Pathobiont-mediated spatial structuring enhances biofilm virulence in childhood oral disease

Hunyong Cho
University of North Carolina-Chapel Hill
Zhi Ren
University of Pennsylvania https://orcid.org/0000-0002-3553-1958
Kimon Divaris (Kimon_Divaris@unc.edu)
University of North Carolina-Chapel Hill https://orcid.org/0000-0003-1290-7251
Jeff Roach
University of North Carolina-Chapel Hill
Bridget Lin
University of North Carolina-Chapel Hill
Chuwen Lin
University of North Carolina-Chapel Hill
Andrea Azcarate-Peril
University of North Carolina
Miguel Simancas-Pallares
University of North Carolina-Chapel Hill
Poojan Shrestha
University of North Carolina-Chapel Hill
Alena Orlenko
Cedars-Sinai
Jeannie Ginnis
University of North Carolina-Chapel Hill
Kari North
UNC Chapel Hill https://orcid.org/0000-0002-8903-0366
Andrea Ferreira Zandona
Tufts University, School of Dental Medicine
Apoena Ribeiro
University of North Carolina-Chapel Hill
Di Wu
University of North Carolina at Chapel Hill https://orcid.org/0000-0001-8331-2357
Hyun Koo
University of Pennsylvania https://orcid.org/0000-0001-9143-2076

Article

Keywords:

Posted Date: June 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1748651/v1

Version of Record: A version of this preprint was published at Nature Communications on May 22nd, 2023. See the published version at https://doi.org/10.1038/s41467-023-38346-3.

Abstract

Microbiome studies are revealing complex microbiota in biofilm-mediated human diseases commonly linked with specific bacterial pathogens. *Streptococcus mutans* has been implicated as the primary pathogen in childhood dental caries (tooth decay). While the role of polymicrobial communities is appreciated, it remains unclear whether other microorganisms are active contributors, inactive cohabitants, or interact with known pathogens such as *S. mutans*. Here, we integrate multi-omics of human dental plaque (biofilm) from two community-based samples of preschool-age children in a discovery-validation pipeline involving bioinformatics, laboratory, and *in vivo* experimental approaches to identify disease-relevant inter-species interactions. In metagenomics and metatranscriptomics analyses among 416 preschool-age children, we identify 16 taxa strongly associated with childhood caries. Using multiscale imaging, virulence assays, microcalorimetry, and computational analyses, we investigate biofilm formation dynamics, microscale spatial arrangement, and metabolic activity by *Selenomonas sputigena, Prevotella salivae* and *Leptotrichia wadei* either individually or with *S. mutans*. Notably, we discover that the flagellated *S. sputigena*, a subgingival anaerobe with previously unknown role in supragingival biofilm virulence, becomes trapped in foreign streptococcal exoglucans, loses its motility but actively proliferates to build a honeycomb-like multicellular superstructure encapsulating *S. mutans* and enhances acidogenesis. Rodent model experiments reveal a previously unrecognized ability of *S. sputigena* to colonize supragingival tooth surfaces. While incapable of causing caries on its own, *S. sputigena* exacerbates the disease severity *in vivo* when co-infected with *S. mutans*, causing extensive lesions on tooth enamel. Our data reveal a pathobiont from a disparate habitat in a prevalent disease that cooperates with a known pathogen to build a unique 3D spatial structure and heighten biofilm virulence—which may be relevant to other polymicrobial diseases.

Introduction

Biofilm-forming pathogens have been implicated in a myriad of human infectious diseases and contamination of biomedical devices and implants[1–3]. Although established pathogens display emergent properties that facilitate biofilm lifestyle and virulence expression, they often reside in polymicrobial communities with increasing evidence of interspecies interactions[4,5]. Among various microbiomes, the human plaque biofilm communities harbor diverse microbiota at oral mucosal barrier and mineralized dental surfaces that play key roles modulating health and disease yet remain underappreciated[6]. Although microbiome-wide association studies have revealed novel microbial species implicated in oral diseases, most evidence has been generated by small samples of participants recruited from clinics or undergoing dental treatment, whereas the identified species' causal roles remain largely undefined. Comprehensive investigations of the taxonomic and functional aspects of microbial interactions in oral biofilms among community-based samples are warranted. Such studies may not only reveal additional pathogens with population relevance but can also shed light on novel pathogenetic mechanisms and interspecies interactions.

Dental caries (tooth decay) is a widespread, biofilm-mediated, and diet-modulated disease that affects approximately 600 million children worldwide and remains a major unresolved public health problem[7]. The disease has a complex etiology, but it is generally understood as a host-diet-dependent process of dental tissue demineralization that relies on dysbiotic community and polymicrobial acidogenesis[4,8]. *Streptococcus mutans* is a gram-positive, biofilm-forming, acidogenic and aciduric bacterium that has been clinically associated with childhood caries[9], has been shown to cause caries in animal models[10], and has been established as a keystone oral pathogen[11]. Although other microbial species on tooth surfaces have been associated with dental caries[12–15], it remains unclear whether they are active contributors, inactive cohabitants, or they interact with *S. mutans* as pathobionts[16] to promote disease development.

Most evidence to-date on the microbial basis of dental caries has been generated from targeted, culture-based, or 16S ribosomal gene-based methods. While providing foundational information, this knowledge base can be substantially augmented using DNAseq-based metagenomics (MTG), as well as RNAseq-based metatranscriptomics (MTX), to gain deeper insights into microbial taxonomy and functional activity. The few studies employing next-generation sequencing have been done in small clinical samples[17,18], and with little or no mechanistic validation. Meanwhile, advances in imaging technology have revealed spatially structured communities in multispecies biofilms, warranting better understanding of the spatial arrangement and positioning or biogeography of the interacting microbes[19,20]. To overcome these limitations and gain fundamental knowledge, we have developed a multi-modality, discovery-validation pipeline integrating clinical data from community-based studies and informatics discovery with *in vitro* and *in vivo* experimental models to study disease-associated microbial taxa and their interactions (Fig. 1 top panel).

Our data highlight *Selenomonas sputigena, Prevotella salivae* and *Leptotrichia wadei* as species with previously unrecognized roles in supragingival biofilms, with functional repertoires and interactions relevant to the pathogenesis of a prevalent childhood dysbiotic disease. We unexpectedly find that the flagellated, gram-negative anaerobic bacterium *S. sputigena*, commonly reported as member of the periodontal niche microbiome[21], has a significant role in the supragingival biofilm microbiome and is strongly associated with childhood dental caries. Here, we show that the motile *S. sputigena* becomes trapped into extracellular glucan-matrix produced by *S. mutans*, loses its motility, but rapidly proliferates to form a multicellular honeycomb-like scaffold that wraps the streptococcal cell clusters and results in enhanced acid production—a key virulence factor in dental caries. We verify experimentally that, although *S. sputigena* alone does not cause caries, it significantly increases disease severity when co-infected with *S. mutans* causing extensive cavitation of tooth enamel in a rodent model. Taken together, our findings reveal a novel pathobiont capable of unique spatial structuring and interspecies cooperation that enhances biofilm virulence—a phenomenon that substantially improves our understanding of childhood dental disease pathogenesis and may have implications for established pathogen and pathobiont interactions in other polymicrobial infections.

Results

Established and novel taxa are associated with childhood caries. We carried out comprehensive taxonomic association analyses in MTG (i.e., whole genome sequencing shotgun) and MTX (i.e., RNAseq) data and identified 16 bacterial species significantly associated with clinically measured childhood caries experience. We used strict, multiple testing-controlled, across-trait criteria to identify this set of 16 disease-associated species; i.e., we required taxa to be FDR-significant for quantitative disease measured both locally (i.e., number of tooth surfaces with caries experience considering tooth surfaces where plaque was harvested from) and at the person-level (i.e., number of tooth surfaces with caries experience considering the entire dentition), in DNA and RNA data, in the discovery sample set (n = 300), and with evidence of replication in an independent sample of similarly-aged children (n = 116, Fig. 2 and Extended Data Table 1). *S. mutans* ($p = 2.6x10^{-6}$) emerged strongly associated with caries experience as the known and expected suspect. Other significantly associated taxa included *P. salivae* ($p = 8.5x10^{-6}$ in the association of DNA-based species taxonomic abundance with localized caries experience), *S. sputigena* ($p = 3.1x10^{-6}$), *L. wadei* ($p = 2.2x10^{-7}$), *Veillonella atypica* ($p = 1.1x10^{-5}$), *Lachnospiraceae bacterium oral taxon 082* ($p = 1.1x10^{-4}$), *Stomatobaculum longum* ($p = 1.6x10^{-3}$), *Lachnoanaerobaculum saburreum* ($p = 9.6x10^{-7}$), and *Centipeda periodontii* ($p = 3.8x10^{-3}$). These associations were consistent in MTG and MTX data (Extended Data Fig. 1), and, as expected, there was a high degree of correlation between the abundance of taxonomically neighboring significant taxa in MTG and MTX (e.g., *Leptotrichia spp.* and *Prevotella spp.*) (Fig. 3). *S. mutans*, being the only *Streptococcus* among the 16 significant species, showed the smallest correlations with the other 15 taxa. For example, *S. mutans'* mean correlations in MTX were 0.09 both in health in disease, whereas *L. wadei's* were 0.50 and 0.49, and *S. sputigena's* 0.45 and 0.38, respectively.

Glycolysis is upregulated in biofilm transcriptomes where novel taxa are implicated. Four pathways emerged as significantly associated with caries experience in MTX: glycolysis IV, pyruvate fermentation to acetate and lactate II, chorismate biosynthesis from 3-dehydroquinate, and UDP-N-acetyl-D-glucosamine biosynthesis I (Extended Data Table 2). Of note, glycolysis IV, one of the most abundant pathways (ranked #3 overall) in expression abundance and an important pathway for acidogenesis implicated in disease, involves *S. sputigena*, *L. wadei* and 4 more significant species. In fact, 7 of these 16 significant species (i.e., *S. mutans*, *P. oulorum, S. sputigena, V. atypica, L. wadei, P. melaninogenica, L. saburreum, and C. periodontii*) were involved in these differentially expressed pathways. In addition, lactose and galactose degradation I, the pathway with the highest proportional representation of significant species (i.e., *L. wadei* and *S. mutans*) was also upregulated in disease (Extended Data Fig. 2).

Three new candidate pathogens nominated for virulence and biofilm studies. Next, we sought to prioritize a subset of species that was practically feasible to carry forward into the multi-modal validation pipeline (Fig. 1 steps 6–8). To select this shortlist of candidates that could be fully characterized in virulence, biofilm, and *in vivo* experiments in this study, we considered statistical evidence of association in the discovery sample (i.e., p-value), evidence independent replication, representation of different genera (i.e.,

one species per genus), and availability of clinical isolates. Based on these criteria, from the list of 16 significant species we prioritized *S. mutans* (the known, well-established pathogen), and *S. sputigena*, *P. salivae*, and *L. wadei* (as new candidates). These four taxa, hereafter referred to as "top species", were carried forward to *in vitro* virulence assessments and biofilm characterization and served as candidates for further *in vivo* colonization and pathogenicity studies.

Interspecies metabolic interactions and acidogenesis. To form the acidic microenvironment that causes demineralization of tooth surfaces, bacterial pathogens within the biofilm microbiota must efficiently metabolize dietary sugars such as sucrose (the primary sugar associated with dental decay) and produce acids (acidogenicity) while surviving in an acidified milieu (aciduricity)[22]. We first sought to determine whether the top species fit the pathogenic profile of being both acidogenic and aciduric.

Glycolytic pH-drop assays for each species individually (Fig. 4-A1) and the new candidates combined with the established *S. mutans* (Fig. 4-A2) showed that all were active acid producers in the presence of sucrose, lowered pH to highly acidic values, i.e., pH 4.3 to 5.5 which can cause enamel demineralization[23], consistent with up-regulation of glycolysis in the disease-associated plaque biofilm transcriptome. *S. mutans* was the most acidogenic, followed by *L. wadei*, *P. salivae*, and *S. sputigena* (Fig. 4-A1). Intriguingly, the combination of *S. mutans* and *S. sputigena* resulted in the highest acid production rate, significantly higher than *S. mutans* alone (Fig. 4-A2) and the most rapid pH drop. These findings are suggestive of a possible co-metabolic relationship between *S. mutans* and *S. sputigena*.

All 4 species could grow in an acidic culture medium (i.e., pH pre-adjusted to 5.5), demonstrative of their ability to tolerate acidic conditions required for the development of caries lesions. Acid tolerance challenges of monocultures showed that all species could grow at an initial pH of 5.5, and further decreased pH to 4.8–4.6 (Fig. 4-B1). Notably, acid tolerance was higher when *S. sputigena*, *L. wadei*, and *P. salivae* were in co-cultures with *S. mutans*, demonstrating ability to grow at a starting pH of 5.2 and further decrease the pH to a range between 3.9 to 4.2 after 48 hours (Fig. 4-B2). Of note, the mixed culture of *S. sputigena* and *S. mutans* achieved the lowest final pH of 3.9 after 48 hours.

S. mutans dominates metabolic activity in top species' mixed cultures. Next, we used real-time isothermal microcalorimetry to characterize top species' sucrose metabolic activity, individually and when the three new candidates were co-cultured with *S. mutans*. In monocultures, *S. mutans* produced the highest peak and rate of metabolic activity, the fastest times to activity and peak, but also the shortest decay time, reflecting a rapid drop in its metabolic activity (Fig. 4-C1). In contrast, *S. sputigena*'s metabolic activity peaked much later (i.e., at 49 hours) and decayed slowly, findings aligned with being a slow growing anaerobic bacterium[24], whereas *L. wadei* presented the second highest peak. Co-cultures of *S. sputigena*, *L. wadei* and *P. salivae* with *S. mutans* had metabolic peaks close to the *S. mutans'* monoculture peak (Fig. 4-C2) indicating the metabolic dominance by the pathogen within co-cultures.

In sum, we found that *S. sputigena*, *L. wadei* and *P. salivae*, alongside the known *S. mutans*, display aciduric and acidogenic properties. All species can metabolize sucrose, the primary disease-associated

dietary sugar, produce acids in the glycolytic assay, and survive in the resulting acidic pH. While the metabolic activity is driven by *S. mutans*, we unexpectedly found that the addition of *S. sputigena* results in faster acid production and increased aciduricity, significantly higher than *S. mutans* alone. This suggests a possible cooperative relationship between *S. mutans* and *S. sputigena* that favors the formation of pathogenic biofilms.

S. mutans enhances S. sputigena colonization and biofilm formation. Dental caries is a prime example of a biofilm-mediated disease that is initiated by colonization of microbes on tooth surfaces and development of structured biofilm communities[25]. We developed an experimental model using saliva-coated hydroxyapatite (sHA, a tooth enamel surrogate) and confocal live-cell imaging to investigate the four top species' biofilm-forming abilities. Because *S. mutans* is a well-characterized pathogen known for its exceptional ability to assemble biofilms on teeth[26], we focused on investigating biofilm forming properties of *L. wadei*, *P. salivae*, and *S. sputigena* alone and in combination with *S. mutans.* We found that the new candidates could co-colonize the sHA surface, with *S. mutans* forming structured but morphologically distinct mixed-species biofilms (Fig. 5A upper panel). In contrast, none of the new species alone could efficiently colonize the surface and develop biofilms, with only single cells (*L. wadei* and *P. salivae*) or small bacterial clusters (*S. sputigena*) on the surface (Fig. 5B). Higher abundances of each of the new species were noted when grown as mixed biofilms with *S. mutans* (Fig. 5A lower panel). These findings were further corroborated by quantifying the surface-bound cell volume (Fig. 5A lower panel and Fig. 5B) using a computational image analysis toolbox optimized for biofilms[27]. Indeed, we confirmed higher amounts of *L. wadei*, *P. salivae*, and *S. sputigena* on the hydroxyapatite surface when co-cultured with *S. mutans* compared to the same species alone (Fig. 5C), indicating that the presence of *S. mutans* can facilitate their surface colonization. Notably, we found that surface-colonized *S. sputigena* displayed the highest (> 10-fold) increase when co-cultured with *S. mutans*, which was higher than that of *L. wadei* or *P. salivae*, suggesting that interspecies interactions may significantly enhance the colonization and biofilm formation by *S. sputigena*.

Interspecies spatial structuring and co-localization within biofilms. Given that *S. mutans* produces extracellular polysaccharides (EPS) that enhance bacterial cell co-adhesion and biofilm accumulation[28], we investigated the spatial localization of the three new candidate species, *S. mutans*, and EPS. High-resolution images of magnified areas showed that most of *L. wadei*, *P. salivae*, and *S. sputigena* cells were spatially located in-between streptococcal clusters (known as microcolonies), rather than intermixing with the *S. mutans* cells within (Fig. 5D, upper panel). Interestingly, we found a densely packed accumulation of *S. sputigena*, whereas *L. wadei* or *P. salivae* sparsely colonized these areas. Using a fluorescent marker that specifically labels *S. mutans*-derived EPS α-glucans[26], we found that most of *L. wadei*, *P. salivae*, or *S. sputigena* cells were co-localized with the extracellular polymeric matrix (Fig. 5D, lower panel).

We performed three-dimensional analyses to characterize the co-localization of two fluorescent signals in relation to each other using Mander's overlap coefficient, a quantitative measure of spatial proximity[29].

For each mixed-species biofilm, we calculated two Manders' coefficients: 1) between each new species and *S. mutans* and 2) between each new species and EPS. We found that *L. wadei* and *S. sputigena* were more proximal to EPS than *S. mutans* (Fig. 5E, left and right panels), indicating that these species may be primarily associated with EPS than to *S. mutans* cells. In contrast, *P. salivae* showed similar proximity to *S. mutans* cells and EPS (Fig. 5E, middle panel). In sum, our findings show that the new candidates form mixed biofilms with *S. mutans*, and this coexistence enhances their surface colonization and co-development of structured biofilms. *S. sputigena* forms a densely populated community interspersed between *S. mutans* clusters, leading to highly cohesive, co-assembled biofilms.

The motile bacterium S. sputigena becomes trapped in streptococcal exoglucans. Motivated by the enhanced surface colonization of *S. sputigena* in the presence of *S. mutans* and the enhanced acidogenicity of this species combination, we sought to characterize the spatial structuring of *S. sputigena* within the mixed-species biofilm. As a backdrop, *S. sputigena* was originally found in the aerodigestive system, the subgingival microbiota of patients affected with periodontitis[30–32], and endodontic infections[33]. *Selenomonas* multispecies communities were common (i.e., among the top 10 abundant taxa) in the biofilms obtained in our recent in vivo experimental study of caries progression and arrest[34]. Its role in supragingival biofilms and how it interacts with other microbes remain unknown.

Unlike most oral microbes, *S. sputigena* has a flagellum, a surface-attached appendage that allows motility in liquid environments. Unexpectedly, we found that *S. sputigena* cells were motile even after their colonization on the sHA, displaying a tumbling, multi-directional motion (Supplemental Material Movie S1) while remaining attached to the hydroxyapatite surface. We used real-time live imaging and computational motion tracking to study the surface motility of *S. sputigena* cells when co-cultured with *S. mutans* (Fig. 6A, Fig. 6B and Supplemental Material Movie S2). The spatial coordinates of individual *S. sputigena* cells were followed at each time frame and a time-resolved trajectory was generated (representative frames shown in Fig. 6B). Our results revealed distinctive motility behaviors of *S. sputigena* cells depending on their spatial location relative to that of *S. mutans*. The upper panels of Fig. 6C (same areas as "Box a", in Fig. 6A) show that *S. sputigena* cells near *S. mutans* cells display no motility (Fig. 6C top). In contrast, *S. sputigena* cells located far away from *S. mutans* remained motile (lower panels of Fig. 6C; same areas as "Box b" in Fig. 6A). We also calculated accumulated displacements (total path length) of individual *S. sputigena* cells relative to their original position. The displacement curves revealed that most *S. sputigena* cells adjacent to *S. mutans* did not move or only moved for short distances (Fig. 6D, left panel), whereas those with no *S. mutans* nearby moved actively for longer distances (Fig. 6D, right panel).

Given the contrasting motility between *S. sputigena* cells at different spatial locations, we hypothesized that *S. sputigena* cells that lost their surface motility might be physically trapped in *S. mutans*-derived EPS. We first tracked *S. sputigena* cells that were co-localized with EPS α-glucans and found that the cells were devoid of mobility (Fig. 6E, upper panels). Then, we performed EPS degradation using glucanohydrolases (mutanase and dextranase) that specifically hydrolyze α-glucans without antibacterial activity[35] and examined the motility behavior of *S. sputigena* in real-time. We found that the surface

motility of *S. sputigena* was recovered following α-glucans degradation (trajectories shown in Fig. 6E, lower panels). These observations were confirmed by computational motility tracking before and after glucanohydrolase treatment (Fig. 6F), indicating that EPS degradation released trapped *S. sputigena* cells restoring their surface motility.

By co-colonizing the surface with *S. mutans*, *S. sputigena* cells appear to lose motility and become immobilized through interspecies cell-glucan matrix interactions yet remain viable and active, which may enhance surface binding and accumulation of these motile bacteria that could influence the scaffolding of the mixed biofilm community. Our biofilm MTX data corroborate these observations. First, we found that *S. mutans' gtfC* (glucosyltransferase GtfC, involved in glucan synthesis) and *glucose-1-phosphate adenylyltransferase* genes were significantly upregulated in disease (Extended Data Table 3). *gtfC* encodes an enzyme that synthesizes both soluble and insoluble α-glucans[28] that are highly susceptible to degradation by dextranase (breaks down soluble glucans) and mutanase (digests insoluble glucans) which were used in this study's motility experiments. Second, we found several negative *S mutans* and *S. sputigena* gene-gene expression interactions. Intriguingly, 11 of 13 gene-gene interactions were negative, and *glucose-1-phosphate adenylyltransferase* demonstrated a strong negative interaction ($p = 3.7x10^{-6}$) with *S. sputigena's* motility-related gene *flagellin* (Extended Data Table 4).

Spatial arrangement of S. sputigena and S. mutans within intact biofilm structure. Considering that *S. mutans* can significantly enhance *S. sputigena* colonization to the tooth-mimetic surface and modulate its surface motility, we investigated how this co-colonizing community mediates biofilm spatial structure (biogeography) at multi-length scales. To this end, we used super-resolution confocal microscopy coupled with computational analyses for mixed biofilm communities[36]. We found a unique 3D multicellular organization composed of densely packed *S. sputigena* cells encasing *S. mutans* microcolonies (Fig. 7A). Using layer-by-layer orthogonal imaging sectioning, we found an orderly arrangement of *S. sputigena* cells across the entire height of the biostructure (Fig. 7B). A representative sectional rendering (z = 15 μm from the surface) of the 3D biostructure revealed a spatially segregated inner core formed primarily by *S. mutans*, and a dense ring-like outer layer of *S. sputigena*, both associated with EPS α-glucans (Fig. 7C). We investigated the composition and the spatial structuring of the microbial and EPS components in relation to the center of mass (referred to as centroid) of the biostructure (see diagram in Fig. 7D). We found that the core near the centroid consisted predominantly of *S. mutans*, whereas the periphery harbored predominantly *S. sputigena* (Fig. 7E, left and middle panels). Notably, *S. mutans*-derived EPS α-glucans were detected throughout the entire biostructure, including the peripheral areas where the abundance of *S. mutans* cells was low (Fig. 7E, right panel), suggesting that the secreted EPS could mediate co-adhesion without direct cellular contact providing scaffolding for the interspecies assembly.

To further assess the structural organization of the biofilm, we performed 3D reconstruction of the confocal images and quantitative analysis of the spatial distribution of the bacterial cells. This analysis revealed a unique honeycomb-like architecture consisting of *S. sputigena* cells forming a multicellular superstructure (Fig. 7F and Supplemental Material Movie S3). To determine the spatial distribution of

bacterial cells, we computationally dissected the large mixed-species biostructure into small cubic volumes in micron scale, and calculated the relative abundance of (i.e., the proportion of the total cellular volume occupied by) *S. sputigena* cells within each cube in correspondence to their 3D spatial coordinates[37]. Using this method which provides local distribution with 3D spatial resolution, we found high *S. sputigena* abundance near the entire outer surface of the biostructure (selected horizontal or vertical planes shown in Fig. 7G). These results indicate a non-random pattern with densely packed yet spatially segregated biogeography between *S. sputigena* and *S. mutans* cells. Taken together, we find evidence for a motile bacterium, *S. sputigena*, becoming immobilized by a foreign EPS produced by a co-colonizing species (*S. mutans*), promoting localized growth and accumulation, and forming a distinctive multicellular honeycomb superstructure guided by the matrix scaffolding.

S. sputigena can colonize tooth surfaces and increases caries experience in vivo. The *in vitro* experimental data presented thus far demonstrate that *S. sputigena* can bind and form highly structured, densely packed biofilms with *S. mutans* on apatitic surface—their combination produces more acid than either species alone. It is thus logical to expect that they can co-colonize tooth surfaces and exhibit enhanced biofilm virulence *in vivo*. To experimentally test this hypothesis, we sought to determine whether *S. sputigena* alone or when co-infected with *S. mutans* (Fig. 8A) can colonize tooth surfaces and cause tooth decay based on an established rodent model that mimics the characteristics of early childhood caries, including exposure to a sugar-rich diet and rampant dental cavitation[38] (Fig. 8B). We found that the animals were infected by *S. mutans*, *S. sputigena*, or both and remained persistently infected (Fig. 8C) as determined by real-time PCR using species-specific probes. The uninfected (control) animals remained free of infection by *S. mutans* or *S. sputigena* (Fig. 8C). All animals maintained good health showing a steady body weight gain with no significant difference between groups (Fig. 8D). These data confirm the ability of *S. sputigena* to co-colonize tooth enamel surfaces with *S. mutans in vivo*.

The impact on caries development and severity was then assessed for each experimental condition. In this model, tooth enamel progressively develops caries lesions (analogous to those observed in humans), proceeding from initial areas of demineralization to moderate lesions and on to extensive (i.e., severe) lesions characterized by enamel structure damage and cavitation. Consistent with previous studies[39,40], *S. mutans* infection led to the onset of dental caries whereas control uninfected animals harboring their natural oral microbiota developed minor tooth demineralization with no severe lesions (Fig. 8E). Infection with *S. sputigena* did not induce development of caries lesions on the tooth surface compared to the uninfected control (Fig. 8E) despite confirmed bacterial colonization, suggesting that, alone, *S. sputigena* has limited cariogenic potential. However, *S. sputigena* co-infected with *S. mutans* resulted in a significant increase of both total caries experience and the severity of caries lesions compared to those infected with either species alone (Fig. 8E), indicating enhanced virulence. Notably, we found that co-infection produced more severe lesions (Fig. 8F), characterized by enamel destruction and frank cavities (Fig. 8B). In sum, we show that *S. sputigena* interacts with *S. mutans* for enhanced co-colonization and the presence of *S. sputigena* leads to the assembly of uniquely structured biofilms with enhanced virulence

that cause the onset of severe dental disease *in vivo*, suggesting a new pathobiont and interspecies spatial structuring to promote disease-causing condition.

Discussion

In this study, we integrated multi-omics of human plaque samples from two community-based clinical cohorts in a discovery-validation pipeline combining informatics, laboratory, and *in vivo* experimental models to identify pathogen candidates and obtain insights into how these previously uncharacterized species interact and form virulent oral biofilms (Fig. 9). Comprehensive taxonomic association analyses and multiple testing-controlled, across-clinical trait criteria identified 16 species significantly associated with early childhood caries, including several new candidates. Among them, *S. mutans* (an established pathogen), and *S, sputigena*, *P. salivae*, and *L. wadei* (as new candidates) were strongly associated with disease and were carried forward into comprehensive biofilm studies. We found that the flagellated and motile *S. sputigena* develops a cooperative interaction with *S. mutans* to exacerbate biofilm virulence. Specifically, *S. sputigena* and *S. mutans* can co-colonize apatitic surfaces, metabolize sugars, and produce and tolerate acids. Together, the two species contribute to significantly enhanced acidogenesis and aciduricity. This interaction is underlaid by *S. sputigena*'s entrapment in streptococcal exoglucans, wherein it loses motility, and builds a densely packed honeycomb-like multicellular structure. Negative interactions between *S. mutans*'s *glucose-1-phosphate adenylyltransferase* and several *S. sputigena*'s genes including *flagellin*[12] in MTX, corroborate the inter-species interactions manifested in the experimental studies. Crucially, while incapable of causing disease on its own, *S. sputigena* significantly increases the biofilm virulence *in vivo* when co-infected with *S. mutans* and augments the severity of dental caries. *S. sputigena* has so far been implicated in periodontal disease[42,43], endodontic infections[33], and is a member of aerodigestive tract microbiome with emerging prognostic significance in cancer[44,45]. Our findings demonstrate that *S. sputigena* acts as a pathobiont outside its typical habitat to modulate biogeography (i.e., spatial structure), metabolic activity, and pathogenicity of supragingival biofilms in the context of a prevalent childhood disease.

We leveraged robust clinical and multi-omics data generated among sizeable community-based samples of young children and used conservative selection criteria to identify bacterial species associated with disease experience and guide the prioritization of candidates for characterization and experimental validation. The large sample size for a multi-omics study of this kind and the application of strict discovery and validation criteria led to a set of high-confidence species associated with caries experience, including known and novel taxa. While we were able to comprehensively test four of these 16 species, we anticipate that this set of high-confidence candidates will be further studied in the near future by us and others. Biofilm transcriptomics analyses affirmed significantly upregulated metabolic pathways (e.g., glycolysis, pyruvate fermentation, and UDP-N-acetylglucosamine) and bacterial genes (e.g., glucose-1-phosphate adenylyltransferase) in disease. These findings are aligned with metabolic interactions between microbiota and dietary sugars expected to be found in dental caries[22]. Sugars fuel the emergence of dysbiosis by favoring proliferation of species that can adapt to ecological changes and

promote acidification of the biofilm microenvironment that underlies destruction of mineralized dental tissues. Enhanced acidogenesis, up-regulation of UDP-N-acetyl-D-glucosamine biosynthesis I, and gene-gene interactions in MTX between the two species strengthen the case for *S. sputigena* being an important species in the caries-associated dental plaque biofilm[34].

A key finding of the present report is that the cooperative interaction between *S. sputigena* and *S. mutans* hinges upon an unexpected phenomenon wherein a motile bacterium that is typically reported as resident of the subgingival (i.e., below gum line) niche is entrapped in an exopolymeric matrix produced by another species residing in the supragingival (i.e., above gum line) area and forms a unique spatial biofilm structure that increases virulence. We show that the dense growth of immobilized *S. sputigena* cells colonizing the apatite surface and surrounding *S. mutans* cell clusters create a cohesively packed community superstructure that cannot be achieved by either species alone. Importantly, we demonstrate that the immobilization of *S. sputigena* and the spatial structuring of the mixed community are associated with *S. mutans*-derived EPS α-glucans. The exact mechanisms by which *S. sputigena* loses motility and forms the densely packed honeycomb structures remain unknown. *S. sputigena* expresses a heavily glycosylated flagellin, building blocks of flagella[41]. It is possible that the glycans on the surface of *S. sputigena* display glucan-binding properties, interacting with streptococcal exoglucans, and thereby immobilize the cell. Additionally, fucose is a major monosaccharide constituent of the O-linked glycans in *S. sputigena* flagellin[41], and our recent metabolomics work among the same discovery sample has demonstrated elevated fucose abundance in early childhood caries[46]. Glucosyltransferases (Gtfs) secreted by *S. mutans* can bind to sugar moieties of different polysaccharides on the surface of other microbes and produce EPS α-glucans *in situ*[47]. It is possible that Gtf can bind to fucose moiety allowing glucan synthesis *in situ* that could contribute to immobilization. Future studies need to investigate the interactions of Gtf with *S. sputigena*, glucan-fucose interactions, and determine whether these interactions are strain-dependent.

S. sputigena has been recalcitrant to genetic manipulation which hinders in-depth molecular studies to further understand its pathogenic roles, such as the functions of flagellin on biofilm formation with other species. However, new microbial genetic engineering methodologies involving restriction modification-silent tools[48] may circumvent this limitation. Aside from *S. sputigena*, our work nominates several new species that are prime targets for additional studies. For example, *L. wadei* and *P. salivae*, both strongly associated with caries experience in our discovery and replication cohorts, may play pathogenic roles in the context of polymicrobial biofilms that are not necessarily associated with *S. mutans* and could be investigated in the future. Another prime target for future studies is *Lachnoanaerobaculum saburreum*, a species that is capable of producing acid from glucose, lactose, and sucrose among other sugars[49] and is involved in pyruvate fermentation to acetate and lactate II—one of the four pathways that we found as significantly differentially expressed in disease. Examination of multi-species combinations of these candidate species may also provide additional insights about the biofilm virulence mechanisms.

A limitation of our current multi-omics pipeline is that it is not optimized for studying inter-kingdom interactions (i.e., viruses and fungi, besides bacteria). For example, the nucleic extraction protocols were not optimal for detecting and analyzing fungi, which are important in the context of childhood caries[50]. This can be overcome with the incorporation of nucleic acid extraction protocols for downstream fungal analyses. At the same time, detection of phages in MTG and MTX data is a rapidly growing area that is likely to provide high yields in the near future[51]. Another limitation is the cross-sectional nature of our human observational data, which limits the samples' inferential potential regarding temporality and causality—it is likely that some taxonomic and functional changes found in the initial discovery of candidate species in human plaque samples are downstream (i.e., consequences) of established disease. This is where the multi-modality aspect of our study is of the essence—the experimental studies of the pipeline are well positioned to help clarify the actual pathogenic potential of candidate species. Nevertheless, community-based longitudinal human studies are warranted to enhance both discovery and validation of new species, not only in the presence of dental caries but also in its onset and progression.

In sum, we employed a multimodal pipeline to gain new knowledge about taxonomic and functional features of the childhood oral microbiome in health and disease from two sizeable discovery and replication community-based samples of human plaque biofilms. We discover a novel inter-species interaction and unique spatial structuring at microscale wherein a motile flagellated species becomes immobilized in the EPS matrix produced by a disease-causing species and proliferates to build a 3D multicellular superstructure with enhanced acidogenesis. We show that the interaction between *S. sputigena* and *S. mutans* augments the severity of dental caries *in vivo*, suggesting a new pathobiont exacerbating biofilm virulence for a common yet unresolved disease. Further understanding of the spatial structuring function and pathobiont-mediated virulence may reveal new mechanisms of biofilm assembly and therapeutic targets, which may be relevant to other polymicrobial infections where other species are interacting with known pathogens in complex biofilms.

Declarations

Acknowledgements

This work was supported by research grants from the National Institutes of Health: National Center for Advancing Translational Sciences National UL1TR001111 (K.D.); National Institute for Dental and Craniofacial Research U01DE025046 (K.D.), R01DE025220 (H.K.), and R03DE028983 (D.W.); Z. R. is supported by the NIDCR Postdoctoral Training Program under award number R90DE031532. Z. R. is a recipient of the Colgate–Palmolive Fellowship. The UNC Microbiome Core is funded in part by the Center for Gastrointestinal Biology and Disease (CGIBD P30 DK034987) and the UNC Nutrition Obesity Research Center (NORC P30 DK056350). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

The authors thank Dr. Patricia V. Basta and her team at the UNC-Chapel Hill Biospecimen Processing facility for the accessioning, storage, and disbursement of the supragingival biofilm microbiome

specimens in the ZOE studies. The authors are thankful to Dr. Justin Merritt at Oregon Health & Science University for kindly providing the *S. mutans* UA159 Ef-Tu-ngGFP strain. We also thank Drs. Lamprini Karygianni and Thomas Thurnheer at University of Zurich for their kind gift of the *S. sputigena* ATCC 35185 strain.

Data availability

Microbiome taxonomy (MTG and MTX) and pathway information for the entire biofilm microbial community, as well as targeted MTX data for the four top species used in this study have been deposited and are freely accessible alongside metadata (i.e., demographic and clinical phenotype information) via the Carolina Digital Repository and accession number 5d86p890x: https://cdr.lib.unc.edu/concern/data_sets/5d86p890x Raw sequence data for ZOE 2.0 have been deposited as part of dbGaP accession phs002232.v1.p1 and are pending release. ZOE pilot study sequence data have been made available via the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) and BioProject accession number: PRJNA843091.

Author Contributions

K.D., K.E.N., A.A.R., D.W. and H.K. designed the study. K.D., M.A.S.P, P.S., J.G. and A.G.F.G participated in data collection and sampling. J.R. and M.A.A.P carried out sample processing and sequencing. H.C., J.R., B.M.L, C.L., A.O. and D.W. conducted bioinformatics analyses. Z.R., A.A.R. and H.K. carried out *in vitro* experiments. Z.R. and H.K carried out *in vivo* experiments. H.C., Z.R., K.D., A.A.R., D.W. and H.K wrote the paper with contributions from all authors.

Competing Interests

The authors declare no competing interests.

Supplementary information

The supplementary data include 5 tables and 5 figures.

References

1. Rossi, E., La Rosa, R., Bartell, J. A., Marvig, R. L., Haagensen, J., Sommer, L. M., Molin, S., & Johansen, H. K. (2021). Pseudomonas aeruginosa adaptation and evolution in patients with cystic fibrosis. Nature reviews. Microbiology, 19(5), 331–342. https://doi.org/10.1038/s41579-020-00477-5
2. Ch'ng, J. H., Chong, K., Lam, L. N., Wong, J. J., & Kline, K. A. (2019). Biofilm-associated infection by enterococci. Nature reviews. Microbiology, 17(2), 82–94. https://doi.org/10.1038/s41579-018-0107-z
3. Lebeaux, D., Ghigo, J. M., & Beloin, C. (2014). Biofilm-related infections: bridging the gap between clinical management and fundamental aspects of recalcitrance toward antibiotics. Microbiology and molecular biology reviews : MMBR, 78(3), 510–543. https://doi.org/10.1128/MMBR.00013-14

4. Lamont, R. J., Koo, H., & Hajishengallis, G. (2018). The oral microbiota: dynamic communities and host interactions. Nature reviews. Microbiology, 16(12), 745–759. https://doi.org/10.1038/s41579-018-0089-x
5. Sadiq, F. A., Burmølle, M., Heyndrickx, M., Flint, S., Lu, W., Chen, W., Zhao, J., & Zhang, H. (2021). Community-wide changes reflecting bacterial interspecific interactions in multispecies biofilms. Critical reviews in microbiology, 47(3), 338–358. https://doi.org/10.1080/1040841X.2021.1887079
6. Marsh P. D. (2018). In Sickness and in Health-What Does the Oral Microbiome Mean to Us? An Ecological Perspective. Advances in dental research, 29(1), 60–65. https://doi.org/10.1177/0022034517735295
7. Uribe, S. E., Innes, N., & Maldupa, I. (2021). The global prevalence of early childhood caries: A systematic review with meta-analysis using the WHO diagnostic criteria. International journal of paediatric dentistry, 31(6), 817–830. https://doi.org/10.1111/ipd.12783
8. Pitts, N. B., Baez, R. J., Diaz-Guillory, C., Donly, K. J., Alberto Feldens, C., McGrath, C., Phantumvanit, P., Seow, W. K., Sharkov, N., Songpaisan, Y., Tinanoff, N., & Twetman, S. (2019). Early Childhood Caries: IAPD Bangkok Declaration. Journal of dentistry for children (Chicago, Ill.), 86(2), 72.
9. Alaluusua, S., & Renkonen, O. V. (1983). Streptococcus mutans establishment and dental caries experience in children from 2 to 4 years old. Scandinavian journal of dental research, 91(6), 453–457. https://doi.org/10.1111/j.1600-0722.1983.tb00845.x
10. Hamada, S., Ooshima, T., Torii, M., Imanishi, H., Masuda, N., Sobue, S., & Kotani, S. (1978). Dental caries induction in experimental animals by clinical strains of Streptococcus mutans isolated from Japanese children. Microbiology and immunology, 22(6), 301–314. https://doi.org/10.1111/j.1348-0421.1978.tb00375.x
11. Clarke J. K. (1924). On the Bacterial Factor in the Ætiology of Dental Caries. British journal of experimental pathology, 5(3), 141–147.
12. Nyvad, B., Crielaard, W., Mira, A., Takahashi, N., & Beighton, D. (2013). Dental caries from a molecular microbiological perspective. Caries research, 47(2), 89–102. https://doi.org/10.1159/000345367
13. Kim, D., Sengupta, A., Niepa, T. H., Lee, B. H., Weljie, A., Freitas-Blanco, V. S., Murata, R. M., Stebe, K. J., Lee, D., & Koo, H. (2017). Candida albicans stimulates Streptococcus mutans microcolony development via cross-kingdom biofilm-derived metabolites. Scientific reports, 7, 41332. https://doi.org/10.1038/srep41332
14. Mira, A., Simon-Soro, A., & Curtis, M. A. (2017). Role of microbial communities in the pathogenesis of periodontal diseases and caries. Journal of clinical periodontology, 44 Suppl 18, S23–S38. https://doi.org/10.1111/jcpe.12671
15. Philip, N., Suneja, B., & Walsh, L. (2018). Beyond Streptococcus mutans: clinical implications of the evolving dental caries aetiological paradigms and its associated microbiome. British dental journal, 224(4), 219–225. https://doi.org/10.1038/sj.bdj.2018.81
16. Hajishengallis, G., & Lamont, R. J. (2016). Dancing with the Stars: How Choreographed Bacterial Interactions Dictate Nososymbiocity and Give Rise to Keystone Pathogens, Accessory Pathogens,

and Pathobionts. Trends in microbiology, 24(6), 477–489. https://doi.org/10.1016/j.tim.2016.02.010

17. Peterson, S. N., Meissner, T., Su, A. I., Snesrud, E., Ong, A. C., Schork, N. J., & Bretz, W. A. (2014). Functional expression of dental plaque microbiota. Frontiers in cellular and infection microbiology, 4, 108. https://doi.org/10.3389/fcimb.2014.00108
18. Kressirer, C. A., Chen, T., Lake Harriman, K., Frias-Lopez, J., Dewhirst, F. E., Tavares, M. A., & Tanner, A. C. (2018). Functional profiles of coronal and dentin caries in children. Journal of oral microbiology, 10(1), 1495976. https://doi.org/10.1080/20002297.2018.1495976
19. Mark Welch, J. L., Ramírez-Puebla, S. T., & Borisy, G. G. (2020). Oral Microbiome Geography: Micron-Scale Habitat and Niche. Cell host & microbe, 28(2), 160–168. https://doi.org/10.1016/j.chom.2020.07.009
20. Azimi, S., Lewin, G. R., & Whiteley, M. (2022). The biogeography of infection revisited. Nature reviews. Microbiology, 10.1038/s41579-022-00683-3. Advance online publication. https://doi.org/10.1038/s41579-022-00683-3
21. Oliveira, R. R., Fermiano, D., Feres, M., Figueiredo, L. C., Teles, F. R., Soares, G. M., & Faveri, M. (2016). Levels of Candidate Periodontal Pathogens in Subgingival Biofilm. Journal of dental research, 95(6), 711–718. https://doi.org/10.1177/0022034516634619
22. Bowen, W. H., Burne, R. A., Wu, H., & Koo, H. (2018). Oral Biofilms: Pathogens, Matrix, and Polymicrobial Interactions in Microenvironments. Trends in microbiology, 26(3), 229–242. https://doi.org/10.1016/j.tim.2017.09.008
23. Stephan, R. M. (1944). Intra-Oral Hydrogen-Ion Concentrations Associated With Dental Caries Activity. Journal of Dental Research, 23(4), 257–266. https://doi.org/10.1177/00220345440230040401
24. Kingsley, V. V., & Hoeniger, J. F. (1973). Growth, structure, and classification of Selenomonas. Bacteriological reviews, 37(4), 479–521. https://doi.org/10.1128/br.37.4.479-521.1973
25. Marsh, P. D., & Zaura, E. (2017). Dental biofilm: ecological interactions in health and disease. Journal of clinical periodontology, 44 Suppl 18, S12–S22. https://doi.org/10.1111/jcpe.12679
26. Xiao, J., Klein, M. I., Falsetta, M. L., Lu, B., Delahunty, C. M., Yates, J. R., 3rd, Heydorn, A., & Koo, H. (2012). The exopolysaccharide matrix modulates the interaction between 3D architecture and virulence of a mixed-species oral biofilm. PLoS pathogens, 8(4), e1002623. https://doi.org/10.1371/journal.ppat.1002623
27. Hartmann, R., Jeckel, H., Jelli, E., Singh, P. K., Vaidya, S., Bayer, M., Rode, D., Vidakovic, L., Díaz-Pascual, F., Fong, J., Dragoš, A., Lamprecht, O., Thöming, J. G., Netter, N., Häussler, S., Nadell, C. D., Sourjik, V., Kovács, Á. T., Yildiz, F. H., & Drescher, K. (2021). Quantitative image analysis of microbial communities with BiofilmQ. Nature microbiology, 6(2), 151–156. https://doi.org/10.1038/s41564-020-00817-4
28. Bowen, W. H., & Koo, H. (2011). Biology of Streptococcus mutans-derived glucosyltransferases: role in extracellular matrix formation of cariogenic biofilms. Caries research, 45(1), 69–86. https://doi.org/10.1159/000324598

29. Wang, W., Zhang, N., Du, Y., Gao, J., Li, M., Lin, L., Czajkowsky, D. M., Li, X., Yang, C., & Shao, Z. (2021). Three-Dimensional Quantitative Imaging of Native Microbiota Distribution in the Gut. Angewandte Chemie (International ed. in English), 60(6), 3055–3061. https://doi.org/10.1002/anie.202010921
30. Drescher, J., Schlafer, S., Schaudinn, C., Riep, B., Neumann, K., Friedmann, A., Petrich, A., Göbel, U. B., & Moter, A. (2010). Molecular epidemiology and spatial distribution of Selenomonas spp. in subgingival biofilms. European journal of oral sciences, 118(5), 466–474. https://doi.org/10.1111/j.1600-0722.2010.00765.x
31. Paster, B. J., Boches, S. K., Galvin, J. L., Ericson, R. E., Lau, C. N., Levanos, V. A., Sahasrabudhe, A., & Dewhirst, F. E. (2001). Bacterial diversity in human subgingival plaque. Journal of bacteriology, 183(12), 3770–3783. https://doi.org/10.1128/JB.183.12.3770-3783.2001
32. Faveri, M., Mayer, M. P., Feres, M., de Figueiredo, L. C., Dewhirst, F. E., & Paster, B. J. (2008). Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. Oral microbiology and immunology, 23(2), 112–118. https://doi.org/10.1111/j.1399-302X.2007.00397.x
33. Rôças, I. N., Siqueira, J. F., Jr, & Debelian, G. J. (2011). Analysis of symptomatic and asymptomatic primary root canal infections in adult Norwegian patients. Journal of endodontics, 37(9), 1206–1212. https://doi.org/10.1016/j.joen.2011.05.026
34. da Costa Rosa, T., de Almeida Neves, A., Azcarate-Peril, M. A., Divaris, K., Wu, D., Cho, H., Moss, K., Paster, B. J., Chen, T., B Freitas-Fernandes, L., Fidalgo, T., Tadeu Lopes, R., Valente, A. P., R Arnold, R., & de Aguiar Ribeiro, A. (2021). The bacterial microbiome and metabolome in caries progression and arrest. Journal of oral microbiology, 13(1), 1886748. https://doi.org/10.1080/20002297.2021.1886748
35. Ren, Z., Kim, D., Paula, A. J., Hwang, G., Liu, Y., Li, J., Daniell, H., & Koo, H. (2019). Dual-Targeting Approach Degrades Biofilm Matrix and Enhances Bacterial Killing. Journal of dental research, 98(3), 322–330. https://doi.org/10.1177/0022034518818480
36. Kim, D., Barraza, J. P., Arthur, R. A., Hara, A., Lewis, K., Liu, Y., Scisci, E. L., Hajishengallis, E., Whiteley, M., & Koo, H. (2020). Spatial mapping of polymicrobial communities reveals a precise biogeography associated with human dental caries. Proceedings of the National Academy of Sciences of the United States of America, 117(22), 12375–12386. https://doi.org/10.1073/pnas.1919099117
37. Hartmann, R., Jeckel, H., Jelli, E., Singh, P. K., Vaidya, S., Bayer, M., Rode, D., Vidakovic, L., Díaz-Pascual, F., Fong, J., Dragoš, A., Lamprecht, O., Thöming, J. G., Netter, N., Häussler, S., Nadell, C. D., Sourjik, V., Kovács, Á. T., Yildiz, F. H., & Drescher, K. (2021). Quantitative image analysis of microbial communities with BiofilmQ. Nature microbiology, 6(2), 151–156. https://doi.org/10.1038/s41564-020-00817-4
38. Bowen W. H. (2013). Rodent model in caries research. Odontology, 101(1), 9–14. https://doi.org/10.1007/s10266-012-0091-0
39. Koo, H., Schobel, B., Scott-Anne, K., Watson, G., Bowen, W. H., Cury, J. A., Rosalen, P. L., & Park, Y. K. (2005). Apigenin and tt-farnesol with fluoride effects on S. mutans biofilms and dental caries. Journal of dental research, 84(11), 1016–1020. https://doi.org/10.1177/154405910508401109

40. Falsetta, M. L., Klein, M. I., Colonne, P. M., Scott-Anne, K., Gregoire, S., Pai, C. H., Gonzalez-Begne, M., Watson, G., Krysan, D. J., Bowen, W. H., & Koo, H. (2014). Symbiotic relationship between Streptococcus mutans and Candida albicans synergizes virulence of plaque biofilms in vivo. Infection and immunity, 82(5), 1968–1981. https://doi.org/10.1128/IAI.00087-14
41. Rath, C. B., Schirmeister, F., Figl, R., Seeberger, P. H., Schäffer, C., & Kolarich, D. (2018). Flagellin Glycoproteomics of the Periodontitis Associated Pathogen Selenomonas sputigena Reveals Previously Not Described O-glycans and Rhamnose Fragment Rearrangement Occurring on the Glycopeptides. Molecular & cellular proteomics : MCP, 17(4), 721–736. https://doi.org/10.1074/mcp.RA117.000394
42. Oliveira, R. R., Fermiano, D., Feres, M., Figueiredo, L. C., Teles, F. R., Soares, G. M., & Faveri, M. (2016). Levels of Candidate Periodontal Pathogens in Subgingival Biofilm. Journal of dental research, 95(6), 711–718. https://doi.org/10.1177/0022034516634619
43. Deng, Z. L., Szafrański, S. P., Jarek, M., Bhuju, S., & Wagner-Döbler, I. (2017). Dysbiosis in chronic periodontitis: Key microbial players and interactions with the human host. Scientific reports, 7(1), 3703. https://doi.org/10.1038/s41598-017-03804-8
44. Chao, X., Lei, Z., Hongqin, L., Ziwei, W., Dechuan, L., Weidong, D., Lu, X., Haitao, C., Bo, Z., Haixing, J., & Qinghua, Y. (2022). Faeces from malnourished colorectal cancer patients accelerate cancer progression. Clinical nutrition (Edinburgh, Scotland), 41(3), 632–644. https://doi.org/10.1016/j.clnu.2022.01.001
45. Xu, J., Xiang, C., Zhang, C., Xu, B., Wu, J., Wang, R., Yang, Y., Shi, L., Zhang, J., & Zhan, Z. (2019). Microbial biomarkers of common tongue coatings in patients with gastric cancer. Microbial pathogenesis, 127, 97–105. https://doi.org/10.1016/j.micpath.2018.11.051
46. Heimisdottir, L. H., Lin, B. M., Cho, H., Orlenko, A., Ribeiro, A. A., Simon-Soro, A., Roach, J., Shungin, D., Ginnis, J., Simancas-Pallares, M. A., Spangler, H. D., Zandoná, A., Wright, J. T., Ramamoorthy, P., Moore, J. H., Koo, H., Wu, D., & Divaris, K. (2021). Metabolomics Insights in Early Childhood Caries. Journal of dental research, 100(6), 615–622. https://doi.org/10.1177/0022034520982963
47. Vacca-Smith, A. M., & Bowen, W. H. (1998). Binding properties of streptococcal glucosyltransferases for hydroxyapatite, saliva-coated hydroxyapatite, and bacterial surfaces. Archives of oral biology, 43(2), 103–110. https://doi.org/10.1016/s0003-9969(97)00111-8
48. Johnston, C. D., Cotton, S. L., Rittling, S. R., Starr, J. R., Borisy, G. G., Dewhirst, F. E., & Lemon, K. P. (2019). Systematic evasion of the restriction-modification barrier in bacteria. Proceedings of the National Academy of Sciences of the United States of America, 116(23), 11454–11459. https://doi.org/10.1073/pnas.1820256116
49. Hedberg, M. E., Moore, E., Svensson-Stadler, L., Hörstedt, P., Baranov, V., Hernell, O., Wai, S. N., Hammarström, S., & Hammarström, M. L. (2012). Lachnoanaerobaculum gen. nov., a new genus in the Lachnospiraceae: characterization of Lachnoanaerobaculum umeaense gen. nov., sp. nov., isolated from the human small intestine, and Lachnoanaerobaculum orale sp. nov., isolated from saliva, and reclassification of Eubacterium saburreum (Prevot 1966) Holdeman and Moore 1970 as

Lachnoanaerobaculum saburreum comb. nov. International journal of systematic and evolutionary microbiology, 62(Pt 11), 2685–2690. https://doi.org/10.1099/ijs.0.033613-0

50. Xiao, J., Huang, X., Alkhers, N., Alzamil, H., Alzoubi, S., Wu, T. T., Castillo, D. A., Campbell, F., Davis, J., Herzog, K., Billings, R., Kopycka-Kedzierawski, D. T., Hajishengallis, E., & Koo, H. (2018). Candida albicans and Early Childhood Caries: A Systematic Review and Meta-Analysis. Caries research, 52(1-2), 102–112. https://doi.org/10.1159/000481833
51. Benler, S., & Koonin, E. V. (2021). Fishing for phages in metagenomes: what do we catch, what do we miss?. Current opinion in virology, 49, 142–150. https://doi.org/10.1016/j.coviro.2021.05.008
52. Divaris, K., Slade, G. D., Ferreira Zandona, A. G., Preisser, J. S., Ginnis, J., Simancas-Pallares, M. A., Agler, C. S., Shrestha, P., Karhade, D. S., Ribeiro, A. A., Cho, H., Gu, Y., Meyer, B. D., Joshi, A. R., Azcarate-Peril, M. A., Basta, P. V., Wu, D., & North, K. E. (2020). Cohort Profile: ZOE 2.0-A Community-Based Genetic Epidemiologic Study of Early Childhood Oral Health. International journal of environmental research and public health, 17(21), 8056. https://doi.org/10.3390/ijerph17218056
53. Ginnis, J., Ferreira Zandoná, A. G., Slade, G. D., Cantrell, J., Antonio, M. E., Pahel, B. T., Meyer, B. D., Shrestha, P., Simancas-Pallares, M. A., Joshi, A. R., & Divaris, K. (2019). Measurement of Early Childhood Oral Health for Research Purposes: Dental Caries Experience and Developmental Defects of the Enamel in the Primary Dentition. Methods Mol Biol 2019;1922:511-523. https://doi.org/10.1007/978-1-4939-9012-2_39
54. Divaris, K., Shungin, D., Rodríguez-Cortés, A., Basta, P. V., Roach, J., Cho, H., Wu, D., Ferreira Zandoná, A. G., Ginnis, J., Ramamoorthy, S., Kinchen, J. M., Kwintkiewicz, J., Butz, N., Ribeiro, A. A., & Azcarate-Peril, M. A. (2019). The Supragingival Biofilm in Early Childhood Caries: Clinical and Laboratory Protocols and Bioinformatics Pipelines Supporting Metagenomics, Metatranscriptomics, and Metabolomics Studies of the Oral Microbiome. Methods Mol Biol. 2019;1922:525-548. https://doi.org/10.1007/978-1-4939-9012-2_40
55. Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome biology, 20(1), 257. https://doi.org/10.1186/s13059-019-1891-0
56. Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science, 3, e104.
57. Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., Lakshmanan, A., & Wade, W. G. (2010). The human oral microbiome. Journal of bacteriology, 192(19), 5002–5017. https://doi.org/10.1128/JB.00542-10
58. Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ, 4, e2584. https://doi.org/10.7717/peerj.2584
59. Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S. T., Methé, B., Schloss, P. D., Gevers, D., Mitreva, M., & Huttenhower, C. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS computational biology, 8(6), e1002358. https://doi.org/10.1371/journal.pcbi.1002358

60. Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife, 10, e65088. https://doi.org/10.7554/eLife.65088
61. Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., & Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature methods, 12(10), 902–903. https://doi.org/10.1038/nmeth.3589
62. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England), 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635
63. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature methods, 14(4), 417–419. https://doi.org/10.1038/nmeth.4197
64. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289-300.
65. Mu, R., Shinde, P., Zou, Z., Kreth, J., & Merritt, J. (2019). Examining the Protein Interactome and Subcellular Localization of RNase J2 Complexes in Streptococcus mutans. Frontiers in microbiology, 10, 2150. https://doi.org/10.3389/fmicb.2019.02150
66. Braissant, O., Bonkat, G., Wirz, D., & Bachmann, A. (2013). Microbial growth and isothermal microcalorimetry: growth models and their application to microcalorimetric data. Thermochimica Acta, 555, 64-71.
67. Braissant, O., Keiser, J., Meister, I., Bachmann, A., Wirz, D., Göpfert, B., Bonkat, G., & Wadsö, I. (2015). Isothermal microcalorimetry accurately detects bacteria, tumorous microtissues, and parasitic worms in a label-free well-plate assay. Biotechnology journal, 10(3), 460–468. https://doi.org/10.1002/biot.201400494
68. Paula, A. J., Hwang, G., & Koo, H. (2020). Dynamics of bacterial population growth in biofilms resemble spatial and structural aspects of urbanization. Nature communications, 11(1), 1354. https://doi.org/10.1038/s41467-020-15165-4
69. Zhang, C., Xing, X. H., & Lou, K. (2005). Rapid detection of a gfp-marked Enterobacter aerogenes under anaerobic conditions by aerobic fluorescence recovery. FEMS microbiology letters, 249(2), 211–218. https://doi.org/10.1016/j.femsle.2005.05.051
70. Tinevez, J. Y., Perry, N., Schindelin, J., Hoopes, G. M., Reynolds, G. D., Laplantine, E., Bednarek, S. Y., Shorte, S. L., & Eliceiri, K. W. (2017). TrackMate: An open and extensible platform for single-particle tracking. Methods (San Diego, Calif.), 115, 80–90. https://doi.org/10.1016/j.ymeth.2016.09.016
71. Klein, M. I., Scott-Anne, K. M., Gregoire, S., Rosalen, P. L., & Koo, H. (2012). Molecular approaches for viable bacterial population and transcriptional analyses in a rodent model of dental caries. Molecular

oral microbiology, 27(5), 350–361. https://doi.org/10.1111/j.2041-1014.2012.00647.x

72. Larson, R. H. (1981). Merits and modifications of scoring rat dental caries by Keyes' method. In Proceedings of" Symposium on animal models in cariology", Spl. Supp. Microbiology Abstracts, 1981 (pp. 195-203).

Methods

Human microbiome studies: context and sampling. Human microbiome studies were done in the context of ZOE 2.0, a genetic epidemiologic study of early childhood oral health among a community-based sample of preschool-age children in North Carolina (NC)[52]. In brief, between 2016 and 2019, 8,059 children ages 36–71 months who attended public preschools in NC were enrolled in the study and 6,404 of them underwent comprehensive clinical examinations by trained and calibrated dental examiners. Data on childhood dental caries experience were collected using modified International Caries Detection and Classification System (ICDAS) criteria[53]. Two supragingival (i.e., “dental plaque”) biofilm samples were collected during the clinical encounters immediately prior to dental examinations, which took place before or at least 30 minutes after snack or breakfast. The plaque sample that was carried forward to microbiome analyses was obtained using sterile toothpicks from the facial/buccal surfaces of primary teeth in the upper-right quadrant, i.e., Universal tooth numbering system: #A, #B, #C, #D, and #E; FDI tooth numbering system: #55, #54, #53, #52, and #51 (Extended Data Fig. 3). Upon collection, plaque samples were placed in RNA*later* TissueProtect 1.5mL tubes and were frozen at -20^{o}C on-site until transferred to university core biospecimen processing facility for further processing or long-term storage at -80^{o}C. Detailed information regarding sample collection, storage, processing, nucleic acid extraction, and sequencing has been reported in a recent protocol publication[54].

Participants and phenotyping. We carried forward to whole genome shotgun sequencing (WGS, metagenomics/MTG) and RNA sequencing (RNA-seq, metatranscriptomics/MTX) supragingival samples of the first 300 ZOE 2.0 participants, 50% with and 50% without person-level dental caries experience defined at the “established” caries lesion detection threshold, ICDAS≥3)[46]. This lesion threshold corresponds to macroscopic tooth structure loss, i.e., a “cavity”. The ZOE 2.0 participants (mean age=52 months) formed the ‘discovery’ sample whereas the ‘replication’ sample comprised 116 similar-aged children (mean age=55 months), members of the same study population (i.e., enrolled in NC public preschools) examined under virtually identical conditions (i.e., by one clinical examiner) during the parent study’s pilot phase. For the purposes of the present study, we quantified caries experience using the most sensitive clinical criteria including the enumeration of early-stage caries lesions (i.e., at the ICDAS≥1 threshold) according to the recent international consensus definition of early childhood caries (ECC)[8]. This was done both locally (i.e., within the five tooth surfaces where plaque biofilm was harvested from) and at the person-level (i.e., in the entire dentition, comprising all 88 primary tooth surfaces). We considered the surfaces where plaque was collected from as the most informative in terms of microbiome taxonomy and thus localized caries experience was the primary clinical trait in all analyses. Nevertheless, we posited that the taxonomy of microbiome biofilm in those areas is likely to also be

informative for the condition of one's entire dentition--therefore, we considered a secondary 'person-level' caries experience trait as a secondary clinical trait. Estimates of these caries experience traits, in the discovery and replication samples, as well as participants' demographic information are present in Extended Data Table 5.

Sequencing and alignment. Total nucleic acid was quantified using QuantIT® PicoGreen®. 5 ng of genomic DNA was processed using the Nextera XT DNA Sample Preparation Kit (Illumina). Target DNA was simultaneously fragmented and tagged using the Nextera Enzyme Mix-containing transposome that fragments the input DNA and adds the bridge PCR (bPCR)-compatible adaptors required for binding and clustering in the flow cell. Next, fragmented and tagged DNA was amplified using a limited-cycle PCR program. In this step index 1(i7) and index 2(i5) was added between the downstream bPCR adaptor and the core sequencing library adaptor, as well primer sequences required for cluster formation. The thermal profile for the amplification had an initial extension step at 72°C for 3 min and initial denaturing step at 95°C for 30 sec, followed by 15 cycles of denaturing of 95°C for 10 seconds, annealing at 55°C for 30 seconds, a 30 second extension at 72°C, and final extension for 5 minutes at 72°C. The DNA library pool was loaded on the Illumina platform reagent cartridge (Illumina) and on the Illumina instrument. Sequencing output from the Illumina HiSeq 4000 2x150 was converted to fastq format and demultiplexed using Illumina Bcl2Fastq 2.20.0 (Illumina, Inc. San Diego, CA, USA.). Quality control of the demultiplexed sequencing reads was verified by FastQC (Babraham Institute. Cambridge, UK). Adapters were trimmed using Trim Galore (Babraham Institute. Cambridge, UK). The resulting paired-end reads were classified with Kraken2[55] and Bracken 2.5[56] using a custom database including human, fungal, bacterial, and the expanded Human Oral Microbiome Database (eHOMD) genomes[57] to produce an initial taxonomic composition profile. All reads identified as 'host' were eliminated. Paired-end reads were joined with vsearch 1.10.2[58]. Any remaining adapter reads were trimmed again using Trim Galore. Estimates of gene family, path abundance, and path coverage were produced from the remaining reads using HUMAnN3[59] based on taxonomic estimates from MetaPhlAn 3[60,61].

RNA isolation was performed using Qiagen RNeasy Mini Kit (Cat. No. / ID: 74104) and RNA was quantified using Nanodrop1000 (ThermoFisher, Waltham, MA). To generate MTX data via RNA-seq., sequencing output from the Illumina HiSeq 4000 2x150 platform was converted to fastq format and demultiplexed using Illumina Bcl2Fastq 2.20.0 (Illumina, Inc. San Diego, CA, USA). Quality control of the demultiplexed sequencing reads was verified by FastQC (Babraham Institute. Cambridge, UK). Adapters were trimmed using Trim Galore (Babraham Institute. Cambridge, UK). The resulting paired-end reads were classified with Kraken2[55] and Bracken 2.5[56] using a custom database including human, fungal, bacterial, and the expanded Human Oral Microbiome Database (eHOMD) genomes[57] to produce an initial taxonomic composition profile. All reads identified as host were eliminated. Paired-end reads were joined with vsearch 1.10.2[58]. The resulting single-end reads were again trimmed of any remaining adapters using Trim Galore. HUMAnN 3.0[59] was used to generate gene family and pathway-level data based on taxonomic estimates from the MetaPhlAn 3[60,61] metagenomic analysis. Additionally, MTX gene expression analyses were performed for the four "top species" that were tested experimentally. Reads

classified by Kraken2 as belonging to each of the four top species and any strain thereof were extracted by species, individually by bacterial species aligned to the relevant reference transcriptome with STAR[62], and subsequently quantified via Salmon[63].

Quality control procedures. An overview of quality control, scaling, and filtering of taxa is presented in Extended Data Fig. 4. After removal of viral sequences, the remaining data were first arranged in a reads-per-kilobase (RPK) format, then rescaled into transcripts-per-million (TPM), and finally rescaled close to the averaged-over-subject RPK level. The average RPK in MTG in the discovery sample was 8,004,958, and the rescaled TPM was set as 8,000,000 for each subject. To facilitate discovery and inference, species with relative abundance less than 10^{-5} or with prevalence rate less than 10% were excluded from all taxonomic discovery analyses. Out of 6,411 non-viral species, 5,990 low-abundance and 2,935 low-prevalence (including 2,932 both low-abundant and low-prevalence) taxa were filtered, retaining 418 taxa for all downstream MTG analyses. MTX data in the discovery sample, as well as MTG and MTX data in the replication sample were all processed in a similar fashion. Total RPKs per subject were 11,014,832, 5,077,759, and 2,739,606 on average, and the total rescaled TPMs per subject were on average 11 million, 5 million, and 3 million, respectively. Numbers of retained taxa after applying abundance and prevalence filters were 385 for the MTX in the discovery sample, 422 and 397 for the MTG and MTX data in the replication sample, respectively. We refer to these sets of retained taxa as "core." TPM normalization was not done for the targeted MTX of the 4 top species, due to limited diversity of the sample. Average RPK per sample was 72,474. We retained genes with average >20 reads for downstream differential expression and gene-gene interaction analyses. Using this filter, out of 9,103 total genes available for these 4 top species, 542 genes were retained: 47 for *S. mutans*, 39 for *S. sputigena*, 8 for *P. salivae*, and 448 for *L. wadei*.

Identification of species significantly associated with caries experience. We used log-normal linear models to test the association between species' differential abundance in MTG and MTX data and quantitative measures of caries experience. The models included terms for phenotypes of interest (i.e., either localized or person-level caries quantitative experience), batch effects (i.e., the first sequencing batch included 52 samples and the second batch the remaining 248 samples), age at enrollment (measured in months), and race/ethnicity (reported by legal guardians and categorized as white-non Hispanic, African American-non-Hispanic, and others including Hispanics), and a unity was added to the rescaled TPM abundance data before log-transformation with base 2. Presence of each species in MTG data was controlled for in models examining differential expression (i.e., abundance in MTX data). A false discovery rate (FDR) correction for multiple testing was applied using the Benjamini-Hochberg procedure[64] for each of the four models (caries traits and abundance / expression).

We used strict criteria for the identification of bacterial species associated with caries experience and required taxa to be FDR-significantly differentially abundant for quantitative disease experience defined both locally and at the person-level, in MTG and MTX data (i.e., 4 models), in the discovery sample (n=300). To reduce the potential for false discovery, we sought for additional evidence of association in an independent sample, i.e., we examined the replication of the previously identified

associations in a sample of 116 similarly aged participants from the same population as the parent study (i.e., public preschools in NC), similar clinical and microbiome data. Thus 4 more models were created, for localized and person-level disease, in MTG and MTX data. As evidence of replication, we considered, in order of ascending importance, directional consistency of the estimate of association, nominal significance, or FDR-level significance in the replication sample. Species that were FDR-significant in all 4 models in the discovery sample and were at least nominally significant for localized disease experience in MTG data were termed “significant species”. This set of species with high-confidence evidence of association from multiple traits, MTG and MTX data, and from all 416 study participants were prioritized for reporting and were candidates for consideration in the experimental validation pipeline.

Inter-species correlations and pathways involving species significantly associated with caries experience. To provide initial insights into the inter-relationships of the 16 significant species, we examined their pairwise correlation patterns in health and disease. For this purpose, we used Pearson correlation coefficients between model residuals that were generated for each of the 16 significant species. These log-normal models had the same specifications with models used in the main analyses (i.e., adjusted for batch effects and age) with the addition of an adjustment for disease experience. Species’ mean correlations with each of the other 15 significant species were examined between health (i.e., no localized disease experience) and disease (i.e., any localized disease experience), and in MTG and MTX.

To identify pathways significantly associated with caries experience in the biofilm transcriptome, we examined pathway and pathway-species MTX data that were prepared in RPK format as described above. Total RPKs per sample were on average 383,119, and the RPK data were transformed into a TPM format with a scale of 400,000. For this analysis, we departed from the top 100 pathways in MTX data in terms of overall abundance and prioritized the top 30 pathways that had representation from at least one of the 16 significant species. Each of these 30 pathways, summed over all species including those unclassified, was tested using the log-normal model with the same set of covariates as in the main discovery analysis and the addition of a unity before $\log_2$ transformation. An FDR correction for testing of 30 pathways was applied using the Benjamini-Hochberg procedure. Using this procedure, we identified pathways whose expression in MTX was significantly associated with caries experience. Additionally, we examined the representation of the 16 significant species in these pathways as percent of significant over all species involved in each pathway.

Shortlisting of species for virulence assays and experimental testing. We departed from a list of 16 species that showed strong and replicable evidence of association with dental caries experience (Extended Data Table 1). To select a shortlist of candidates that could be fully characterized in virulence, biofilm, and possibly in vivo studies in this study, we considered species’ statistical evidence of association in the discovery sample (i.e., p-value), evidence replication in the independent sample, representation of different genera (i.e., selected one species per genus), and availability of clinical isolates. Based on these criteria, from the list of 16 significant species we prioritized *Streptococcus mutans* (the known, well-established pathogen), and *Selenomonas sputigena*, *Prevotella salivae*, and

Leptotrichia wadei (as 3 new candidates). These 4 taxa, referred to as “top species”, were carried forward to in vitro virulence assessments and biofilm characterization and served as candidates for further in vivo colonization and cariogenicity studies.

Microorganisms and growth conditions used in laboratory validation. *Streptococcus mutans* UA159 (ATCC 700610), *Selenomonas sputigena* ATCC 35185, *Leptotrichia wadei* JCM 16777, *Prevotella salivae* JCM 12084 were used in *in vitro* and *in vivo* studies. Bacteria were grown in Brain Heart Infusion Supplemented (BHIS) medium (Brain Heart Infusion broth supplemented with 5 g/L yeast extract, 5 mg/L Hemin, 1 mg/L Vitamin K1, and 0.5 g/mL L-cysteine) at 37 °C to exponential phase in an anaerobic chamber (Anaerobe Systems). For *in vitro* biofilm studies, a saliva-coated hydroxyapatite disc (2.7 ± 0.2 cm^2; Clarkson Chromatography Products) was placed in a vertical position to mimic the tooth-enamel surfaces of human teeth. Single or mixed biofilms were grown on the apatitic surfaces in BHIS medium supplemented with 1% sucrose, the primary dietary sugar associated with tooth decay, to simulate the cariogenic condition in ECC patients. Biofilms were grown on saliva-coated hydroxyapatite surfaces for 24 h to allow the establishment of single or mixed-species communities. For real-time live imaging, a green fluorescent protein (GFP)-tagged *S. mutans* UA159 strain (*S. mutans* UA159 Ef-Tu-ngGFP)[65] was used. Biofilm EPS glucan matrices were labeled via supplementing the culture medium with 1 μM Alexa Fluor 647 dextran conjugate (Molecular Probes) during biofilm growth. This labeling method is highly specific for *S. mutans*-derived α-glucans since the fluorescently labeled dextrans serve as primers for *streptococcal* glycosyltransferases and are directly incorporated into glucans during biofilm EPS synthesis.

Top species’ virulence and metabolic profile assessment

Acid tolerance and acidogenicity. Acid tolerance tests were performed using Brain Heart Infusion (BHI) broth (Anaerobe Systems) with pre-adjusted pH. Lactic acid (13.42 M) was used to adjust the pH of the medium (ranging between 2.9 and 6.5). Each of the top species was serially diluted to 10^7 cells (CFU)/mL. One hundred microliters (100 μL) of bacterial suspension were transferred to 900 μL pH adjusted medium to study the growth of single species under acidic conditions. In parallel, each of the new species (*S. sputigena*, *L. wadei*, or *P. salivae*) was mixed with *S. mutans* (equal proportion of single species) to examine the growth of mixed species combinations. The cultures were incubated in the anaerobic chamber at 37 °C for 48 h. Bacterial growth was quantified using a microplate reader (SpectraMax M2e) by measuring optical density values at 610 nm, whereas the pH values were measured using a pH meter (FiveEasy Benchtop F20 pH/mV Meter, Mettler Toledo). Both the lowest pH values at which each bacterial species survived (detectable growth) and the final pH values of the culture (after 48h) were recorded to assess acid tolerance. For acidogenicity (acid production), we used standard glycolytic pH-drop assay. Bacterial cells (single or mixed-species, as described above) were incubated in salt solution (50 mM KCl, 1mM $MgCl_2.6H_2O$) with 1% glucose and allowed to produce acids over time. The decrease in pH was measured using a pH meter every 1 h over a period of 12 h and the proton

production rate was calculated to compare acidogenicity of different species either alone or in combination with *S. mutans*.

Metabolic profiling. Top species' metabolic profiles were measured using real-time isothermal microcalorimetry[66]. *S. mutans*, *S. sputigena*, *L. wadei*, *P. salivae,* and *P. oulorum* were transferred to BHI broth and incubated in the anaerobic chamber at 37^0C for 48 hours. Each culture was serially diluted to 10^7 cells, and 60 µl of each bacterial suspension was transferred, in triplicate, in titanium vials (calWell; SymCel) containing 540 µl of pre-reduced BHI. Real-time heat production, proxying metabolic activity, was measured using a calScreener™ microcalorimeter (SymCel Sverige AB, Stockholm, Sweden) in a 48-well plate (calPlate™) as previously described[67]. The full processing of the samples and plate preparation were performed inside the anaerobic chamber, to maintain the anaerobic condition for optimal bacterial growth. Heat and corresponding energy data were quantified with calView™ software (Version 1.0.33.0, © 2015 SymCel, Sverige AB). The instrument was set and calibrated at 37 °C, with all handling and set-up done according to the manufacturer's recommendations.

Biofilm live imaging. Biofilm live imaging was performed based on our established fluorescence labeling and confocal imaging protocols optimized for oral biofilms with some modifications[68]. Biofilms were dip-washed three times in phosphate buffered saline (PBS, pH 7.1) to remove any loosely bound microbes from the surface. To enhance the GFP fluorophore development in GFP-tagged *S. mutans* cells, we performed aerobic fluorescence recovery immediately before imaging, following the protocol previously reported[69]. Biofilms were counterstained with Syto 60 (Molecular Probes), which labeled all bacterial cells. Super-resolution live imaging was conducted at 37 °C using a 40× (numerical aperture = 1.2) water immersion objective on a Zeiss LSM800 upright confocal microscope with Airyscan. For real-time imaging of bacterial motility, multi-channel confocal images were taken at a 2.6-second interval for 30 s.

Computational biofilm image analysis. We then used a fluorescence subtraction method to analyze the biofilm spatial structuring (positioning of different species and EPS across the 3D biofilm structure) as detailed previously[36]. In brief, we applied channel subtraction using the Image Calculator in ImageJ Fiji (https://imagej.net/Fiji) to calculate the fluorescence signal from the new species in the mixed-species biofilm, using the following equation: $Ch_{New\ species} = Ch_{All\ bacteria} - Ch_{S.\ mutans}$, where $Ch_{All\ bacteria}$ is the channel of all bacteria (SYTO60), and $Ch_{S.\ mutans}$ is the channel of *S. mutans* (GFP). Computational image processing and quantitative analysis were performed using BiofilmQ software (https://drescherlab.org/data/biofilmQ), an image analysis toolbox optimized for biofilms[37]. After image segmentation using Otsu algorithm, we conducted a cube-based object declumping in BiofilmQ that dissected the entire biofilm into small cubic volumes (cube size = 2 µm). This function allows analyses of structural properties within biofilm subdomains with spatial resolution since each cube has a unique spatial coordinate in 3D. Parameters including local shape volume, relative abundance, and intensity were calculated for each channel within the cubes. For co-localization analysis, we used Mander's overlap coefficient in BiofilmQ to quantify the degree of spatial proximity of two different fluorescence signals in relation to each other. Two Manders' coefficients were calculated: 1) the new species with *S. mutans* and

2) the new species with EPS. To track single-cell bacterial motility in real time, we performed computational single-particle tracking and generated time-resolved trajectories using the TrackMate plugin in ImageJ Fiji[70]. Biofilm visualization was performed using maximum intensity projection and 3D surface rendering in ImageJ Fiji.

Targeted bacterial gene expression and gene-gene interaction analyses. We used a targeted MTX approach to test *S. mutans, S. sputigena*, *L. wadei*, and *P. salivae* gene expression. Reads were aligned to these species reference genomes using STAR-Salmon[62,63]. Where multiple strains were available, results were merged to the species level. The association of gene expression with caries experience was then tested for each gene using log-normal models adjusting for batch effects, age, and race/ethnicity, and applying an FDR multiple testing correction for each species. Additionally, we additionally examined gene-gene interactions between *S. mutans* and *S. sputigena*—the two species that demonstrated enhanced acidogenesis and unique biofilm structure when co-cultured. To this end, the same log-normal models including gene-gene interactions were used including pairwise combinations of all filtered genes from these two species, applying an FDR multiple testing correction.

***In vivo* rodent model of childhood caries.** Bacterial colonization on teeth and their impacts on disease onset were assessed on an established rodent caries model as detailed elsewhere with some modifications[40,71]. In brief, 15-d old female Sprague-Dawley rat pups (specific-pathogen-free grade) were purchased with dams (8 pups per dam) from Harlan Laboratories (Indianapolis, IN, USA). Upon arrival, animals were pre-screened for oral infection of *S. mutans* and *S. sputigena* by oral swabs and real-time polymerase chain reaction (qPCR) and were determined not to be infected with either organism. Oral swabs (FLOQSwab, COPAN Diagnostics, Murrieta, CA, USA) were taken from each of the animals and bacterial DNA were extracted using DNeasy PowerLyzer Microbial Kit (Qiagen, Valencia, CA, USA). Species-specific qPCR primer sets were used for microorganism detection as follows: *S. mutans:* Forward 5' ACCAGAAAGGGACGGCTAAC 3', Reverse 5' TAGCCTTTTACTCCAGACTTTCCTG 3'; *S. sputigena:* Forward 5' GGTCAGCCTTATCAGTTCCGTT 3', Reverse 5' GGCGAGCTTTCAGCAATCTTAG 3'; all bacteria: Forward 5' TCCTACGGGAGGCAGCAGT 3', Reverse 5' GGACTACCAGGGTATCTAATCCTGTT 3'.

The pups were randomly assigned into four groups that received different bacterial infections and were housed separately: (1) *S. mutans* alone; (2) *S. sputigena* alone; (3) *S. mutans* plus *S. sputigena;* (4) control without *S. mutans* or *S. sputigena* infection. Animals were inoculated daily using cotton oral swabs with actively growing cultures of *S. mutans* and/or *S. sputigena* (~10^8 CFU/mL in BHIS) at the age of 19-23 d (five doses in total). Each of the infected groups were confirmed for its designated microbial infection at 21 d, 24 d and 30 d by oral swabs and qPCR, while the control group remained free of either *S. mutans* or *S. sputigena*. No cross-contamination was detected throughout the experiment. All animals were provided with the National Institutes of Health cariogenic diet #2000 and 5% sucrose water *ad libitum* to mimic the cavity-promoting diet in childhood caries. The experiment proceeded for three weeks, at the end of which the animals were euthanized using carbon dioxide. The jaws were dissected and processed for caries scoring of teeth according to Larson's modification of Keyes' system[72]. Caries scores

were determined by a calibrated examiner. Investigators were masked to experimental group (i.e., infection) allocations during the infection, sampling, and assessment stages, by using color-coded samples. *In vitro* and *in vivo* experimental data were presented as mean ± standard deviation. Data were subjected to Student's pairwise *t*-test or analysis of variance (ANOVA) with post-hoc Tukey HSD test to account for multiple comparisons. Differences between groups were considered statistically significant when $p < 0.05$.

Ethical approval. Human observational data and analyses received approval (#14-1992) from the University of North Carolina-Chapel Hill Office of Human Research Ethics Institutional Review Board on September 18, 2014. Legal guardians of all children provided written informed consent for participation in the study. The *in vivo* experimental study was reviewed and approved by the Institutional Animal Care and Use Committee of the University of Pennsylvania (IACUC#805735). All research was performed in accordance with the Declaration of Helsinki.

Reporting summary. Further information on research design is available in the Reporting Summary linked to this article.

Data availability

Microbiome taxonomy (MTG and MTX) and pathway information for the entire biofilm microbial community, as well as targeted MTX data for the four top species used in this study have been deposited and are freely accessible alongside metadata (i.e., demographic and clinical phenotype information) via the Carolina Digital Repository and accession number 5d86p890x: https://cdr.lib.unc.edu/concern/data_sets/5d86p890x Raw sequence data for ZOE 2.0 have been deposited as part of dbGaP accession phs002232.v1.p1 and are pending release. ZOE pilot study sequence data have been made available via the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) and BioProject accession number: PRJNA843091.

Code availability

The scripts used to perform this analysis can be found at https://github.com/Hunyong/ZOE_metagenomics_2022

Figures

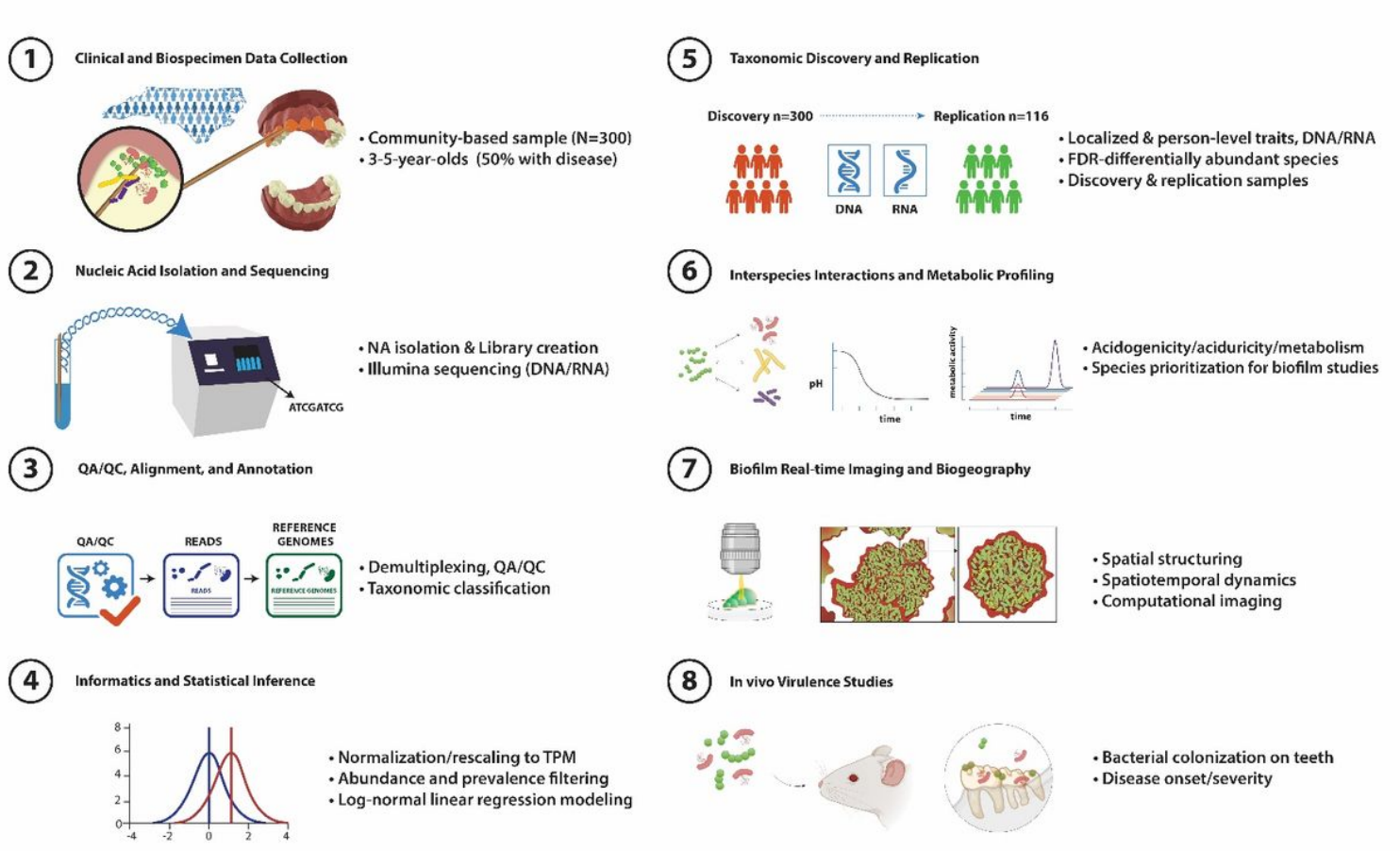


Figure 1

Overview of the design of the multi-method, discovery-validation pipeline and its application in the present study.

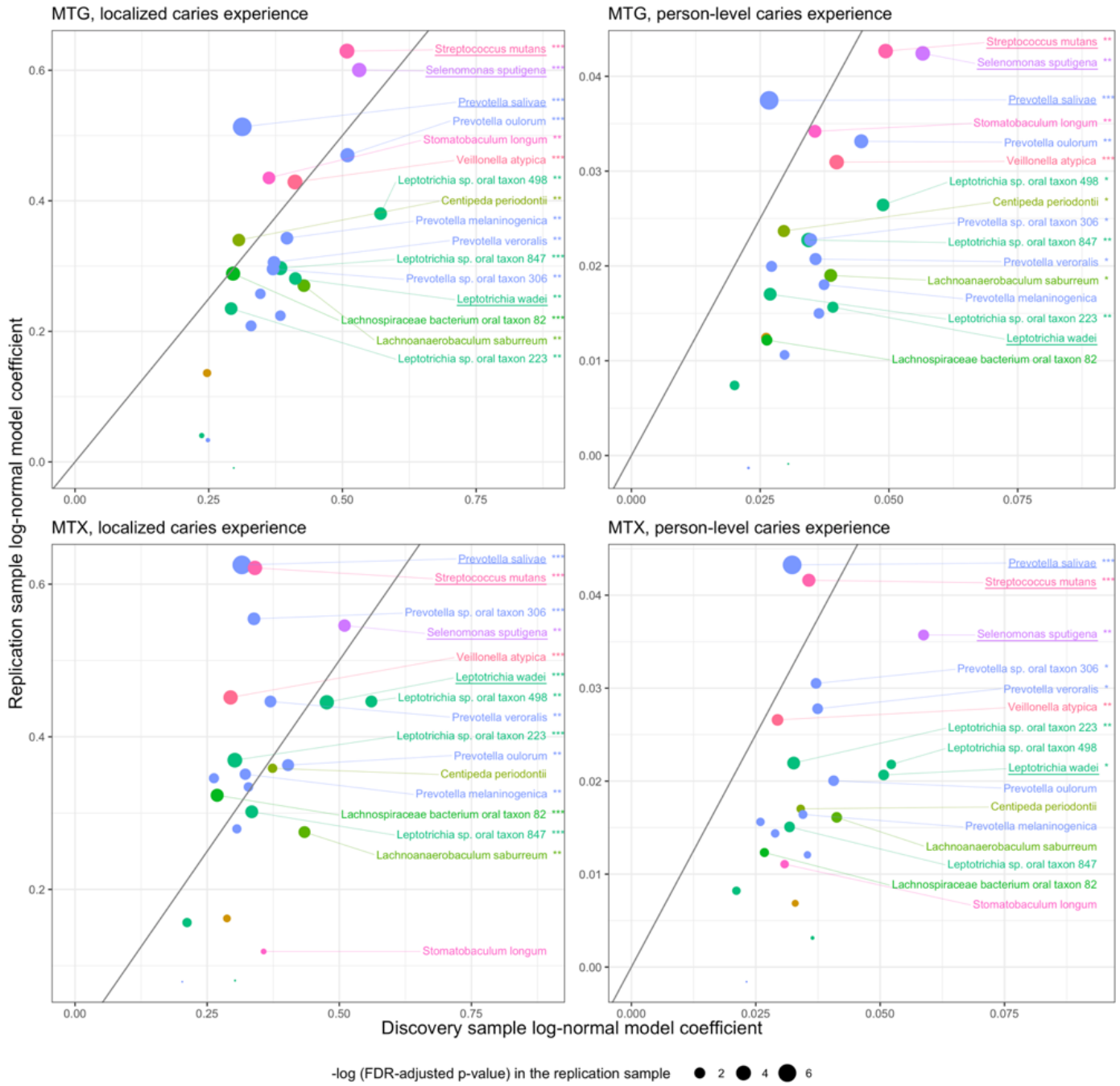


Figure 2

Estimates of association between dental caries experience and the abundance of 23 species that were significantly associated with caries experience in MTG and MTX analyses in the discovery (n=300) sample. 16 of those species showed evidence of association in the replication sample (n=116) and are annotated by name in the figure. *, **, and *** denote FDR-adjusted p-values of less than 0.1, 0.05, 0.01, respectively in the replication sample. The diagonal lines represent $y = x$. Similarly colored taxa are members of the same genus. The 4 underlined taxa are the “top species” that were prioritized for in vitro virulence assays, biofilm studies, and were candidates for *in vivo* experiments.

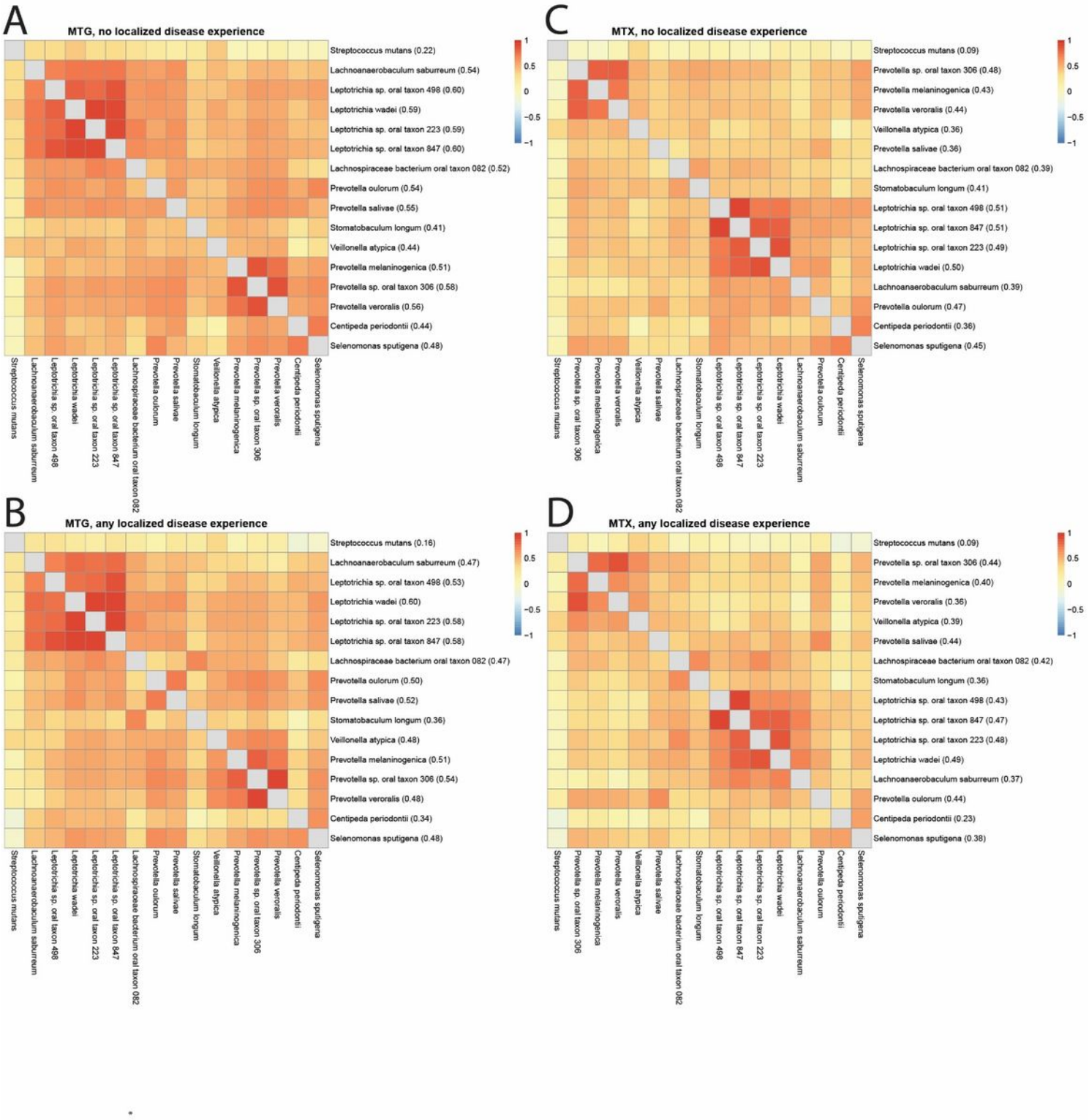


Figure 3

Pairwise correlations (i.e., Pearson correlations of residuals generated from a log-normal model for quantitative localized caries experience including terms for batch effects and age) between the 16 significant species' abundance in MTG (A, B) and MTX (C, D), in dental health (i.e., no localized caries experience) and disease (i.e., any localized caries experience). Parenthesized values are individual species' mean correlations with the other 15.

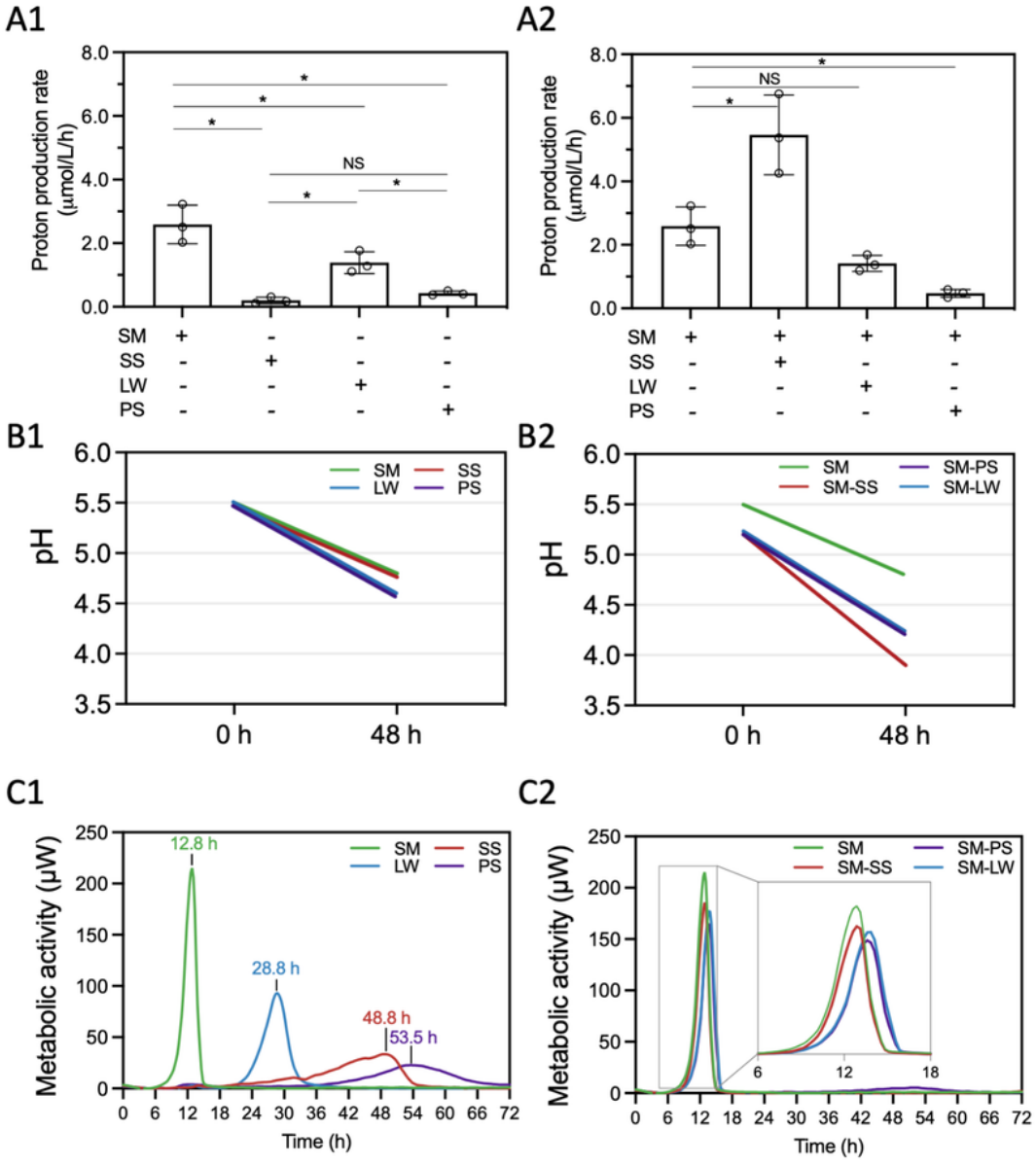


Figure 4

Metabolic and acidogenic profile of the top species. Top species include *S. mutans*, *S. sputigena*, *L. wadei*, and *P. salivae*. Each species was grown alone or in combination with *S. mutans*. (A1) Proton production rate of mono-species culture. (A2) Proton production rate of mixed-species cultures (*S. mutans* + new species) and *S. mutans* alone. (B1) Acid tolerance presented as the lowest starting pH that each of the species could survive and the final pH (after 48h) (B2) lowest starting pH that each of the

mixed-species could survive, and the final pH (after 48h). When paired with *S. mutans, S. sputigena* results in the lowest final pH compared to all other combinations. (C1) Curves of the metabolic activity (μW) of mono-species cultures. Data above individual peaks indicate the peak metabolic activity time of each species. (C2) Curves of the metabolic activity (μW) of mixed-species cultures (*S. mutans* + new species) and that of *S. mutans* alone. Inset, metabolic curves between 6-18 h. SM, *S. mutans*; SS, *S. sputigena*; LW, *L. wadei*; PS, *P. salivae*. For A1 and A2, data are plotted as mean ± standard deviation from three independent experiments. *, $p<0.05$ derived by one-way ANOVA with post hoc Tukey HSD test. NS stands for difference not statistically significant ($p>0.05$).

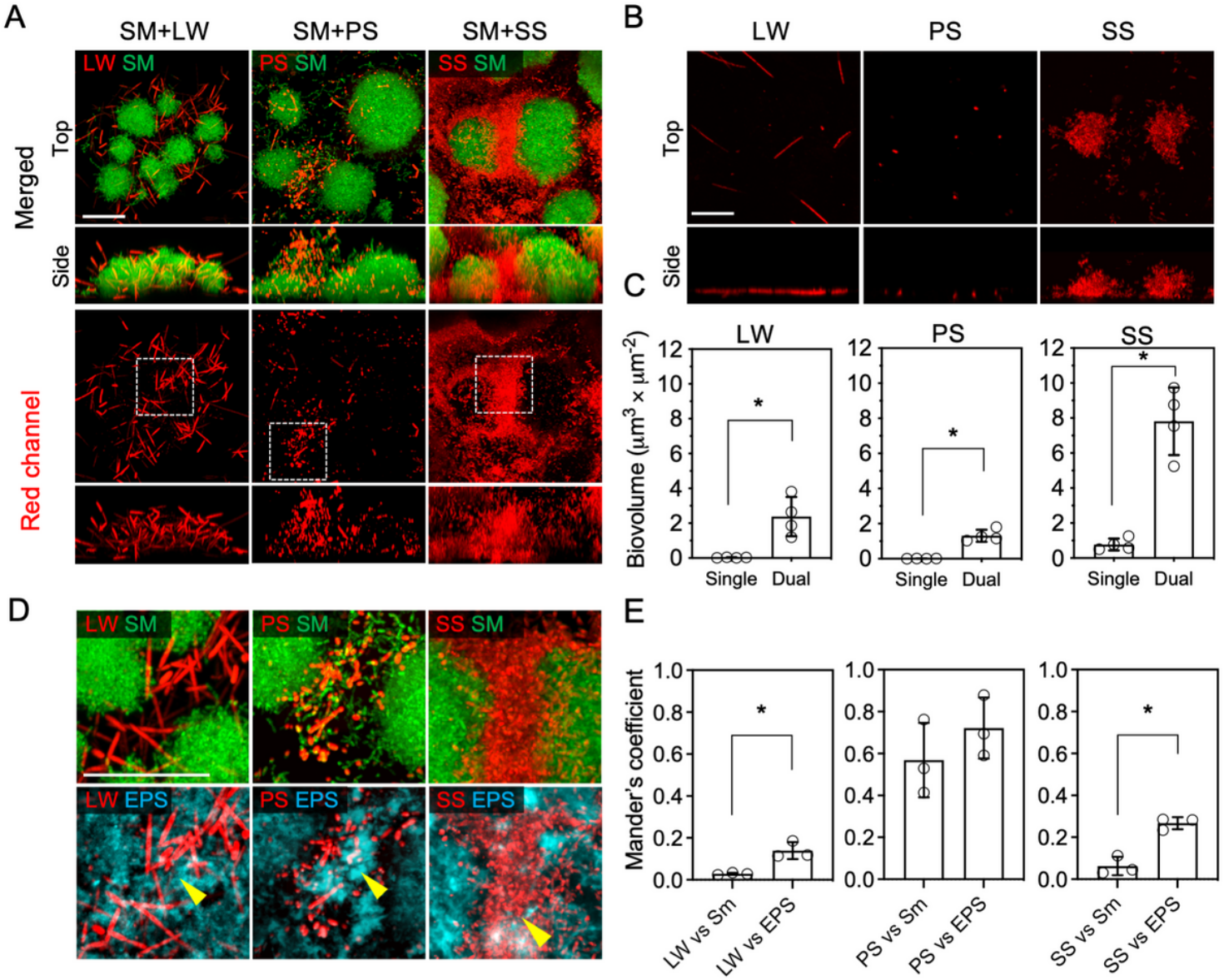
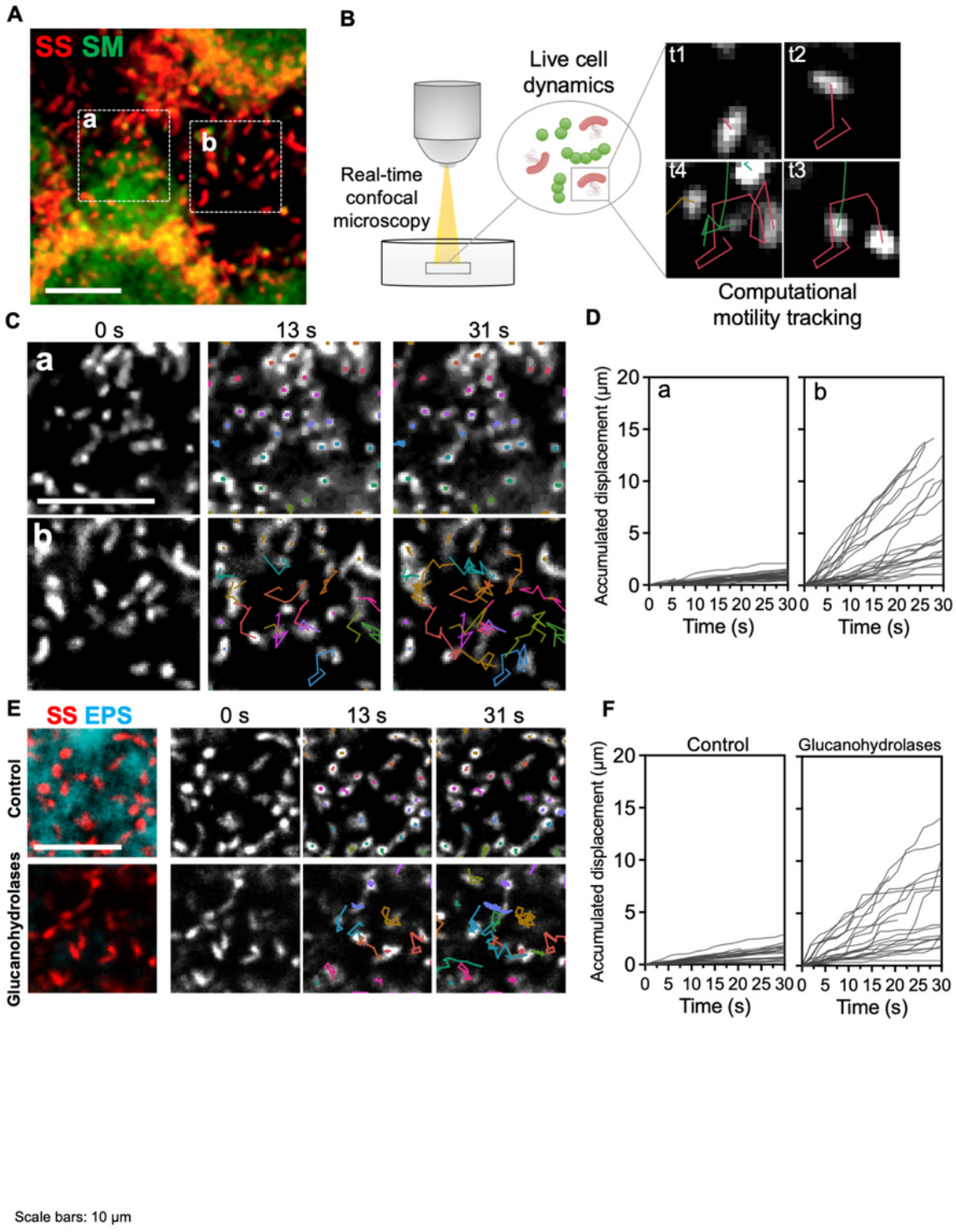


Figure 5

Biofilm formation and spatial structuring on tooth-mimetic surface. (A) Confocal images (top and side views) of mixed-species biofilms on saliva-coated hydroxyapatite surfaces formed by each of the new species together with *S. mutans.* Upper panel, *merged* image demonstrating the spatial structuring of *S.*

mutans (in green) and the new species (in red) within mixed biofilms. Lower panel, red channel only. Dotted box, magnified area as shown in panel D. (B) Mono-species biofilms formed by the top species (top and side views) (C) Biovolume of the new species within single and mixed biofilms, based on computational image analysis of panels A and B. (D) Top, magnified confocal images of mixed species biofilms; bottom, the new species physically interacting with extracellular α-glucan matrix (EPS, in cyan). (E) Computational colocalization analysis of the new species versus *S. mutans* cells or EPS. SM, *S. mutans*; SS, *S. sputigena*; LW, *L. wadei*; PS, *P. salivae.* Scale bars, 20 μm. For C and E, data are plotted as mean ± standard deviation from three independent experiments. *, $p<0.05$ derived from Student's *t*-test.

Figure 5. *S. sputigena*, a motile bacterium, becomes trapped in extracellular glucans.



Scale bars: 10 μm

Figure 6

***S. sputigena*, a motile bacterium, becomes trapped in extracellular glucans.** (A) Confocal image of *S. sputigena-S. mutans* mixed biofilm formed on the surface. Red, *S. sputigena*; green, *S. mutans*. Dotted boxes, biofilm subareas with *S. sputigena* physically associated with *S. mutans* clusters (a) or surface-attached *S. sputigena* alone (b). (B) A diagram of the experimental. Surface-attached microbes were visualized by fluorescence staining and real-time confocal microscopy. Individual *S. sputigena* cells (red

channel in panel A) were tracked computationally over time to generate spatiotemporal trajectories. (C) Trajectories of surface-attached *S. sputigena* cells in Areas (a) and (b), as shown in panel A. Colors indicate trajectories that originated from individual cells. (D) Accumulated *S. sputigena* cell displacement (total path length) relative to the initial position. Left, accumulated displacement of *S. sputigena* cells in Area (a); right, in Area (b). (E) Immobilized *S. sputigena* cells trapped by *S. mutans*-derived α-glucan matrix. Red, S. sputigena; cyan, α-glucan matrix. Colors indicate trajectories that originated from individual cells. Top panel, *S. sputigena* cells trapped by S. mutans-derived α-glucans showed no mobility; bottom panel, upon EPS degradation using glucanohydrolases (dextranase and mutanase), surface mobility of *S. sputigena* cells could be recovered. (F) Accumulated *S. sputigena* cell displacement (total path length) relative to the initial position. Left, accumulated displacement of *S. sputigena* cells in the control group; right, after EPS degradation using glucanohydrolases. SM, *S. mutans*; SS, *S. sputigena.* Scale bars, 10 μm.

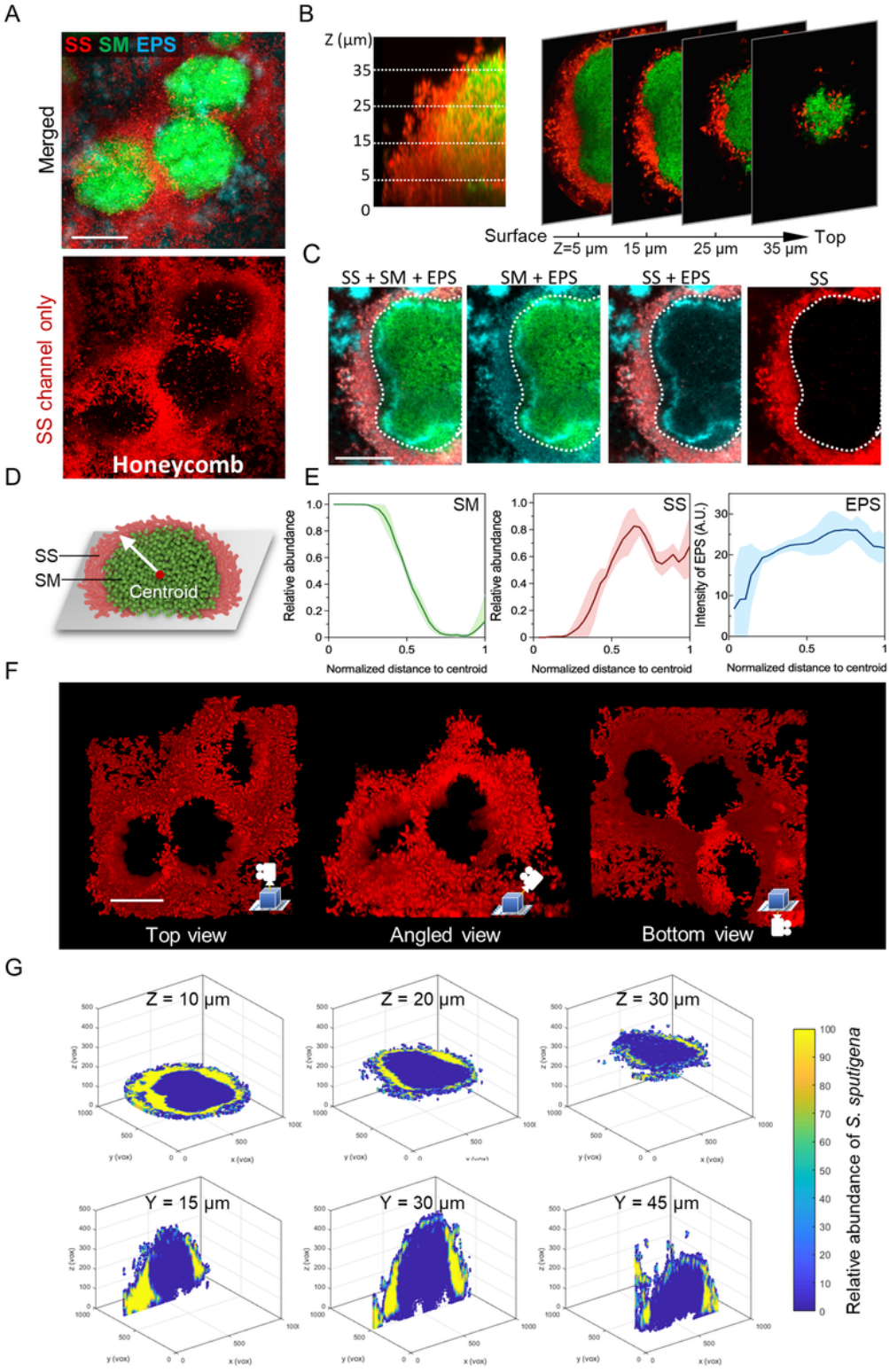


Figure 7

Biogeography of *S. sputigena-S. mutans* biofilm. (A) *S. sputigena-S. mutans* mixed biofilm featured by a honeycomb-like structure formed by *S. sputigena* (in red) encapsulating cell clusters of *S. mutans*. Both species are embedded by EPS α-glucan matrix (in cyan) (B) Optical sectioning of a mixed-species microcolony formed by *S. sputigena* and *S. mutans* reveals a corona-like segregation of *S. mutans* (green) and *S. sputigena* (red) within a mixed microcolony. (C) Cross-sectional confocal image

demonstrating the spatial arrangement of the *S. sputigena, S. mutans* and EPS. *S. mutans*-derived EPS (cyan) distribute both inside and outside of the *S. mutans* cluster (green) in the core; S. *sputigena* cells distribute predominantly outside, embedded in EPS while encapsulating *S. mutans* cluster (D) A diagram of *S. sputigena*-*S. mutans* microcolony structure and the centroid of mass used for spatial analysis. (E) Spatial distribution of *S. sputigena, S. mutans* and EPS relative to the distance to the centroid. Lines correspond to mean and shaded regions to standard deviation from three independent experiments. (F) three-dimensional rendering of the honeycomb structure. (G) Spatial distribution (relative abundance) of *S. sputigena/S. mutans* within the microcolony at different cross-sectional (top) and coronal planes (bottom). SM, *S. mutans*; SS, *S. sputigena*. Scale bars, 20 μm. (Figure is placed in next page)

Figure 7. *In vivo* study of *S. sputigena-S. mutans* co-infection.

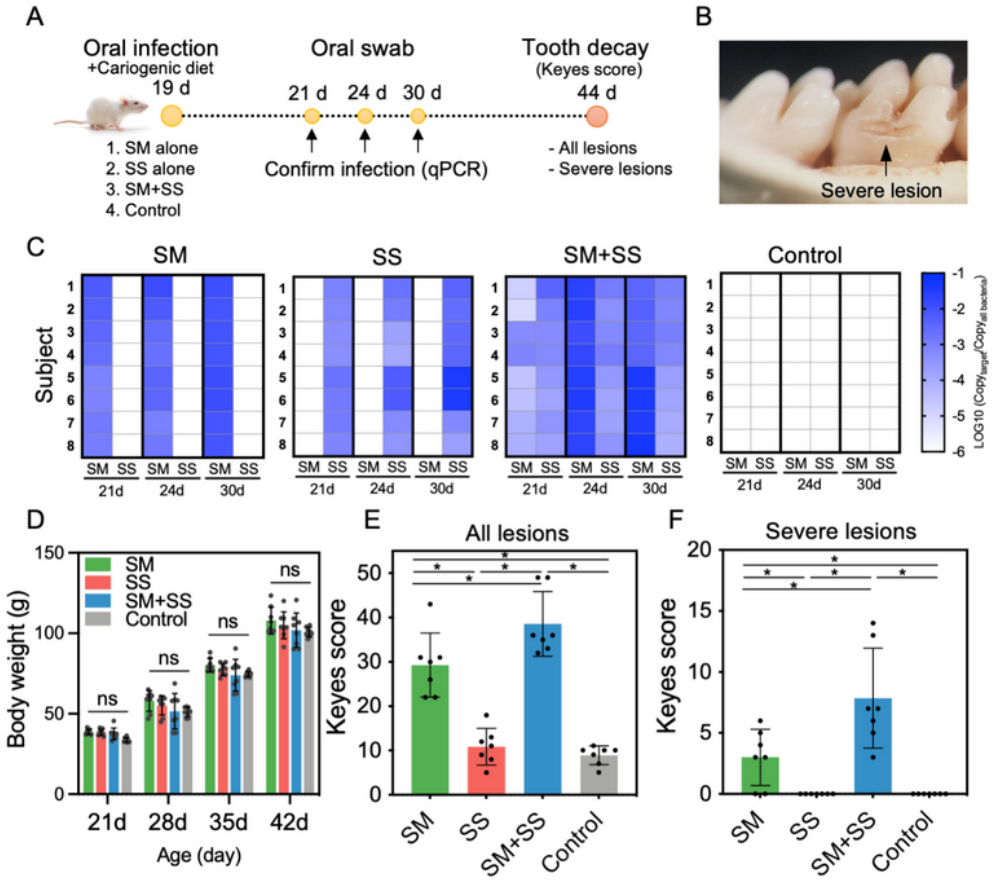


Figure 8

***In vivo* study of *S. sputigena-S. mutans* co-infection.** (A) A diagram of the experimental design. (B) A demonstration of cavitated (severe) carious lesions developed on animal teeth, similar to those found clinically in severe childhood tooth-decay. (C) Confirmation of co-infection by qPCR. Oral swabs were taken on Day 21, Day 24 and Day 30 and were subject to qPCR analysis using species specific probes. (D) Body weight of the animals during the experiment period. Body weights of animals were measured

weekly to monitor the systematic impact of the bacterial infection on animal health. No significant difference was observed between the body weights in the uninfected control versus experimental groups. (E) Keyes scoring of total carious lesions (tooth-decay) developed on the smooth surfaces. (F) Keyes Scoring of cavitated (severe) lesions developed on the smooth surfaces. Caries scores were recorded as stages and extent of carious lesion severity according to Larson's modification of Keyes' scoring system. Data are presented as mean ± standard deviation. (n = 8). *, p<0.05, derived from one-way ANOVA with post hoc Tukey HSD test.

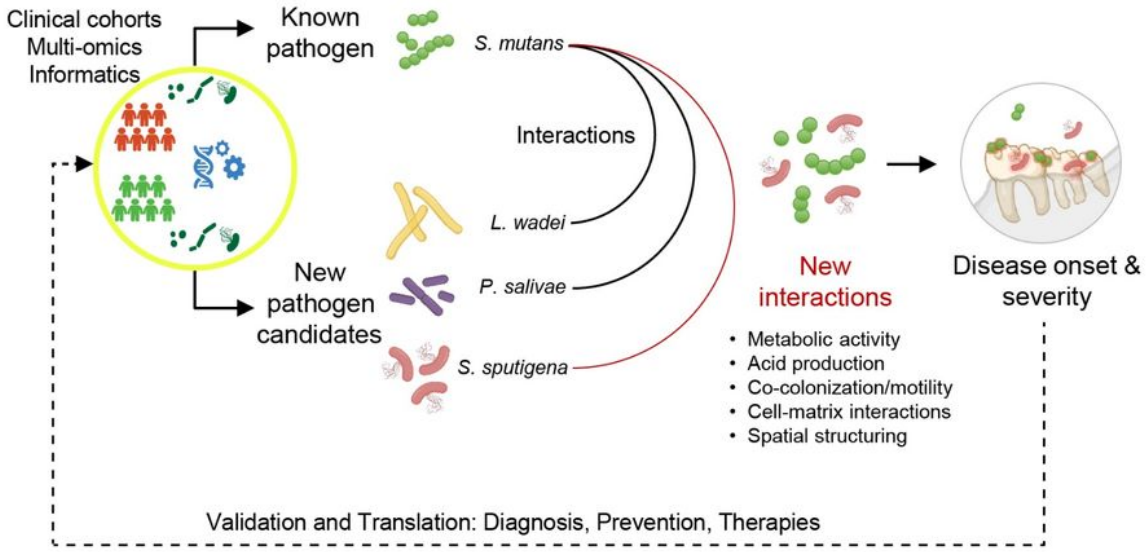


Figure 9

Application of the multi-method, discovery-validation pipeline in the present study.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ExtendedData.docx
- 02.pdf
- MovieS1SS.avi
- MovieS2SMSS.avi
- MovieS3honeycomb.avi
- checklists.pdf