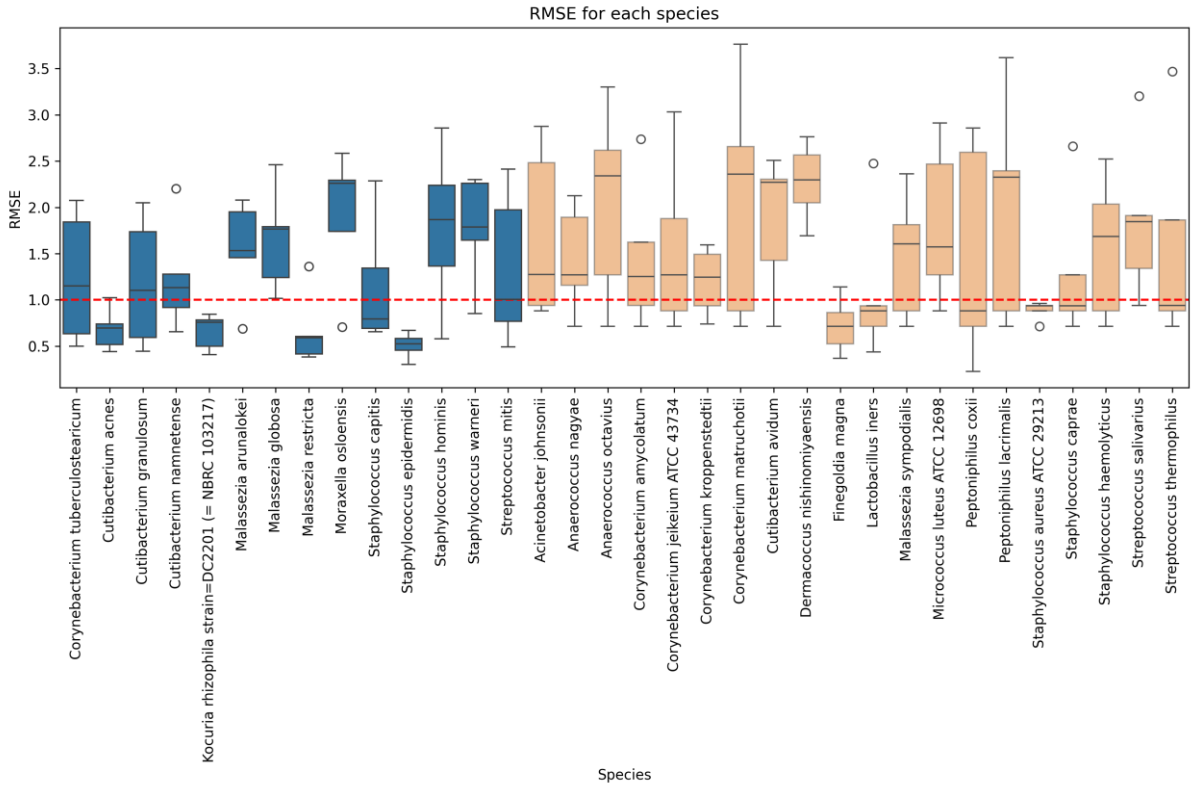
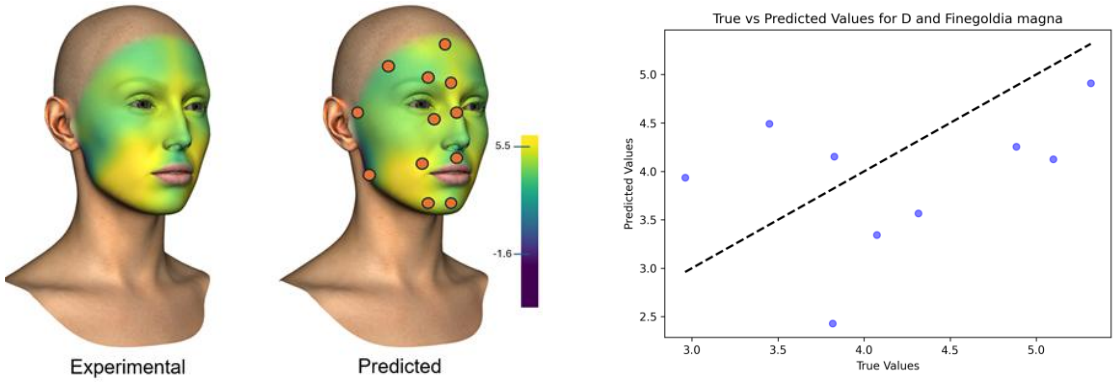


Figure 3 Model performance aggregated by species. The boxplot show the distribution of the RMSE across the 5 subject (when the subject is held-out for testing). The blue boxplots represent species used for training whereas the examples depicted in orange were held out for testing. The red dotted-line represent the target RMSE=1.



(a) (b)

Figure 4. Example of model application on *F. magna* on a single subject. (a) Predicted vs experimental facial map using 12 sites for prediction (filled orange dots). (b) Experimental vs Predicted values in CLR transformed space, the dashed diagonal represents an ideal fit (RMSE = 0.86). The results are presented in the transformed space to prevent data leakage by introducing real microbial load.

The resulting model can be applied to generate facial maps (Figure 4).

4. Discussion

The skin microbiome environment can be modeled for various species, with model performance largely dependent on data quality and sparsity. Common machine learning methods have been validated to predict microbial distribution on the face, although prediction quality declines with very low microbial abundance. Without analytical replicates, it is challenging to determine noise levels from the sampling process.

Despite this limitation, this study is unique in its large cohort and extensive site collection. We observed significant inter-personal differences in the microbiome, yet this did not hinder the modeling effort. Model performance was mainly affected by data sparsity, particularly in low abundance scenarios. For species with higher abundance, such as C. acnes, the model effectively learned the underlying distribution. The LOSO strategy suggests that these results can be replicated in future studies. However, model performance on held-out species was expectedly worse due to their sparse and low abundance nature. For some species, like F. Magnia, the model effectively predicted distribution, supporting the hypothesis of common distribution patterns across the face.

Interestingly, the sites with the lowest RMSE predictions were at the edges (Figure 4a), likely due to their high variance. Selecting these sites as inputs allowed the model to suppress greater prediction errors. With technological improvements, system variance is expected to decrease, necessitating further training iterations to select optimal prediction sites. The selection of the site under the eye as an input was expected given its centrality within the PCoA space, potentially making it the most representative site in our cohort. The absence of cheek and forehead sites, despite their centrality in the PCoA, suggests that collinearity degree influences site preference for achieving the lowest RMSE.

Future iterations will refine site selection and further reduce prediction error, enhancing our understanding of microbial distribution and informing personalized skincare and targeted therapies.

Limitations of splining for facial map generation and modeling approach

While we are able to generate facial plots using TPS, we do not have a complete ground truth of the facial microbiome. As a reference, we use the interpolated facial map generated from all available measurement points. The density and distribution of such measurements can significantly impact the obtained representations and radically change local maxima and minima. This should be taken into account when using facial maps for microbiome interpretation. This effect may compound further with the prediction model proposed in this work.

5. Conclusion

The study presents the application of machine learning techniques to predict microbial distributions on facial skin. Understanding their behavior can aid in developing effective treatments. Currently, this research offers a framework for modeling the facial skin microbiome and practical methods for data handling. The approach results in a notable reduction in experimental costs by minimizing the required sampling sites.

Future efforts will focus on enhancing these models and further studying the implications of microbial distributions. Alternative preprocessing strategies and regression models will be

considered to reduce prediction error. With additional data and improved analytical methods, the framework could yield more accurate predictions, including total microbial load, further reducing experimental costs.

These advancements will deepen our understanding of skin microbiomes. Examining the interaction between microbes and environmental factors may provide valuable insights into personalized skincare and targeted therapies. This study establishes a framework for modeling facial skin microbiomes and offers practical data handling methods. Reducing sampling sites without compromising data quality represents significant progress in microbiome research, enabling more cost-effective and scalable future studies.

6. References

[1] E. A. Grice and J. A. Segre, “The skin microbiome,” *Nat Rev Microbiol*, vol. 9, no. 4, pp. 244–253, Apr. 2011, doi: 10.1038/nrmicro2537.

[2] W. Zhou *et al.*, “Skin microbiome attributes associate with biophysical skin aging,” Genomics, preprint, Feb. 2023. doi: 10.1101/2023.01.30.526239.

[3] M. P. Pagac, M. Stalder, and R. Campiche, “Menopause and facial skin microbiomes: a pilot study revealing novel insights into their relationship,” *Front. Aging*, vol. 5, p. 1353082, Mar. 2024, doi: 10.3389/fragi.2024.1353082.

[4] R. Voegeli, A. V. Rawlings, P. Seroul, and B. Summers, “A novel continuous colour mapping approach for visualization of facial skin hydration and transepidermal water loss for four ethnic groups,” *International Journal of Cosmetic Science*, 2015.

[5] H. Lee *et al.*, “Comparative analysis of human facial skin microbiome between topical sites compared to entire face,” *Genes Genom*, vol. 43, no. 12, pp. 1483–1495, Dec. 2021, doi: 10.1007/s13258-021-01180-2.

[6] Jai Ram Rideout *et al.*, *scikit-bio/scikit-bio: scikit-bio 0.6.3*. (Jan. 13, 2025). Zenodo. doi: 10.5281/ZENODO.14640761.

[7] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” Jun. 05, 2018, *arXiv*: arXiv:1201.0490. doi: 10.48550/arXiv.1201.0490.

[8] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[9] J. Aitchison, “The Statistical Analysis of Compositional Data,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 44, no. 2, pp. 139–160, Jan. 1982, doi: 10.1111/j.2517-6161.1982.tb01195.x.