UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Unique ecology of co-occurring functionally and phylogenetically undescribed species in the infant oral microbiome

Pucci, N.; Kaan, A.M.; Ujčič-Voortman, Joanne; Verhoeff, A.P.; Zaura, E.; Mende, Daniel R.

DOI
10.1101/2025.06.03.657575

Publication date
2025

Document Version
Final published version

License
CC BY

Link to publication

Citation for published version (APA):
Pucci, N., Kaan, A. M., Ujčič-Voortman, J., Verhoeff, A. P., Zaura, E., & Mende, D. R. (2025). *Unique ecology of co-occurring functionally and phylogenetically undescribed species in the infant oral microbiome*. (v1 ed.) BioRxiv. https://doi.org/10.1101/2025.06.03.657575

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 11月 2026

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Unique ecology of co-occurring functionally and phylogenetically undescribed species in the infant oral microbiome

Running title: Infant oral microbiome species ecology and functional adaptations

Nicholas Pucci[1], Amke Marije Kaan[2], Joanne Ujčič-Voortman[3], Arnoud P. Verhoeff[3,4], Egija Zaura[2] and Daniel R. Mende[1,5]

[1]*Department of Medical Microbiology & Infectious Diseases, Amsterdam UMC, Location AMC at University of Amsterdam, Amsterdam, The Netherlands*

[2]*Department of Preventive Dentistry, Academic Center for Dentistry Amsterdam, University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.*

[3]*Sarphati Amsterdam, Public Health Service Amsterdam, Amsterdam, The Netherlands*

[4]*Department of Sociology, University of Amsterdam, Amsterdam, The Netherlands*

[5]*Human Biology-Microbiome-Quantum Research Center (WPI-Bio2Q), Keio University, Tokyo, Japan*

Corresponding Author:

Daniel R. Mende

Keio University, Tokyo 108-8345, Japan

Email address: d.r.mende@amsterdamumc.nl

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Abstract

Early-life oral microbiome development is a complex community assembly process that influences long-term health outcomes. Nevertheless, microbial functions and interactions driving these ecological processes remain poorly understood. In this study, we analyze oral microbiomes from a longitudinal cohort of 24 mother-infant dyads at 1 and 6 months postpartum using shotgun metagenomics. We identify two previously undescribed *Streptococcus* and *Rothia* species to be among the most prevalent, abundant and strongly co-occurring members of the oral microbiome of six-month-old infants. To explore the underlying genomic traits enabling this unique lifestyle, we leveraged metagenome-assembled genomes (MAGs) and genome-scale metabolic models (GEMS). Comparative analyses revealed specific genomic and functional characteristics relative to other closely related species and highlighted unique functional features, including genes encoding adhesins and carbohydrate-active enzymes (CAZymes). The observed co-occurrence patterns were further supported by predicted metabolic interactions within a network of co-occurring oral taxa. Metabolic modeling identified potential exchange of key nutrients, particularly malate and lysine, between these species, suggesting metabolic cross-feeding interactions that may explain their co-occurrence across infant oral microbiomes. Overall, this study provides key insights into the functional adaptations and microbial interactions shaping early colonization in the oral cavity.

Author Summary

We investigated how infant oral bacterial communities develop during their first six months of life, with the aim to understand which microbes colonize, how they establish themselves and why they succeed together. Using high throughput DNA sequencing techniques, we analyzed oral samples from 24 mother-infant pairs at one and six months after birth.

We found two abundant, but previously unknown bacterial species (one Streptococcus and one Rothia) at six months of age. These bacteria consistently appear together across different babies, suggesting they may depend on each other for survival and growth.

By reconstructing the genomes of these bacteria directly from our samples, we discovered specific genetic features that help explain their success in the infant mouth. Streptococcus carries genes that

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

allow it to break down nutrients from breast milk. Rothia has genes that help it rapidly build cell membranes and protect against harmful molecules, while producing nutrients that Streptococcus needs. We predict these bacteria exchange key nutrients like malate and lysine, creating a mutually beneficial partnership. The bacteria seem to cooperate and are predicted to exchange key nutrients like malate and lysine which may help maintain a healthy oral environment by regulating acidity levels, potentially protecting against tooth decay.

Introduction

The first bacteria (‘animalcules’) Antonie van Leeuwenhoek observed in the 17th century were oral microbes. In centuries afterwards, oral microbes were found to be involved in a range of conditions, such as dental caries, periodontal diseases and oral candidosis (1). An understanding of the beneficial role that the microbes constituting the oral microbiome play was realized much later. A balanced oral microbiome promotes oral health by serving as a protective barrier, preventing infections and colonization of oral surfaces by harmful pathogens as well as regulating oral pH (2,3). Beyond oral health, the oral microbiome has significant implications for systemic health (4), contributing to vasodilation and blood pressure regulation through nitrogen and enterosalivary nitrate metabolism (5). Though its importance is undeniable, our knowledge about the oral microbiome, particularly its development during infancy, remains limited, and public metagenomics datasets remain scarce.

Most of the existing knowledge about the infant oral microbiome stems from 16S rRNA gene amplicon surveys, which revealed the genus composition of the oral microbiome development through a predictable succession of microbial colonization, where early colonizers such as *Streptococcus* and *Actinomyces* adhere to oral tissues thanks to their adhesin repertoire (6–11). These pioneers subsequently form a foundation that supports the establishment of secondary colonizers, including *Veillonella* and *Rothia* (12). Developmental milestones, such as primary dentition, and environmental exposures are often accompanied by further compositional shifts (8). While these coarse grained taxonomic succession patterns have been described, the functional potential and interactions within these microbial communities, particularly during the formative months of life, remain poorly characterized. Yet,

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

such knowledge about the developing oral microbiome would be crucial given the potential persisting effects on oral and systemic health later in life (8,13,14).

Microbial interactions fundamentally shape microbiome structure and function (12,15), yet little is known about their contribution to the stability and resilience of the developing oral microbiome. Typically, these interactions are a combination of spatial organization (16,17), complex metabolic dependencies and cross-feeding mechanisms (18) leading to mutualistic, competitive or commensal relationships (3,19–21), while driving community-level processes such as pH regulation (22) and nitrate reduction (23). Understanding the genomic basis of these synergistic activities, including metabolic pathways and genetic determinants, is crucial for deciphering how microbial communities establish and maintain functional stability during infant oral microbiome development.

Methodological advances allowing for the reconstruction of genomes directly from metagenomics, and their interrogation through comparative approaches (metapangenomics) (24) have the potential to provide insight into the functional potential and metabolic interactions of the developing microbiome. For example, genome-resolved metagenomic approaches allowed Utter et al. (24) to have strain-level resolution and identification of accessory genes conferring adaptive advantages to *Rothia* and *Haemophilus* spp. to specific oral niches in healthy adults. Further, novel tools allow for the generation of accurate metabolic models directly from genomes (25,26), which allows the prediction of substrate utilization and metabolite production. Community models combining the metabolic models of multiple co-occurring microbes can provide further insight into metabolic interactions and cross-feeding, therefore elucidating the metabolic basis for the synergistic activities and successional patterns observed during oral microbiome development (27–29).

Here, we performed shotgun metagenomics on a time series of 24 infants and their mother's oral microbiome and identified novel, co-occurring species using a meta-pangenomics approach. Analysis of a time-series of tongue dorsum and dental biofilm samples collected from children up to six months of age and their mothers as part of the prospective Amsterdam Infant Microbiome Study (AIMS) revealed previously undescribed *Streptococcus* and *Rothia* species to be predominant members of the infant oral cavity. By leveraging metagenome-assembled genomes (MAGs), we characterize genomic features and functional attributes of these species to explain their adaptations to the infant oral cavity. In order to

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

interpret their co-occurrence patterns we further decipher their metabolic interactions within an oral species network. Overall, our findings highlight important microbial ecological dynamics in the developing oral ecosystem of infants and establish a foundation for understanding how early microbial interactions help shape the infant oral microbiome.

Results

Undescribed *Streptococcus* and *Rothia* species emerge as predominant in six-month infant oral microbiome

We investigated the oral microbiome of 24 mother-infant pairs from the AIMS cohort, focusing on full-term, vaginally-born infants with no reported antibiotic exposure (Table S1). By six months, 87% of AIMS infants (21 out of 24) had obtained their first teeth, allowing collection of dental biofilm. Shotgun metagenomics sequencing (mean depth tongue biofilm: 20.5M reads per sample; mean depth dental biofilm: 15.7M reads per sample; Figure S1) was performed on samples collected from tongue dorsum swabs from mothers (successfully sequenced samples: n=22) and infants both 1 (n=9) and 6 months (n=18), and dental biofilm from mothers (n=21) and infants at 6 months (n=14). All analyzed one-month-old infants were exclusively breastfed. By six months, we observed diverse feeding patterns including breast- (n=8), mixed-(n=7) and formula feeding (n=5), with 95% (21 of 22) of infants being introduced to solid foods (Table S1).

Metagenomic profiling revealed the *Streptococcus* genus to be the most abundant and prevalent taxonomic group in the oral microbiome of AIMS infants at both 1 and 6 months of age. However, we observed a notable species-level compositional shift over time, with previously undescribed species emerging by six months. At one month, infant tongue dorsum hosted an average of 19 (±6) bacterial species (Fig. 1B) and was predominantly composed of *Streptococcus* (relative abundance±standard deviation, SD 57.0±6.8%), *Veillonella* (9.0±2.0%), *Lactobacillus* (6.1±3.3%) and *Rothia* (4.9±1.8%) species (Fig. 1A). By six months, we observed an increase in bacterial species richness (tongue dorsum 43±17, dental biofilm 43±14) with an overall shift in tongue microbiome composition relative to one month (Fig. 1A-C). Notably, previously undescribed *Rothia* (GTDB: sp902373285: relative abundance=9.2±2.2%, prevalence=61%) and *Streptococcus* (GTDB: sp000187445: 5.6±1.9%, 66%)

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

species, hereon referred to as *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2, emerged as abundant and prevalent bacteria in the oral cavity, particularly in the tongue biofilm (Fig. 1B). Dental biofilm composition at six months (despite the emergence of teeth as a new niche) largely resembled the tongue biofilm composition (Fig. 1C). Milk feeding type did not significantly influence community composition (Fig. S3). Maternal dental and tongue biofilm communities were markedly different (Fig. 1C) and strain sharing was found to be rare with only two events found across the whole cohort (Table S2), both between mothers and their respective one-month-old infants.

Figure 1. Oral microbiome abundance and composition of AIMS mothers and their infants. A) Relative abundance of the 5 most abundant bacterial genera detected in the tongue dorsum and dental plaque of mothers (34wks gestation) and their infants (one and six months of age). **B)** Relative abundance (mean±SD) of the 15 most abundant bacterial species detected in the tongue dorsum and dental plaque of mothers (34wks gestation) and their infants (one and six months of age). Species displayed include all taxa that ranked among the 15 most abundant in at least one sample group or time point. Bars are coloured by genus. **C)** Principal coordinate analysis (PCoA) of tongue dorsum (circles) and dental plaque (triangles) microbiomes from AIMS mothers (34wks gestation=brown) and their infants (one month=green; six months=red).

Streptococcus AIMSoral1 and *Rothia* AIMSoral2 show significant co-occurrence in infant oral ecosystems

To elucidate potential species interactions, we inferred pairwise co-occurrence patterns across ages by combining hierarchical clustering and Spearman's rank correlations for tongue and dental biofilm samples, separately. As the most abundant species in the infant oral microbiomes were almost all absent from the maternal oral microbiomes (Fig. 1B), we present co-occurrence network results for infant samples only. In one-month-old edentate infants, we detected 31 species pairs with significant correlations involving 28 (of 32) abundant (relative abundance≥1%) tongue biofilm species (Fig. S4A, Table S3). Among these correlations, we found few involving predominant tongue species (relative

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

abundance≥5%, prevalence>30%). *L. paragasseri* showed a significant pairwise correlation with *R. mucilaginosa_A,* only (Spearman's rank correlation: R=0.87, *p*=0.002). Similarly, *S. oralis*_BN showed a significant, but negative correlation with *V. nakazawae* (Spearman's rank correlation: R=-0.88, *p*=0.002), while *S. salivarius* showed no significant correlations.

By six months of age, we detected 24 and 21 significant pairwise abundance correlations involving 14 (out of 17) and 17 (out of 24) species across tongue and dental biofilms, respectively. *Rothia* AIMSoral2 and *Streptococcus* AIMSoral1 displayed the strongest positive co-occurrence among abundant bacteria in the infant tongue biofilm (Spearman's rank correlation: R=0.68, p=0.00008; Fig. S4B, Table S3). *Rothia* AIMSoral2 and *Streptococcus* AIMSoral1 significantly co-occurred with nine and four other species, respectively (Fig. S4B). Notably, both species exhibited significant abundance correlations with undescribed *Veillonella* (GTDB: sp018367495) and *Pauljensenia* (GTDB: sp900541895) species on the tongue dorsum, hereon referred to as *Veillonella* AIMSoral3 and *Pauljensenia* AIMSoral4. These species also correlated significantly with each other, forming a co-occurrence network of four species. Neither of these species was detected in maternal or one month oral samples.

In dental biofilm, the significant co-occurrence between *Rothia* AIMSoral2 and *Streptococcus* AIMSoral1 persisted, but was weaker (R=0.40, p=0.04). Both species also showed strong correlations with *Veillonella* AIMSoral3 (*Rothia* AIMSoral2: R=0.54 , p=0.004; *Streptococcus* AIMSoral1: R=0.54, p=0.01) (Fig. S4C, Table S3). Overall, abundance of these species was not significantly influenced by feeding type (Fig. S4).

Genomic relatedness analyses reveal undescribed *Streptococcus* and *Rothia* to be novel species

Among identified species in the developing oral microbiome, the undescribed *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2 were selected for functional characterization due to their high abundance (≥5%) and prevalence (>60%) across infant oral cavities. To confirm metagenomic taxonomic profiling of these species, we reconstructed metagenome-assembled genomes (MAGs).

Metagenomic binning from maternal and infant oral samples yielded 20 *Streptococcus* (completeness, mean±SD=64.15±12.17%; contamination, mean±SD=2.11±2.84%) and 51 *Rothia* (91.24±13.32%; 0.91±1.27%) medium and high-quality MAGs (>50% completeness and <10%

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

194 contamination), of which five were taxonomically assigned to *Streptococcus* AIMSoral1 and 15 to *Rothia*
195 AIMSoral2 using GTDB-tk (Table S5). To verify these taxonomic assignments, we reconstructed
196 phylogenomic trees based on single-copy core gene protein sequences (Fig. 2), incorporating reference
197 genomes from public databases (Table S5).
198
199 **Figure 2. Phylogenomic trees of *Streptococcus salivarius* clade and *Rothia* genus, highlighting**
200 **the undescribed species *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2. A)** Phylogenomic tree of
201 the *salivarius* clade of the *Streptococcus* genus containing 33 genomes (reference genomes=27,
202 MAGs=6). **B)** Phylogenomic tree of the *Rothia* genus containing 63 genomes (reference genomes=13,
203 MAGs=50). Genome names are colored by species identity as defined by genetic relatedness analyses.
204 Rings indicate the time point (mother 34wks gestation, infant one and six months) and sampling location
205 (tongue dorsum and dental plaque). Red circles indicate bootstrap values≥0.9.
206
207 Phylogenomic analysis of 89 *Streptococcus* species (110 genomes) placed *Streptococcus*
208 AIMSoral1 within the *salivarius* group with high confidence (Fig. S5). To refine this placement, we
209 employed a pangenomic approach focusing specifically on the *salivarius* group, constructing a
210 phylogenomic tree from single-copy core genes (Fig. 2A). This targeted analysis included 27 reference
211 genomes and six MAGs: *S. salivarius* (n=6), *S. vestibularis* (n=5), *S. thermophilus* (n=5) and
212 *Streptococcus* AIMSoral1 (n=15). The phylogenetic analysis revealed *Streptococcus* AIMSoral1 to form a
213 distinct, well-supported monophyletic clade, representing a sister clade to the *salivarius* group. This
214 observation was supported by average nucleotide identity (ANI, species threshold=95%) (32) analyses,
215 confirming that *Streptococcus* AIMSoral1 represents a species-level entity, with 97.3% to 98.6% ANI
216 within its cluster, but lower ANI to phylogenetically neighboring species (*S. salivarius*: 88.6-89.3%, *S.*
217 *vestibularis*: 88.3-88.5%, *S. thermophilus*: 87.1-87.4%; Fig. S6A).

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

The *Rothia* phylogenomic tree included 63 genomes (reference genomes=13, MAGs=50) and placed the undescribed *Rothia* AIMSoral2 in a monophyletic clade, closely related to *R. mucilaginosa*_A and *R. mucilaginosa* (33) (Fig. 2B).

For *Rothia*, ANI analyses also supported *Rothia* AIMSoral2 as a species-level entity, showing 97.0% to 99.9% ANI within its cluster, but markedly lower ANI with phylogenetically neighboring *R. mucilaginosa*_A (90.2-91.4%) and *R. mucilaginosa* (86.9-88.5% ANI) (Fig. S6B).

Adhesion and carbohydrate utilization differentiate *Streptococcus* AIMSoral1 from other infant-associated species

To uncover functional traits driving the abundance and persistence of *Streptococcus* AIMSoral1 in the infant oral cavity, we compared its pangenome, metabolic capacity, and adhesion potential to other infant-associated *Streptococcus* species (see Methods: Public reference genomes acquisition).

The *Streptococcus* pangenome included 57 genomes (Table S5) from six infant-associated species: *Streptococcus* AIMSoral1 (n=14), *S. lactarius* (n=12), *S. salivarius* (n=6), *S. oralis* (n=9), *S. mitis* (n=12) and *S. peroris* (n=3). *Streptococcus* AIMSoral1 genomes averaged 1.9±0.9Mb (average±SD), carrying roughly 1930±98 genes. Gene-cluster (GC) analysis showed largely species-driven clustering, with *Streptococcus* AIMSoral1 displaying distinct, but close genomic similarity to *S. salivarius* (Fig. S7). The pangenome was further analyzed to reveal metabolic capabilities of *Streptococcus* AIMSoral1 supporting its predominance during infant weaning.

Adhesion potential, assessed via lectin-encoding gene copy number, largely mirrored species identity (Fig. 3). The adhesin repertoire of *Streptococcus* AIMSoral1 showed high similarity to *S. salivarius* and *S. lactarius*, including enrichment in *fimA* lipoprotein genes (present in *Streptococcus* AIMSoral1 and *S. lactarius*, but absent in other species; Fig. 3, Table S10).

Figure 3. Overview comparative genomics analyses outputs for *Streptococcus* AIMSoral1 and infant-associated *Streptococcus* species. Phylogenomics tree based on single-copy core genes of 57 infant-associated *Streptococcus* genomes, color-coded by species annotation: *Streptococcus* AIMSoral1, *S. salivarius*, *S. lactarius*, *S. peroris*, *S. oralis* and *S. mitis*. Genome completeness and contamination

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

metrics are displayed for each genome, together with sampling location and time point of origin. Heatmaps below the tree show for each genome comparative genomics analyses results, highlighting metabolic modules/genes significantly enriched in *Streptococcus* AIMSoral1 relative to other infant-associated *Streptococcus* species. From top to bottom, panels display: metabolic module abundance by broad metabolic pathway (Metabolism=gray palette), metabolic modules (light gray), CAZyme- (orange palette) and lectin/adhesin- (red palette) encoding genes enriched in *Streptococcus* AIMSoral1 relative to other infant-associated *Streptococcus* spp.

Carbohydrate-active enzymes (CAZymes) analysis revealed *Streptococcus* AIMSoral1 harbors significantly higher copy numbers of eight glycosyltransferases (GTs), including α-N-acetylglucosaminyltransferase (GT45) and starch phosphorylases (GT35), as well as four glycoside hydrolases (GHs), such as GH66, GH68, and GH87 families, which were shared between *Streptococcus* AIMSoral1 and *S. salivarius* (Fig. 3, Table S10). Additionally, *Streptococcus* AIMSoral1 had significantly more GH8, GH36, and GT2_Glyco_trans_2_3 genes compared to *S. salivarius*, suggesting distinct carbohydrate-utilization strategies. Overall, *Streptococcus* AIMSoral1 and *S. salivarius* have fewer genes involved in carbohydrate metabolism than other infant-associated species (Fig. 3).

Metabolic enrichment analysis showed *Streptococcus* AIMSoral1 and *S. salivarius* to be significantly enriched in genes for amino acid metabolism compared to other infant-associated species (Fig. 3). Notably, only *Streptococcus* AIMSoral1 and *S. salivarius* encoded the full arginine biosynthesis pathway from glutamate (M00028, M00844; Table S6), a downstream metabolite of the citrate cycle (also enriched; M00010; Table S6) and major component of breast/formula milk. No metabolic module was exclusively enriched in *Streptococcus* AIMSoral1.

Carbohydrate utilization and fatty acid biosynthesis differentiate *Rothia* AIMSoral2 from other *R. mucilaginosa* species

We compared genomic traits of *Rothia* AIMSoral2 to infant-associated *Rothia* species to uncover metabolic adaptations supporting its predominance across 6-month-old infant oral cavities.

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

The *Rothia* pangenome included 53 MAGs and reference genomes, from three species (as annotated by GTDB-tk), retrieved from infants in our study: *R. mucilaginosa* (30 MAGs, 2 reference genomes), *R. mucilaginosa_A* (5 MAGs, 1 reference genome), and *Rothia* AIMSoral2 (15 MAGs, no reference genomes available). *Rothia* AIMSoral2 genomes averaged 2.1±0.1Mb and carried 1760 (±37) genes. In line with phylogenomic analysis, these genomes form a consistent clade based on gene presence/absence, showing greater similarity to *R. mucilaginosa*_A than other *R. mucilaginosa* strains (Fig. S8).

All infant-associated *Rothia* species carried genes-encoding for CBM50 domains (also known as LysM; Table S10), associated with binding bacterial peptidoglycans (34). No matches were found in the LectomExplore database for *Rothia* lectins, suggesting limited host-surface adhesion potential. Further CAZyme analysis showed *Rothia* AIMSoral2 to be enriched in genes encoding for, among others, GT2, GT4, and GT11, involved in EPS production, biofilm formation and surface adhesion (35–39) (Table S10). Additionally, *Rothia* AIMSoral2 carried significantly higher copy-numbers of redox-active enzyme-encoding genes (Auxillary Activities: AA1 and AA1_1/2; Fig. 4 and Table S10), co-acting with CAZymes.

Figure 4. Overview comparative genomics analyses outputs for *Rothia* AIMSoral2 and infant-associated *Rothia* species. Phylogenomics tree based on single-copy core genes of 53 infant-associated *Rothia* genomes, color-coded by species annotation: *Rothia* AIMSoral2, *R. mucilaginosa* and *R. mucilaginosa_A*. Genome completeness and contamination metrics are displayed for each genome, together with sampling location and time point of origin. Heatmaps below the tree show for each genome comparative genomics analyses results, highlighting metabolic modules/genes significantly enriched in *Rothia* AIMSoral2 relative to other infant-associated *Rothia* species. From top to bottom, panels display: metabolic module abundance by broad metabolic pathway (gray palette), metabolic modules (light gray), CAZyme-encoding (orange palette) and enterosalivary pathway-associated (red palette) genes enriched in *Rothia* AIMSoral2 relative to other infant-associated *Rothia* spp.

We further investigated genes involved in the nitrate-nitrite-nitric oxide (NO_3-NO_2-NO) enterosalivary pathway in infant-associated *Rothia* species. All infant-associated *Rothia* species carried

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

genes involved in the respiratory denitrification of NO_3 to NO and, in part, intracellular ammonia (NH_3) production. Further functional analyses revealed *Rothia* AIMSoral2 to carry significantly higher numbers of genes associated with cofactors and vitamin biosynthesis compared to *R. mucilaginosa,* but not *R. mucilaginosa_A* (Fig. 4). Notably, the key process of fatty acid biosynthesis initiation (M00082) was significantly enriched in *Rothia* AIMSoral2 (Fig. 4; Table S7).

Streptococcus AIMSoral1 shows higher metabolic complementarity to _Rothia_ AIMSoral2 than other _Streptococcus_ species

To investigate whether the co-occurrence of *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2 at six months of age corresponded with their metabolic interdependence, we reconstructed GEMS for infant-associated *Streptococcus* and *Rothia* species using the PhyloMInt tool (Methods: Metabolic complementarity/interaction potential).

Overall, *Streptococcus* species exhibited comparatively greater metabolic complementarity ($MI_{complementarity}$) with *Rothia* than viceversa (Fig. S9A, Table S13), though we found significant differences among individual species. *Rothia* AIMSoral2 exhibited, comparatively, the highest $MI_{complementarity}$ with *S. lactarius* and *Streptococcus* AIMSoral1. Despite *Streptococcus* AIMSoral1 and *S. salivarius* having similar metabolic profiles, *S. salivarius* showed the lowest $MI_{complementarity}$ with all *Rothia* species. While *Rothia* AIMSoral2 and *R. mucilaginosa_A* had comparable $MI_{complementarity}$ patterns, *R. mucilaginosa_A* showed the highest $MI_{complementarity}$ with all *Streptococcus* species. Notably, *Streptococcus* AIMSoral1 (and *S. peroris*) exhibited significantly greater $MI_{complementarity}$ with all *Rothia* species than other *Streptococcus* species.

Using the PhyloMInt PTM tool, we identified metabolites each species could produce and utilize, allowing us to assess potential differences in metabolic interactions between *Rothia* AIMSoral2 and *Streptococcus* AIMSoral1 compared to their interactions with other infant-associated species. We predicted multiple potential metabolic interactions between *Rothia* AIMSoral2 and *Streptococcus* AIMSoral1, but these were often shared with other species (Fig. S9B, Table S14).

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Genome-scale predictions of metabolic interactions in an oral bacterial network

We hypothesized that metabolic interactions facilitate the co-occurrence network comprising S*treptococcus* AIMSoral1, *Rothia* AIMSoral2, *Veillonella* AIMSoral3 and *Pauljensenia* AIMSoral4 (Fig. 5A). To test this hypothesis we predicted metabolic interactions among these species. Although we lacked sufficient *Veillonella* AIMSoral3 and *Pauljensenia* AIMSoral4 MAGs for comprehensive comparative genomics, metabolic prediction was still possible using available reference genomes. Using GEMS, we inferred $MI_{complementarity}$ based on the highest quality available MAG or reference genome (*see* Methods: Prediction of metabolic complementarity and interaction potentials; Table S13). Comparatively, all species exhibited the lowest $MI_{complementarity}$ to *Streptococcus* AIMSoral1, while the latter exhibited the highest $MI_{complementarity}$ with *Pauljensenia* AIMSoral4 and *Rothia* AIMSoral2 (Fig. 5B). Both *Rothia* AIMSoral2 and *Pauljensenia* AIMSoral4 were comparatively highly complementary to *Veillonella* AIMSoral3. Notably, *Veillonella* AIMSoral3 showed the strongest $MI_{complementarity}$ with *Rothia* AIMSoral2, suggesting a key metabolic relationship between these species.

Figure 5. Metabolic complementarity and interactions within an infant tongue microbial co-occurrence network. A) Microbial network of four species (*Streptococcus* AIMSoral1=blue, *Rothia* AIMSoral2=pink, *Veillonella* AIMSoral3=green, *Pauljensenia* sp9002372285=yellow), sharing significant abundance correlations among each other on the tongue surface of six-month-old infants, and their relative abundances in this niche at this age. Significant correlations were calculated using Spearman's rank correlation. Edges width indicate the strength of the significance. *$p \leq 0.05$, $p \leq 0.01$,***$p \leq 0.001$, ****$p \leq 0.0001$. **B)** Bargraphs show the metabolic complementarity index for pairwise interactions among species in the tongue microbial network. Each facet title indicates the reference species for which the index is calculated. **C)** Heatmap showing bacterial-metabolite bipartite interactions (production=red, utilization=blue, both=purple, no interaction=white) potentially occurring among species in the tongue network. Metabolites are classified by their superclasses.

GEMS predicted distinct, species-level patterns of metabolite production and utilization, which are summarized in Fig. 5C and Table S14. Within the co-occurrence network, *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2 were predicted to interact through 27 metabolites (Fig. 5C). Nine were potentially

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

356 produced by *Streptococcus* AIMSoral1 and required by *Rothia* AIMSoral2, including malate. Malate
357 exchange was supported by malate efflux carriers in *Streptococcus* AIMSoral1 (BLAST protein
358 identity=95.2±1.6%; Table S11) and malate symporters in *Rothia* AIMSoral2 (51.6±0.2%; Table S12). The
359 remaining 18 metabolites, including ethanol and lysine, were potentially produced by *Rothia* AIMSoral2
360 and required by *Streptococcus* AIMSoral1. Lysine exchange was supported by lysine exporters (LysE,
361 K06895, Table S8) and amino acid permeases (98.8±1.5%; Table S11) in all *Rothia* AIMSoral2 and 93%
362 of *Streptococcus* AIMSoral1 genomes, respectively. Although no metabolic interactions were exclusive to
363 *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2, both species exhibited unique interactions with other
364 members of the network. Notably, *Veillonella* AIMSoral3 produced agmatine, utilizable only by *Rothia*
365 AIMSoral2, while *Streptococcus* AIMSoral1 could use 2-oxo-3-hydroxy-4-phosphobutanoate produced by
366 *Pauljensenia* AIMSoral4.
367

368 **DISCUSSION**

369 In this study, we performed shotgun metagenomic sequencing on oral samples from 24 mother-infant
370 pairs from the prospective AIMS cohort, enabling finer taxonomic resolutions than previous studies. This
371 approach allowed the identification of previously undescribed species in the developing oral ecosystem
372 and reconstruction of their genomes, revealing functional potential and metabolic interactions within the
373 infant oral microbiome.

374 Our findings highlight the early dominance of *Streptococcus* spp. and increasing bacterial
375 richness over time, including the emergence of the *Rothia* genus. Species-level changes in these genera,
376 resulted in the establishment of *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2 as predominant
377 members of the infant oral microbiome. These genus-level transitions likely reflect adaptations to dietary
378 changes during weaning and structural changes from dentition (40,41) aligning with previous 16S rRNA
379 gene amplicon sequencing studies (6,8,40,41). However, our species-level observations diverge from
380 previous findings, which suggested *R. mucilaginosa* and *S. salivarius* were highly prevalent after first
381 dentition (6). Conversely, six-months-old AIMS infants lacked previously reported *S. salivarius* group
382 species and had comparatively low prevalence of *R. mucilaginosa*. This discrepancy likely stems from
383 lower 16S rRNA resolution, potentially misannotating *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2 as

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

384 *S. salivarius* and *R. mucilaginosa*, respectively, while our analysis shows these as novel species. We
385 found no significant evidence that milk or formula feeding influenced the abundance of these species. To
386 understand the ecological significance of these taxonomic shifts, we examined the functional capabilities
387 of these predominant species, beginning with *Streptococcus* AIMSoral1.
388 *Streptococcus* AIMSoral1 displayed an enriched repertoire of genes for amino acid metabolism,
389 particularly arginine and glutamate derivatives biosynthesis, suggesting adaptations to milk-derived
390 nutrients as well as a role in oral nitrogen cycling (42–44). Additionally, dextran-degrading enzymes
391 (GH66) may confer competitive advantages by breaking down EPS from other *Streptococcus* spp.
392 (45,46), potentially facilitating colonization of mucosal and dental surfaces. The presence of the fimA
393 adhesin suggests an increased specificity for the oral cavity (47). Notably, this gene was absent in *S.*
394 *salivarius* and other infant-associated streptococci, except *S. lactarius*, the predominant streptococcal
395 species in six-month-old infants' oral cavities.
396 Our genomic analyses of *Rothia* AIMSoral2 suggest a metabolic strategy focused on rapid
397 membrane synthesis (48) through enrichment of genes involved in fatty acid biosynthesis initiation
398 compared to other infant-associated *Rothia* species. This, may in turn support environmental adaptation
399 within the developing oral cavity (49). Genes encoding for GT11 cazymes (α(1,2/3)-fucosyltransferases)
400 were also found to be enriched across *Rothia* AIMSoral2 genomes, potentially playing a role in host
401 surface attachment (50,51). Together, these genomic features suggest *Streptococcus* AIMSoral1 and
402 *Rothia* AIMSoral2 are adapted to the developing oral cavity, with distinct metabolic strategies providing
403 ecological advantages during early colonization.
404 The strong positive correlation between *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2 at six
405 months suggests a potential cooperative relationship. Genomic and functional analyses suggest
406 *Streptococcus* AIMSoral1 to create niches supporting *Rothia* AIMSoral2 colonization, modifying dextran-
407 rich biofilm matrices. Meanwhile, *Rothia* AIMSoral2, encodes laccase-like multicopper oxidases
408 (AA1_1/2), which can neutralize reactive oxygen species (ROS) (52). This capability may confer
409 resilience against oxidative stress generated by hydrogen peroxide (H_2O_2) produced by cohabiting
410 *Streptococcus* species (53) during biofilm formation (54,55). The presence of CBM50 domains further

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

411 suggests that *Rothia* AIMSoral2 can effectively interact with other bacterial cell walls (56), consistent with
412 Rothia's role as an initiator of cell-cell interactions in early oral biofilms (57).

413 Metabolic modeling supports a cooperative relationship between *Streptococcus* AIMSoral1 and
414 *Rothia* AIMSoral2. In our bacterial co-occurrence network analysis, *Streptococcus* AIMSoral1 exhibited
415 the comparatively lowest metabolic complementarity with the other members, but highest
416 complementarity with *Rothia* AIMSoral2 and *Pauljensenia* AIMSoral4. While previous studies have shown
417 beneficial interactions between these genera (17,58,59), interactions between these undescribed species
418 in the infant oral cavity remained uninvestigated. Metabolic modeling identified 27 potential interactions
419 between *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2, including malate and lysine, supported by
420 genes encoding enzymes for their transport. These metabolic exchanges may underpin their stable co-
421 occurrence in the infant oral microbiome, serving as energy source (60) and supporting essential cellular
422 processes (61).

423 The identification of novel, prominent *Streptococcus* and *Rothia* species with specialized
424 functional traits has significant implications for understanding early oral microbiome development and
425 health outcomes. Arginine production by *Streptococcus* AIMSoral1 may modulate oral pH, as arginine
426 catabolism by other bacteria generates ammonia (62,63), potentially neutralizing acids from cariogenic
427 bacteria and reducing caries risk (64,65). Similarly, *Rothia* AIMSoral2 and *Veillonella* AIMSoral3 were
428 predicted to produce ammonia, suggesting a broader microbial strategy for counteracting acidification and
429 maintaining a balanced oral environment (5,66). Additionally, the exchange of lysine and agmatine within
430 the network could influence pH regulation through polyamine (e.g. putrescine) metabolism, reinforcing
431 metabolic interdependencies among these species. Beyond oral health, nitrate metabolism by *Rothia*
432 species is linked to NO production through the enterosalivary pathway, a process benefiting systemic
433 cardiovascular health (5,23). Disruptions in ammonia-producing pathways, due to antibiotics or diet, could
434 therefore increase risk of caries and other oral diseases.

435 Genome-resolved metagenomics is a powerful tool for identifying the functional elements of the
436 microbiome and provide insights into potential ecological interactions of *Streptococcus* and *Rothia*
437 species during early oral microbiome development. However, MAGs represent incomplete genomes,
438 potentially misrepresenting gene content. Most *Streptococcus* AIMSoral1 MAGs had medium-quality,

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

which might have hindered functional annotations and downstream interpretations. Nevertheless, these MAGs clustered consistently with high-quality reference genomes in both gene presence/absence patterns and phylogenomic relationships, supporting our observations. Despite limitations, MAG reconstruction and metabolic modelling revealed the functional potential of undescribed oral species, generating testable hypotheses about species interactions for future infant oral microbiome studies.

Identifying novel species and their metabolic interactions has implications for understanding early-life oral microbiome assembly and health. Future research should prioritize the isolation of *Streptococcus* AIMSoral1 and *Rothia* AIMSoral2 and validate their predicted interactions, in particular regarding their pH homeostasis capabilities. Additionally, longitudinal studies that track these species from infancy through childhood will help clarify their long-term health impacts while revealing factors that influence their abundance dynamics. By deepening our knowledge of these ecological interactions, we ultimately can facilitate the development of precision-based interventions that promote both oral and systemic health.

Material and methods

Amsterdam Infant Microbiome Study (AIMS) cohort

AIMS (https://aimsonderzoek.nl/en/) (67) is a microbiome-focused multi-ethnic, prospective birth cohort study by the GGD Amsterdam (Public Health Service of Amsterdam, the Netherlands) within the Sarphati Cohort (68). The study was approved by the Medical Ethical Examination Committee of Amsterdam University Medical Center (METC Amsterdam UMC, Reference NL64399.018.17), with written informed consent obtained from all participants, and for infants, parental consent was secured after birth.

Study population and sample collection

We collected 83 non-invasive oral biosamples (tongue swab: n=45, dental plaque: n=38) for metagenomic shotgun sequencing from 24 AIMS mother-infant pairs. Mothers self-collected both their samples (at 34 weeks pregnancy) and their infants' (at 1 and 6 months postpartum). Tongue samples were collected by stroking a synthetic swab (FLOQSwabs®, Copan Italia s.p.a.) across the tongue dorsum for 30 seconds. Dental plaque was collected by using a sterile toothbrush (without toothpaste) to

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

brush the buccal, lingual and occlusal surfaces of right lower molars for mothers and incisors in infants. Samples were home-stored in Liquid Amies Medium (eSwab®) at -18/-20°C, transported at -4°C, and stored at -70°C. Questionnaires tested in a face validity study (69) (Table S1) gathered metadata on infant milk-feeding type, antibiotic use, health status, and lifestyle factors (Table S1).

DNA extraction and sequencing

DNA was extracted using the ZY-D4306 ZymoBIOMICS-96 Magbead DNA Kit (with the use of the KingFisher™ Flex Purification System for high throughput) following BaseClear protocols. Of 73 samples, 57 yielded sufficient DNA for sequencing: maternal (tongue=23, dental plaque=21), 1-month infants (tongue=9, dental plaque=0), and 6-month infants (tongue=18, dental plaque=14). Metagenomic libraries were sequenced on an Illumina NovaSeq platform, yielding 2x150bp paired-end reads.

Pre-processing of metagenomic reads

Raw metagenomics reads were pre-processed using BBDuk and BBMap (70). Raw reads were trimmed to remove adapters, poor quality bases (Q15) and reads shorter than 45bp (minlength=45). BBMap was used to discard human reads (minID=0.95) and microbial reads were merged into longer single reads using BBMerge (min overlap=16). The SPAdes (71) assembler (metagenomic mode) was used for metagenomic assembly.

MAG reconstruction

We used SemiBin for multi-sample binning to generate MAGs (SemiBin 1.0.0) (72). For each mother-infant pair, we mapped short reads to a set of assembled contigs (500bp) from the same mother-infant pair via Bowtie2 (73). Samples were binned in multi-sample binning mode (*SemiBin multi_easy_bin*).

MAG quality was estimated using CheckM (73,74) and GUNC v.0.1 (75), and genomes were taxonomically classified using the GTDB database. High-quality *Streptococcus* and *Rothia* MAGs were further passed through the *anvi-estimate-genome-completeness* module in Anvi'o v.8 (76) to estimate their completeness and redundancy; draft genomes were considered suitable for downstream analyses

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

when displaying ≥50% completeness & ≤10% redundancy (i.e medium, medium-high and high quality).

The complete list of *Streptococcus* and *Rothia* MAGs used in this study are provided in Table S5.

Public reference genome acquisition

We conducted comparative genomic analyses on MAGs and publicly available *Streptococcus* and *Rothia* genomes from Progenomes3 (77) and GTDB (78). To ascertain the taxonomic placement of the undescribed *Streptococcus* species, we retrieved reference sequences from the *salivarius* group (*S. salivarius*: n=6, *S. vestibularis*: n=5, *S. thermophilus*: n=5). For *Rothia*, we used type strains sequences or GTDB representatives (Table S5). To perform comparative genomic analyses, we included only *Streptococcus* and *Rothia* species considered infant-associated, defined as those with average relative abundance of ≥5% and a prevalence of ≥30% (Figure S2). After quality filtering, 33 *Streptococcus* and 63 *Rothia* MAGs and genomes (hereon referred to as genomes) were used for genetic relatedness analysis, while 57 *Streptococcus* and 53 *Rothia* genomes comprised the final dataset for phylogenomics, pangenomics and metabolic enrichment analyses.

Metagenomic species-profiling and strain tracking

In order to assess the relative abundance of bacterial taxonomic groups per sample, concatenated reads were mapped onto reference genomes from the GTDB R220 database (79) by using the computational tool sylph v.0.6.1 (80).

Single nucleotide variant (SNV) calling was used to determine overall strain-tracking between mothers and 1mo infants (i.e. vertical transmission) and mothers and 6mo infants (i.e. persisted vertical transmission or horizontal transmission). Reference-based strain tracking was performed using *inStrain* v1.0.0 (81) with reference genomes from the UHGG database, which we used to recruit quality-controlled merged reads using Bowtie2. Scaffold-level microdiversity metrics were calculated using the *inStrain profile* module. Pairwise species comparisons were performed using the *inStrain compare* module to calculate genome-level differences between samples (average nucleotide identity threshold = 0.999).

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Phylogenomic and genetic relatedness analyses

Phylogenomic analyses were carried out in Anvi'o v8. Amino acid sequences of single-copy core genes (https://merenlab.org/2017/06/07/phylogenomics/) were extracted from the genus set of 110 and 63 *Streptococcus* and *Rothia* genomes (see Methods: Public reference genomes acquisition), respectively, and aligned using the *anvi-get-sequences-for-hmm-hits* program. We further retrieved and aligned single-copy core genes from 33 genomes from the *salivarius* group for a targeted phylogenomic analysis of these *Streptococcus* species. We visualized the phylogenomics tree by using the resulting concatenated amino acid alignment as input for the *anvi-gen-phylogenomics-tree* program.

Genetic relatedness analyses were conducted for *Rothia* and the *salivarius* lineage within *Streptococcus* genus using the *anvi-compute-genome-similarity* module, which uses FastANI to calculate the average nucleotide identity (ANI). Genomes with an ANI ≥95% were classified as the same species.

***Streptococcus* and *Rothia* pangenome**

Pangenome analyses were carried out in Anvi'o v8. Genomes were converted into contig databases (*anvi-gen-contig-database*) and functionally annotated from KEGG databases (*anvi-hmm* and *anvi-run-kegg-kofams*). Functionally annotated contig databases were stored (*anvi-gen-genomes-storage*) and used for pangenome reconstruction. Using the *anvi-pangenome* and *anvi-interactive* modules we reconstructed the *Streptococcus* and *Rothia* genus pangenome and subsequently visualize the hierarchical clustering of genomes based on gene presence/absence (Method: Ward; Distance: Euclidean; Fig. S7 and Fig. S8).

Metabolic enrichment analyses

We analyzed high- and medium-quality genomes to identify metabolic differences among *Streptococcus* and *Rothia* species. Functional annotations were used to predict metabolic pathways (KEGG MODULE resource) for *Streptococcus* and *Rothia* species. For this purpose, we used the *anvi-estimate-metabolism* module, which predicts both presence and completeness of the detected metabolic modules . Module abundance was determined as detailed in Watson et al. 2023 and the anvi'o tutorial

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

548 https://anvio.org/m/anvi-estimate-metabolism (82). Enrichment of metabolic modules across species was
549 estimated using the *anvi-compute-metabolic-enrichment* module, resulting in enrichment scores and
550 adjusted p-values (q-value) of modules found in *Streptococcus* and *Rothia* genomes (83).
551
552 **Profiling of carbohydrate-active enzymes and adhesion potential**
553 High-quality infant-type *Streptococcus* and *Rothia* genomes were annotated to the Carbohydrate-Active
554 enZYmes (CAZy) database using Anvi'o v8 (run_dbCAN2) (84). For adhesion potential, CAZy annotation
555 identified carbohydrate-binding modules (CBMs), while lectin-encoding genes were determined using
556 known *Streptococcus* adhesins from literature and LectomeXplore (85). BLASTp (sequence identity≥40%,
557 bitscore≥50, e-value≤0.001) (86) aligned genomes against lectin-sequences. Both CAZy and adhesin
558 profiles were analyzed in R using hierarchical clustering (euclidean distance, Ward.D2 method) on scaled
559 absolute gene counts. Differences in gene abundance across clusters were tested using Kruskal-Wallis
560 followed by pairwise Wilcoxon Signed-Rank tests (*pairwise.wilcox.test*) for post-hoc comparisons.
561
562 **Metabolic complementarity/interaction potential**
563 To test whether species co-occurrence was associated with metabolic interactions, we reconstructed
564 phylogenetically-adjusted GEMS using PhyloMInt (v0.1.0) (87). Briefly, PhyloMInt employs CarveMe (26)
565 to reconstruct GEMS and calculates (asymmetric) metabolic complementarity indices ($MI_{complementarity}$) for
566 genomes pairs based on the metabolites required or produced by their metabolic modules. We calculated
567 the average $MI_{complementarity}$ for all species pairs tested differences using pairwise-wilcoxon tests. To predict
568 metabolic exchanges, we used the PhyloMInt PTM tool in iNAP 2.0 (88).
569 To support predicted metabolic interactions, we determined the presence of specific transporters
570 (Table S9) for metabolites of interest using BLASTp (v2.6.0+) filtered by bitscore (≥50), e-value (≤0.001),
571 sequence identity (≥40%) and best hit.
572
573 **Data analysis and statistics**

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

574 Data analyses and statistical procedures were performed using a combination of RStudio
575 (v2022.02.0+443) and Anvi'o (v8). Pangenomic hierarchical clustering, metabolic enrichment analyses
576 using anvi-modules were performed in the Anvi'o 8.0 platformwhile other statistical tests were conducted
577 in RStudio. False discovery rate ('fdr') correction was applied for multiple testing.
578

579 **Acknowledgements**

580 We thank all the families who participate in the Amsterdam Infant Microbiome Study (AIMS) cohort and
581 the entire research team responsible for the recruitment of the cohort and collection of samples and
582 questionnaires.
583

584 **Study funding**

585 This project was supported by the GGD Amsterdam, the University of Amsterdam - Research Priority
586 Area Personal Microbiome Health (RPA-PMH), the Stichting Orale Biologie and the Dutch Research
587 Council (NWO) - MetaHealth project (NWA.1389.20.080). The funders had no role in study design, data
588 collection and analysis, decision to publish, or preparation of the manuscript. Human Biology-Microbiome-
589 Quantum Research Center (Bio2Q) is supported by the World Premier International Research Center
590 Initiative (WPI), MEXT, Japan.
591

592 **Conflict of Interest**

593 The authors declare no conflict of interest.
594

595 **Data availability**

596 The raw-reads for each AIMS biosample were submitted to the European Nucleotide Archive (ENA) at
597 EMBL-EBI repository and the National Center for Biotechnology and Information (NCBI with BioProject
598 accession number PRJEB88622). The dataset will be publicly available on 2025-09-01.
599 The code generating the main figures and underlying data are available at GitHub:
600 https://github.com/nicholaspucci/infant-oral-microbiome-strep-rothia.
601

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

References

1. Sedghi L, DiMassa V, Harrington A, Lynch SV, Kapila YL. The oral microbiome: Role of key organisms and complex networks in oral health and disease. Periodontol 2000. 2021 Oct;87(1):107–31.

2. Kilian M, Chapple ILC, Hannig M, Marsh PD, Meuric V, Pedersen AML, et al. The oral microbiome - an update for oral healthcare professionals. Br Dent J. 2016 Nov 18;221(10):657–66.

3. Rosier BT, Marsh PD, Mira A. Resilience of the Oral Microbiota in Health: Mechanisms That Prevent Dysbiosis. J Dent Res. 2018 Apr;97(4):371–80.

4. Kapila YL. Oral health's inextricable connection to systemic health: Special populations bring to bear multimodal relationships and factors connecting periodontal disease to systemic diseases and conditions. Periodontol 2000. 2021 Oct;87(1):11–6.

5. Rosier BT, Takahashi N, Zaura E, Krom BP, MartÍnez-Espinosa RM, van Breda SGJ, et al. The Importance of Nitrate Reduction for Oral Health. J Dent Res. 2022 Jul;101(8):887–97.

6. Dashper SG, Mitchell HL, Lê Cao KA, Carpenter L, Gussy MG, Calache H, et al. Temporal development of the oral microbiome and prediction of early childhood caries. Sci Rep. 2019 Dec 24;9(1):19732.

7. Sulyanto RM, Thompson ZA, Beall CJ, Leys EJ, Griffen AL. The Predominant Oral Microbiota Is Acquired Early in an Organized Pattern. Scientific Reports. 2019 Jul 22;9(1):1–8.

8. Dzidic M, Collado MC, Abrahamsson T, Artacho A, Stensson M, Jenmalm MC, et al. Oral microbiome development during childhood: an ecological succession influenced by postnatal factors and associated with tooth decay. ISME J. 2018 Sep;12(9):2292–306.

9. Mergenhagen SE, Sandberg AL, Chassy BM, Brennan MJ, Yeung MK, Donkersloot JA, et al. Molecular basis of bacterial adhesion in the oral cavity. Rev Infect Dis. 1987 Sep-Oct;9 Suppl 5:S467–74.

10. Nobbs AH, Lamont RJ, Jenkinson HF. Streptococcus adherence and colonization. Microbiol Mol Biol Rev. 2009 Sep;73(3):407–50, Table of Contents.

11. Bagg J. Essentials of Microbiology for Dental Students. Oxford University Press, USA; 1999. 326 p.

12. Marsh PD, Zaura E. Dental biofilm: ecological interactions in health and disease. J Clin Periodontol. 2017 Mar;44 Suppl 18:S12–22.

13. Pattanaporn K, Saraithong P, Khongkhunthian S, Aleksejuniene J, Laohapensang P, Chhun N, et al. Mode of delivery, mutans streptococci colonization, and early childhood caries in three- to five-year-old Thai children. Community Dent Oral Epidemiol. 2013 Jun;41(3):212–23.

14. Li Y, Caufield PW, Dasanayake AP, Wiener HW, Vermund SH. Mode of delivery and other maternal factors influence the acquisition of Streptococcus mutans in infants. J Dent Res. 2005 Sep;84(9):806–11.

15. Valiei A, Dickson A, Aminian-Dehkordi J, Mofrad MRK. Metabolic interactions shape emergent biofilm structures in a conceptual model of gut mucosal bacterial communities. NPJ Biofilms Microbiomes. 2024 Oct 2;10(1):99.

16. Kolenbrander PE, Palmer RJ Jr, Rickard AH, Jakubovics NS, Chalmers NI, Diaz PI. Bacterial interactions and successions during plaque development. Periodontol 2000. 2006;42:47–79.

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

17. Wilbert SA, Mark Welch JL, Borisy GG. Spatial Ecology of the Human Tongue Dorsum Microbiome. Cell Rep. 2020 Mar 24;30(12):4003–15.e3.

18. San Roman M, Wagner A. An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. PLoS Comput Biol. 2018 Jul;14(7):e1006340.

19. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol. 2012 Jul 12;8(7):e1002606.

20. Díaz R, Garrido D. Screening competition and cross-feeding interactions during utilization of human milk oligosaccharides by gut microbes. Microbiome Res Rep. 2024 Jan 1;3(1):12.

21. Bowen WH, Burne RA, Wu H, Koo H. Oral Biofilms: Pathogens, Matrix, and Polymicrobial Interactions in Microenvironments. Trends Microbiol. 2018 Mar;26(3):229–42.

22. Nascimento MM, Alvarez AJ, Huang X, Hanway S, Perry S, Luce A, et al. Arginine Metabolism in Supragingival Oral Biofilms as a Potential Predictor of Caries Risk. JDR Clin Trans Res. 2019 Jul;4(3):262–70.

23. Rosier BT, Johnston W, Carda-Diéguez M, Simpson A, Cabello-Yeves E, Piela K, et al. Publisher Correction: Nitrate reduction capacity of the oral microbiota is impaired in periodontitis: potential implications for systemic nitric oxide availability. Int J Oral Sci. 2024 Jan 26;16(1):8.

24. Utter DR, Borisy GG, Eren AM, Cavanaugh CM, Mark Welch JL. Metapangenomics of the oral microbiome provides insights into habitat adaptation and cultivar diversity. Genome Biol. 2020 Dec 16;21(1):293.

25. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. 2010 Sep;28(9):977–82.

26. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 2018 Sep 6;46(15):7542–53.

27. Proffitt C, Bidkhori G, Lee S, Tebani A, Mardinoglu A, Uhlen M, et al. Genome-scale metabolic modelling of the human gut microbiome reveals changes in the glyoxylate and dicarboxylate metabolism in metabolic disorders. iScience. 2022 Jul 15;25(7):104513.

28. Pinto F, Medina DA, Pérez-Correa JR, Garrido D. Modeling Metabolic Interactions in a Consortium of the Infant Gut Microbiome. Front Microbiol. 2017 Dec 14;8:2507.

29. Giordano N, Gaudin M, Trottier C, Delage E, Nef C, Bowler C, et al. Genome-scale community modelling reveals conserved metabolic cross-feedings in epipelagic bacterioplankton communities. Nat Commun. 2024 Mar 28;15(1):2721.

30. Gao XY, Zhi XY, Li HW, Klenk HP, Li WJ. Comparative genomics of the bacterial genus Streptococcus illuminates evolutionary implications of species groups. PLoS One. 2014 Jun 30;9(6):e101229.

31. Kilian M, Slotved HC, Fuursted K, D'Mello A, Tettelin H. Re-evaluation of boundaries of and and demonstration of multiple later synonyms of and description of subsp. subsp. nov. and emended description of. Int J Syst Evol Microbiol [Internet]. 2025 Mar;75(3). Available from: http://dx.doi.org/10.1099/ijsem.0.006704

32. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007 Jan;57(Pt 1):81–91.

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

33. Gilroy R, Adam ME, Kumar B, Pallen MJ. An initial genomic blueprint of the healthy human oesophageal microbiome. Access Microbiol [Internet]. 2023 Jun 26;5(6). Available from: http://dx.doi.org/10.1099/acmi.0.000558.v3

34. Buist G, Steen A, Kok J, Kuipers OP. LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol. 2008 May;68(4):838–47.

35. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem. 2008;77:521–55.

36. Wang G, Ge Z, Rasko DA, Taylor DE. Lewis antigens in Helicobacter pylori: biosynthesis and phase variation. Mol Microbiol. 2000 Jun;36(6):1187–96.

37. Elola MT, Capurro MI, Barrio MM, Coombs PJ, Taylor ME, Drickamer K, et al. Lewis x antigen mediates adhesion of human breast carcinoma cells to activated endothelium. Possible involvement of the endothelial scavenger receptor C-type lectin. Breast Cancer Res Treat. 2007 Jan;101(2):161–74.

38. Moran AP. Relevance of fucosylation and Lewis antigen expression in the bacterial gastroduodenal pathogen Helicobacter pylori. Carbohydr Res. 2008 Aug 11;343(12):1952–65.

39. Wang G, Boulton PG, Chan NWC, Palcic MM, Taylor DE. Novel Helicobacter pylori alpha1,2-fucosyltransferase, a key enzyme in the synthesis of Lewis antigens. Microbiology (Reading). 1999 Nov;145 (Pt 11):3245–53.

40. Mason MR, Chambers S, Dabdoub SM, Thikkurissy S, Kumar PS. Characterizing oral microbial communities across dentition states and colonization niches. Microbiome. 2018 Apr 10;6(1):67.

41. Xiao J, Fiscella KA, Gill SR. Oral microbiome: possible harbinger for children's health. International Journal of Oral Science. 2020 Apr 30;12(1):1–13.

42. Poulsen KO, Meng F, Lanfranchi E, Young JF, Stanton C, Ryan CA, et al. Dynamic Changes in the Human Milk Metabolome Over 25 Weeks of Lactation. Front Nutr. 2022 Jul 14;9:917659.

43. Smilowitz JT, O'Sullivan A, Barile D, German JB, Lönnerdal B, Slupsky CM. The human milk metabolome reveals diverse oligosaccharide profiles. J Nutr. 2013 Nov;143(11):1709–18.

44. Rosier BT, Buetas E, Moya-Gonzalvez EM, Artacho A, Mira A. Nitrate as a potential prebiotic for the oral microbiome. Sci Rep. 2020 Jul 30;10(1):12895.

45. Suzuki N, Kim YM, Fujimoto Z, Momma M, Okuyama M, Mori H, et al. Structural elucidation of dextran degradation mechanism by streptococcus mutans dextranase belonging to glycoside hydrolase family 66. J Biol Chem. 2012 Jun 8;287(24):19916–26.

46. Mahmoud S. Dextran degradation and biofilm inhibition using microbial dextranases. Merit Medical Journal. 2024 Jan 1;1(1):17–20.

47. Oligino L, Fives-Taylor P. Overexpression and purification of a fimbria-associated adhesin of Streptococcus parasanguis. Infect Immun. 1993 Mar;61(3):1016–22.

48. Parsons JB, Rock CO. Bacterial lipids: metabolism and membrane homeostasis. Prog Lipid Res. 2013 Jul;52(3):249–76.

49. Fozo EM, Kajfasz JK, Quivey RG Jr. Low pH-induced membrane fatty acid alterations in oral bacteria. FEMS Microbiol Lett. 2004 Sep 15;238(2):291–5.

50. Wong EHJ, Ng CG, Chua EG, Tay ACY, Peters F, Marshall BJ, et al. Comparative Genomics Revealed Multiple Helicobacter pylori Genes Associated with Biofilm Formation In Vitro. PLoS One.

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

2016 Nov 21;11(11):e0166835.

51. Shao C, Sun Y, Wang N, Yu H, Zhou Y, Chen C, et al. Changes of proteome components of Helicobacter pylori biofilms induced by serum starvation. Mol Med Rep. 2013 Dec;8(6):1761–6.

52. Sitthisak S, Howieson K, Amezola C, Jayaswal RK. Characterization of a multicopper oxidase gene from Staphylococcus aureus. Appl Environ Microbiol. 2005 Sep;71(9):5650–3.

53. Zhu L, Kreth J. The role of hydrogen peroxide in environmental adaptation of oral microbial communities. Oxid Med Cell Longev. 2012 Jul 16;2012:717843.

54. García-Mendoza A, Liébana J, Castillo AM, de la Higuera A, Piédrola G. Evaluation of the capacity of oral streptococci to produce hydrogen peroxide. J Med Microbiol. 1993 Dec;39(6):434–9.

55. Liu X, Ramsey MM, Chen X, Koley D, Whiteley M, Bard AJ. Real-time mapping of a hydrogen peroxide concentration profile across a polymicrobial bacterial biofilm using scanning electrochemical microscopy. Proc Natl Acad Sci U S A. 2011 Feb 15;108(7):2668–73.

56. Mesnage S, Dellarole M, Baxter NJ, Rouget JB, Dimitrov JD, Wang N, et al. Molecular basis for bacterial peptidoglycan recognition by LysM domains. Nat Commun. 2014 Jun 30;5:4269.

57. Palmer RJ Jr, Shah N, Valm A, Paster B, Dewhirst F, Inui T, et al. Interbacterial Adhesion Networks within Early Oral Biofilms of Single Human Hosts. Appl Environ Microbiol [Internet]. 2017 Jun 1;83(11). Available from: http://dx.doi.org/10.1128/AEM.00407-17

58. Giacomini JJ, Torres-Morales J, Dewhirst FE, Borisy GG, Mark Welch JL. Site Specialization of Human Oral Species. Microbiol Spectr. 2023 Feb 14;11(1):e0404222.

59. Mashima I, Nakazawa F. The interaction between Streptococcus spp. and Veillonella tobetsuensis in the early stages of oral biofilm formation. J Bacteriol. 2015 Jul;197(3):2104–11.

60. Schubert C, Unden G. C-Dicarboxylates as Growth Substrates and Signaling Molecules for Commensal and Pathogenic Enteric Bacteria in Mammalian Intestine. J Bacteriol. 2022 Apr 19;204(4):e0054521.

61. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev. 1973 Jun;37(2):258.

62. Schneider BL, Kiupakis AK, Reitzer LJ. Arginine catabolism and the arginine succinyltransferase pathway in Escherichia coli. J Bacteriol. 1998 Aug;180(16):4278–86.

63. Bradshaw DJ, Marsh PD. Analysis of pH-driven disruption of oral microbial communities in vitro. Caries Res. 1998;32(6):456–62.

64. Vaziriamjad S, Solgi M, Kamarehei F, Nouri F, Taheri M. Evaluation of L-arginine supplement on the growth rate, biofilm formation, and antibiotic susceptibility in Streptococcus mutans. Eur J Med Res. 2022 Jul 2;27(1):108.

65. Carda-Diéguez M, Moazzez R, Mira A. Functional changes in the oral microbiome after use of fluoride and arginine containing dentifrices: a metagenomic and metatranscriptomic study. Microbiome. 2022 Sep 28;10(1):159.

66. Yamamoto Y, Washio J, Shimizu K, Igarashi K, Takahashi N. Inhibitory Effects of Nitrite on Acid Production in Dental Plaque in Children. Oral Health Prev Dent. 2017;15(2):153–6.

67. AIMS [Internet]. 2023 [cited 2025 Mar 20]. About AIMS. Available from: https://aimsonderzoek.nl/en/about-aims/

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

766 68. Sarphati Amsterdam [Internet]. 2023 [cited 2025 Mar 20]. Sarphati Cohort. Available from:
767 https://sarphati.amsterdam/en/sarphati-cohort/

768 69. Haring LV, Hall JT, Janssen A, Johannes JM, Verhoeff AP, Ujcic-Voortman JK. Developing a digital
769 informed consent app: opportunities and challenges of a new format to inform and obtain consent in
770 public health research. BMC Med Ethics. 2023 Nov 8;24(1):97.

771 70. BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014.

772 71. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler.
773 Curr Protoc Bioinformatics. 2020 Jun;70(1):e102.

774 72. Pan S, Zhu C, Zhao XM, Coelho LP. A deep siamese neural network improves metagenome-
775 assembled genomes in microbiome datasets across different environments. Nat Commun. 2022 Apr
776 28;13(1):2326.

777 73. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar
778 4;9(4):357–9.

779 74. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of
780 microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015
781 Jul;25(7):1043–55.

782 75. Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, et al. GUNC: detection of
783 chimerism and contamination in prokaryotic genomes. Genome Biol. 2021 Jun 13;22(1):178.

784 76. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated,
785 reproducible multi-omics with anvi'o. Nat Microbiol. 2021 Jan;6(1):3–6.

786 77. Fullam A, Letunic I, Schmidt TSB, Ducarmon QR, Karcher N, Khedkar S, et al. proGenomes3:
787 approaching one million accurately and consistently annotated high-quality prokaryotic genomes.
788 Nucleic Acids Res. 2023 Jan 6;51(D1):D760–6.

789 78. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with
790 the genome taxonomy database. Bioinformatics. 2022 Nov 30;38(23):5315–6.

791 79. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. Author Correction: A
792 complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol. 2020
793 Sep;38(9):1098.

794 80. Shaw J, Yu YW. Rapid species-level metagenome profiling and containment estimation with sylph.
795 Nat Biotechnol [Internet]. 2024 Oct 8; Available from: http://dx.doi.org/10.1038/s41587-024-02412-y

796 81. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles
797 population microdiversity from metagenomic data and sensitively detects shared microbial strains.
798 Nat Biotechnol. 2021 Jun;39(6):727–36.

799 82. Veseli I, Chen YT, Schechter MS, Vanni C, Fogarty EC, Watson AR, et al. Microbes with higher
800 metabolic independence are enriched in human gut microbiomes under stress. Elife [Internet]. 2025
801 May 16;12. Available from: http://dx.doi.org/10.7554/eLife.89862

802 83. Anvi'o Home [Internet]. [cited 2025 Apr 3]. anvi-compute-metabolic-enrichment [program]. Available
803 from: https://anvio.org/help/7.1/programs/anvi-compute-metabolic-enrichment/

804 84. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated
805 carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018 Jul 2;46(W1):W95–101.

806 85. Bonnardel F, Mariethoz J, Pérez S, Imberty A, Lisacek F. LectomeXplore, an update of UniLectin for

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

the discovery of carbohydrate-binding proteins based on a new lectin classification. Nucleic Acids Res. 2021 Jan 8;49(D1):D1548–54.

86. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.

87. Lam TJ, Stamboulian M, Han W, Ye Y. Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. PLoS Comput Biol. 2020 Oct;16(10):e1007951.

88. Peng X, Feng K, Yang X, He Q, Zhao B, Li T, et al. iNAP 2.0: Harnessing metabolic complementarity in microbial network analysis. Imeta. 2024 Oct;3(5):e235.

79. Schleifer KH & Kandler OJBR. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological reviews* 1972, **36**: 407-477.

Captions

Figure S1. Overview of the number of metagenomics reads retained per sample. For each bar, colors indicate the proportion of reads fromA) tongue and B) tooth biofilm samples that passed (green) and failed (red) quality control as well as reads mapping to human DNA (blue). Failed and human reads were subsequently filtered out the dataset.

Figure S2. Abundance and prevalence thresholds used for inclusion in the Spearman's rank correlations and to define infant-associated Streptococcus species in our study. Spearman's rank correlation analyses were performed separately on bacterial species pairs on the tongue dorsum of A) 1- and B) 6-month-old infants and the C) dental plaque of 6-month-old infants that had a mean relative abundance of ≥1% (black dotted line). For the comparative genomics analyses, we defined infant-associated Streptococcus species as those showing a mean relative abundance ≥5% (horizontal red dotted line) and prevalence across AIMS infants ≥30% (vertical red dotted line) on the tongue dorsum of A) 1- and/or B) 6-month-old infants and/or in the C) dental plaque of 6-month-old infants.

Figure S3. Oral microbiota of AIMS infants and corresponding milk feeding information. A) PCoA of tongue dorsum (circles) and dental plaque (triangles) microbiomes from AIMS infants (one

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

834 month=small; six months=large). For each sample, the corresponding milk feeding type is reported
835 (breastmilk=pink, mixed=red,formula=green). PERMANOVA analyses showed no significant clustering
836 based on milk feeding within the six months samples (adonis: R^2=0.07, F=1.10, p=0.32). **B)** Relative
837 abundances of Streptococcus AIMSoral1, Rothia AIMSoral2, Veillonella AIMSoral3 and Pauljensenia
838 AIMSoral4 found in infants breast-, formula- and mixed milk-feeding. Significance among groups was
839 tested for each species by pairwise Wilcoxon's signed rank tests. No significant differences were
840 detected. **C)** Correlation plot of Streptococcus AIMSoral1 and Rothia AIMSoral2 showed no association
841 between their abundance and milk-feeding type.

842 **Figure S4. Heatmaps showing Spearman's correlation coefficients for infant oral species as 1 and**
843 **6 months of age.** Correlations between oral species found to be abundant (mean relative abundance
844 ≥1%) in the tongue biofilms of A) 1 and B) 6-months-old and C) dental biofilm of six-month-old AIMS
845 infants are colored in red (positive) or blue (negative). Significance is indicated by asterisks: *p≤0.05,
846 **≤0.01, ***≤0.001, ****≤0.0001. Self-self correlations are shown in grey.

847 **Figure S5. Phylogenomic tree of the *Streptococcus* genus.** Phylogenomic analysis of 110
848 *Streptococcus* genomes representing 89 species, including 89 reference genomes and 21 metagenome-
849 assembled genomes (MAGs). *Lactococcus lactis* was included as an outgroup. Red circles indicate
850 bootstrap values≥0.9.

851 **Figure S6. Heatmaps of Average Nucleotide Identity (ANI) of Streptococcus (salivarius group) and**
852 **Rothia genomes.** Pairwise ANI comparisons between A) Streptococcus (salivarius group) and B) Rothia
853 MAGs (white) and reference genomes (red). For each genome, the GTDB/AIMSoral and final species
854 assignments are provided.

855 **Figure S7. Pangenome of infant-associated Streptococcus species in the oral cavity of AIMS**
856 **infants.** The figure shows the pangenomic characteristics of infant-associated Streptoccoccus species
857 sorted by gene presence/absence. The pangenome includes MAGs reconstructed from AIMS oral
858 samples (sampling location: tongue dorsum = purple, tooth plaque = orange) collected at different

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

timepoints (infant 1 month=green, infant 6 months = red). No high- and medium-quality MAGs from infant-associated species were retrieved from maternal oral samples. Reference genomes are given in grey. Species annotations as well as bars showing genome/MAG total length, completion and redundancy are shown for each strain. For each gene cluster, metrics such as number of genomes where a gene cluster is present ('num of contributing genomes) and various functional annotations are shown.

Figure S8. Pangenome of infant-associated Rothia species in the oral cavity of AIMS infants. The figure shows the pangenomic characteristics of infant-associated Rothia species sorted by gene presence/absence. The pangenome includes MAGs reconstructed from AIMS oral samples (sampling location: tongue dorsum=purple, tooth plaque=orange) collected at different timepoints (mother 34wks gestation = brown, infant 1 month =green, infant 6 months = red). Reference genomes are given in grey. Species annotations as well as bars showing genome/MAG total length, completion and redundancy are shown for each strain. For each gene cluster, metrics such as number of genomes where a gene cluster is present ('num of contributing genomes) and various functional annotations are shown.

Figure S9. Metabolic complementarity and interactions among infant-associated Streptococcus and Rothia species in our study. A) Bargraphs show the average metabolic complementarity index (±SD) for pairwise interactions among infant-associated Streptococcus and Rothia, and vice versa. For each complementarity species pair, the reference species is shown on the x-axis, while the other species is shown in the facet on top of the bar. B) Heatmap showing bacterial-metabolite bipartite interactions production = red, utilization = blue, both = purple, no interaction = white) potentially occurring among Streptococcus AIMSoral1, Rothia AIMSoral2 and their closely related infant-associated species in our study. Metabolites are classified by their superclasses.

Figure S10. Principal coordinate analysis (PCoA) of tongue dorsum microbiomes from AIMS mothers (34wks gestation=brown) and their infants (one month=green; six months=red).

Figure S11. Analysis of 165 rRNA sequences from S. salivarius and Streptococcus AIMSoral1.

Figure S12. Abundance of Streptococcus AIMSoral1 and Rothia AIMSoral2 across infant oral samples from Ferretti et al. 2018.

Table S1. Metadata for metagenomics samples retrieved from AIMS mothers (34wks gestation) and their infants (one and six months) available for this study.

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

887 **Table S2. Overview strain tracking output (inStrain).**

888 **Table S3. Spearman rank-order R-values from pairwise abundance correlations between abundant (≥1%)**

889 **species in the tongue dorsum and dental plaque samples of AIMS infants (one and six months).**

890 **Table S4. Spearman rank-order *p*-values from pairwise abundance correlations between abundant (≥1%)**

891 **species in the tongue dorsum and dental plaque samples of AIMS infants (one and six months).**

892 **Table S5. Overview of medium- and high-quality metagenome-assembled genomes (MAGs) and**

893 **genomes used for this study.**

894 **Table S6. Output of the metabolic enrichment analysis performed among infant-associated**

895 ***Streptococcus* species.**

896 **Table S7. Output of the metabolic enrichment analysis performed among infant-associated *Rothia***

897 **species.**

898 **Table S8. Matrices of KEGG Orthologs (KOs) copy numbers per infant-associated *Streptococcus***

899 **and *Rothia* genomes/MAGs.**

900 **Table S9. Overview reference proteins used for genomic analyses of infant-associated**

901 ***Streptococcus* spp. adhesin repertoire and validation of malate and lysine transporters from**

902 ***Streptococcus AIMSoral*1 and *Rothia AIMSoral*2.**

903 **Table S10. Overview pairwise Wilcoxon test outputs on differences in gene copy-number among**

904 **infant-associated *Streptococcus* and *Rothia* genomes/MAGs.**

905 **Table S11. NCBI BLASTp summary for malate and lysine transporter/permease proteins across**

906 **infant-associated *Streptococcus* genomes/MAGs.**

907 **Table S12. NCBI BLASTp summary for malate and lysine uptake proteins across infant-associated**

908 ***Rothia* genomes/MAGs.**

909 **Table S13. PhyloMInt complementarity indices for pairwise comparisons between infant-**

910 **associated *Streptococcus* and *Rothia* genomes and *Veillonella* AIMSoral3 and *Pauljensenia***

911 ***AIMSoral*4.**

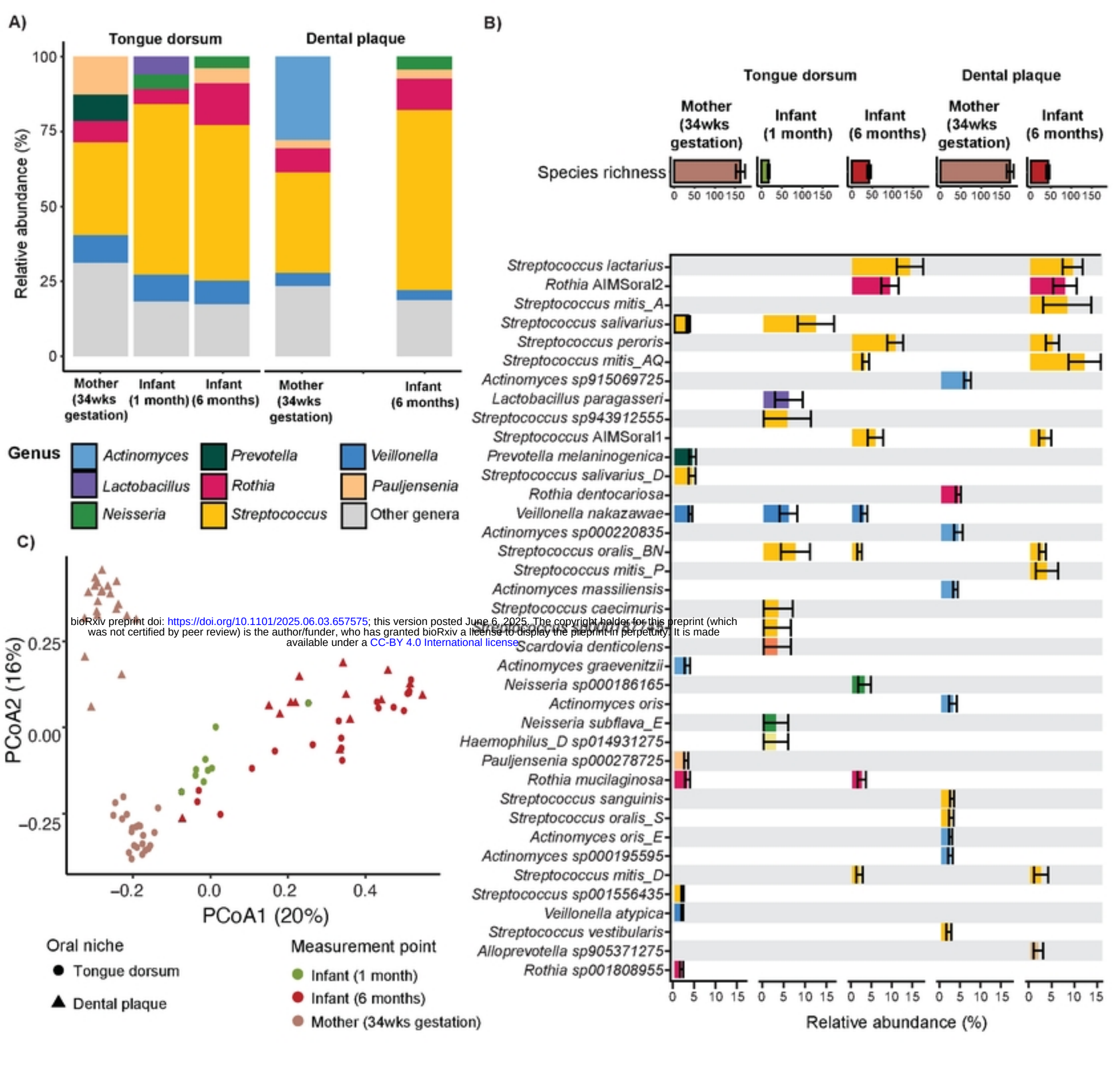
912 **Table S14. Overview metabolites predicted to be produced/utilized by *S.* salivarius, *Streptococcus***

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

913 **AIMSoral1, *R.* mucilaginosa, *R.* mucilaginosa_A, *Rothia* AIMSoral2, *Veillonella* AIMSoral3 and**

914 ***Pauljensenia* AIMSoral4.**

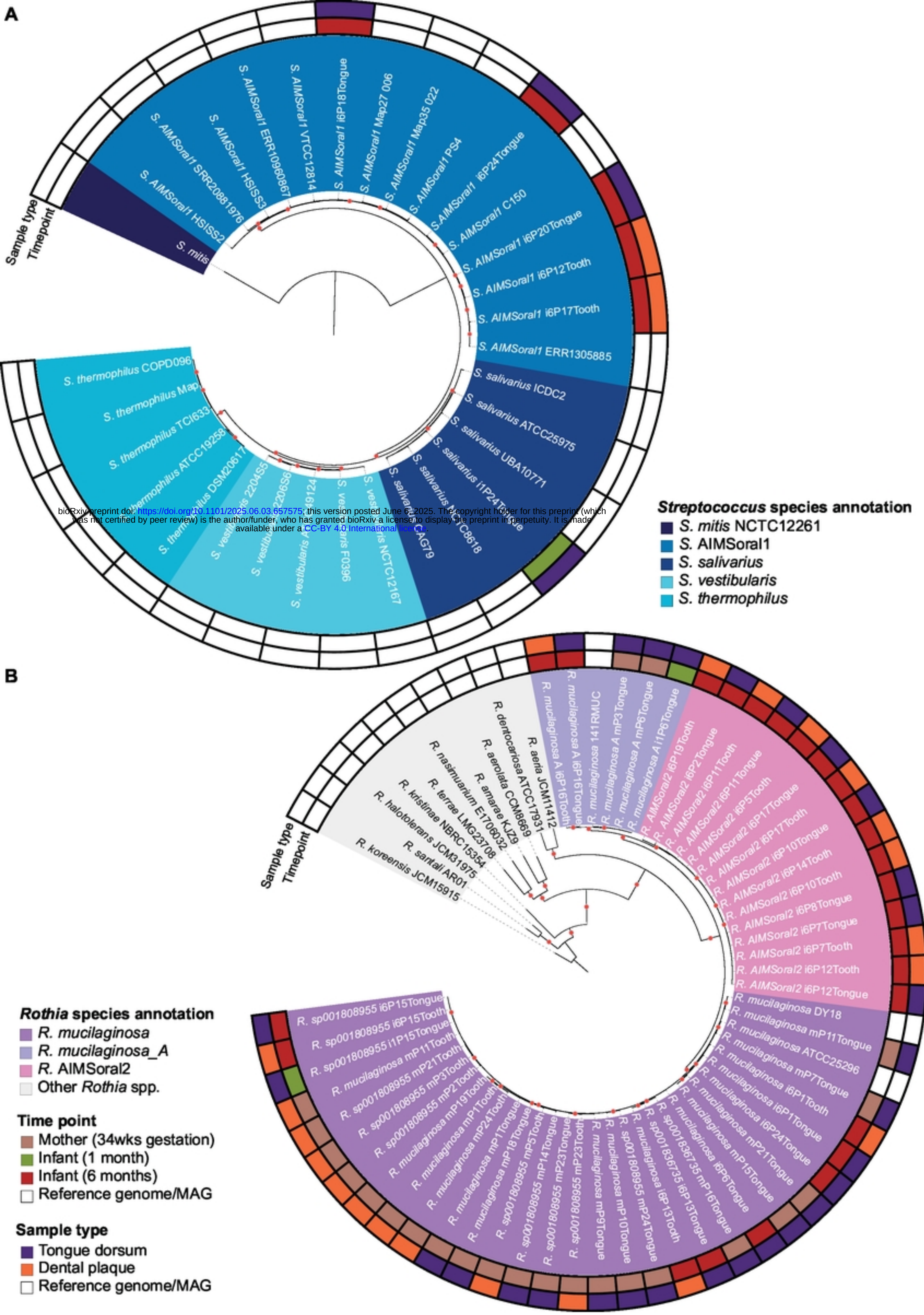
915 **Table S15. AIMS oral microbiome data.**



bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

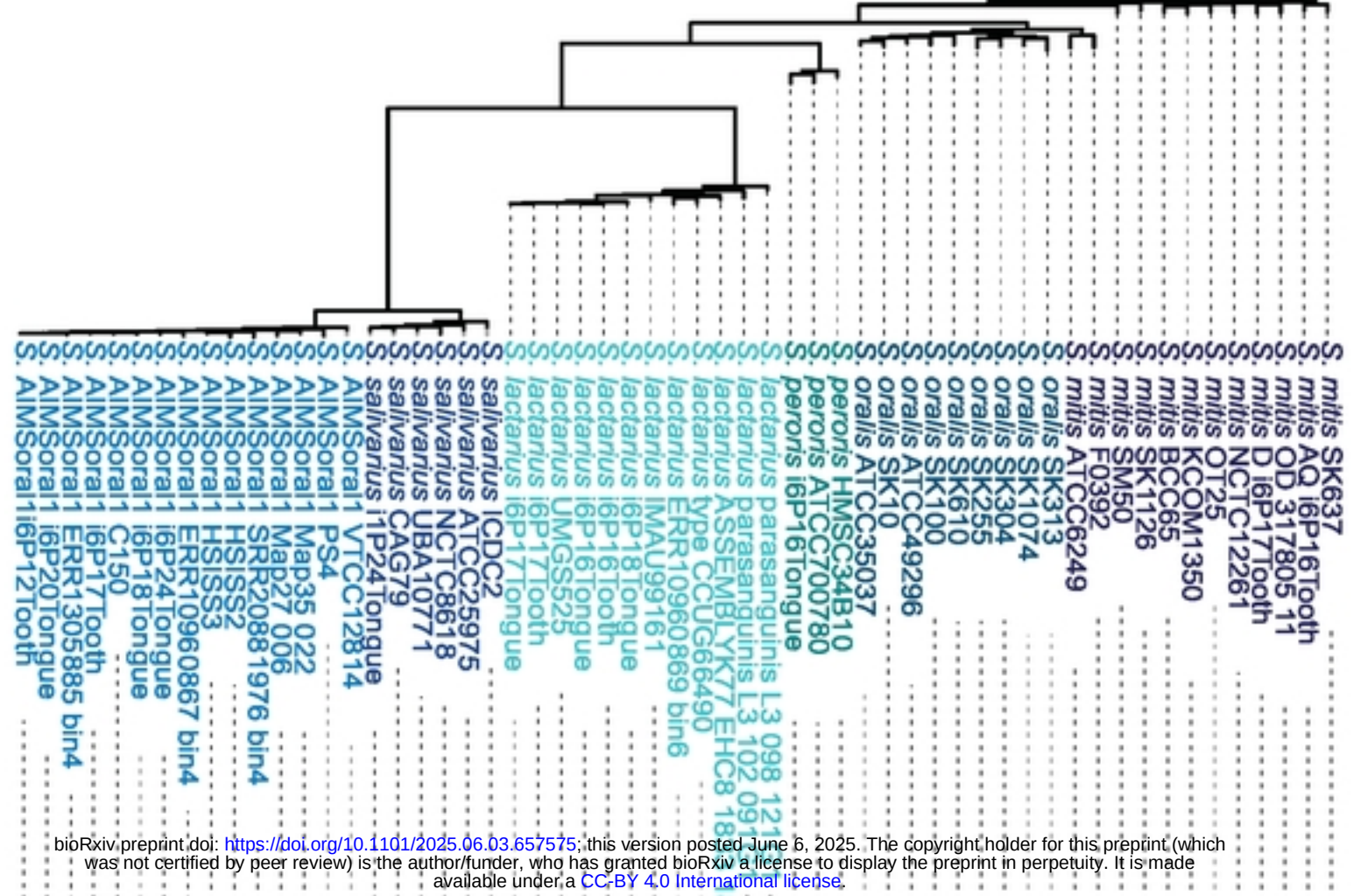
Figure

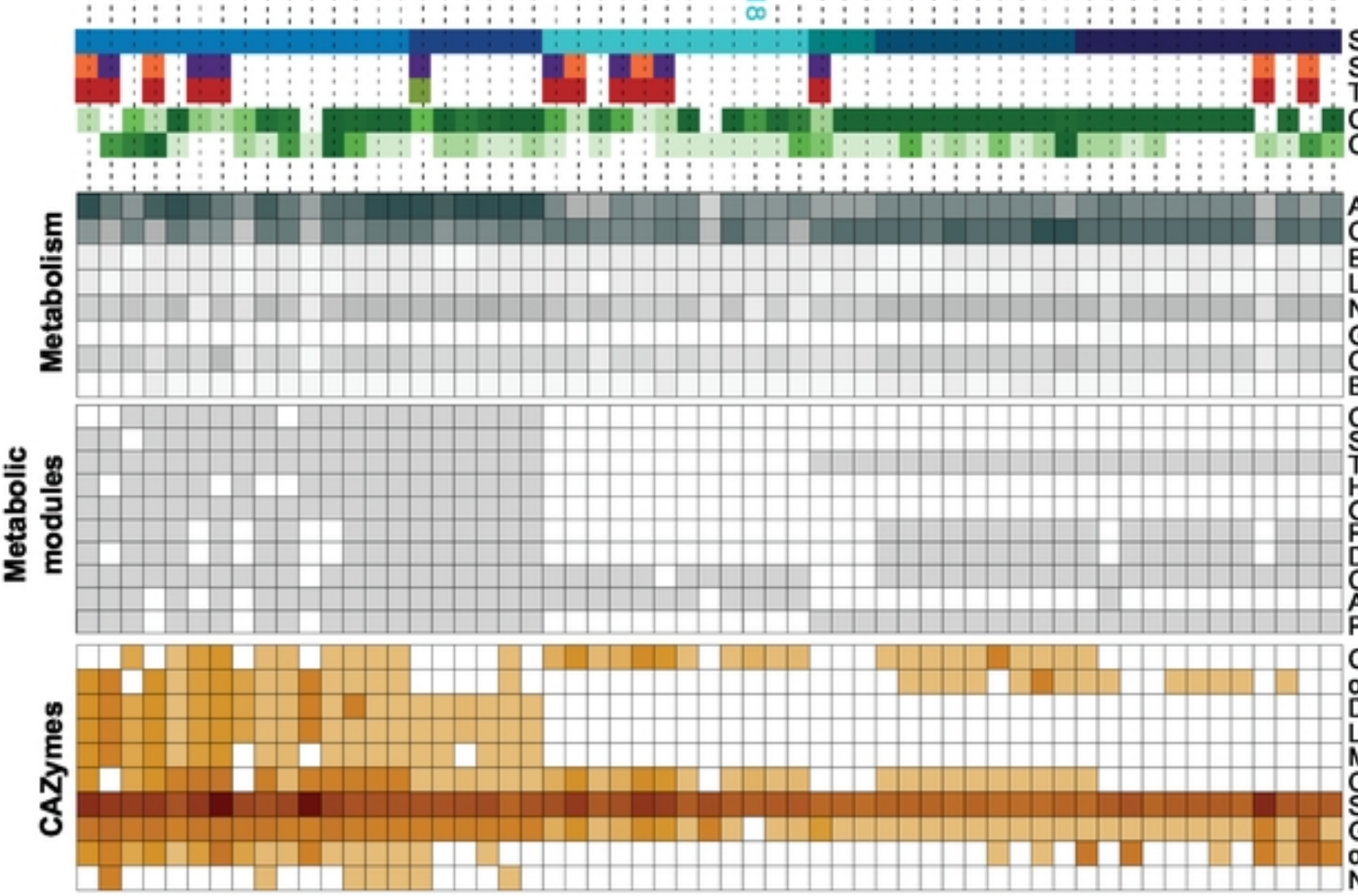
bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

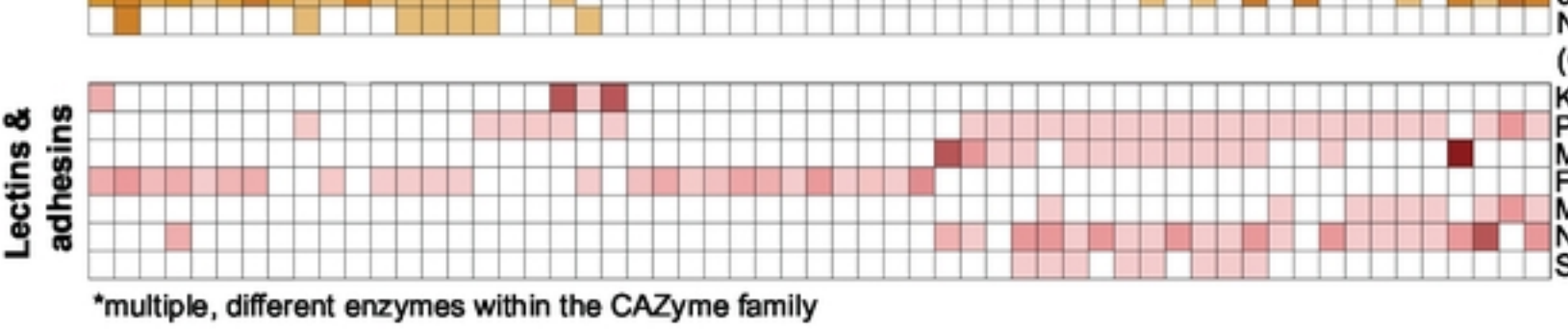


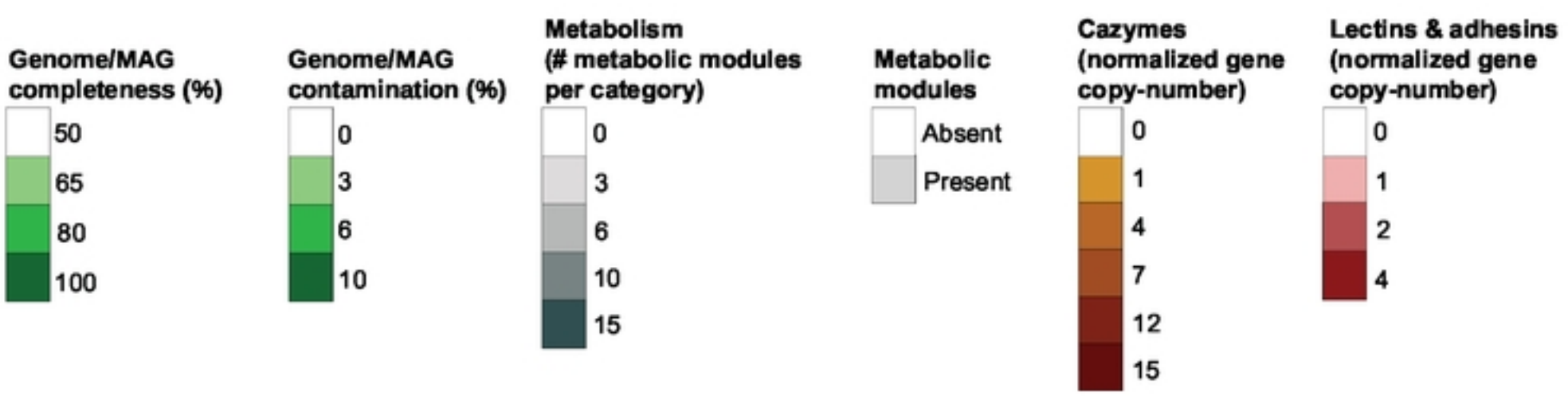
Figure

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.



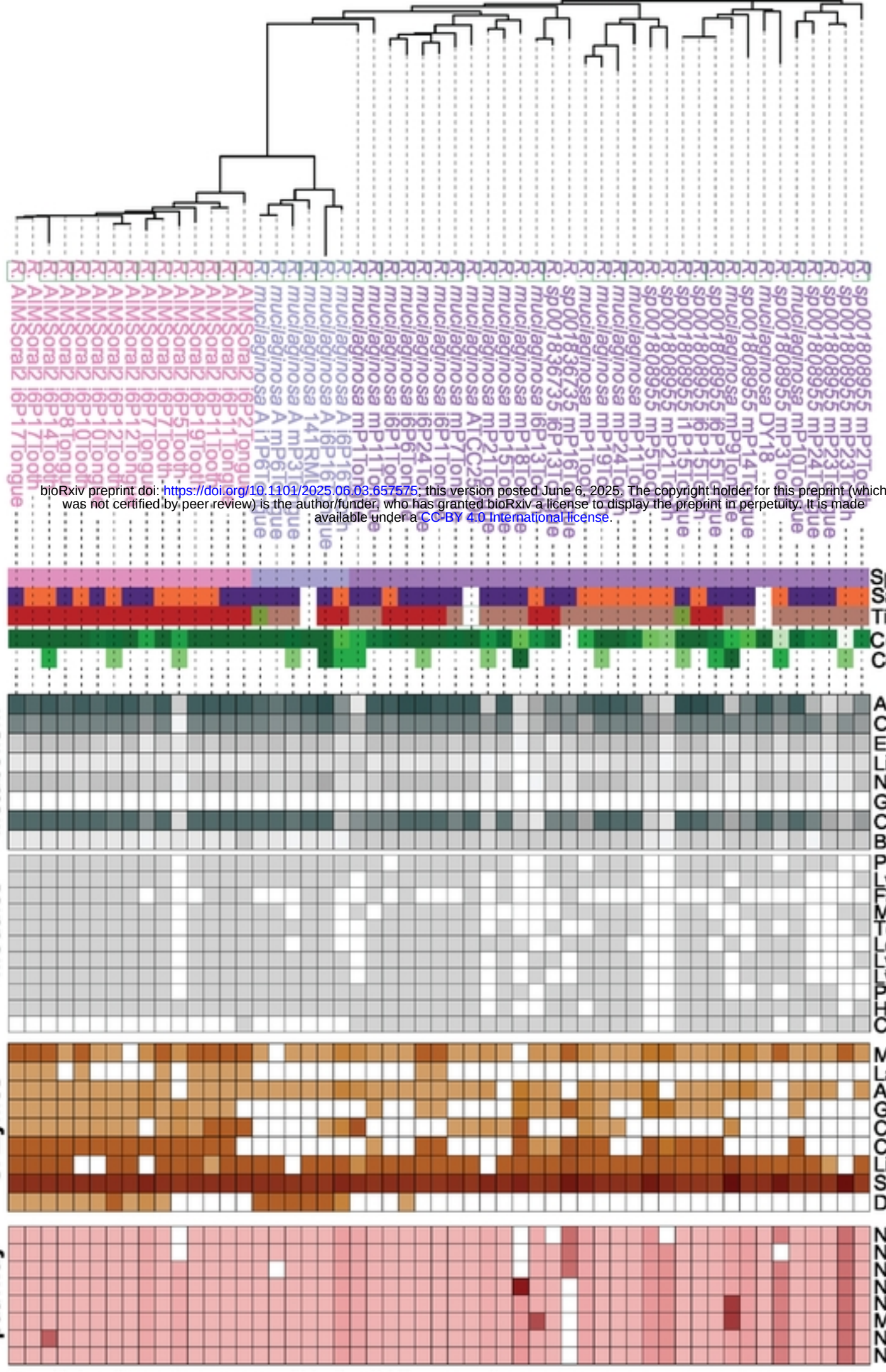


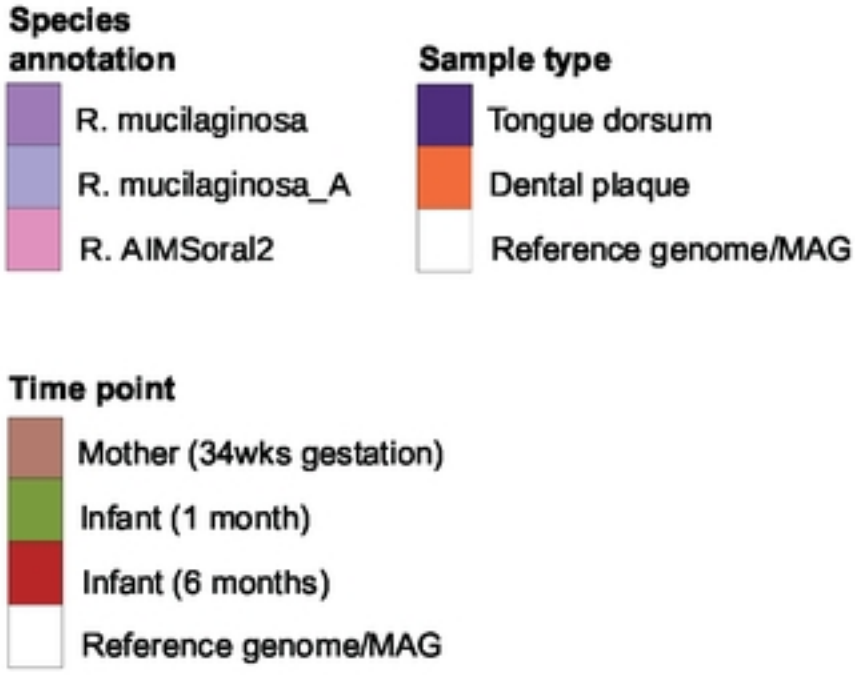


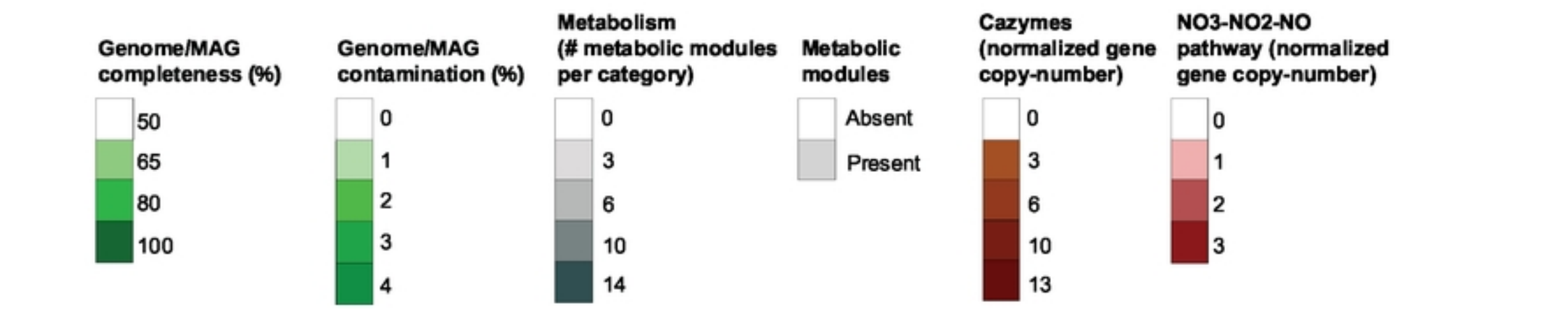


Figure

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.657575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

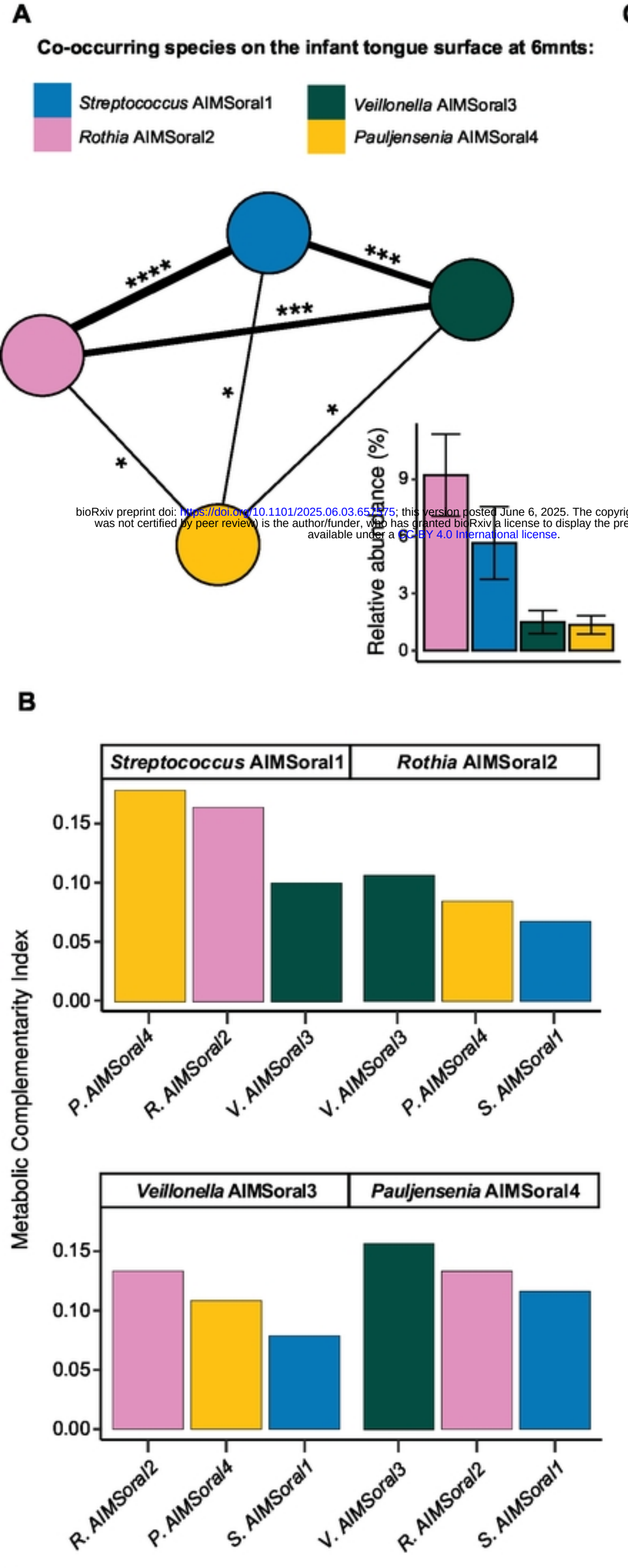






Figure

bioRxiv preprint doi: https://doi.org/10.1101/2025.06.03.652575; this version posted June 6, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.



Figure