Tracing of streptococcal strains from infant stool across human body sites links gut-specificity to adhesins

Ida Ormaasen[a]#, Morten Kjos[a], Melanie Rae Simpson[b], Torbjørn Øien[b], Lars Snipen[a] and Knut Rudi[a]

[a]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

[b]Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway

#Address correspondence to Ida Ormaasen, ida.ormaasen@nmbu.no

Abstract

Streptococcal species are human commensals known to colonize multiple body sites. Despite being early gut colonizers, we lack strain-level information about their origin and persistence in the gut. To gain more insight into the habitats of the streptococci present in the infant gut, we did a systematic study where mother-infant pairs were sampled from multiple body sites (stool, oral cavity, vagina, breast milk). We performed whole metagenome sequencing and isolated streptococci from 100 infant stool samples (collected at 10 days of age). To trace the streptococci at the strain level, we designed selective qPCR primers for seven streptococcal strains, and these were later utilized to screen corresponding samples from multiple body sites of the infants and their mothers. We found that two of the strains (one *Streptococcus parasanguinis* and one *Streptococcus vestibularis*) were highly prevalent in stool samples, both from infants during their first 2 years of life and from their mothers, indicating that these strains are adapted to the gut environment. Interestingly, another *S. parasanguinis* strain, closely related to the gut-prevalent strain, showed a completely different prevalence pattern, and was mainly detected in vaginal swabs, breast milk and oral swabs. Comparisons of their genomes revealed major differences in genes encoding adhesins, suggesting that host surface attachment could be a key factor for the observed differences in body site specificity. Together, our extensive tracing of streptococci across body sites of 100 infants and their mothers, provide strain-level information of prevalence patterns and reveal the presence of gut-specific streptococci.

Importance

Streptococci thrive on the mucosal surfaces covering the human body and colonize multiple body sites. To determine the distinct streptococcal composition in each habitat and to evaluate their presence in different habitats, strain-level characterization is crucial. We show that two closely related strains, both isolated from stool, are distributed differently across the human body, with one of them prevalent in the stool samples and the other more prevalent in other samples. This emphasizes the necessity of strain-level analysis for the identification of true colonizers of a habitat.

Introduction

After birth, the infant's body undergoes an immediate bacterial colonization. While most colonizers are transient, other persist (1, 2). The *Streptococcus* genus is a diverse group of both beneficial and harmful species that can colonize multiple sites of the human body (3). Shortly after birth, streptococci colonize the infant gut (4-6), with high abundance the first weeks (7-9). These bacteria remain as members of the gut microbiota throughout life, although their relative abundance significantly decreases after a few months post-delivery (10-12). Studies suggest that streptococci can be transmitted to the infant gut from maternal reservoirs (13-15). Although mechanisms for gut-bacterial sharing early in life have been studied (16), we still lack knowledge about the gut microbiota establishment at the strain level. To identify gut-colonizing streptococci and their potential maternal reservoirs, it is crucial to trace individual isolates and further characterize these for phenotypic traits that can explain the colonization.

More than 30 streptococcal species have been isolated from the human body (17). Although some are pathogenic (18), most species are commensals (17, 19) or even display probiotic effects (20, 21). As facultative anaerobes, streptococci obtain energy by fermenting simple sugars such as glucose and lactose into lactate, both in the presence or absence of oxygen (22). Such versatility is important, since oxygen is depleted in the infant gut the first days after birth, leading to an anaerobic gut environment and subsequent establishment of obligate anaerobes (4). Some studies indicate that streptococci may have a key role in the adult gut, more specifically in the small intestine. It has been reported that *Streptococcus* is one of the most abundant genera in the ileal microbiota (11, 12, 23), and omics analyses of the ileal streptococci and their metabolic activity imply that these species are involved in the carbohydrate degradation that takes place in the small intestine (19, 24, 25). Considering that streptococci may have an important role in the small intestine microbiota, more knowledge on the gut streptococci is needed.

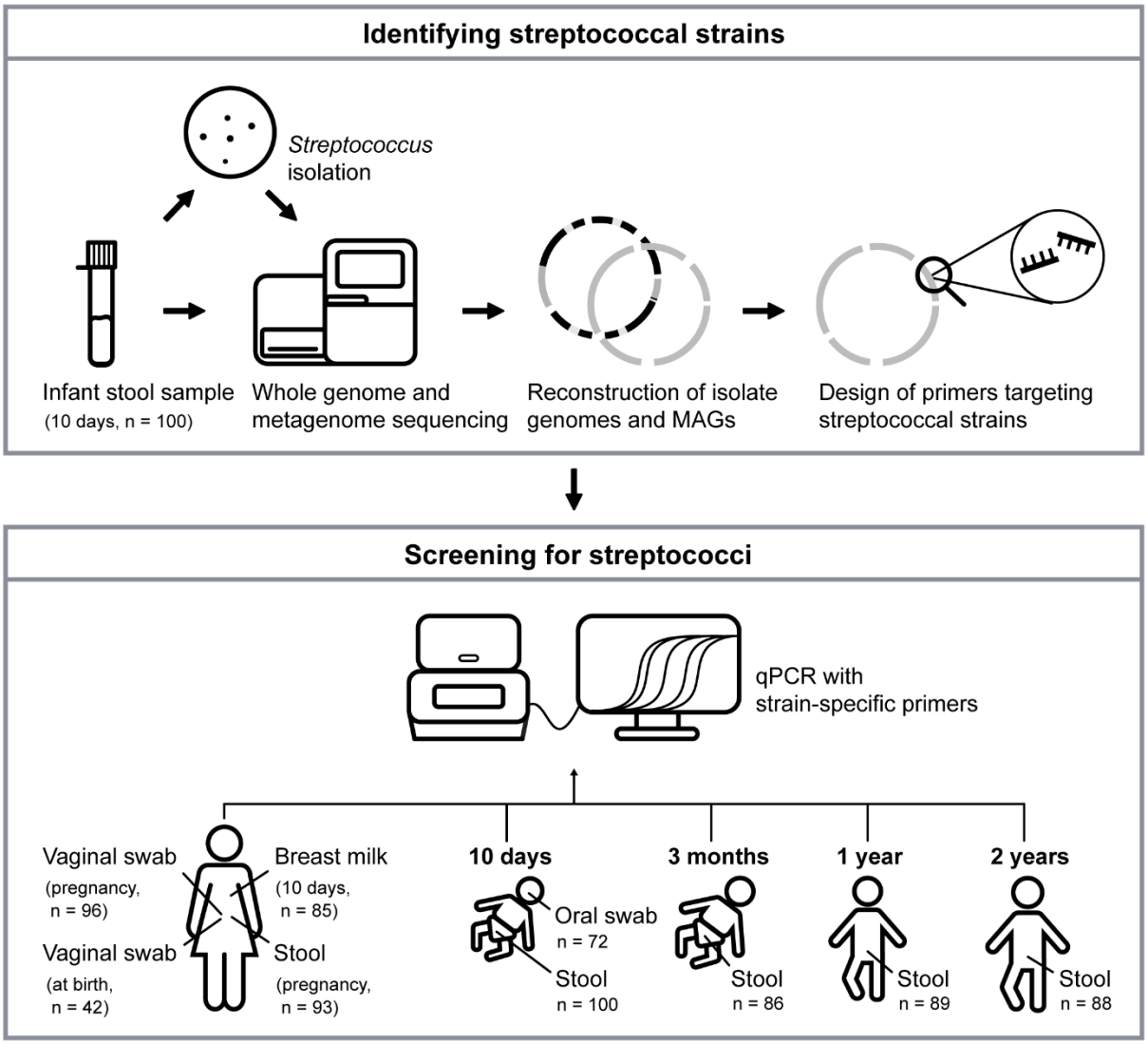
Streptococci are commonly found in the oral cavity, upper respiratory tract, skin and genitourinary system in addition to the gut (26). The oral cavity is colonized by an abundant streptococcal population (27). Reported members of the oral microbiota include *Streptococcus parasanguinis*, *Streptococcus gordonii, Streptococcus salivarius* and *Streptococcus vestibularis*. The three latter species are also associated with the gastrointestinal tract (28-30). Other common human streptococcal species include

71 *Streptococcus pneumoniae* in the upper respiratory tract (31) and *Streptococcus*
72 *agalactiae* in the genitourinary system (32). Moreover, various streptococcal species like
73 *S. parasanguinis, S. salivarius and S. vestibularis* have been found in breast milk (33-35).
74 Different genetic and environmental factors probably influence the body site preference
75 of streptococci, including nutrient availability. Importantly, streptococci produce various
76 adhesins used in host surface attachment (26). These cell surface-associated proteins
77 recognize and bind to receptors found in the different tissues on the human body (36).
78 Therefore, the colonization sites depend on the adhesin targets, in addition to
79 environmental conditions like oxygen and pH levels (26).

80 Given the widespread presence of the *Streptococcus* genus across the human body, strain-
81 level characterization of the streptococcal populations in the different habitats is crucial.
82 Whole metagenome sequencing has made it possible, to some degree, to characterize the
83 most abundant bacterial species in a metagenome at the strain level, making it suitable
84 for determining prevalence and abundance of specific strains in a collection of samples
85 (37). However, further phenotypic characterization of strain-specific traits, beyond what
86 metagenomics can provide, demands possession of strain isolates (38). Furthermore, the
87 possession of strain isolates enables reconstruction of their genomes, which again allows
88 large scale tracing of strains by sensitive and low-cost methods such as real-time
89 quantitative PCR (qPCR) screenings (39, 40). This underlines the importance of bacterial
90 cultivation in combination with metagenomic analyses.

91 Here, we investigated the prevalence of streptococcal isolates across different body sites
92 in 100 mother-infant pairs from the Probiotics in Prevention of Allergy among Children
93 in Trondheim (ProPACT) study (41). The experimental setup is illustrated in Fig. 1. By
94 performing whole genome sequencing of streptococcal isolates derived from infant stool
95 samples collected at 10 days of age, combined with whole metagenomic sequencing of the
96 same samples, we designed primers specific for each strain among the isolates. The strain-
97 specific primers allowed us to perform a comprehensive qPCR-based screen of different
98 samples from the mothers (stool, vagina, breast milk) and their infants (stool, oral cavity),
99 aiming to determine the overall prevalence of the streptococcal strains. Firstly, we
100 present the relative distribution of seven streptococcal strains, in which two are
101 prominent gut-associated strains. Secondly, we suggest that the distinct habitat pattern

102 observed for two closely related strains is partly due to difference in proteins related to
103 host surface attachment.



104

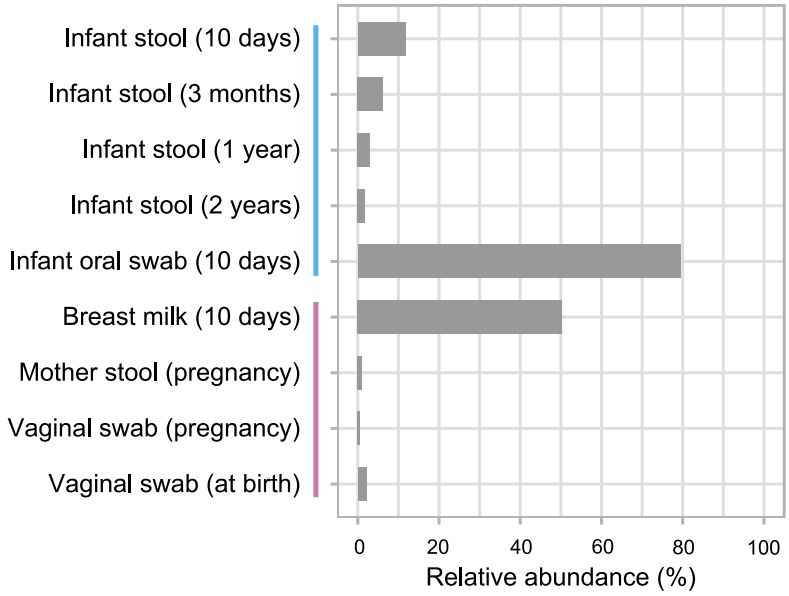
105 **Fig. 1. Experimental setup of the streptococcal isolation and screening.** Stool
106 samples, collected from 10-day-old infants (n=100), were cultivated on streptococcal-
107 selective agar plates to obtain streptococcal isolates. Whole metagenome sequencing of
108 the stool samples and whole genome sequencing of the isolates resulted in the
109 reconstruction of the isolate genomes and streptococcal metagenome-assembled
110 genomes (MAGs). These were used in the design of primers that are specific to each of the
111 streptococcal strains isolated from the stool samples. The designed primers were applied
112 in a qPCR screening to enable the detection of the streptococcal strains and subsequent
113 tracing of these across samples from the infants and their mothers (stool, oral swabs,
114 vaginal swabs, breast milk). The different samples had been collected in week 30 of

115 pregnancy and at birth, as well as 10 days, 3 months, 1 year and 2 years after birth, as
116 shown in the illustration.

117 Results

118 The overall distribution of *Streptococcus* in the different sample types

119 To investigate the presence of streptococci in the collection of samples, we analyzed
120 previously generated 16S rRNA gene sequencing data (41-44). The analysis identified
121 *Streptococcus* species in all the sample types (Fig. 2). Overall, the highest average relative
122 abundances of streptococci were found in the infant oral cavity (79%) and breast milk
123 (50%). In the infant stool samples, the highest average relative abundance was observed
124 in the stool collected at 10 days of age (12%), followed 3 months samples (6%). In the
125 stool samples collected at 1 year and 2 years of age, the average relative abundances had
126 decreased to 3% and 2%, respectively.



127

128 **Fig. 2. Relative abundance of *Streptococcus*.** The figure shows the average relative
129 abundance of Streptococcus in the different sample types, determined by 16S rRNA gene
130 analysis. The time points of sample collection are given in parentheses. The vertical lines
131 on the y-axis indicate the samples from the infants (blue) and the mothers (pink).

132 **Streptococcal diversity in the infant gut microbiota at 10 days of age**

133 Since the 16S rRNA gene analysis revealed that the highest average streptococcal relative
134 abundance in the infant stool samples was at 10 days of age, we selected these for
135 streptococcal-selective cultivation. Out of the 100 samples, 69 samples displayed growth
136 on the streptococcal-selective agar. A total of 177 bacterial isolates were obtained from
137 53 different samples. By Sanger sequencing of the full length 16S rRNA gene, the most
138 abundant genera among the isolates were identified as enterococci (56%), lactobacilli
139 (18%) and streptococci (9%). Further taxonomic classification of the streptococcal
140 isolates (n = 16) was achieved by whole genome sequencing. This revealed that the
141 isolates belonged to seven species, as defined by GTDB taxonomy: *Streptococcus*
142 *parasanguinis_D* (n = 2), *Streptococcus parasanguinis_I* (n = 2), *Streptococcus*
143 *parasanguinis_F* (n = 1), *S.* sp900766505 (n = 3), *Streptococcus agalactiae* (n = 5),
144 *Streptococcus vaginalis* (n = 2), and *Streptococcus vestibularis* (n = 1) (Table S1).

145 Complementary to the cultivation-based approach, we performed whole metagenome
146 sequencing of the same 100 infant stool samples and subsequent assembly, binning and
147 genome reconstruction to identify *Streptococcus* spp.. The sequencing resulted in 112
148 metagenome-assembled genomes (MAGs), of which 14 were classified as streptococci
149 (Table S2). Among the streptococcal MAGs, several represented the same species as the
150 isolates (*S. parasanguinis_D* (n = 1), *S. agalactiae* (n = 1), *S.* sp900766505 (n = 1), *S.*
151 *vaginalis* (n = 1), and *S. vestibularis* (n = 1)). The remaining streptococcal MAGs were
152 assigned to the following species according to the GTDB taxonomy: *S. parasanguinis_S* (n
153 = 1), *S. salivarius* (n = 1), *S. infantis_M* (n = 1), *S. lactarius* (n = 1), *S*. sp001556435 (n = 1),
154 *S*. sp000187445 (n = 1), and *Streptococcus* spp. (n = 3). To investigate the phylogenetic
155 relationship between the streptococcal isolates and MAGs, we determined the estimated
156 evolutionary distances between them. As a reference, we also included complete genomes
157 from the GTDB database that belonged to the same species as the isolates and MAGs. The
158 estimated evolutionary distances, illustrated in the dendrogram in Fig. S1, show that the
159 isolate assemblies cluster according to their assigned species. The same can be seen for
160 most of the MAGs. Furthermore, some of the isolates, including *S. vaginalis* isolate 14, *S.*
161 *agalactiae* isolate 13, *S. parasanguinis* isolates 8 and 6, and *S. parasanguinis* isolates 5 and
162 3 (Fig. S1) seem to be identical to a MAG within the same cluster, since the corresponding
163 estimated evolutionary distances are close to zero. Based on the estimated evolutionary

164 distances between the 16 isolates, we concluded that these represent nine different
165 strains, including two *S. agalactiae*, three *S. parasanguinis* (*S. parasanguinis_D*, *S.*
166 *parasanguinis_I* and *S. parasanguinis_F* according to GTDB taxonomy), one *S.*
167 sp900766505, two *S. vaginalis* and one *S. vestibularis*. These will be referred to as strains
168 below. To distinguish between the strains belonging to the same species, a number was
169 assigned to their names (Table S1).

170 We investigated how abundant the nine streptococcal strains and the streptococcal MAGs
171 were across the infant stool samples at 10 days of age (see Fig. S2 for total bacterial
172 composition in the infant stool samples). On average, the relative abundance of
173 streptococci in these samples, determined by whole metagenome sequencing, was 6.3%,
174 and the highest occurring abundance in a sample was 73.3% (Fig. 3). Both the strains and
175 the streptococcal MAGs not represented by any of the strains contributed substantially to
176 the relative abundance in the samples (Fig. 3). Notably, the *S. vestibularis* strain was
177 substantially more prevalent and abundant than the other strains, contributing to at least
178 1% of the bacterial composition in 28% of the samples (Fig. 3).

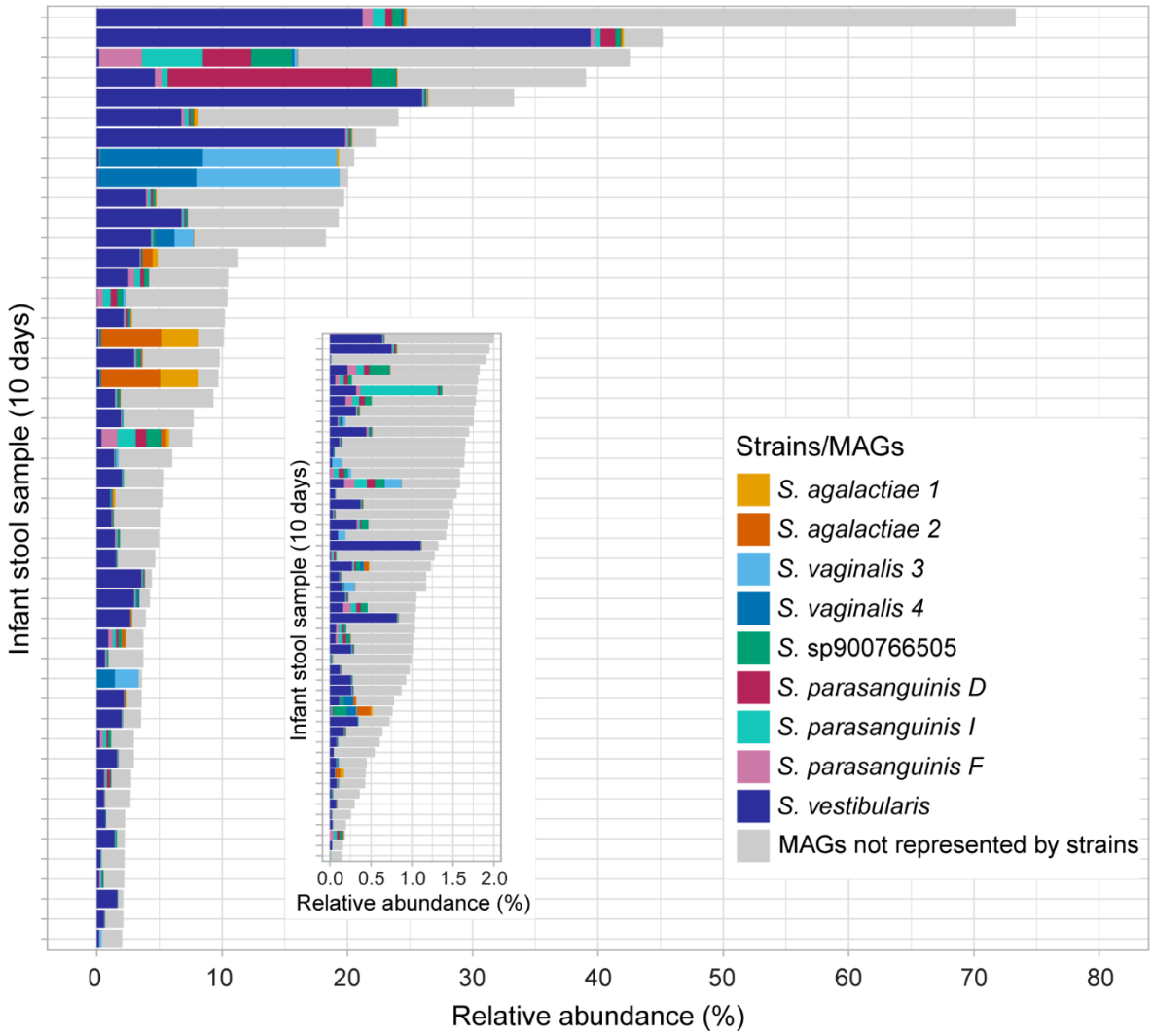


Fig. 3. Streptococci in the infant gut at 10 days of age. The relative abundance of streptococci, as determined by whole metagenome sequencing, is shown. The relative abundances of the nine streptococcal strains in the infant stool samples are indicated in colors. The relative abundances of the streptococcal MAGs with a different taxonomic classification than the nine strains are shown in grey. Note that the stool samples with a streptococcal relative abundance less than 2% are shown in a different scale.

Streptococcal screening of infant- and mother-derived samples reveals strain-specific prevalence patterns

To study the prevalence of the strains across the samples collected from the mother-infant pairs, we performed a screening with strain-specific primers (Fig. 1). The primers were thoroughly designed to target nucleotide sequences unique to each strain. Systematic testing of the primers showed that 7 out of the 9 designed primers could successfully be used to specifically identify the targeted strain (we were not able to make strain-specific

193 primers for the strains *S.* sp900766505 and *S. parasanguinis D*). These seven primer pairs
194 were used further in the qPCR-based tracing of strains. The results from the primer design
195 and testing can be found in Table S3.

196 In line with the metagenome data (Fig. 3), the qPCR screening for the streptococcal strains
197 across the samples revealed that the *S. vestibularis* strain was by far the most prevalent
198 strain in the infant stool samples collected at 10 days of age, being detected in 84% of the
199 samples (Fig. 4). Further, the *S. vestibularis* strain was frequently detected in the infant
200 oral samples (44%), the breast milk samples (32%) and the mothers' stool samples (44%)
201 (Fig. 4), making it the most prevalent strain across the whole sample collection (27%).
202 The second most overall prevalent strain was *S. parasanguinis F*, which was found to be
203 present in 19% of the sample collection. This strain was detected across most of the
204 sample types, but it was particularly prevalent in breast milk (53%) and vaginal samples
205 (pregnancy) (51%) (Fig. 4). In contrast, the closely related strain *S. parasanguinis I*
206 displayed a completely different detection pattern than *S. parasanguinis F* (Fig. 4). *S.*
207 *parasanguinis I* was detected mainly in stool samples, both in mothers' stool and infant
208 stool collected at 10 days, 1 year and 2 years of age (38%, 14%, 35% and 19%,
209 respectively). When it comes to the two *S. vaginalis* strains, they shared a similar pattern
210 of detection across the samples (Fig. 4). They were detected mainly in mothers' stool
211 samples, vaginal samples during pregnancy and infant stool samples (10 days). The two
212 *S. agalactiae* isolates were the least prevalent strains in the screening, with only six
213 detections in total.

214 Detections of the same strain in both infant- and mother-derived samples within one
215 mother-infant pair were frequently made (Fig. S3). There was a statistically significant
216 association between *S. vestibularis* detection in the infant oral cavity and breast milk
217 (Fisher's test, $p < 0.005$). The odds ratio of the statistical test indicated that it was 10.8
218 times more likely that *S. vestibularis* was detected in the breast milk if it was detected in
219 the infant oral cavity as well. When testing for associations with strains detected in infant
220 stool samples, no statistically significant associations were found. The p-values and odds
221 ratios for all tested associations can be found in Table S4.

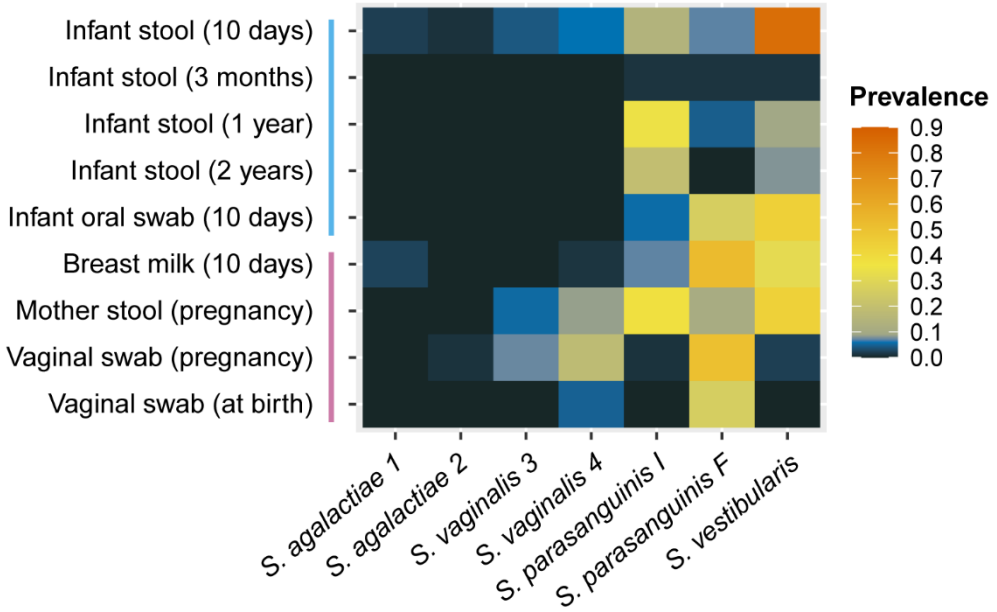


Fig. 4. Strain prevalence. Prevalence of the strains in all the samples from maternal and infant body sites at several time points, based on a qPCR screen with strain-specific primers. The time points of sample collection are given in parentheses. The vertical lines on the y-axis indicate the samples from the infants (blue) and the mothers (pink). A detailed overview of the strain detections in each sample is given in Fig. S3.

Phenotypic and genotypic comparison of the strains

To investigate whether the genetically different strains also exhibited phenotypic distinct traits, their capability to utilize various substrates, including carbohydrates, was tested using the Rapid ID 32 Strep test. Although many of the strains displayed enzymatic activity towards the same substrates, there were differences in substrate utilization among the strains, even between the strains belonging to the same species or closely related species (Fig. S4). The differences seen for the closely related *S. parasanguinis F* and *I* strains were the latter's additional capability to utilize D-ribose D-mannitol, D-trehalose, D-melibiose, pullulan, resorufin-ßD-glucopyranoside and methyl-ßD-glucopyranoside. Among all the strains, *S. vestibularis* was the only strain capable of utilizing urea.

The strikingly opposite prevalence patterns of the two closely related *S. parasanguinis* strains prompted us to further investigate these genomes in more detail. We compared their genomes to identify dissimilar genes (see Methods). We anticipated some gene hits (instances where a query from the one strain aligns, to some degree, with a position on a

243 contig from the other strain) to be the same, regardless of the strain used as query
244 (whether *S. parasanguinis I* was blasted against *S. parasanguinis F* or vice versa). The
245 comparison resulted in 28 gene hits for *S. parasanguinis I* and 32 for *S. parasanguinis F*, in
246 which 16 of the hits were the same for both strains (Table S5). Notably, 6 of these 16 hits
247 were genes that encode surface proteins and putative adhesins potentially involved in
248 colonization (Table S5). In addition, two adhesin genes were only found in the *S.*
249 *parasanguinis I* strain, and three other adhesin genes were detected solely in the *S.*
250 *parasanguinis F* strain. A closer inspection of the sequence alignments and domain
251 organization in these putative adhesins (11 in total) revealed sequence differences and
252 variable numbers of domain repeats (Table S6). For example, we found dissimilarities in
253 surface attachment-related protein domain regions in the genes encoding the GbpC/Spa
254 domain-containing protein, the SspB-related isopeptide-forming adhesin and the
255 CshA/CshB family fibrillar adhesin-related protein, as illustrated in Fig. S5.

256 Discussion

257 Streptococcal species are known to colonize a number of body sites; however, strain-level
258 information is mostly lacking. Here we performed a large-scale tracing of individual
259 streptococcal strains isolated from stool samples of 10-day-old infants across body sites
260 of 100 mother-infant pairs. Most prominently, this strain-level mapping revealed gut-
261 adapted streptococcal strains, suggested that the maternal gut microbiota serves as a
262 reservoir for these, and pointed towards genes related to host surface attachment as one
263 of the key drivers for the body site preference of streptococcal strains.

264 The two strains *S. parasanguinis I* and *S. vestibularis* were highly prevalent in the infant
265 stool samples collected at different time points up to 2 years. Our strain-specific tracing
266 revealed that these strains were also the most prevalent in the stool samples of the
267 mothers, indicating that they are adapted to the human gut environment. In line with
268 these results, Ferretti et al. also found that bacterial strains present in mothers' stool
269 samples were more persistent in the infant gut compared to strains from other sample
270 types (15). Together with our results, this strongly suggest that the maternal gut
271 microbiota serves as an important reservoir for gut-colonizing streptococcal strains in the
272 infant. The ecological niches of streptococci in the gut are not well elaborated, but
273 streptococci have been suggested to thrive within the small intestine (11, 19, 24, 25). We
274 specifically looked for the *S. parasanguinis I* and *S. vestibularis* strains in stool samples,

275 however, their representative species have been found in effluents from the small
276 intestine (19, 24). This indicates that these strains might be able to occupy niches in the
277 small intestine.

278 The two closely related strains *S. parasanguinis I* and *F* displayed distinct prevalence
279 patterns across the samples from different body sites (Fig. 4). This indicates that despite
280 their genetic similarity, these strains colonize different habitats. *The S. parasanguinis I*
281 strain occurred mainly in the stool samples, while the *S. parasanguinis F* strain was
282 frequently found in the other sample types, including the infant oral samples. Recognized
283 as a common member of the oral microbiota (27), the species *S. parasanguinis* has been
284 speculated to appear in the gut due to a translocation from the mouth (45), in which the
285 oral-adapted species only transiently appears (46, 47). Our finding suggests that there is
286 a more nuanced explanation for *S. parasanguinis* presence in the gut than previously
287 proposed, involving gut colonization of specific *S. parasanguinis* strains.

288 The habitat differences seen for the *S. parasanguinis I* and *F* strains were further
289 investigated by substrate utilization tests and by searching for strain-specific variation in
290 their genomes. The two strains displayed differences in substrate utilization capabilities,
291 and this may be a contributing factor to the habitat preferences of the strains. Moreover,
292 the genome comparison also revealed that a major fraction of the genetic variation
293 between the strains were found in genes encoding adhesins or adhesin-associated
294 proteins, such as the GbpC/Spa domain-containing protein, the SspB-related isopeptide-
295 forming adhesin and the CshA/CshB family fibrillar adhesin-related protein (Table S6).
296 Genetic variations were found in regions encoding protein domains involved in surface
297 attachment. An example of this was observed in the GbpC/Spa domain-containing protein,
298 where the number of repeats in the cell surface antigen I/II A repeat domain varied.
299 Streptococci are known to produce a great variety of adhesins (48), such as the antigen
300 I/II (AgI/II) polypeptide family adhesins (49). Orthologous AgI/II proteins are found
301 throughout the *Streptococcus* genus. Species-specific variations among these include
302 different number of repeats in the alanine-rich repeats (A region) and the proline-rich
303 repeats (P region), as well as dissimilarities in the flanking variable region (V region) (49,
304 50). Adhesin genes from streptococcal strains within one species have been reported to
305 display high sequence conservation (51, 52). Based on this, it has been suggested that the
306 divergence seen in orthologous AgI/II adhesins is due to niche adaptations evolved in the

307 various species (49, 50). In contrast, our results showing differences in the prevalence
308 patterns between *S. parasanguinis I* and *S. parasanguinis F* suggest that distinct sets of
309 adhesins used for attachment to host surfaces may also play an important role for strain-
310 level variation in colonization.

311 Three out of seven strains in the screening were not detected in the samples from which
312 they were originally isolated, demonstrating that the strict criteria affected the strain
313 detection sensitivity. This is partly caused by the chosen study design, in which strict
314 terms were set for the screening analysis to ensure high specificity. Furthermore, the
315 streptococcal-selective cultivation resulted in the revival of only 16 streptococcal isolates,
316 belonging to nine strains. Although the whole metagenome sequencing identified
317 additional species, the distribution of the streptococcal strains and MAGs in the samples
318 shows that the isolated strains are representative for the streptococcal population in the
319 infant gut 10 days after birth in our study population. Notably, while the same strains
320 were also detected in the infant stool after 1 and two years, few detections of these strains
321 were made in the infant stool samples collected at 3 months of age. The 16S rRNA gene
322 sequencing results showed that *Streptococcus* was present in the gut at this time point.
323 The reason for this discrepancy is not known. A possible explanation for this may be that
324 the strains found to be prevalent in the stool samples at the other time points are
325 outperformed by other *Streptococcus* species or strains at this stage in the gut microbiota
326 development, before reaching a new bloom within 1 year of age. However, this possible
327 temporal variation needs to be further investigated.

328 In summary, this comprehensive screening of samples from infant and maternal body
329 sites reveals streptococcal species that are likely to thrive in the human gut environment.
330 Further, distinct habitats, where the gut habitat was one of them, were observed for
331 closely related species, possibly due to niche adaptations evolved in genes encoding
332 proteins that facilitate host surface attachment. Further research, performed on a larger
333 cohort and additional streptococcal isolates, is needed to determine phenotypic traits
334 involved in gut microbiota colonization. Particularly, mechanistic studies on adhesins are
335 essential to determine their contribution to body site-specificity among closely related
336 commensal streptococci.

337 # Materials and Methods

338 ## Sample description

339 The 751 samples in this study were collected from 100 mother-infant pairs participating
340 in the ProPACT study (41, 53, 54). The majority of the infants were born at term and
341 breastfed during the first 3 months post-delivery (41). The infant stool samples were
342 collected 10 days after birth (n = 100), 3 months after birth (n= 86), 1 year after birth (n
343 = 89) and 2 years after birth (n = 88). The infant oral swabs (n = 72) were collected 10
344 days after birth. Stool samples (n = 93) and vaginal swabs (n = 96) from the mothers were
345 collected in week 30 of pregnancy. In addition, 42 vaginal swabs were collected at birth.
346 Breast milk samples (n = 85) were collected 10 days after birth. Out of the 100 mother-
347 infant pairs, 20 pairs had a complete set of samples. The samples were stored at -40°C
348 (vaginal swabs and breast milk) and -80°C (stool) before the analysis.

349 The study was approved by the Regional Committee for Medical Research Ethics Central
350 Norway (Ref. 2012/2123).

351 ## Isolation of *Streptococcus* from infant stool samples collected at 10 days of age

352 For selection of streptococci, Streptococcal Selective Agar C.O.B.A. plates (Thermo
353 Scientific Oxoid) were used for cultivation. The 100 infant stool samples, stored in Cary-
354 Blair transport medium (St. Olavs Hospital, Norway) at -80°C, were thawed on ice before
355 50 µl from each sample was spread onto C.O.B.A. plates. The plates were incubated
356 anaerobically, using a GasPak system and AnaeroGen sachets (Oxoid, Thermo Scientific),
357 at 37°C for 2 days. After the incubation, a maximum of 5 colonies with unique
358 morphologies were picked from each plate and inoculated in 5 ml Todd Hewitt broth
359 (Millipore), followed by an anaerobic incubation in the GasPak system at 37°C overnight.
360 Glycerol stocks from each culture with visible growth were made and stored at -80°C. To
361 identify the isolated bacteria, Sanger sequencing of the 16S rRNA gene was performed on
362 amplicons generated with the cover all primer pair from Genetic Analysis (55).

363 ## DNA extraction

364 DNA extraction of the infant oral swabs, the infant stool samples collected 3 months, 1
365 year and 2 years after birth, and the mothers' stool samples were performed prior to this
366 study (41, 43), using the LGC Mag Midi DNA extraction kit (LGC Genomics, UK). DNA
367 extraction of the infant stool samples (10 days), vaginal samples and breast milk samples,

368 as well as the streptococcal isolates and a ZymoBIOMICS Microbial Community Standard
369 (Zymo Research, USA), was performed using the Zymo Research Quick-DNA Fecal/Soil
370 Microbe 96 MagBead kit (Zymo Research, USA). The samples were extracted following the
371 manufacturer's recommendations, with some exceptions elaborated below. Before the
372 extraction, the sample material was concentrated. Specifically, for the infant stool samples
373 (10 days), 500 µl of sample was centrifuged at 13000 rpm for 5 min at 4°C, before the
374 supernatant was removed. Similarly, 500 µl from the vaginal and breast milk samples
375 were centrifuged at 21 500 x g for 30 min at 4°C before the supernatant was removed. The
376 pellets from all the sample types were resuspended in 350 µl lysis buffer (Bashing Bead
377 Buffer) and transferred to FastPrep tubes containing acid-washed glass beads of three
378 sizes; 0.2 g of the <106 µm beads, 0.2 g of the 425-600 µm beads and 2 of the 2.5-3.5 mm
379 beads (Sigma-Aldrich, Germany). *E. coli* 11775 was used as a positive DNA extraction control
380 and lysis buffer (Bashing Bead Buffer) was used as a negative DNA extraction control. The
381 samples were treated in a TissueLyser II (Qiagen) for 2 x 2.5 min at 30 Hz and centrifuged
382 at 13000 x g for 1 min. The DNA was eluted in 30 µl elution buffer to ensure adequate DNA
383 concentrations and quantified with the Qubit dsDNA HS assay kit (Thermo Fisher
384 Scientific, USA).

385 **Library preparations for whole genome and metagenome sequencing**

386 Library preparation of the stool samples (10 days) was performed with the Illumina DNA
387 Prep kit (Illumina, USA), adding 2-250 ng DNA. Extracted DNA from the ZymoBIOMICS
388 Microbial Community Standard and ZymoBIOMICS Microbial Community DNA Standard
389 (Zymo Research, USA) was included. The pooled library was sequenced on NovaSeq6000
390 (Illumina, USA) at Novogene (UK), and 30 million reads per sample (2x150 paired-end
391 reads) were requested. Library preparation of the streptococcal isolates, and the
392 following sequencing on NovaSeq 6000 (Illumina, USA) (2x150 paired-end reads) was
393 performed by Novogene (UK).

394 **Whole genome and metagenome data processing and analysis**

395 The metagenome-assembled genomes (MAGs) were constructed as follows: The
396 metagenomic raw reads were filtered and trimmed with BBDuk
397 (https://sourceforge.net/projects/bbmap/) and mapped to the human genome using
398 Bowtie 2 (56) for decontamination of human DNA. The decontaminated reads were
399 assembled with SPAdes (57). The binning of assembled contigs was done twice using

400 MaxBin2 (58) and MetaBAT2 (59). The MAGs were taxonomically classified using the
401 newest release (R220) of the GTDB database (60).

402 Reconstruction of the isolate genomes included filtering, trimming, assembly of reads and
403 taxonomic classification of assembled contigs, using the same tools and databases as
404 mentioned for the MAGs. To establish the phylogenetic relationship between the isolates,
405 estimated evolutionary distances were calculated using the MASH tool (61). The distances
406 revealed that some of the isolates were phylogenetically identical. Based on this, a new
407 assembly was done where the reads from the identical isolates were combined to increase
408 coverage, resulting in 9 strain assemblies.

409 The phylogenetic relationship between the isolate assemblies, streptococcal MAGs and
410 genomes in the GTDB database was estimated with MASH distances. From the GTDB
411 database, all genomes from each of the isolate/MAG species were selected. Of these, only
412 the genomes with the highest available assembly quality were used further. The GTDB
413 genomes were clustered based on the MASH distances to find one representative genome
414 for each species. A hierarchical clustering of these representative genomes, isolate
415 assemblies and MAGs was done based on the MASH distances.

416 The strain assemblies of *S. parasanguinis I* and *S. parasanguinis F* were compared, aiming
417 to find genetic differences. A blast database was made from the *S. parasanguinis I*
418 assembly using BLAST+ (62). Then, the genes in the *S. parasanguinis F* assembly were
419 identified with Prodigal (63) and blasted against the generated database using blastn. The
420 same process was done the other way around, making a database from *S. parasanguinis* F
421 and blasting the genes from *S. parasanguinis* I against it. The blast tables were filtered,
422 keeping the hit with the best bit score in each case. Looking for dissimilar genes, the hits
423 with alignment length to query length ratio lower than 0.75 were kept. The hit query
424 sequences were translated to proteins, and a blastp search against the nr-database at
425 NCBI (64) was done for identification. Protein sequence alignments between the matching
426 genes were performed with Clustal Omega (1.2.4) (65, 66), and the protein domains were
427 identified using InterProScan (67) with the InterPro database (68).

428 **Data availability**

429 The whole genome sequencing data of the 16 streptococcal isolates are available in the
430 NCBI SRA database, with the BioProject ID PRJNA1168056.

431 The 112 MAGs and their corresponding coverage table are available in
432 https://arken.nmbu.no/~larssn/MiDiv/ida/.

433 **Primer design and testing**

434 The nine streptococcal strain assemblies were used as template sequences, and the RUCS
435 tool was used for primer generation (69). The specificity of a primer-pair was ensured by
436 specifying for each chosen positive genome, a set of negative genomes where potential
437 primers could not produce matches. For each assembly used as positive, the other
438 assemblies, as well as all the MAGs from the infant stool samples, were set as negatives.
439 In the cases where a MAG was very similar to an isolate, the MAG was set as positive as
440 well.

441 The primers were tested *in silico* against all the strain assemblies and MAGs, using
442 Geneious Prime 2023.1.2 (https://www.geneious.com), to validate their specificity.
443 Similarly, potential binding towards common gut bacterial genomes from the NCBI
444 Reference Sequence Database was investigated (Table S7). Two primer pairs for each
445 strain were chosen based on their specificity and low hairpin melting temperature. These
446 were finally tested with Primer-BLAST (70).

447 The specificities of the chosen primers were tested further using qPCR, by investigating
448 their ability to amplify DNA from all the 16 streptococcal isolates. The annealing
449 temperature of the primers was set to the same as their melting temperature (Tm = 60°C)
450 to keep the occurrence of unspecific binding to a minimum. The reaction mix and
451 amplification program used were the same as described for the streptococcal screening
452 (see below). The best primer pair for each strain was determined by its specificity in the
453 qPCR test run. The primer pair that showed the largest number of cycles between the true
454 signal from the strain in question and signals from DNA amplification from other strains
455 was used further. The characteristics of each primer can be found in Table S4.

456 **Screening of streptococcal strains by qPCR**

457 The 751 samples were screened for the streptococcal strains. The total bacterial load in
458 the samples was determined by amplification of the V3-V4 region of the 16S rRNA gene,
459 using the primer pair PRK341F (5'-CCTACGGGRBGCASCAG-3') and PRK806R (5'-
460 GGACTACYVGGGTATCTAAT-3') (71). The master mix for all qPCR runs contained 1x Hot
461 FirePol EvaGreen qPCR Supermix (Solis Biodyne, Estonia) and 0.2 μM of each primer, in a

462 reaction volume of 20 µl. In each run, DNA from the isolate in question was included as a
463 positive control, and nuclease free water was used as a negative control. The amplification
464 was carried out in a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA). The
465 initial denaturation step was executed at 95°C for 15 min, followed by 40 cycles of
466 denaturation at 95°C for 30 sec and amplification and elongation at 60°C for 60 sec
467 (streptococcal strain-specific primers) or amplification at 55°C for 30 sec and elongation
468 at 75°C for 45 sec (16S rRNA gene primers). A melt curve analysis was included at the end
469 of each run.

470 The qPCR results were analyzed with LinRegPCR (2014.x) (72) for baseline correction. To
471 distinguish true qPCR signals from unspecific amplification, two criteria were set in the
472 processing of the qPCR results; (1) the melting temperature had to match the melting
473 temperature of the amplicon, and (2) the Cq value had to be two cycles lower than the
474 negative control. These criteria are rather strict, lowering the streptococcal strain
475 detection sensitivity, but ensuring a high specificity.

476 When a strain was detected in several samples from two different body sites, Fisher's
477 exact tests were used to test for associations between the samples from the different body
478 sites, i.e. if there is a non-random co-occurrence of the strain in the two body sites. The p-
479 values were corrected with a false discovery rate (FDR).

480 **Characterization of substrate utilization capability**

481 The isolates were grown on different substrates in a Rapid ID 32 Strep test (bioMérieux,
482 France). Prior to the test, isolate cultures were grown in Todd Hewitt broth (Millipore)
483 and incubated anaerobically at 37 °C overnight. The cultures were centrifuged to pellet
484 the cells, which then were washed with 1X PBS and centrifuged once more. The cells were
485 resuspended in MilliQ water to a turbidity of approximately McFarland standard 4.
486 Further, the test was carried out according to the manufacturer's recommendations, and
487 the results were read manually after incubation for 4 hours and overnight. For each
488 isolate, parallel tests were performed.

489 Acknowledgments

490 We thank Ole Hjermundrud (master student, Norwegian University of Life Sciences) for
491 contributing to the primer testing and Kaja Sundquist Wilhelmson (master student,
492 Norwegian University of Life Sciences) for contributing to the qPCR screening.

This research was funded by the Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, NORWAY and the Research Council of Norway through the project UnveilMe: Unveiling the role of microbial metabolites in human infant development (project number 301364).

497 # References

498 1. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community
499 variation in human body habitats across space and time. Science 326:1694-7.
500 2. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. 2007. Development of the human infant
501 intestinal microbiota. PLoS Biol 5:e177.
502 3. Yadav P, Verma S, Bauer R, Kumari M, Dua M, Johri AK, Yadav V, Spellerberg B. 2020.
503 Deciphering Streptococcal Biofilms. Microorganisms 8.
504 4. Bezirtzoglou E. 1997. The Intestinal Microflora During the First Weeks of Life. Anaerobe 3:173-
505 177.
506 5. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010.
507 Delivery mode shapes the acquisition and structure of the initial microbiota across multiple
508 body habitats in newborns. Proc Natl Acad Sci U S A 107:11971-5.
509 6. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, Belzer C, Delgado Palacio S,
510 Arboleya Montes S, Mancabelli L, Lugli GA, Rodriguez JM, Bode L, de Vos W, Gueimonde M,
511 Margolles A, van Sinderen D, Ventura M. 2017. The First Microbial Colonizers of the Human
512 Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiol
513 Mol Biol Rev 81.
514 7. Reddel S, Pascucci GR, Foligno S, Del Chierico F, Vernocchi P, Marzullo A, Pattumelli MG, Palma
515 P, Salvatori G, Putignani L. 2022. A Parallel Tracking of Salivary and Gut Microbiota Profiles Can
516 Reveal Maturation and Interplay of Early Life Microbial Communities in Healthy Infants.
517 Microorganisms 10.
518 8. Solís G, de Los Reyes-Gavilan CG, Fernández N, Margolles A, Gueimonde M. 2010.
519 Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-
520 milk and the infant gut. Anaerobe 16:307-10.
521 9. Hill CJ, Lynch DB, Murphy K, Ulaszewska M, Jeffery IB, O'Shea CA, Watkins C, Dempsey E,
522 Mattivi F, Tuohy K, Ross RP, Ryan CA, O' Toole PW, Stanton C. 2017. Evolution of gut microbiota
523 composition from birth to 24 weeks in the INFANTMET Cohort. Microbiome 5:4.
524 10. Avershina E, Storrø O, Øien T, Johnsen R, Pope P, Rudi K. 2014. Major faecal microbiota shifts
525 in composition and diversity with age in a geographically restricted cohort of mothers and their
526 children. FEMS Microbiol Ecol 87:280-90.
527 11. Booijink CCGM, El-Aidy S, Rajilić-Stojanović M, Heilig HGHJ, Troost FJ, Smidt H, Kleerebezem
528 M, De Vos WM, Zoetendal EG. 2010. High temporal and inter-individual variation detected in
529 the human ileal microbiota. Environmental Microbiology 12:3213-3227.
530 12. Villmones HC, Haug ES, Ulvestad E, Grude N, Stenstad T, Halland A, Kommedal Ø. 2018. Species
531 Level Description of the Human Ileal Bacterial Microbiota. Scientific Reports 8:4736.
532 13. Biagi E, Quercia S, Aceti A, Beghetti I, Rampelli S, Turroni S, Faldella G, Candela M, Brigidi P,
533 Corvaglia L. 2017. The Bacterial Ecosystem of Mother's Milk and Infant's Mouth and Gut. Front
534 Microbiol 8:1214.
535 14. Fehr K, Moossavi S, Sbihi H, Boutin RCT, Bode L, Robertson B, Yonemitsu C, Field CJ, Becker AB,
536 Mandhane PJ, Sears MR, Khafipour E, Moraes TJ, Subbarao P, Finlay BB, Turvey SE, Azad MB.
537 2020. Breastmilk Feeding Practices Are Associated with the Co-Occurrence of Bacteria in
538 Mothers' Milk and the Infant Gut: the CHILD Cohort Study. Cell Host Microbe 28:285-297.e4.
539 15. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, Armanini F, Truong DT, Manara S, Zolfo
540 M, Beghini F, Bertorelli R, De Sanctis V, Bariletti I, Canto R, Clementi R, Cologna M, Crifò T,
541 Cusumano G, Gottardi S, Innamorati C, Masè C, Postai D, Savoi D, Duranti S, Lugli GA,
542 Mancabelli L, Turroni F, Ferrario C, Milani C, Mangifesta M, Anzalone R, Viappiani A, Yassour
543 M, Vlamakis H, Xavier R, Collado CM, Koren O, Tateo S, Soffiati M, Pedrotti A, Ventura M,
544 Huttenhower C, Bork P, Segata N. 2018. Mother-to-Infant Microbial Transmission from
545 Different Body Sites Shapes the Developing Infant Gut Microbiome. Cell Host Microbe 24:133-
546 145.e5.

547 16. Browne HP, Shao Y, Lawley TD. 2022. Mother–infant transmission of human microbiota.
548 Current Opinion in Microbiology 69:102173.
549 17. Baty JJ, Stoner SN, Scoffield JA. 2022. Oral Commensal Streptococci: Gatekeepers of the Oral
550 Cavity. J Bacteriol 204:e0025722.
551 18. Krzyściak W, Pluskwa KK, Jurczak A, Kościelniak D. 2013. The pathogenicity of the
552 Streptococcus genus. Eur J Clin Microbiol Infect Dis 32:1361-76.
553 19. den Bogert Bv, Erkus O, Boekhorst J, Goffau Md, Smid EJ, Zoetendal EG, Kleerebezem M. 2013.
554 Diversity of human small intestinal Streptococcus and Veillonella populations. FEMS
555 Microbiology Ecology 85:376-388.
556 20. Power DA, Burton JP, Chilcott CN, Dawes PJ, Tagg JR. 2008. Preliminary investigations of the
557 colonisation of upper respiratory tract tissues of infants using a paediatric formulation of the
558 oral probiotic Streptococcus salivarius K12. European Journal of Clinical Microbiology &
559 Infectious Diseases 27:1261-1263.
560 21. Burton JP, Drummond BK, Chilcott CN, Tagg JR, Thomson WM, Hale JDF, Wescombe PA. 2013.
561 Influence of the probiotic Streptococcus salivarius strain M18 on indices of dental health in
562 children: a randomized double-blind, placebo-controlled trial. Journal of Medical Microbiology
563 62:875-884.
564 22. Willenborg J, Goethe R. 2016. Metabolic traits of pathogenic streptococci. FEBS Lett 590:3905-
565 3919.
566 23. Hayashi H, Takahashi R, Nishi T, Sakamoto M, Benno Y. 2005. Molecular analysis of jejunal,
567 ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and
568 terminal restriction fragment length polymorphism. Journal of Medical Microbiology 54:1093-
569 1101.
570 24. Van den Bogert B, Boekhorst J, Herrmann R, Smid EJ, Zoetendal EG, Kleerebezem M. 2013.
571 Comparative genomics analysis of Streptococcus isolates from the human small intestine
572 reveals their adaptation to a highly dynamic ecosystem. PLoS One 8:e83418.
573 25. Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CC, Troost FJ, Bork P, Wels M,
574 de Vos WM, Kleerebezem M. 2012. The human small intestinal microbiota is driven by rapid
575 uptake and conversion of simple carbohydrates. Isme j 6:1415-26.
576 26. Nobbs AH, Lamont RJ, Jenkinson HF. 2009. Streptococcus adherence and colonization.
577 Microbiol Mol Biol Rev 73:407-50, Table of Contents.
578 27. Abranches J, Zeng L, Kajfasz JK, Palmer SR, Chakraborty B, Wen ZT, Richards VP, Brady LJ, Lemos
579 JA. 2018. Biology of Oral Streptococci. Microbiol Spectr 6.
580 28. Couvigny B, de Wouters T, Kaci G, Jacouton E, Delorme C, Doré J, Renault P, Blottière HM,
581 Guédon E, Lapaque N. 2015. Commensal Streptococcus salivarius Modulates PPARγ
582 Transcriptional Activity in Human Intestinal Epithelial Cells. PLoS One 10:e0125371.
583 29. Park OJ, Kwon Y, Park C, So YJ, Park TH, Jeong S, Im J, Yun CH, Han SH. 2020. Streptococcus
584 gordonii: Pathogenesis and Host Response to Its Cell Wall Components. Microorganisms 8.
585 30. Gunathilake M, Lee J, Choi IJ, Kim YI, Yoon J, Sul WJ, Kim JF, Kim J. 2020. Alterations in Gastric
586 Microbial Communities Are Associated with Risk of Gastric Cancer in a Korean Population: A
587 Case-Control Study. Cancers (Basel) 12.
588 31. Weiser JN, Ferreira DM, Paton JC. 2018. Streptococcus pneumoniae: transmission,
589 colonization and invasion. Nat Rev Microbiol 16:355-367.
590 32. Laupland KB, Ross T, Pitout JDD, Church DL, Gregson DB. 2007. Community-onset Urinary Tract
591 Infections: A Population-based Assessment. Infection 35:150.
592 33. Heikkilä MP, Saris PEJ. 2003. Inhibition of Staphylococcus aureus by the commensal bacteria
593 of human milk. Journal of Applied Microbiology 95:471-478.
594 34. Jeurink PV, van Bergenhenegouwen J, Jiménez E, Knippels LMJ, Fernández L, Garssen J, Knol J,
595 Rodríguez JM, Martín R. 2013. Human milk: a source of more life than we imagine. Beneficial
596 Microbes 4:17-30.
597 35. Lopez Leyva L, Gonzalez E, Solomons NW, Koski KG. 2022. Human milk microbiome is shaped
598 by breastfeeding practices. Front Microbiol 13:885588.

36. Moubareck CA. 2021. Human Milk Microbiota and Oligosaccharides: A Glimpse into Benefits, Diversity, and Correlations. Nutrients 13.
37. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. 2017. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res 27:626-638.
38. Li Y, Jin Y, Zhang J, Pan H, Wu L, Liu D, Liu J, Hu J, Shen J. 2021. Recovery of human gut microbiota genomes with third-generation sequencing. Cell Death & Disease 12:569.
39. Hernández I, Sant C, Martínez R, Fernández C. 2020. Design of Bacterial Strain-Specific qPCR Assays Using NGS Data and Publicly Available Resources and Its Application to Track Biocontrol Strains. Front Microbiol 11:208.
40. Kralik P, Ricchi M. 2017. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. Front Microbiol 8:108.
41. Avershina E, Lundgård K, Sekelja M, Dotterud C, Storrø O, Øien T, Johnsen R, Rudi K. 2016. Transition from infant- to adult-like gut microbiota. Environ Microbiol 18:2226-36.
42. Avershina E, Slangsvold S, Simpson MR, Storrø O, Johnsen R, Øien T, Rudi K. 2017. Diversity of vaginal microbiota increases by the time of labor onset. Scientific Reports 7:17558.
43. Isaksen SM. 2018. Maternal probiotic consumption and its effects on children's oral microbiota. Master. Norwegian University of Life Sciences.
44. Simpson MR, Avershina E, Storrø O, Johnsen R, Rudi K, Øien T. 2018. Breastfeeding-associated microbiota in human milk following supplementation with Lactobacillus rhamnosus GG, Lactobacillus acidophilus La-5, and Bifidobacterium animalis ssp. lactis Bb-12. Journal of Dairy Science 101:889-899.
45. Kageyama S, Sakata S, Ma J, Asakawa M, Takeshita T, Furuta M, Ninomiya T, Yamashita Y. 2023. High-Resolution Detection of Translocation of Oral Bacteria to the Gut. J Dent Res 102:752-758.
46. Songjinda P, Nakayama J, Kuroki Y, Tanaka S, Fukuda S, Kiyohara C, Yamamoto T, Izuchi K, Shirakawa T, Sonomoto K. 2005. Molecular monitoring of the developmental bacterial community in the gastrointestinal tract of Japanese infants. Biosci Biotechnol Biochem 69:638-41.
47. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, Izard J, Garrett WS, Chan AT, Huttenhower C. 2014. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci U S A 111:E2329-38.
48. Moschioni M, Pansegrau W, Barocchi MA. 2010. Adhesion determinants of the Streptococcus species. Microb Biotechnol 3:370-88.
49. Manzer HS, Nobbs AH, Doran KS. 2020. The Multifaceted Nature of Streptococcal Antigen I/II Proteins in Colonization and Disease Pathogenesis. Front Microbiol 11:602305.
50. Brady LJ, Maddocks SE, Larson MR, Forsgren N, Persson K, Deivanayagam CC, Jenkinson HF. 2010. The changing faces of Streptococcus antigen I/II polypeptide family adhesins. Mol Microbiol 77:276-86.
51. Ma JK, Kelly CG, Munro G, Whiley RA, Lehner T. 1991. Conservation of the gene encoding streptococcal antigen I/II in oral streptococci. Infect Immun 59:2686-94.
52. Rego S, Heal TJ, Pidwill GR, Till M, Robson A, Lamont RJ, Sessions RB, Jenkinson HF, Race PR, Nobbs AH. 2016. Structural and Functional Analysis of Cell Wall-anchored Polypeptide Adhesin BspA in Streptococcus agalactiae. J Biol Chem 291:15985-6000.
53. Dotterud CK, Storrø O, Johnsen R, Øien T. 2010. Probiotics in pregnant women to prevent allergic disease: a randomized, double-blind trial. British Journal of Dermatology 163:616-623.
54. Dotterud CK, Avershina E, Sekelja M, Simpson MR, Rudi K, Storrø O, Johnsen R, Øien T. 2015. Does Maternal Perinatal Probiotic Supplementation Alter the Intestinal Microbiota of Mother and Child? Journal of Pediatric Gastroenterology and Nutrition 61:200-207.
55. Casén C, Vebø HC, Sekelja M, Hegge FT, Karlsson MK, Ciemniejewska E, Dzankovic S, Frøyland C, Nestestog R, Engstrand L, Munkholm P, Nielsen OH, Rogler G, Simrén M, Öhman L, Vatn MH,

Rudi K. 2015. Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. Aliment Pharmacol Ther 42:71-83.
56. Langmead B, Wilks C, Antonescu V, Charles R. 2018. Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics 35:421-432.
57. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70:e102.
58. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605-607.
59. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359.
60. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research 50:D785-D794.
61. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 17:132.
62. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
63. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
64. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res 50:D20-d26.
65. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.
66. Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 27:135-145.
67. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236-1240.
68. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD. 2020. The InterPro protein families and domains database: 20 years on. Nucleic Acids Research 49:D344-D354.
69. Thomsen MCF, Hasman H, Westh H, Kaya H, Lund O. 2017. RUCS: rapid identification of PCR primers for unique core sequences. Bioinformatics 33:3917-3921.
70. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134.
71. Yu Y, Lee C, Kim J, Hwang S. 2005. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. Biotechnology and Bioengineering 89:670-679.
72. Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, van den Hoff MJ, Moorman AF. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 37:e45.