Received: 2 January 2025 | Revised: 21 July 2025 | Accepted: 25 July 2025

DOI: 10.1002/jper.70000

JOURNAL OF Periodontology

ORIGINAL ARTICLE

Diversity and random forest models of oral microbiomes in periodontal health using publicly available data

Alba Regueira-Iglesias[1] | **Berta Suárez-Rodríguez**[1] | **Triana Blanco-Pintos**[1] | **Alba Sánchez-Barco**[1] | **Marta Relvas**[2] | **Carlos Balsa-Castro**[1] | **Inmaculada Tomás**[1]

[1]Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialties, School of Medicine and Dentistry, Universidade de Santiago de Compostela, Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain

[2]Oral Pathology and Rehabilitation Research Unit (UNIPRO), University Institute of Health Sciences (IUCS-CESPU), Gandra, Portugal

Correspondence
Inmaculada Tomás and Triana Blanco-Pintos, School of Medicine and Dentistry, Universidade de Santiago de Compostela, 15872 Santiago de Compostela, Spain.
Email: inmaculada.tomas@usc.es and triana.blanco.pintos@usc.es

Funding information
Instituto de Salud Carlos III, Grant/Award Number: PI24/00222

Abstract

Background: Evidence on the 16S metabarcoding of supragingival, subgingival, and salivary microbiomes in periodontal health remains limited. We aimed to analyze the diversity and potential of machine-learning models of supragingival, subgingival, and salivary microbiomes in periodontal health.

Methods: A total of 848 samples (supragingival = 210; subgingival = 155; saliva = 483) from 491 periodontally healthy subjects were included. Publicly available Illumina sequences were processed with mothur, and taxonomy was assigned using an oral-specific database. Random forest (RF) models were built on the training set (2/3 of the samples) using a 3-fold cross-validation. They were tested on the test set (1/3).

Results: A total of 121 amplicon sequence variants (ASVs) presented with differential abundances between the two types of plaque, 212 between the supragingival and saliva samples, and 160 between the subgingival and saliva ($p < 0.01$). Furthermore, the supragingival versus subgingival model consisted of five ASVs. The performance parameters on the test set were area under the curve (AUC) = 0.908, accuracy (ACC) = 84.30%, sensitivity = 95.71%, and specificity = 68.63%. Both the supragingival and subgingival versus saliva models also had five ASVs. These two models revealed similar performance (AUC = 0.992 and 0.986, ACC > 95%, sensitivity > 90%, specificity > 95%).

Conclusion: Although supragingival and subgingival bacterial profiles diverged only modestly, primarily due to taxa with small effect sizes, they were both compositionally distinct from the salivary microbiome. RF models accurately classified samples by niche, with higher performance in distinguishing saliva from plaques. Specific ASVs from *Escherichia*, *Fusobacterium*, *Granulicatella*, *Treponema*, *Peptostreptococcaceae* [XI][G-9], and *Prevotella* were identified in subgingival plaque, while *Oribacterium* and *Solobacterium* were identified in

saliva, indicating potential niche-specific microbial signatures in periodontal health.

KEYWORDS
16S rRNA gene, dental plaque, machine learning, microbiome, periodontal health, saliva, sequencing

Plain Language Summary

Mapping oral microbes in relation to periodontal health is essential for microbiome-based diagnostics and the development of new preventive/therapeutic strategies. Our two-by-two predictive models demonstrated that a small set of bacterial ASVs can accurately classify periodontally healthy samples according to their oral niche. Notably, models distinguishing saliva from dental plaques achieved superior performance compared to those discriminating between plaques. This likely reflects the greater resemblance in dominant microbial taxa between the two plaque niches. These findings underscore the potential of machine-learning approaches to identify key microbial signatures and highlight the predictive ASVs as promising biomarkers for characterizing oral niches in periodontal health.

1 | INTRODUCTION

The symbiosis between the mouth microbes is essential for maintaining health.[1] Conversely, oral-microbiome dysbiosis plays a crucial role in developing globally prevalent mouth and systemic diseases.[2] One of the most widely used techniques for studying microbial communities is sequencing the 16S rRNA gene. This technology has described the microbiome's diversity in oral and general health using supragingival,[3] subgingival,[4] or salivary[5] samples. Investigations involving healthy specimens from the three niches are minimal, with only two articles in the literature.[6,7]

Nevertheless, 16S gene-sequencing studies typically have sample-size limitations.[3–5] This makes it difficult to define health-related profiles and differentiate between niches, as individual variations can be mistaken for real biological differences.[8] Moreover, methodological shortcomings in the sequencing workflow are frequent.[9,10] Common issues include using 454-pyrosequencing, clustering amplicons into operational taxonomic units, or ignoring microbiome compositionality.[6,7] Lastly, no publication assesses the predictive ability of taxa to distinguish between oral niches in periodontal health.

In recent years, the rise of artificial intelligence has impacted microbiome research,[11] with unsupervised machine learning (ML) methods like the principal component analysis (PCA) in regular use.[3,5,6] Supervised approaches, which accurately classify samples[11] and complement and enhance differential abundance analyses in identifying class-associated variables,[11,12] are uncommon.

The attainment of reliable, non-overfitted, and generalizable results requires a large sample size that also allows for the validation of the models obtained.[13] Concerning this, heterogeneity within each class that the model attempts to distinguish is fundamental.[14] Variability among sample donors in demographic aspects allows for the capture of relevant differences, thereby enhancing the generalizability of the diagnostic model.[14]

Given the above, this study aims to: (1) analyze the diversity of the supragingival, subgingival, and salivary microbiomes of periodontally healthy subjects; and (2) develop supervised ML models to classify healthy samples according to their niche. A multi-batch approach was adopted to achieve a large sample size,[9] and the sequences from publicly available Illumina V3-V4 bioprojects were processed together. Finally, scientific evidence was considered to apply the best 16S gene-sequencing methodological practices.[9,10]

2 | MATERIALS AND METHODS

Figure 1 illustrates the methodological workflow followed in this observational cross-sectional study. This investigation is part of a large-scale project of our research team.[12] We include publicly available metadata and Illumina V3–V4 sequence data of periodontally healthy adults acquired

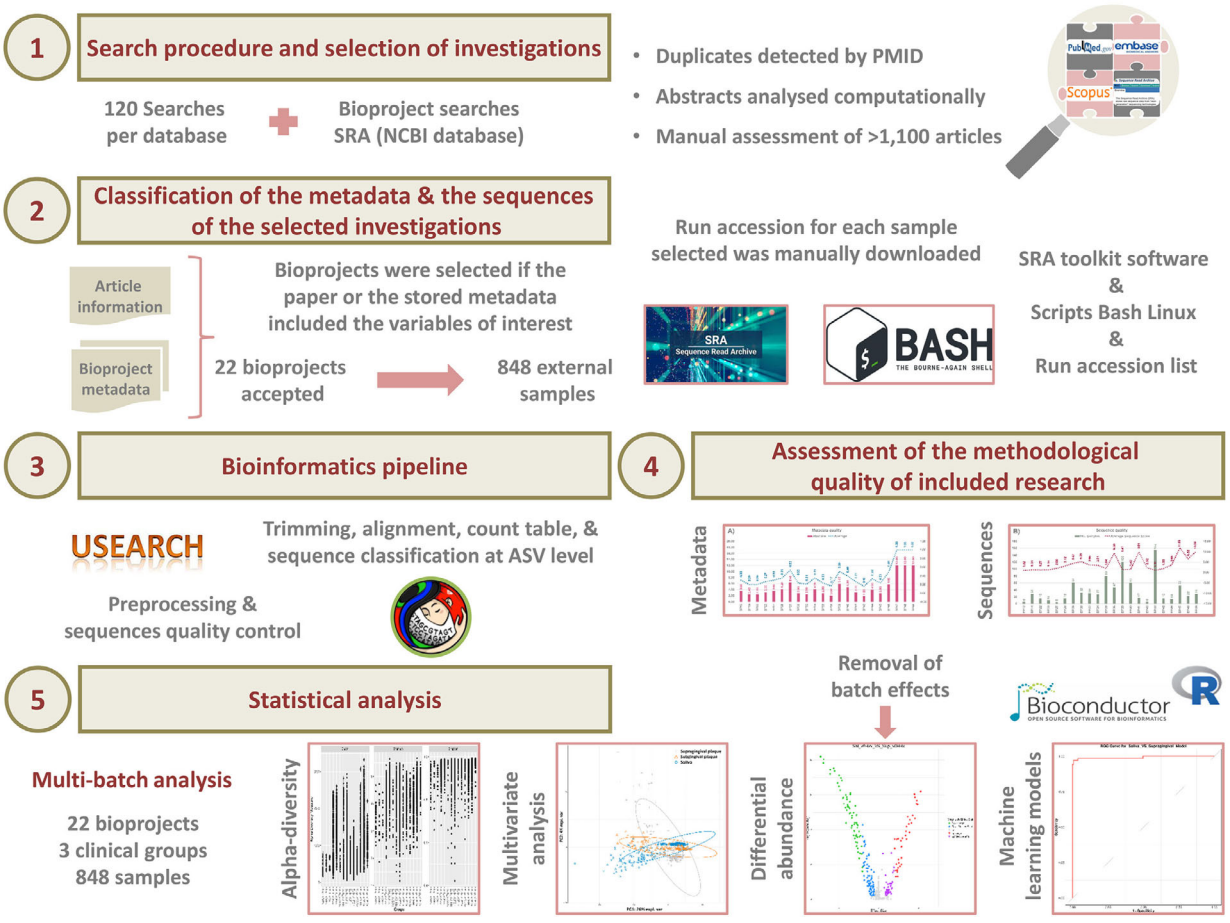

FIGURE 1 Methodological workflow of the present multi-batch study.

from previous studies on supragingival, subgingival, and salivary microbiomes.

2.1 | Inclusion and exclusion criteria

Studies were included if they used a reference standard for periodontal diagnosis based on clinical or clinical and radiographic parameters. Those without a diagnostic reference or not evaluating periodontal status with at least one clinical parameter were excluded.

Selected bioprojects needed properly assigned metadata in the repository, ensuring that each sample could be correctly classified as periodontal health. Among other standards (Appendix S1), the sequences had to have a minimum contig length ≥ 350 base pairs (bps), and primers had to align with the *Escherichia coli* J01859.1 16S rRNA gene.

2.2 | Searches procedure

In May 2023, 120 searches per database were conducted in PubMed, Scopus, and Embase using terms related to periodontal conditions, oral niches, and microbiota. Found abstracts were computationally and manually evaluated (Appendix S1).

2.3 | Classification of data from included research

The information in the sequence read archive (SRA)[15] was accessed using the bioproject identifiers of the accepted research. A new metadata table was created for each bioproject, including information from articles/authors. Sequence data were downloaded using the SRA Toolkit.

2.4 | Bioinformatics pipeline

The bioinformatic analysis was carried out according to a protocol described in detail in Appendix S1. Briefly, sequences were pre-processed and quality assessed with USEARCH,[16] discarding those with a length < 300 bps. The mothur pipeline[17] was employed for amplicon sequence variants (ASVs) inference. Sequences were

not clustered, ensuring maximum identification and classification at the ASV level, where only 100% identical sequences were grouped as the same ASV. The oral-specific database of Escapa et al.,[18] an expanded and curated version of the expanded Human Oral Microbiome Database (eHOMD),[19] was used for the taxonomic assignment.

2.5 | Methodological quality of included research

The metadata and sequence data were evaluated to inform readers about the quality of the bioprojects included. The lack of standardized tools for this purpose led to the development of our methodology.

Metadata quality was assessed using a 12-variable checklist (Appendix S1). Each variable was scored from 1.00 (specified in the metadata table) to 0.00 (unavailable). The mean score of applicable variables was used to categorize the quality of the metadata as low (0.00–0.33), medium (0.34–0.66), or high (0.67–1.00).

The quality of sequence data was evaluated using the number of samples/bioproject and the average sequence score (ASS, calculated as the average number of high-quality sequences/sample divided by 10,000). ASS was categorized as very low (< 0.25), low (0.25–0.75), acceptable (0.75–1.00), high (1.00–2.00), and very high (> 2.00).

2.6 | Statistical analysis

The statistical analysis was performed using R-Bioconductor.[20] Samples with $< 2{,}500$ sequences were excluded. This left 848 specimens (subgingival [sub] = 155, supragingival [sup] = 210, and saliva [sal] = 483) provided by 491 periodontally healthy adults. ASVs with abundance ≤ 10 counts and present in ≤ 2 samples were excluded, meaning 10,577 ASVs remained. Throughout the study, ASVs will be designated by their corresponding genus and species names, as well as their ASV identifiers in the taxonomic database. The original count matrix was transformed using the centered log-ratio (CLR) method for differential abundance analysis and random forest modeling.

2.6.1 | Analysis of alpha-diversity

The observed ASVs, 95% coverage index, Shannon index, and Pielou index were calculated using phyloseq and microbiome packages.[21,22] The comparative analyses of alpha-diversity estimators among study groups were performed using the Mann–Whitney *U* test.

2.6.2 | Analysis of microbial community structure

A PCA was employed to visualize the clustering of the healthy samples according to their niche. A non-parametric permutational multivariate analysis of variance (PERMANOVA) was employed to assess community-level differences between groups. Both were performed with vegan.[23]

2.6.3 | Differential abundance analysis: Elimination of batch effects

Batch effects (BEs) were removed, replicating Wang and Lê Cao,[24] and this was done just before the differential abundance analysis[12] (Appendix S1). Once completed, the difference in median abundance between all the ASVs in the study groups was assessed with the Mann–Whitney–Wilcoxon test. The *p*-value was adjusted with the Benjamini–Hochberg correction using mutoss.[25] We obtained the corresponding effect size for each ASV, including its confidence intervals and magnitude, using Cohen's *d* and Hedges' *g* statistics from the effsize package.[26] ASVs with an adjusted *p*-value < 0.01 and large, medium, or small effect sizes were determined to be differentially abundant.

2.6.4 | Random forest modeling

We used mixOmics[27] to conduct a sparse Partial Least Squares Discriminant Analysis (sPLS-DA) to facilitate the two-by-two categorization of the study groups and to identify the ASVs that best distinguished them (Appendix S1). The adapted genetic algorithm process was initialized with the GA package.[28] The 25 variables with the largest weights in the sPLS-DA were selected to perform 20 processes for each classification. Each of these was initialized with two random variables from the previous 25. The first model was created by random forest (RF) with caret[29] using the training data (2/3 of samples: sup = 140; sub = 104; sal = 322 and a 3-fold cross-validation to control for overfitting. The performance of this model was calculated using the test set (1/3: sup = 70, sub = 51, sal = 161).

Applying the RF procedure explained, multiple models were trained (training data), and the last model generated was compared with its immediate predecessor. The models were compared and selected using an overall fitness score

(test data) (Appendix S1). The optimum model for each of the 20 processes was chosen, and the best model for each classification was selected.

Finally, we calculated the performance parameters on the test data: area under the curve (AUC), accuracy (ACC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Receiver operating characteristic (ROC) and precision-recall curves were constructed.

3 | RESULTS

3.1 | Bioprojects obtained in the search process

Following the computational analysis of 30,389 studies, 1,202 were selected for the manual assessment (Appendix S2). Twenty-six articles were full-text reviewed, and four were removed during the metadata and sequence evaluations. Twenty-two bioprojects were included in our analysis.

3.2 | Characteristics of the selected studies

Around 20% of the investigations established periodontal diagnosis using the 2018 Classification of Periodontal and Peri-implant Diseases and Conditions[30] (Appendix S3). Overall, periodontal health was defined as (1) probing pocket depth (PPD) ≤ 3 mm (15/22; 68.18%); (2) bleeding on probing (BOP) < 10% (7/22; 31.82%), < 20% (5/22; 22.73%), or no BOP (3/22; 13.64%); and (3) no clinical attachment loss (CAL) of < 1 mm (11/22; 50.00%), or CAL < 3 mm (1/22; 4.55%). Importantly, the latter study contributed just two samples to the total of 848 specimens included in our analysis, thus exerting minimal influence on the overall results.

3.3 | Methodological quality of the included research

The quality of the metadata was high (range = 1.00 to 0.67) in three out of 22 included studies (13.64%), medium in six (range = 0.66 to 0.34; 27.27%), and low in 13 (range = 0.33 to 0.00; 59.09%) (Appendix S4). All the bioprojects had the basic information required to be part of our analysis: type of oral sample and periodontal health status of the specimen donor.

However, papers with medium- and low-quality metadata lacked information on each sample's ethnicity, number of teeth, or periodontal parameters. Although the mean estimates for all participants were given in the paper, the values of, for example, the mean full-mouth PPD associated with each patient sample were not specified. Articles with low quality also failed to specify each participant's age or sex.

As for the quality of the sequences, it was also evaluated on the data included in our analysis and was done after applying the quality filters indicated in the Materials and Methods section. Ten bioprojects (45.45%) achieved an ASS between 1.00 and 2.00, demonstrating a high quantity. Twelve (54.55%) had an ASS > 2.00, representing research with a very high number of sequences.

3.4 | Analysis of alpha-diversity

In periodontal health, all the richness estimators were higher in the supragingival plaque than in subgingival plaque and saliva (ASVs observed = 612.00 versus 479.00 and 445.00, respectively; 95% coverage = 224.00 *vs*. 171.00 and 172.00; $p < 0.05$ in all comparisons). The 95% coverage index was lower in subgingival plaque than in saliva ($p < 0.01$; Appendix S5).

In contrast, diversity and evenness were higher in the subgingival and salivary niches than in the supragingival (Shannon = 4.15 and 4.11 vs. 4.07; Pielou = 0.65 both vs. 0.62). Differences were significant for the Shannon index between the supragingival and saliva samples ($p < 0.05$) and for Pielou's between the supragingival and the other two niches ($p < 0.01$). No differences were found between subgingival plaque and saliva (Appendix S5).

3.5 | Analysis of the microbial community structure

The PCA revealed a clustering of periodontally healthy samples according to the mouth niche to which they belonged (Appendix S6). The PERMANOVA test confirmed this visual observation ($p < 0.0001$ for the three groups and the two-by-two comparisons).

3.6 | Differential abundance analysis

The comparison of supragingival vs. subgingival plaque in periodontal health identified abundance differences between the niches in 121 ASVs (39.93% of those with an abundance $\geq 0.05\%$ in at least one of the groups being compared). Only 10 (3.30%) had the most remarkable differences (effect size ranges: sup = 1.20 to 0.87; sub = −0.82 to −1.42). Additionally, most differentially abundant ASVs

TABLE 1 Number of ASVs with differential abundance between the oral niches in periodontal health.

Groups	No. of ASVs (%) with each effect size magnitude			
	All	Large	Medium	Small
Supra vs. Sub	121 (39.93)	10 (3.30)	38 (12.54)	73 (24.09)
Abundant in supra	90 (29.70)	4 (1.32)	34 (11.22)	52 (17.16)
Abundant in sub	31 (10.23)	6 (1.98)	4 (1.32)	21 (6.93)
Supra vs. Saliva	212 (68.83)	107 (34.74)	44 (14.29)	61 (19.81)
Abundant in supra	72 (23.38)	36 (11.69)	16 (5.19)	20 (6.49)
Abundant in saliva	140 (45.45)	71 (23.05)	28 (9.09)	41 (13.31)
Sub vs. Saliva	160 (51.61)	95 (30.65)	32 (10.32)	33 (10.65)
Abundant in sub	42 (13.55)	23 (7.42)	9 (2.90)	10 (3.23)
Abundant in saliva	118 (38.06)	72 (23.23)	23 (7.42)	23 (7.42)

Note: ASVs with an adjusted *p*-value < 0.01 and large, medium, or small effect sizes were determined to be differentially abundant.

The percentages of ASVs are calculated with respect to the number of ASVs with abundance ≥ 0.05% in the supragingival vs. subgingival comparison (303) and ≥ 0.404% in the supragingival and subgingival vs. saliva comparisons (308 and 310, respectively).

Abbreviations: ASVs, amplicon sequence variants; No., number; Sub, subgingival plaque; Supra, supragingival plaque.

were more present in the supragingival plaque (sup = 90; sub = 31; Table 1).

In the comparison of supragingival plaque vs. saliva, there were differential abundances in 212 ASVs (68.83% of those with abundance ≥0.04% in at least one of the groups). Of these, 107 (34.74%) had the largest effect sizes (sup = 2.11 to 0.80; sal = −0.80 to −2.91). Most differentially abundant ASVs were more present in saliva (sup = 72; sal = 140; Table 1).

For the subgingival plaque vs. saliva, there were 160 differentially abundant ASVs (51.61% of those with abundance ≥ 0.04% in at least one of the groups). Ninety-five (30.65%) had the most outstanding differences (effect size ranges: sub = 2.48 to 0.81; sal = −0.80 to −3.33). Again, most of the different ASVs were more present in saliva (sub = 42; sal = 118; Table 1).

3.6.1 | ASVs associated with periodontal health in each mouth niche

Figure 2 lists the main ASVs with differential abundance in the group comparisons. In the first contrast, sorted by the absolute value of effect size (≥ 1.00), *Actinomyces* HMT448-AV384, *Parvimonas* HMT110-AV21, and *Solobacterium moorei*-AV197 were identified as being associated with supragingival plaque. Meanwhile, *Rothia mucilaginosa*-AV40, 48, and 54 were strongly related to the subgingival plaque (≤ −1.00).

In the second comparison, *Actinomyces massiliensis*-AV206, *Corynebacterium durum*-AV102, *Actinomyces* HMT169-AV34, *Kingella oralis*-AV66, and *C. durum*-AV199 were associated with supragingival plaque (≥ 1.80, taxonomy defined at species level). For its part, *Prevotella melaninogenica*-AV7, *Oribacterium sinus*-AV117, *Fusobacterium periodonticum*-AV11, *Haemophilus parainfluenzae*-AV109, and *Veillonella rogosae*-AV84 had a remarkably high association with saliva (≤ -2.00).

For subgingival plaque vs. saliva, the ASVs most associated with the former (≥ 1.50) were *Pseudopropionibacterium propionicum*-AV429, *Veillonella dispar*-AV5, *E. coli*-AV116, *Leptotrichia hongkongensis*-AV24, *Pseudomonas fluorescens*-AV108, and *A.* HMT169-AV34. Prominent in saliva (≤ −2.50) were *Campylobacter concisus*-AV80, *H. parainfluenzae*-AV3, *Veillonella atypica*-AV91, *Streptococcus* HMT057-AV87, *Selenomonas* HMT136-AV223 and 686, *Streptococcus vestibularis*-AV27, and *Megasphaera micronuciformis*-AV409. Appendix S7 contains all ASVs with differential abundance between study groups.

3.7 | Random Forest modeling

The model for classifying the healthy supragingival and subgingival plaque samples consisted of five ASVs (4.17% of the preselected variables). Performance analysis with the test set provided an AUC of 0.908, ACC of 84.30%, a sensitivity of 95.71%, and a PPV of 80.72% for supragingival plaque, and a specificity of 68.63% and an NPV of 92.11% for subgingival samples.

The model to distinguish supragingival plaque from saliva consisted of five ASVs (2.50%). All the performance values on the test set were higher than those obtained for the differentiation between the two plaques (AUC = 0.992; ACC = 98.70%; sup: sensitivity = 97.14%, PPV = 98.55%; sal: specificity = 99.38%, NPV = 98.77%).

Lastly, five ASVs (2.50%) were used to classify the subgingival and saliva samples. Testing with the test set provided similar performance estimators to those in the previous model (Table 2). The sensitivity and PPV values for subgingival plaque were lower, at 90.20%. The ROC and precision-recall curves of the models are depicted in Figure 3.

3.7.1 | Predictive ASVs of each oral niche in periodontal health

Table 3 lists the ASVs constituting the models described above. No ASVs were predictors of supragingival niche in any model.

JOURNAL OF Periodontology

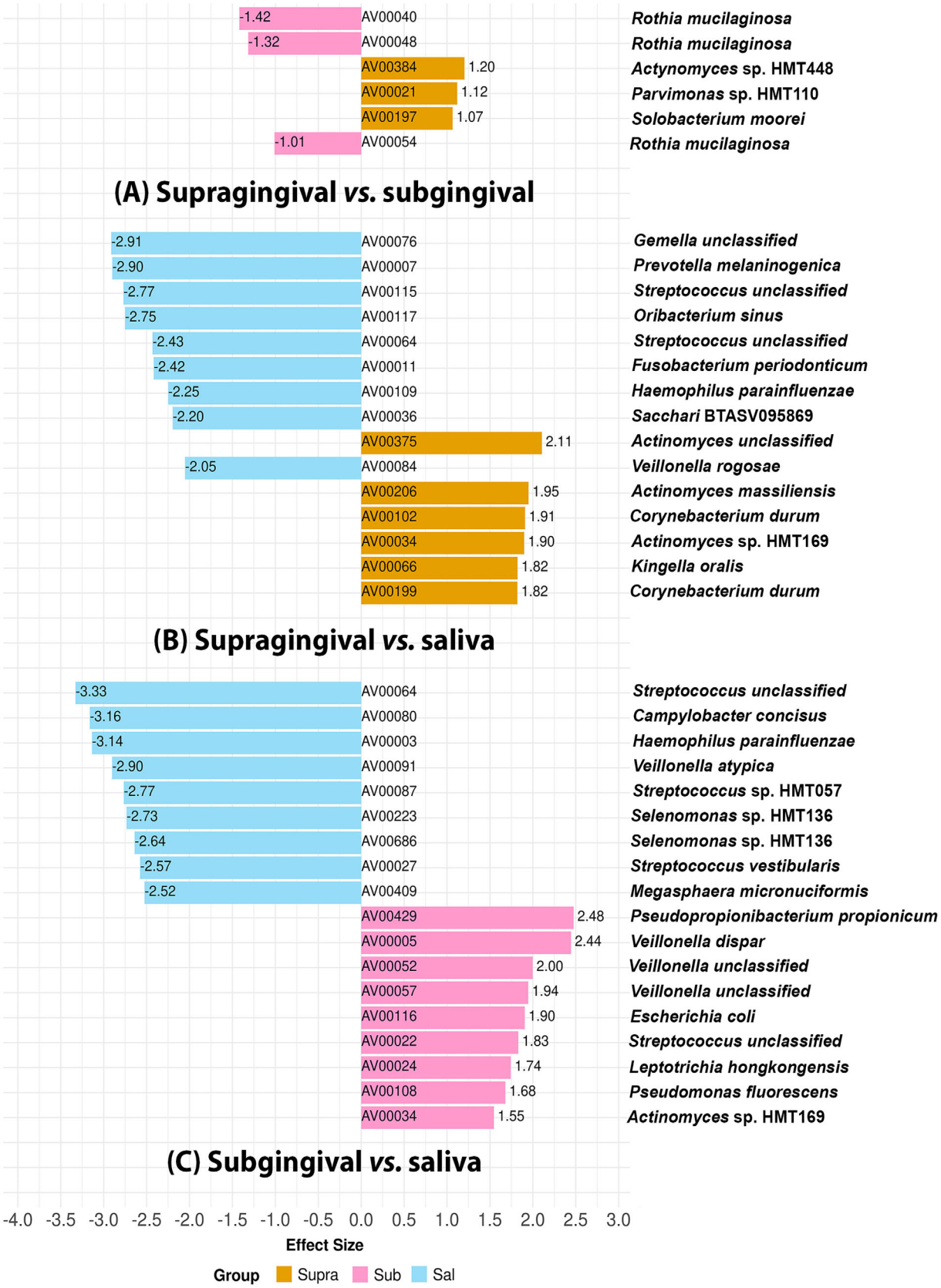

FIGURE 2 Main ASVs with differential abundance and large effect size in the two-by-two comparisons between the study groups. ASV, amplicon sequence variant.

TABLE 2 Random Forest models for distinguishing between the oral niches in periodontal health.

Model	No. ASVs (%)	ACC	Sensitivity	Specificity	PPV	NPV	AUC
Supra vs. sub	5 (4.17)	84.30%	95.71%	68.63%	80.72%	92.11%	0.908
Supra vs. saliva	5 (2.50)	98.70%	97.14%	99.38%	98.55%	98.77%	0.992
Sub vs. saliva	5 (2.50)	95.28%	90.20%	96.89%	90.20%	96.89%	0.986

Note: The percentages of ASVs are calculated with respect to the number of ASVs pre-selected within the sparse Partial Least-Squares Discriminant Analysis (120 for supragingival vs. subgingival, 200 for supragingival and subgingival vs. saliva). In the supragingival plaque vs. subgingival plaque model, the sensitivity and positive predictive value corresponded to supragingival plaque, and the specificity and negative predictive value to subgingival plaque. In the supragingival plaque vs. saliva model, the sensitivity and positive predictive value corresponded to supragingival plaque and the specificity and negative predictive value to saliva. Finally, in the subgingival plaque vs. saliva model, the sensitivity and positive predictive value corresponded to subgingival plaque and the specificity and negative predictive value to saliva.

Abbreviations: ACC, accuracy; ASVs, amplicon sequence variants; AUC, area under the curve; No., number; NPV, negative predictive value; PPV, positive predictive value; Sub, subgingival plaque; Supra, supragingival plaque.

The main subgingival predictor ASVs were *E. coli*-AV116, *Fusobacterium nucleatum vincentii*-AV10, *Granulicatella elegans*-AV207, *Treponema*-AV195 (sup vs. sub), *Gemella morbillorum*-AV136, *Peptostreptococcaceae* [XI][G-9] brachy-AV51, and *Streptococcus intermedius*-AV62 (sub vs. saliva). P. HMT110-AV21 was a predictor of subgingival plaque in both models. For its part, *Capnocytophaga gingivalis*-AV93, *K. oralis*-AV66, *O. sinus*-AV117, *Rothia dentocariosa*-AV2, *Streptococcus*-AV4 (sup vs. sal), and *S. moorei*-AV247 (sub vs. sal) were saliva-predictive ASVs in periodontal health.

Overall, the predictor-ASVs showed concordance between the niche in which they are highly abundant and the niche they were predicted to inhabit (Table 3). Specifically, *O. sinus*-AV117 and *S. moorei*-AV247 were strongly abundant and predictors of saliva. However, *K. oralis*-AV66, *R. dentocariosa*-AV2, and *Streptococcus*-AV4 were more abundant in the niche opposite (supragingival) to the one they predicted (saliva).

4 | DISCUSSION

Comprehensive mapping of the healthy mouth's microbes is crucial for advancing precision dentistry.[31] As the mouth consists of different niches with distinct microbial communities, no universal oral sample represents the entire ecosystem.[32] Identifying the microbes linked to each niche in health would aid in developing diagnostic, preventive, and therapeutic approaches targeted at the location of the pathology.

The only two articles[6,7] that described 16S metabarcoding of supragingival, subgingival, and salivary microbiomes in oral health have methodological shortcomings.[9] Furthermore, none of these studies identified taxa that can discriminate between different sample types based on periodontal health status.

We applied a multi-batch analysis strategy to obtain ~850 periodontally healthy samples from the three oral niches, processing the Illumina V3-V4 sequences from distinct publicly available bioprojects together. A compositional approach was adopted to assess the three microbiomes at the ASV level, focusing on diversity and RF models. Although metadata was low or medium quality in ~86% of the bioprojects analyzed, the strict filter applied to the 16S sequences guaranteed the high quality of the sequence data.[12] The high number of sequences per sample (ASS > 1.00 for all bioprojects) also ensured the robustness of our findings.

4.1 | Methodology of abundance differential and Random Forest modeling

In this study, a limited but expected heterogeneity was observed in the diagnostic criteria used to define periodontal health, as reported in the original research. This reflects the inherent variability of multi-bioproject investigations. Importantly, such variability remained within the accepted clinical spectrum of health, as subjects did not meet diagnostic criteria for periodontitis and presented absent or minimal gingival inflammation. Additionally, there was variability in aspects known to affect the composition of the oral microbiome, such as the subject's age[33] or country of origin.[34] Differentiation between health niches was also favored because the supragingival, subgingival, and salivary samples were not from the same donors.

On the one hand, the heterogeneity arising from sources unrelated to the variable of interest must be controlled before differential abundance analyses using robust statistical approaches, such as the BEs removal strategy proposed by Wang and Lê Cao.[24] If not, the number of differentially abundant bacteria between groups may be artificially inflated, increasing the risk of false positives. Indeed, we have previously demonstrated that the number of differentially abundant ASVs between health and periodontitis decreased from 265 before BE removal to 190

JOURNAL OF Periodontology

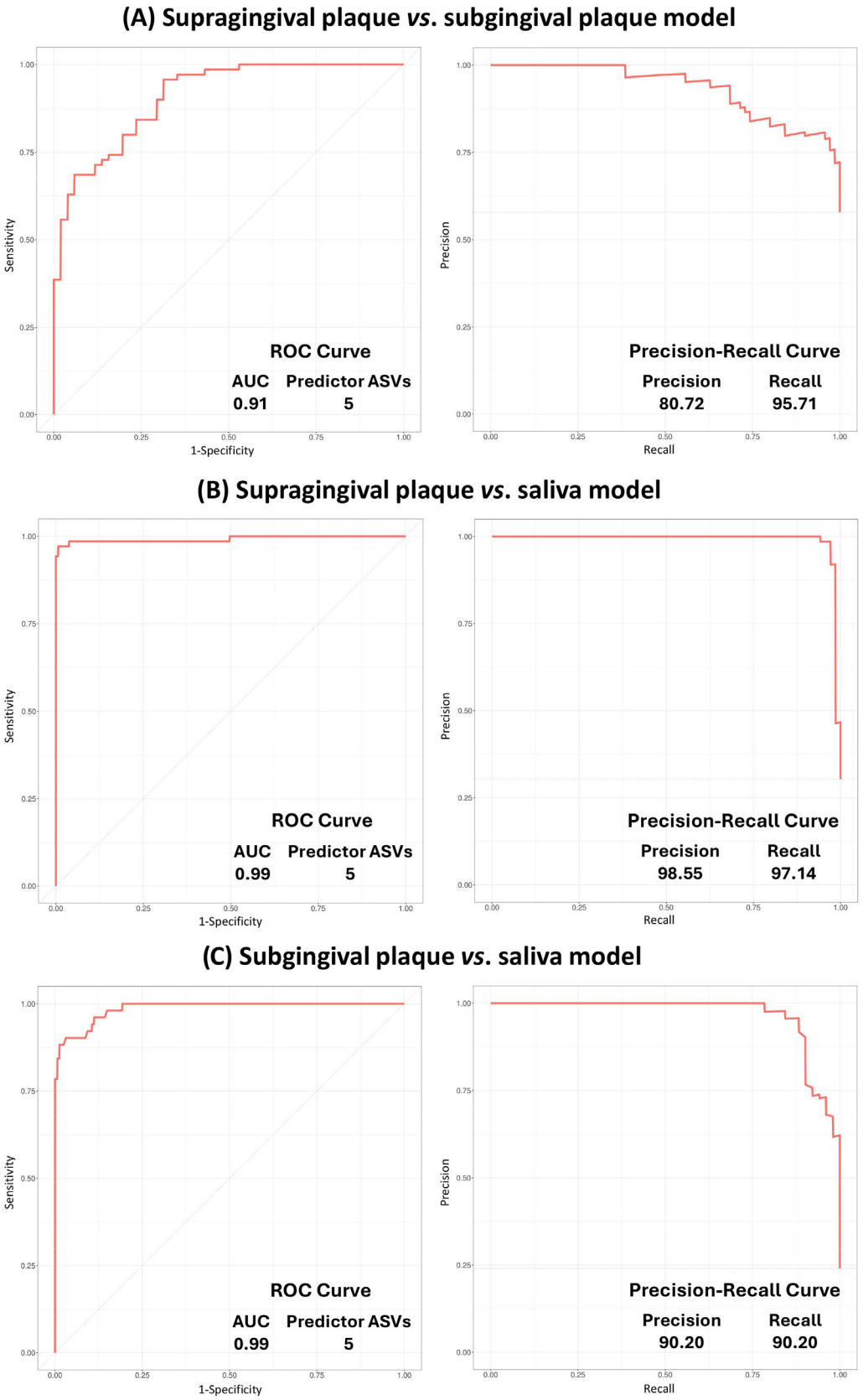

FIGURE 3 Potential of the supragingival, subgingival, and salivary microbiomes to categorize samples from each oral niche in periodontal health.

TABLE 3 Amplicon sequence variants forming the most accurate Random Forest models.

ASVid	Genus	Species	ASV	Supragingival vs. subgingival	Supragingival vs. saliva	Subgingival vs. saliva
AV00093	*Capnocytophaga*	*gingivalis*	Unclassified		0.19[B]	
AV00116	*Escherichia*	*coli*	BTASV133996	1.00[B]		
AV00010	*Fusobacterium*	*nucleatum* subsp. *vincentii*	Unclassified	0.16[A]		
AV00136	*Gemella*	*morbillorum*	Unclassified			0.15[A]
AV00207	*Granulicatella*	*elegans*	Unclassified	0.16[A]		
AV00066	*Kingella*	*oralis*	BTASV175208		0.37[B]	
AV00117	*Oribacterium*	*sinus*	BTASV107685		1.00[A]	
AV00021	*Parvimonas*	sp. HMT110	Unclassified	0.16[A]		0.18[A]
AV00051	*Peptostreptococcaceae* [XI][G-9]	*brachy*	BTASV129419			0.21[A]
AV00002	*Rothia*	*dentocariosa*	BTASV138915		0.21[B]	
AV00247	*Solobacterium*	*moorei*	Unclassified			1.00[A]
AV00062	*Streptococcus*	*intermedius*	BTASV162089			0.19[A]
AV00004	*Streptococcus*	unclassified	Unclassified		0.46[B]	
AV00195	*Treponema*	unclassified	Unclassified	0.18[A]		

Note: The cells are colored according to the conditions predicted by a given ASV in each predictive model. Thus, pink corresponds to subgingival plaque-predictive ASV and blue to saliva-predictive ASV. The weight of the variables in the variable pre-selection model (sparse Partial-Least Squares Discriminant Analysis) performed in mixOmics is indicated in each case. ASVs with values closer to 1.00 show a higher weight in the models. Codes A and B reflect different relationships between ASV percentage abundance (% prior to CLR transformation) and its predictive role. In case A, the ASV is more abundant in the niche it predicts, while in case B, it is more abundant in the opposite niche. These situations may arise from complex interactions within the model and highlight the need to interpret predictive value and percentage abundance values as complementary but not necessarily aligned.
Abbreviations: ASV, amplicon sequence variant; ASVid, amplicon sequence variant identifier.

after, with approximately 45% of the initially identified ASVs not retained after correction.[12]

Moreover, the tools commonly employed for differential abundance analyses in microbiome studies (i.e., LEfSe, DESeq2)[6,7,35] ignore microbiome data compositionality and are sensitive to zero inflation. As these limitations can also result in inflated false-positive rates,[10,36] we addressed this challenge by performing a centered log-ratio (CLR) transformation.[24]

On the other hand, our methodological decision not to remove BEs prior to predictive modeling was deliberate. Population heterogeneity is a crucial criterion that must be met when building accurate and generalizable two-class classification models that are not over-fitted.[14] Failure to consider this diversity may not have captured relevant patterns and led to a decrease in the diagnostic performance of the models.[14] In this sense, we have shown how BEs elimination before predictive modeling can increase the number of predictor ASVs by up to 12-fold (16 before vs. 200 after removal).[12] This is a critical consideration, as simpler models using fewer variables are more practical and cost-effective for clinical implementation.

Furthermore, the methodological requirements applied to control for structural confounding factors ensure the validity of our results. The sequences evaluated belonged to the same technology and gene region and had a minimum length requirement. This, together with the use of an extension of the eHOMD database adapted to ASVs,[18] controlled bacterial taxonomic identification.

4.2 | Analysis of alpha-diversity and microbial community structure

As previously described,[35] the supragingival microbiome was richer than the subgingival and salivary microbiomes. In contrast, the subgingival samples revealed a more even distribution, while those of saliva were both more diverse and more even than the supragingival specimens. Our results confirmed the findings of Segata et al.[7] on evenness but differed from those of Chen et al.[35] on diversity.

The differences above can be explained by the distinct biological characteristics of the niches, even in health. The greater richness of supragingival plaque could be due to its increased access to dietary nutrients compared to subgingival plaque[37] and the retention areas present on tooth surfaces. Meanwhile, reduced oxygen availability in the subgingival niche encourages the emergence of anaerobes.[38] This may lead to the displacement of other microbes, resulting in a more even community. For its part, saliva bathes all oral tissues, favoring the greater diversity of transient organisms.[39]

On the other hand, the structure of the healthy communities differed in the three sample types evaluated, as reported for the two dental plaques vs. saliva.[7] These discrepancies could be due to local selective pressures on community composition in health.[7] This pattern is consistent with Chen et al.,[35] who also observed that differences between oral habitats outweighed those among health states, highlighting the dominant role of site-specific factors in shaping bacterial communities.

4.3 | Differential abundance analysis

Previously undescribed in the context of periodontal health, we observed that supragingival and subgingival plaque were more similar to each other than to saliva in terms of the differential abundance of their ASVs (121 vs. 212 and 160, respectively). This discrepancy was more pronounced for large effect-size ASVs, both in number (10 vs. 107 and 95) and concerning the maximum/minimum effect (sup vs. sub = $\sim \pm 1.30$; sup and sub vs. sal = ~ 2.30 and ~ -3.00).

Additionally, most of the different ASVs in the plaques vs. saliva comparatives were more present in the latter (sup = 140 of 212; sub = 118 of 160). Although, a priori, supragingival plaque is affected more by external factors (i.e., chemical agents in toothpaste) than the subgingival,[40] our findings suggest that both niches are equally different compared to saliva.

4.4 | Random Forest modeling

Large-scale datasets allow researchers to use sophisticated analysis methods, including ML.[41] This is the first article to describe the use of RF algorithms to classify samples from different oral niches in periodontal health. As detailed below, this analytical approach is essential for identifying biomarkers.[11,12] Defining site-specific biomarkers in health enables the early detection of microbial deviations within the niche where disease may develop. Moreover, it can aid the development of pro-, pre-, or symbiotic formulations tailored to reinforce the native microbiota of distinct oral sites, contributing to preventive and therapeutic strategies.

According to experts in predictive diagnostic,[42,43] the capacity of a small number of ASVs (5) to categorize the samples per niche membership was outstanding (AUC $\geq$ 90%) and achieved excellent sensitivity (> 90%). Despite the performance being only fair (70%–79%) in distinguishing supragingival specimens from subgingival, the plaque vs. saliva models produced excellent specificity outcomes (> 90%).

The modeling findings are in line with those on differential abundance. The slightly lower performance parameters for the two-plaque model can be explained by their anatomical proximity, only separated by the gum line.[44] This makes them more similar to each other and more difficult to distinguish from saliva. The supragingival/saliva classification was particularly remarkable, with all the performance estimators achieving outcomes > 97%. Despite the close contact between these two niches, the presence of microbes from other oral tissues seems to affect the salivary microbiome in periodontal health.

4.5 | Microbial signatures of oral niches in periodontal health

Differential abundance analyses are valuable for detecting changes in community composition. They can, however, associate a particular ASV with distinct conditions[12] or niches. Examples are *A*. HMT169-AV34 or *V. dispar*-AV5 related to both plaques, and *A. massiliensis*-AV206 or *R. mucilaginosa*-AV88 associated with plaque and saliva.

Supervised ML in microbiome research is often used to predict host phenotypes, mainly pathologies.[41] They also provide us with a more refined understanding and selection capability of the potential microbial biomarkers than differential abundance analyses.[11,12] In our analysis, 9 out 14 ASVs showed agreement between their higher abundance in a niche and their contribution in the model. For the remaining five, the opposite pattern was found. This may be to non-linear effects, interactions between ASVs, the compositional nature of microbiome data, or correlations between taxa, which can lead the model to rely on patterns beyond simple abundance differences.

Although no ASV was a predictor of supragingival plaque, the ML algorithms applied enabled us to identify 14 genetic sequences as potential biomarkers of subgingival and salivary niches in periodontal health. These ASVs belonged to health-related genera in both dental plaque and saliva, such as *Capnocytophaga*, *Fusobacterium*, *Granulicatella*, or *Rothia*.[45] Still, species identification is highly desirable in 16S sequencing,[18] as different species from the same genus have been linked to distinct oral conditions.[46] In this research, ~86% of the predictor ASVs could be classified at the species level. This allowed better biological interpretability of the results than if only the genus level had been reached.

Among our subgingival plaque biomarkers, *G. morbillorum*[47] and *S. intermedius*[48] have also been associated with this niche in health. This was also the case for *K. oralis*, *O. sinus*, *R. dentocariosa*, and *S. moorei*, described in previous articles as predictors of periodontal health in saliva.[12]

Conversely, our work describes for the first time that *E. coli*, *F. nucleatum vincentii*, *P.* HMT110, *P. brachy* (subgingival), and *C. gingivalis* (saliva) act as variables that are part of predictive models in health. Of them, *E. coli* is typically related to gut diseases,[49] the following three are recognized periodontopathogens,[12] and *C. gingivalis* is a tumor promoter.[50] Therefore, the pathogenic role of the above taxa might be modulated by the surrounding microbial community in periodontal health.

These findings corroborate that microbes traditionally considered pathobionts are also present in the oral cavity in periodontal health. Interestingly, and as a novel contribution of this study, we demonstrate that in the absence of disease, these microorganisms are part of the models that enable the distinction between supragingival plaque, subgingival plaque, and saliva, suggesting that their ecological role extends beyond pathogenicity and contributes to niche-specific microbial signatures.

Importantly, the ASVs identified as predictors were consistently detected across individuals with diverse demographic characteristics. This suggests that, despite population heterogeneity, these microbial signatures remain stable, reinforcing the robustness and broader applicability of our models.

The main limitation of this study was the lack of detailed metadata in the included public bioprojects, which prevented the inclusion of relevant covariates such as age, sex, and clinical characteristics. This hindered the development of more refined statistical analyses, particularly in diagnostic predictive modeling, where such variables are key to enhancing model performance and clinical relevance. We encourage the scientific community to improve metadata completeness in public repositories to support more robust and translational research.

5 | CONCLUSIONS

In periodontal health, the supragingival plaque presents higher richness than both the subgingival plaque and saliva. In contrast, subgingival plaque displays greater diversity, and saliva is characterized by both higher diversity and evenness compared to the supragingival microbiome. The structure of the microbial communities also differs in the three sample types evaluated.

Although supragingival and subgingival bacterial profiles diverged only modestly, primarily due to taxa with small effect sizes, they were both compositionally distinct from the salivary microbiome. RF models accurately classified samples by niche, with higher performance in distinguishing saliva from plaque samples. Several ASVs emerged as potential niche-specific signatures in periodontal health by their predictive capacity. They belonged to the genera *Escherichia*, *Fusobacterium*, *Gemella*, *Granulicatella*, *Treponema*, *Peptostreptococcaceae* [XI][G-9], *Prevotella*, and *Streptococcus* in subgingival plaque and *Capnocytophaga*, *Kingella*, *Oribacterium*, *Rothia*, *Streptococcus*, and *Solobacterium* in saliva.

AUTHOR CONTRIBUTIONS

Inmaculada Tomás and Alba Regueira-Iglesias contributed to the conception and design of the research and critically reviewed the manuscript. Berta Suárez-Rodríguez, Triana Blanco-Pintos, Alba Sánchez-Barco, and Marta Relvas performed the searches and selected the studies employed. Carlos Balsa-Castro and Inmaculada Tomás conducted the bioinformatic and biostatistical analyses. Inmaculada Tomás, Alba Regueira-Iglesias, Berta Suárez-Rodríguez, Triana Blanco-Pintos, and Alba Sánchez-Barco interpreted the data and drafted the manuscript. All the authors have given their final approval to the paper and agree to be responsible for every aspect of the work, thus ensuring that any issues regarding its accuracy or completeness will be investigated and resolved appropriately.

ACKNOWLEDGMENTS

This study has been funded by the Instituto de Salud Carlos III (ISCIII) through project PI24/00222 and co-funded by the European Union (EU). The funders had no role in the study's design, data collection and analysis, publication choice, or manuscript preparation.

CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest.

ETHICS STATEMENT

Not applicable—the patients' metadata and sequence data were obtained from publicly accessible databases.

DATA AVAILABILITY STATEMENT

The principal data generated or analyzed during this study are included in this published article.

REFERENCES

1. Kilian M, Chapple IL, Hannig M, et al. The oral microbiome—an update for oral healthcare professionals. *Br Dent J*. 2016;221(10):657-666. doi:10.1038/sj.bdj.2016.865
2. Zhang Y, Wang X, Li H, Ni C, Du Z, Yan F. Human oral microbiota and its modulation for oral health. *Biomed Pharmacother*. 2018;99:883-893. doi:10.1016/j.biopha.2018.01.146
3. Lundtorp Olsen C, Markvart M, Vendius VFD, Damgaard C, Belstrøm D. Short-term sugar stress induces compositional changes and loss of diversity of the supragingival microbiota. *J Oral Microbiol*. 2023;15(1):2189770. doi:10.1080/20002297.2023.2189770
4. Bamashmous S, Kotsakis GA, Jain S, Chang AM, McLean JS, Darveau RP. Clinically healthy human gingival tissues

show significant inter-individual variability in GCF chemokine expression and subgingival plaque microbial composition. *Front Oral Health*. 2021;2:689475. doi:10.3389/froh.2021.689475

5. Zhu C, Yuan C, Wei FQ, Sun XY, Zheng SG. Intraindividual variation and personal specificity of salivary microbiota. *J Dent Res*. 2020;99(9):1062-1071. doi:10.1177/0022034520917155
6. Mason MR, Chambers S, Dabdoub SM, Thikkurissy S, Kumar PS. Characterizing oral microbial communities across dentition states and colonization niches. *Microbiome*. 2018;6(1):67. doi:10.1186/s40168-018-0443-2
7. Segata N, Haake SK, Mannon P, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*. 2012;13(6):R42. doi:10.1186/gb-2012-13-6-r42
8. Meuric V, Le Gall-David S, Boyer E, et al. Signature of microbial dysbiosis in periodontitis. *Appl Environ Microbiol*. 2017;83(14):e00462-17. doi:10.1128/AEM.00462-17
9. Regueira-Iglesias A, Balsa-Castro C, Blanco-Pintos T, Tomás I. Critical review of 16S rRNA gene sequencing workflow in microbiome studies: from primer selection to advanced data analysis. *Mol Oral Microbiol*. 2023;38(5):347-399. doi:10.1111/omi.12434
10. Nearing JT, Douglas GM, Hayes MG, et al. Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun*. 2022;13(1):777. doi:10.1038/s41467-022-28034-z
11. Hernández Medina R, Kutuzova S, Nielsen KN, et al. Machine learning and deep learning applications in microbiome research. *ISME Commun*. 2022;2(1):98. doi:10.1038/s43705-022-00182-9
12. Regueira-Iglesias A, Suárez-Rodríguez B, Blanco-Pintos T, et al. The salivary microbiome as a diagnostic biomarker of periodontitis: a 16S multi-batch study before and after the removal of batch effects. *Front Cell Infect Microbiol*. 2024;14:1405699. doi:10.3389/fcimb.2024.1405699
13. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-W73. doi:10.7326/M14-0698
14. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd ed. Springer; 2019:558.
15. Leinonen R, Sugawara H, Shumway M. International nucleotide sequence DC. The sequence read archive. *Nucleic Acids Res*. 2011;39:D19-D21. doi:10.1093/nar/gkq1019
16. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460-2461. doi:10.1093/bioinformatics/btq461
17. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537-7541. doi:10.1128/AEM.01541-09
18. Escapa IF, Huang Y, Chen T, et al. Construction of habitat-specific training sets to achieve species-level assignment in 16S rRNA gene datasets. *Microbiome*. 2020;8(1):65. doi:10.1186/s40168-020-00841-w
19. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New insights into human nostril microbiome from the expanded Human Oral Microbiome Database (eHOMD): a resource for the microbiome of the human aerodigestive tract. *mSystems*. 2018;3(6):e00187-18. doi:10.1128/mSystems.00187-18
20. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80. doi:10.1186/gb-2004-5-10-r80
21. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217. doi:10.1371/journal.pone.0061217
22. Lahti L, Shetty S. Tools for Microbiome Analysis in R. Microbiome Package. Version 1.26.0. R Foundation for Statistical Computing; 2019. http://microbiome.github.com/microbiome
23. Oksanen J, Simpson G, Blanchet F, et al. Vegan: Community Ecology Package. R Package. Version 2.6-8. R Foundation for Statistical Computing; 2024. https://cran.r-project.org/web/packages/vegan/index.html
24. Wang Y, Lê Cao KA. PLSDA-batch: a multivariate framework to correct for batch effects in microbiome data. *Brief Bioinform*. 2023;24(2):bbac622. doi:10.1093/bib/bbac622
25. MuToss Coding Team, Blanchard G, Dickhaus T, et al. Unified Multiple Testing Procedures. R Package. Version 0.1-13. R Foundation for Statistical Computing; 2023. https://cran.r-project.org/web/packages/mutoss/index.html
26. Torchiano M. effsize: Efficient Effect Size Computation. R Package. Version 0.8.1. R Foundation for Statistical Computing; 2020. https://cran.r-project.org/web/packages/effsize/index.html
27. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13(11):e1005752. doi:10.1371/journal.pcbi.1005752
28. Scrucca L. GA: a package for genetic algorithms in R. *J Stat Soft*. 2013;53(4):1-37. 10.18637/jss.v053.i04
29. Kuhn M. Building predictive models in R using the caret package. *J Stat Soft*. 2008;28(5):1-26. 10.18637/jss.v028.i05
30. Papapanou PN, Sanz M, Buduneli N, et al. Periodontitis: consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J Periodontol*. 2018;89(Suppl 1):S173-S182. doi:10.1002/JPER.17-0721
31. Siddiqui R, Badran Z, Boghossian A, Alharbi AM, Alfahemi H, Khan NA. The increasing importance of the oral microbiome in periodontal health and disease. *Future Sci OA*. 2023;9(8):FSO856. doi:10.2144/fsoa-2023-0062
32. Zaura E, Pappalardo VY, Buijs MJ, Volgenant CMC, Brandt BW. Optimizing the quality of clinical studies on oral microbiome: a practical guide for planning, performing, and reporting. *Periodontol 2000*. 2021;85(1):210-236. doi:10.1111/prd.12359
33. Takeshita T, Kageyama S, Furuta M, et al. Bacterial diversity in saliva and oral health-related conditions: the Hisayama Study. *Sci Rep*. 2016;6:22164. doi:10.1038/srep22164
34. Gupta VK, Paul S, Geography DC. Ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front Microbiol*. 2017;8:1162. doi:10.3389/fmicb.2017.01162
35. Chen H, Liu Y, Zhang M, et al. A filifactor alocis-centered co-occurrence group associates with periodontitis across different oral habitats. *Sci Rep*. 2015;5:9053. doi:10.1038/srep09053

36. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol*. 2017;8:2224. doi:10.3389/fmicb.2017.02224
37. Jakubovics NS. Saliva as the sole nutritional source in the development of multispecies communities in dental plaque. *Microbiol Spectr*. 2015;3(3). doi:10.1128/microbiolspec.MBP-0013-2014
38. Willis JR, Gabaldón T. The human oral microbiome in health and disease: from sequences to ecosystems. *Microorganisms*. 2020;8(2):308. doi:10.3390/microorganisms8020308
39. Proctor GB. The physiology of salivary secretion. *Periodontol 2000*. 2016;70(1):11-25. doi:10.1111/prd.12116
40. Jin Y, Yip HK. Supragingival calculus: formation and control. *Crit Rev Oral Biol Med*. 2002;13(5):426-441. doi:10.1177/154411130201300506
41. Namkung J. Machine learning methods for microbiome studies. *J Microbiol*. 2020;58(3):206-216. doi:10.1007/s12275-020-0066-8
42. Hosmer DJ, Lemeshow S, Sturdivant R. *Applied Logistic Regression*. 3rd ed. John Wiley & Sons, Inc.; 2013:528.
43. De Luca Canto G, Pachêco-Pereira C, Aydinoz S, Major PW, Flores-Mir C, Gozal D. Diagnostic capability of biological markers in assessment of obstructive sleep apnea: a systematic review and meta-analysis. *J Clin Sleep Med*. 2015;11(1):27-36. doi:10.5664/jcsm.4358
44. Mark Welch JL, Ramírez-Puebla ST, Borisy GG. Oral microbiome geography: micron-scale habitat and niche. *Cell Host Microbe*. 2020;28(2):160-168. doi:10.1016/j.chom.2020.07.009
45. Zaura E, Keijser BJF, Huse SM, Crielaard W. Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol*. 2009;9:259-259. doi:10.1186/1471-2180-9-259
46. Relvas M, Regueira-Iglesias A, Balsa-Castro C, et al. Relationship between dental and periodontal health status and the salivary microbiome: bacterial diversity, co-occurrence networks and predictive models. *Sci Rep*. 2021;11(1):929. doi:10.1038/s41598-020-79875-x
47. Torres-Morales J, Mark Welch JL, Dewhirst FE, Borisy GG. Site-specialization of human oral *Gemella* species. *J Oral Microbiol*. 2023;15(1):2225261. doi:10.1080/20002297.2023.2225261
48. Del Pilar Angarita-Díaz M, Fong C, Medina D. Bacteria of healthy periodontal tissues as candidates of probiotics: a systematic review. *Eur J Med Res*. 2024;29(1):328. doi:10.1186/s40001-024-01908-2
49. Kitamoto S, Nagao-Kitamoto H, Hein R, Schmidt TM, Kamada N. The bacterial connection between the oral cavity and the gut diseases. *J Dent Res*. 2020;99(9):1021-1029. doi:10.1177/0022034520924633
50. Zhu W, Shen W, Wang J, et al. Capnocytophaga gingivalis is a potential tumor promotor in oral cancer. *Oral Dis*. 2024;30(2):353-362. doi:10.1111/odi.14376

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Regueira-Iglesias A, Suárez-Rodríguez B, Blanco-Pintos T, et al. Diversity and random forest models of oral microbiomes in periodontal health using publicly available data. *J Periodontol*. 2025;1-14. https://doi.org/10.1002/jper.70000