



The Influence of *FUT2* and *FUT3* Polymorphisms and Nasopharyngeal Microbiome on Respiratory Infections in Breastfed Bangladeshi Infants from the Microbiota and Health Study

Aristea Binia,^a Léa Siegwald,^a Shamima Sultana,^b Maya Shevlyakova,^a Gregory Lefebvre,^a Francis Foata,^a Séverine Combremont,^a Aline Charpagne,^{a*} Karine Vidal,^a Norbert Sprenger,^a Mahbubar Rahman,^b Albert Palleja,^c Aron C. Eklund,^c Henrik Bjørn Nielsen,^c Harald Brüßow,^{a§} Shafiqul Alam Sarker,^b  Olga Sakwinska^a

^aNestlé Research, Lausanne, Switzerland

^bInternational Centre for Diarrheal Diseases Research, Bangladesh (icddr,b), Dhaka, Bangladesh

^cClinical Microbiomics, Copenhagen, Denmark

Aristea Binia, Léa Siegwald, and Shamima Sultana contributed equally to this work. Author order was determined alphabetically.

ABSTRACT Acute respiratory infections (ARIs) are one of the most common causes of morbidity and mortality in young children. The aim of our study was to examine whether variation in maternal *FUT2* (α 1,2-fucosyltransferase 2) and *FUT3* (α 1,3/4-fucosyltransferase 3) genes, which shape fucosylated human milk oligosaccharides (HMOs) in breast milk, are associated with the occurrence of ARIs in breastfed infants as well as the influence of the nasopharyngeal microbiome on ARI risk. Occurrences of ARIs were prospectively recorded in a cohort of 240 breastfed Bangladeshi infants from birth to 2 years. Secretor and Lewis status was established by sequencing of *FUT2/3* genes. The nasopharyngeal microbiome was characterized by shotgun metagenomics, complemented by specific detection of respiratory pathogens; 88.6% of mothers and 91% of infants were identified as secretors. Maternal secretor status was associated with reduced ARI incidence among these infants in the period from birth to 6 months (incidence rate ratio [IRR], 0.66; 95% confidence interval [CI], 0.47 to 0.94; $P = 0.020$), but not at later time periods. The nasopharyngeal microbiome, despite precise characterization to the species level, was not predictive of subsequent ARIs. The observed risk reduction of ARIs among infants of secretor mothers during the predominant breastfeeding period is consistent with the hypothesis that fucosylated oligosaccharides in human milk contribute to protection against respiratory infections. However, we found no evidence that modulation of the nasopharyngeal microbiome influenced ARI risk.

IMPORTANCE The observed risk reduction of acute respiratory infections (ARIs) among infants of secretor mothers during the predominant breastfeeding period is consistent with the hypothesis that fucosylated oligosaccharides in human milk contribute to protection against respiratory infections. Respiratory pathogens were only weak modulators of risk, and the nasopharyngeal microbiome did not influence ARI risk, suggesting that the associated protective effects of human milk oligosaccharides (HMOs) are not conveyed via changes in the nasopharyngeal microbiome. Our observations add to the evidence for a role of fucosylated HMOs in protection against respiratory infections in exclusively or predominantly breastfed infants in low-resource settings. There is no indication that the nasopharyngeal microbiome substantially modulates the risk of subsequent mild ARIs. Larger studies are needed to provide mechanistic insights on links between secretor status, HMOs, and risk of respiratory infections.

KEYWORDS *FUT2*, microbiome, respiratory infections, secretor status

Citation Binia A, Siegwald L, Sultana S, Shevlyakova M, Lefebvre G, Foata F, Combremont S, Charpagne A, Vidal K, Sprenger N, Rahman M, Palleja A, Eklund AC, Nielsen HB, Brüßow H, Sarker SA, Sakwinska O. 2021. The influence of *FUT2* and *FUT3* polymorphisms and nasopharyngeal microbiome on respiratory infections in breastfed Bangladeshi infants from the Microbiota and Health Study. *mSphere* 6:e00686-21. <https://doi.org/10.1128/mSphere.00686-21>.

Editor Robert A. Britton, Baylor College of Medicine

Copyright © 2021 Binia et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Olga Sakwinska, olga.sakwinska@rdls.nestle.com.

*Present address: Aline Charpagne, University of Lausanne, Lausanne, Switzerland.

§Present address: Harald Brüßow, Department of Biosystems, Laboratory of Gene Technology, Leuven, Belgium.

Received 7 August 2021

Accepted 21 October 2021

Published 10 November 2021

Acute respiratory infections (ARIs) are a leading cause of morbidity and mortality in early childhood (1). Infant feeding plays a decisive role in modifying susceptibility to ARIs; notably, the absence of breastfeeding is associated with a higher risk of infection, in particular in low-resource settings during early infancy (2–4). Human milk oligosaccharides (HMOs) are the third most abundant solid component of breast milk (5, 6). They have been hypothesized to be one of the key milk components contributing to its protective effect regarding respiratory infections (7, 8).

HMO content of breast milk is known to vary. In particular, FUT2 (α 1,2-fucosyltransferase 2) and FUT3 (α 1,3/4-fucosyltransferase 3) enzymes are needed for production of α 1-2 and α 1-3/4 fucosylated HMOs, respectively (9). *FUT2/3*-inactivating genetic variants, present in up to 30% of the population, lead to nonsecretor/Lewis-negative phenotypes which considerably influence breast milk composition and result in differences in overall quantity and concentrations of individual HMOs (10, 11). Breast milk of nonsecretor (*FUT2*-negative) individuals does not contain α 1-2 fucosylated HMOs, for example 2' fucosyllactose (2'FL), which is the most abundant HMO in secretor milk, or lacto-*N*-fucopentaose I and III (LNFP-I and LNFP-III). On the other hand, HMOs such as lacto-*N*-fucopentaose II (LNFP-II) and 3-fucosyllactose (3FL) are absent or very low in breast milk from Lewis-negative (*FUT3*-negative) mothers. Secretor and Lewis status thus highly correlate with both individual HMOs but also HMO milk groups (11) and can therefore be used as good proxies of HMO content when milk analysis is not feasible.

The variation in composition and concentration of HMOs has previously been associated with risk of infection in breastfed infants. In a cohort of predominantly Mexican mothers and their breastfed infants, the concentrations of 1,2-fucosylated HMOs such as 2'FL and lacto-*N*-di-fucohexaose (LNDFH-I) measured in milk were associated with a lower risk of moderate-to-severe diarrhea caused by calicivirus and *Campylobacter* (12, 13). A study in Gambia identified that the relative amount of LNFP-I and LNFP-III, but not 2'FL, in breast milk was associated with reduced parent-reported overall morbidity at 4 months postpartum (14). A recent randomized clinical trial examined the safety of feeding formula containing two HMOs, 2'FL and lacto-*N*-neo-tetraose (LNnT), in infants from approximately 2 weeks to 6 months of life (15). As a secondary outcome of this trial, reduced risk of lower respiratory tract infections (LRTIs) and decreased need for antibiotic and antipyretic use was observed in the group receiving HMOs compared to control (standard formula). This provided support for the hypothesis that fucosylated HMOs, among which 2'FL is the most abundant, could provide protection from respiratory infections.

One of the proposed mechanisms for enhancing protection against infections stipulates that HMOs act as decoys by binding and sequestering pathogens and facilitating their elimination, thus preventing pathogen attachment and invasion (16). This hypothesis is supported by some evidence from *in vitro* studies (17–19). While the attention has been recently more focused on the impact of HMOs on the gut microbiome and gut pathogens, earlier studies had indicated a potential effect on key respiratory pathogens (20). Further, it has been suggested that the carriage of respiratory pathogens, such as *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis*, could constitute a risk factor for subsequent respiratory infections (21). Could keystone pathogens residing in the respiratory tract of infants be influenced by breast milk HMOs delivered to the infant's gastrointestinal tract? This may seem difficult to envisage; however, *Bifidobacterium*, a key infant gut species highly adapted to utilization of milk, has been observed at low abundance but 50% prevalence in the neonatal airway microbiome (22). This suggests that in young infants, milk could enter airways during breastfeeding and impact the nasopharyngeal niche. Therefore, the natural variation in HMO content may influence the risk of respiratory infections in breastfed infants via an effect on facultative pathogens.

Moreover, respiratory pathogens should not be considered in isolation. They form a part of the respiratory microbiome comprising pathogenic, commensal, and potentially beneficial organisms mutually interacting. Accordingly, the respiratory microbiome has

been proposed to influence the susceptibility to respiratory infections, potentially through its antagonistic interactions with facultative respiratory pathogens (23).

Our study aimed to interrogate the link between maternal secretor and Lewis status, as a proxy for HMO composition in milk, and risk of acute respiratory infections in infants. Further, we evaluated whether the protective effect is at least partially conveyed via the effect of HMOs on the nasopharyngeal microbiome. Finally, we have tested the associations between maternal secretor and Lewis status and respiratory infections beyond the period of exclusive breastfeeding, as well as the impact of infant secretor and Lewis status on the risk of infections.

To test our hypotheses, we have used data from the Microbiota and Health Study (24), a prospective cohort study conducted in Dhaka, Bangladesh, featuring high breastfeeding rates and active surveillance of infectious episodes throughout infancy.

RESULTS

Characteristics of the study population. The Microbiota and Health Study was designed to explore the interplay of the respiratory and gut microbiome and multiple factors, including environmental, maternal, and genetic variables in modulating the risk of respiratory and gastrointestinal infections in infants. The present analysis focused on a specific hypothesis linking genetic risk factors and the nasopharyngeal microbiome with ARIs during the predominantly exclusive breastfeeding period of early infancy. For that purpose, we selected mother-infant pairs for which the infants were breastfed up to at least 4 months of age. A total of 240 infants were included (91.0% of the original study; Vidal et al., in preparation). Table 1 summarizes the baseline characteristics of the population; 89.2% of infants were exclusively breastfed, and 10.8% were partially breastfed at 4 months of age. Nearly all (96.7%) were still at least partially breastfed at 6 months of age.

Incidence of acute respiratory infections (ARIs). A total of 644 episodes of ARI were recorded during the follow-up period of 24 months. ARI was defined as one or more of the following symptoms: cough, runny nose, nasal congestion, ear discharge, and rapid breathing. The number of infants decreased from an initial $n = 264$ to $n = 204$ at 24 months, and data for 240 infants were available from birth to 6 months. The peak incidence of ARI was observed between 2 and 6 months of age, with 263 episodes (Fig. 1).

Characterization of *FUT2* and *FUT3* secretor- and Lewis-defining status variants. *FUT2* and *FUT3* exome sequences were obtained for 213 mothers and 212 infants in order to identify missense/nonfunctional single-nucleotide polymorphisms (SNPs) known to affect the secretor and Lewis status, as described in Materials and Methods. After quality control, 19 variants for the mothers and 18 variants for the infants passed the filters, including 7 missense SNPs in the total population which defined secretor and Lewis status (Table S1).

In order to define secretor status, we considered all missense *FUT2* SNPs known to result in a nonfunctional enzyme (11, 25–27). In our study population, three *FUT2* missense SNPs (rs601338, rs1047781, and rs200157007) were present. Missense SNP rs602662 was in perfect linkage disequilibrium with rs601338; thus, we considered it redundant in our selection (data not shown). Based on these three SNPs, 24 mothers (11.4%) and 17 infants (8.2%) were identified as nonsecretors (Se⁻). Similarly, according to four missense *FUT3* SNPs present in the study population (rs3745635, rs28362459, rs3894326, and rs812936) (11) and used to define the Lewis (positive/negative) status, 13 mothers (6.2%) and 8 infants (3.8%) were identified as Lewis negative (Le⁻) (Table 1). Only two mothers were both *FUT2* and *FUT3* negative. No significant differences were observed for any of the maternal and infant variables between secretors and nonsecretors, as well Lewis positive compared to Lewis negative (Fisher's exact test, χ^2 test).

Maternal secretor status was associated with lower ARI risk in infants. Infants of secretor-positive mothers (mSe⁺) had a significantly lower incidence of ARI than those of secretor-negative mothers (mSe⁻) during the first 6 months of life, with an incidence

TABLE 1 Baseline characteristics and breastfeeding status of the study population

Variable	Category	All, n (%)	Secretor mothers (mSe+), n (%)	Nonsecretor mothers (mSe-), n (%)	Lewis-positive mothers, n (%)	Lewis-negative mothers, n (%)
Age of the mother (yrs)	>22	141 (58.8)	105 (56.5)	15 (62.5)	113 (57.4)	7 (53.8)
	≤22	99 (41.3)	81 (43.5)	9 (37.5)	84 (42.6)	6 (46.2)
Mode of delivery	Vaginal	178 (74.2)	140 (75.3)	17 (70.8)	148 (75.1)	9 (69.2)
	Cesarean section	62 (25.8)	46 (24.7)	7 (29.2)	49 (24.9)	4 (30.8)
Place of delivery	Health facility	103 (42.9)	77 (41.4)	10 (41.7)	81 (41.1)	6 (46.2)
	Home	137 (57.1)	109 (58.6)	14 (58.3)	116 (58.9)	7 (53.8)
Season at time of delivery	Premonsoon (dry) hot (Mar–May)	79 (32.9)	64 (34.4)	3 (12.5)	64 (32.5)	3 (23.1)
	Rainy monsoon (Jun–Oct)	103 (42.9)	77 (41.4)	15 (62.5)	85 (43.1)	7 (53.8)
	Cool dry winter (Nov–Feb)	58 (24.2)	45 (24.2)	6 (25.0)	48 (24.4)	3 (23.1)
Infant sex	Female	124 (51.7)	92 (49.5)	15 (62.5)	103 (52.3)	4 (30.8)
	Male	116 (48.3)	94 (50.5)	9 (37.5)	94 (47.7)	9 (69.2)
No. of siblings	Not available ^a	3 (1.3)	3 (1.6)		3 (1.5)	
	0: no sibling	97 (40.4)	77 (41.4)	8 (33.3)	77 (39.1)	8 (61.5)
	1: ≥1 sibling(s)	140 (58.3)	106 (57.0)	16 (66.7)	117 (59.4)	5 (38.5)
Gestational age (wks)	0: ≥37 (at term)	223 (92.9)	172 (92.5)	23 (95.8)	182 (92.4)	13 (100.0)
	1: <37 (preterm)	17 (7.1)	14 (7.5)	1 (4.2)	15 (7.6)	
Infant birth wt (kg)	Not available ^a	4 (1.7)	4 (2.2)		4 (2.0)	
	0: >2.5	167 (69.6)	134 (72.0)	14 (58.3)	138 (70.1)	10 (76.9)
	1: ≤2.5	69 (28.8)	48 (25.8)	10 (41.7)	55 (27.9)	3 (23.1)
Infant secretor status	Not available ^a	32 (13.3)	3 (1.6)		3 (1.5)	
	Secretors (iSe+)	191 (79.6)	173 (93.0)	17 (70.8)	179 (90.9)	11 (84.6)
	Nonsecretors	17 (7.1)	10 (5.4)	7 (29.2)	15 (7.6)	2 (15.4)
Infant Lewis status	Not available ^a	31 (12.9)	2 (1.1)		2 (1.0)	
	Lewis positive	201 (83.8)	176 (94.6)	24 (100.0)	188 (95.4)	12 (92.3)
	Lewis negative	8 (3.3)	8 (4.3)		7 (3.6)	1 (7.7)
Breastfeeding at 4 mo	Exclusively breastfed	214 (89.2)	163 (87.6)	23 (95.8)	175 (88.8)	11 (84.6)
	Partially breastfed	26 (10.8)	23 (12.4)	1 (4.2)	22 (11.2)	2 (15.4)
Breastfeeding at 6 mo	Not available ^a	7 (2.9)	2 (1.1)		2 (1.0)	
	Exclusively breastfed	172 (71.7)	134 (72.0)	21 (87.5)	146 (74.1)	9 (69.2)
	Partially breastfed	60 (25.0)	49 (26.3)	3 (12.5)	48 (24.4)	4 (30.8)
	Not breastfed	1 (0.4)	1 (0.5)		1 (0.5)	
Breastfeeding >6 mo	No	8 (3.3)	2 (1.1)		2 (1.0)	
	Yes	232 (96.7)	184 (98.9)	24 (100.0)	195 (99.0)	13 (100.0)
Totals		240 (100.0)	186 (100.0)	24 (100.0)	197 (100.0)	13 (100.0)

^aThe information could not be collected or sample not analyzed.

rate ratio (IRR) of 0.66 (95% confidence interval [CI], 0.47 to 0.94; $P = 0.020$) (Fig. 2a). Maternal Lewis status was not associated with ARI incidence during the first 6 months of life (Fig. 2b). No significant associations were found between maternal secretor or Lewis status and incidence of ARI in infants during later time points (Fig. 2a and b).

We conducted additional *post hoc* analyses focusing on the nearly exclusive breastfeeding period coinciding with the peak of ARIs (Fig. 1). The statistical model employing IRR did not converge for the period of 0 to 4 months. Therefore, the analyses over this period were performed using the occurrence of at least one ARI episode, employing odds ratios (OR). As observed for the IRR of ARIs during the first 6 months, infants of secretor-positive mothers (mSe+) had a lower ARI occurrence in the period of 0 to 4 months (Fig. 2c), with an OR of 0.34 (95% CI, 0.13 to 0.92; $P = 0.033$).

Regarding the infection susceptibility linked to an infants' secretor and Lewis status, exploratory analysis employing OR for the 0- to 4-month period revealed a significantly

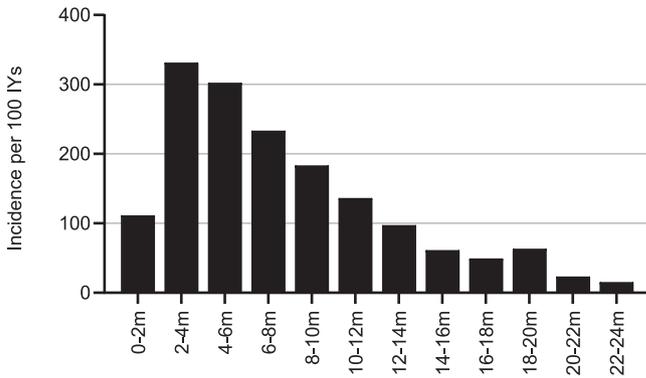


FIG 1 Incidence of acute respiratory infections (ARIs) by bi-monthly intervals. Incidence is expressed as the number of episodes per 100 infant-years of observation.

increased risk of ARIs for secretor-positive infants only, with and without accounting for the maternal secretor status (OR = 3.41; 95% CI, 1.11 to 10.43; $P = 0.031$) (Fig. 2c). However, this association was not detected using IRR for the period of 0 to 6 months (IRR = 1.21; 95% CI, 0.76 to 1.92; $P = 0.45$) (Fig. S1). Single SNP associations were finally tested for the number of ARI episodes, but none of the results reached the level of statistical significance after adjustment for multiple testing with a 0.05 false-discovery rate (FDR) (data not shown).

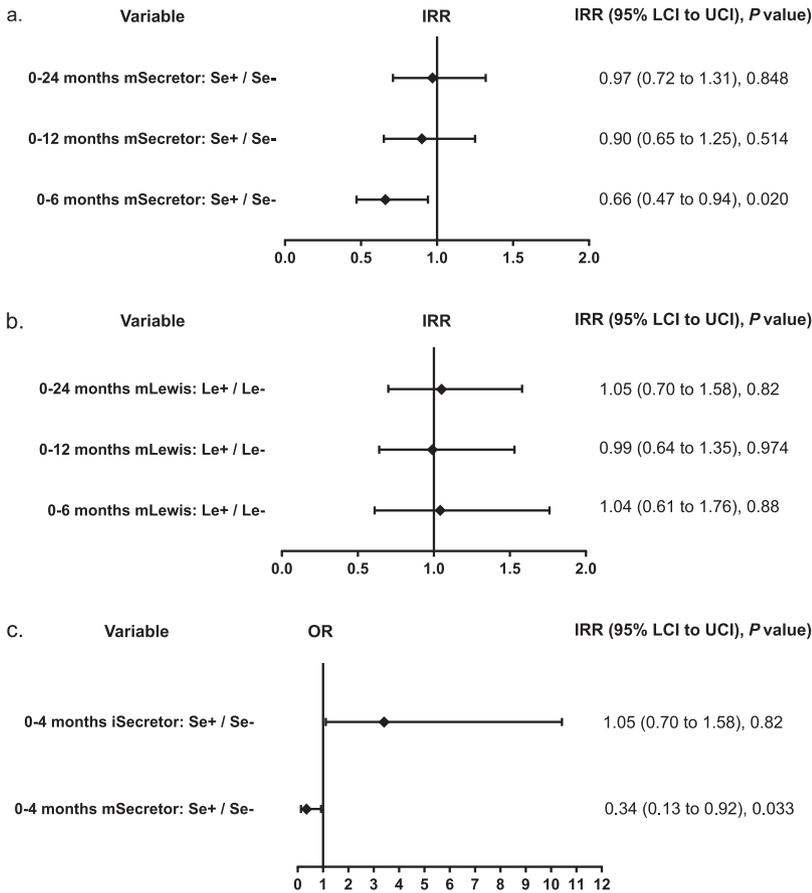


FIG 2 Forest plots for IRR or OR (95% lower confidence interval [LCI] to upper confidence interval [UCI]) represent ARI risk in relation to maternal secretor status (mSecretor) from 0 to 24 months (a), Lewis maternal status (mLewis) from 0 to 24 months (b), and maternal and infant secretor status (mSecretor and iSecretor) from 0 to 4 months (c). Status could be either positive (Se+, Le+) or negative (Se-, Le-). The negative status was considered a reference category. These results derive from the regression model considering both maternal and infant secretor/Lewis status.

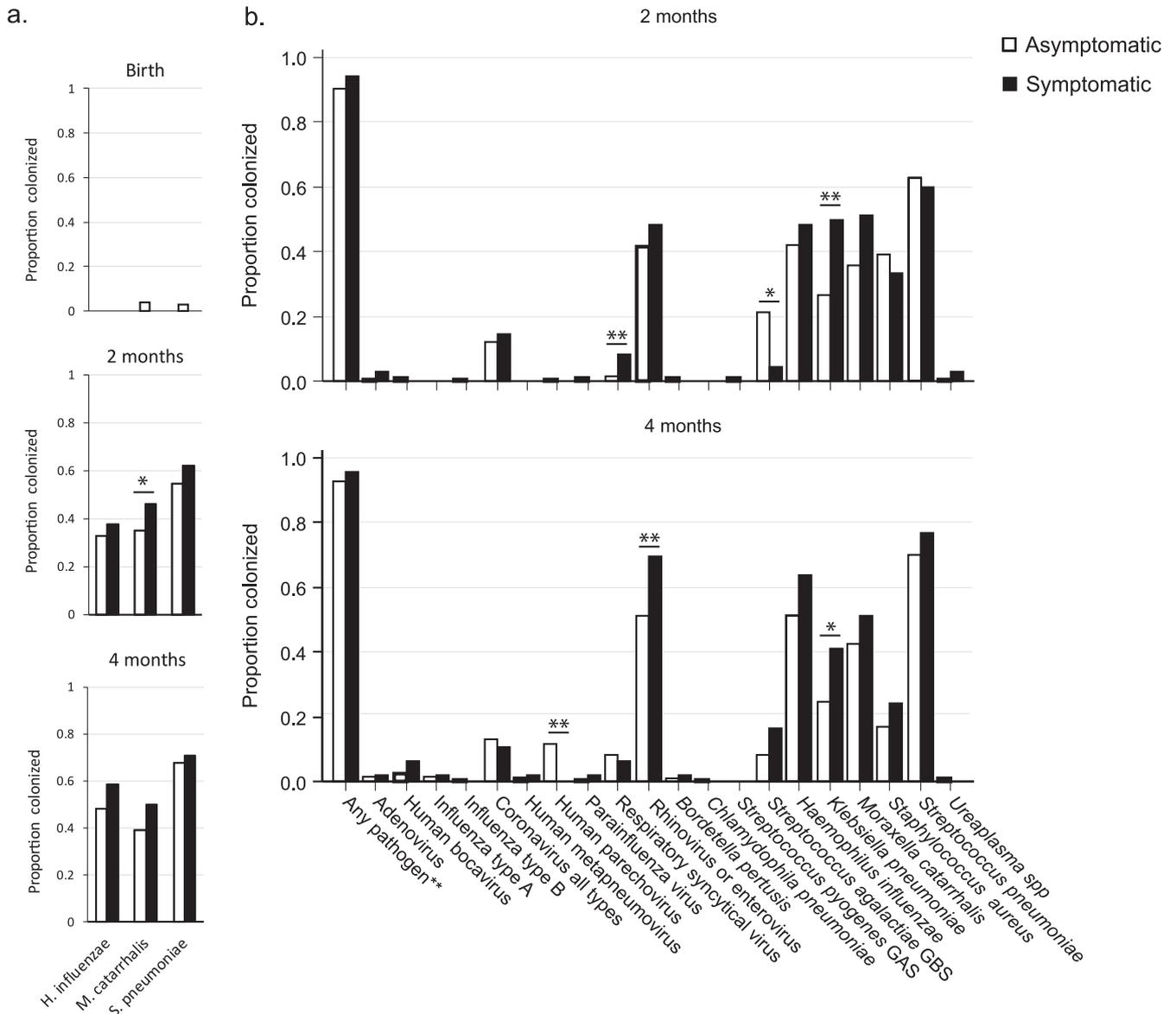


FIG 3 Nasopharyngeal colonization pathogens in infants asymptomatic and symptomatic for ARIs at the time of sample collection assessed by culture (a) and TaqMan array cards (b). One star (*) denotes a trend ($0.05 < P < 0.1$) and ** denotes a significant ($P < 0.05$) difference between symptomatic and asymptomatic infants.

Nasopharyngeal colonization by respiratory pathogens. To examine the hypothesis that the maternal secretor and Lewis status affected the risk of ARI in infants by modulating respiratory pathogen colonization, we examined their presence using both culture and molecular methods. A large proportion of infants were colonized by facultative bacterial respiratory pathogens, as assessed by culture (*S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*) (Fig. 3A). Molecular detection of bacterial and viral respiratory targets at 2 and 4 months of age confirmed the frequent colonization by bacterial pathogens (Fig. 3B). In addition to *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*, colonization with *Klebsiella pneumoniae* and *Staphylococcus aureus* was common. Among viruses, rhinovirus was the most frequently detected, notably in asymptomatic infants showing 42% prevalence at 2 months of age that increased to 53% at 4 months of age.

As expected, symptomatic infants tended to have a higher risk of being colonized, when assessed by both culture and molecular methods (Fig. 3). However, the effect size was small, with a trend for more prevalent colonization significant only for *K. pneumoniae* among bacterial pathogens. Regarding viruses, rhinovirus was significantly

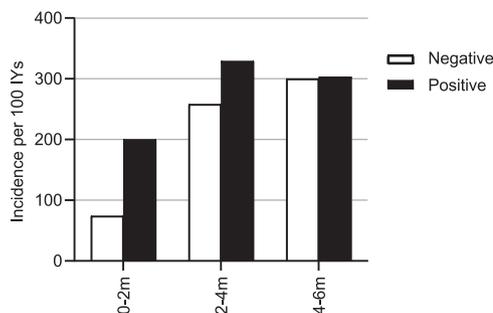


FIG 4 ARI incidence in the 2-month period following the sample collection for asymptomatic infants colonized (positive) by facultative bacterial respiratory pathogens (*M. catarrhalis*, *S. pneumoniae*, *H. influenzae*) and those who were not colonized by these pathogens (negative) at the beginning of the observation period.

more common in symptomatic infants at 4 months of age, with the same trend at 2 months. Respiratory syncytial virus was also overrepresented in symptomatic infants at 2 months, although this trend was no longer significant at 4 months. Surprisingly, human parechovirus was more often detected in asymptomatic infants at 4 months, but not at 2 months (Fig. 3B).

Effect of secretor and Lewis status on nasopharyngeal colonization by facultative bacterial respiratory pathogens. Infants of Lewis-positive mothers were less likely to be colonized by *M. catarrhalis* in the period from birth to 6 months (OR = 0.42; CI, 0.22 to 0.83; $P = 0.036$), and there was a trend for less colonization by *H. influenzae* for infants of secretor-positive mothers in the same period (OR = 0.55; CI; 0.29 to 1.05; $P = 0.083$; only detectable when infant secretor status was included in the model). No other significant associations were observed.

Nasopharyngeal colonization by facultative bacterial respiratory pathogens in asymptomatic infants was weakly associated with subsequent ARIs. The data on colonization by facultative pathogens were collected at bi-monthly intervals; therefore, the influence of the colonization status on risk of subsequent ARIs was analyzed using a repeated measures model. The selection of covariates for this analysis was based on the identification of the significant predictors of ARIs in the entire cohort (Vidal et al. in preparation). The colonization by facultative bacterial pathogens (*M. catarrhalis*, *S. pneumoniae*, *H. influenzae*) was associated with marginally higher infection risk in the following 2-month period (IRR = 1.59; CI, 1.01 to 2.53; $P = 0.047$) (Fig. 4, Table S2), apparently driven by *M. catarrhalis* (IRR = 1.64; CI, 1.12 to 2.40; $P = 0.011$) (Table S2). Although the above-described P values refer to the entire analysis period, numerically, the largest difference in risk was observed for the colonization at birth for the period 0 to 2 months (Fig. 4) despite low rates of colonization at birth (Fig. 3A).

We also evaluated the impact of colonization of a wide range of bacterial and viral pathogens detected by molecular methods on cumulative ARI infection rate through a Kaplan-Meier analysis, available at 2 and 4 months. No differences were observed in cumulative ARI infection rate according to infant colonization status (Fig. S2).

Characterization of the nasopharyngeal microbiome using shotgun metagenomics. We hypothesized that the overall nasopharyngeal microbiome could significantly contribute to subsequent respiratory outcomes and potentially convey the effect of maternal secretor status. Therefore, we characterized the nasopharyngeal microbiome of infants at 2 and 4 months of age using shotgun metagenomics, focusing on asymptomatic infants to consider the microbiome as a potential cause and not as a consequence of disease.

A total of 422 samples were sequenced on the Illumina HiSeq platform with an average of 38.1 million (M) read pairs per sample. As expected for a low-biomass microbiome (28), most samples were dominated by human host reads, which were filtered out as described in Supplemental Text S1. By assembling the remaining reads and clustering the genes called from the contigs, a nonredundant nasopharyngeal microbial catalog of 3,110,772 genes was built. By using the coabundance principle across

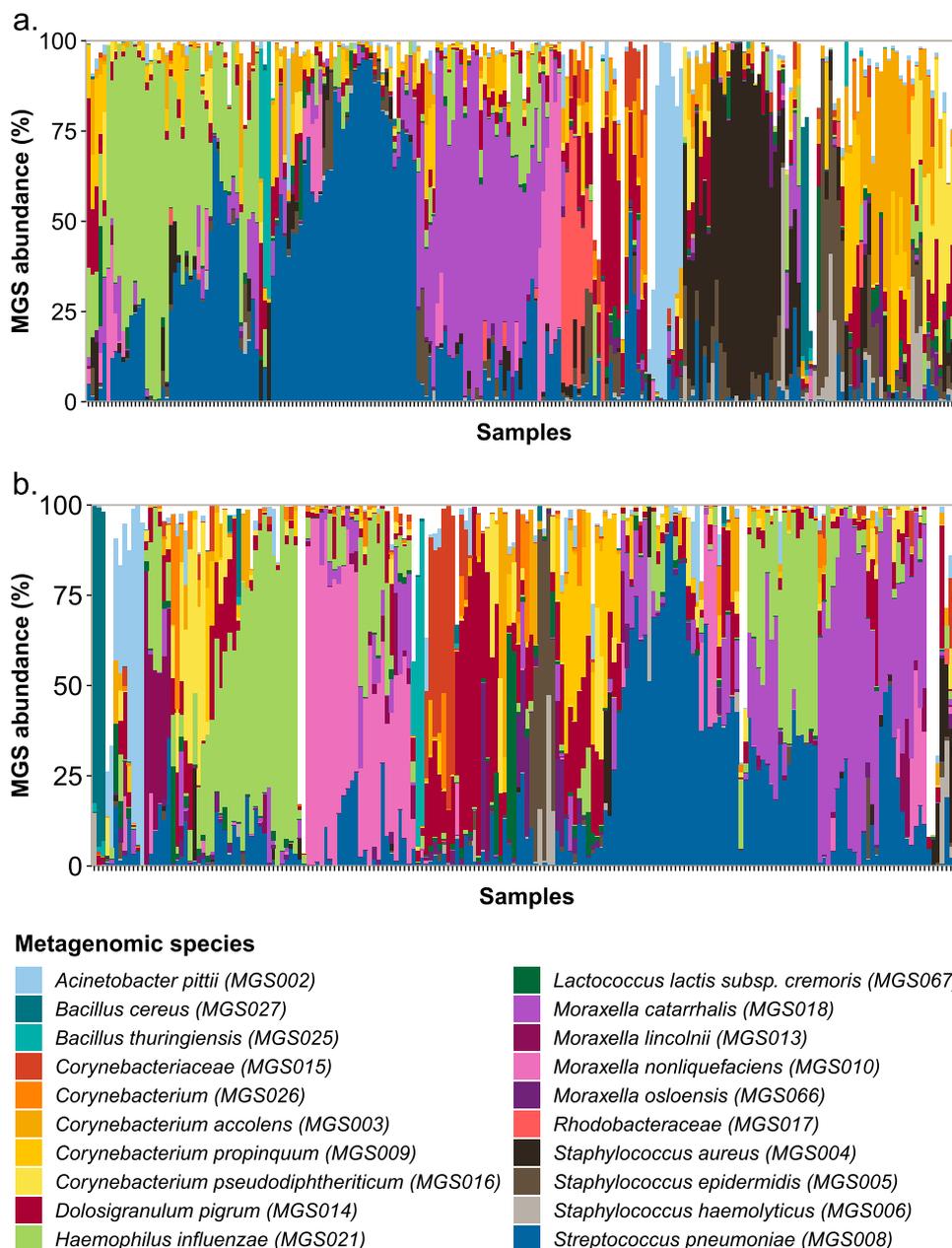


FIG 5 The relative abundance of the top 20 most abundant metagenomic species (MGS) in nasopharyngeal samples from (a) 2- and (b) 4-month-old asymptomatic infants. Samples were ordered according to hierarchical clustering based on Bray-Curtis dissimilarities among the samples (clustering and associated ordering is time point-specific and therefore different between panels a and b).

samples, 55 metagenomic species (MGS) were identified, of which 53 were annotated up to the species level. On average, 94.7% of the genes within each MGS mapped to the same reference genome, suggesting a reliable high-quality MGS taxonomy.

The observed nasopharyngeal microbiome at both time points was composed of both commensal and facultative pathogens (Fig. 5). The most abundant species included *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Dolosigranulum pigrum*, and *Staphylococcus aureus*; the detailed description of MGS abundance and full taxonomical characterization can be found in the Supplemental Data Set S1.

The differentiation among individuals was striking; the variation attributable to individuals accounted for most of the variance (permutational multivariate analysis of variance [PERMANOVA] test; $R^2 = 62\%$, $P = 0.0005$). The effect of time was much less pronounced, and no clear separation between samples of the two age groups was observed (Fig. S3), even though the shift in composition was statistically significant

TABLE 2 Influence of extrinsic factors on the nasopharyngeal microbiome at 2 and 4 months of age analyzed by PERMANOVA tests with 10,000 permutations^a

Clinical factor	Df	R ²	P
2 mo			
Mode of delivery	1	0.010	0.065
Maternal secretor	1	0.004	0.63
Maternal Lewis	1	0.006	0.41
Infant secretor	1	0.006	0.29
Infant Lewis	1	0.005	0.57
Breastfeeding at 2 mo	2	0.007	0.90
Prior infection	1	0.005	0.55
Season at sampling	2	0.048	0.0001
4 mo			
Mode of delivery	1	0.005	0.53
Maternal secretor	1	0.003	0.90
Maternal Lewis	1	0.010	0.070
Infant secretor	1	0.009	0.12
Infant Lewis	1	0.005	0.62
Breastfeeding at 4 mo	2	0.009	0.77
Prior infection	1	0.007	0.31
Season at sampling	2	0.033	0.0003

^aSignificant factors are highlighted in bold. Df, degrees of freedom; P, P value in pseudo-F test.

(PERMANOVA test; R² = 1.5%, P = 0.0001). MGS richness slightly decreased from 2 months (average, 8; range, 1 to 19) to 4 months (average, 7.4; range, 2 to 16) (two-sided Wilcoxon signed-rank test, P = 0.027), while there was no difference in alpha diversity (Shannon diversity index = 1.16 and 1.11, at 2 and 4 months, respectively).

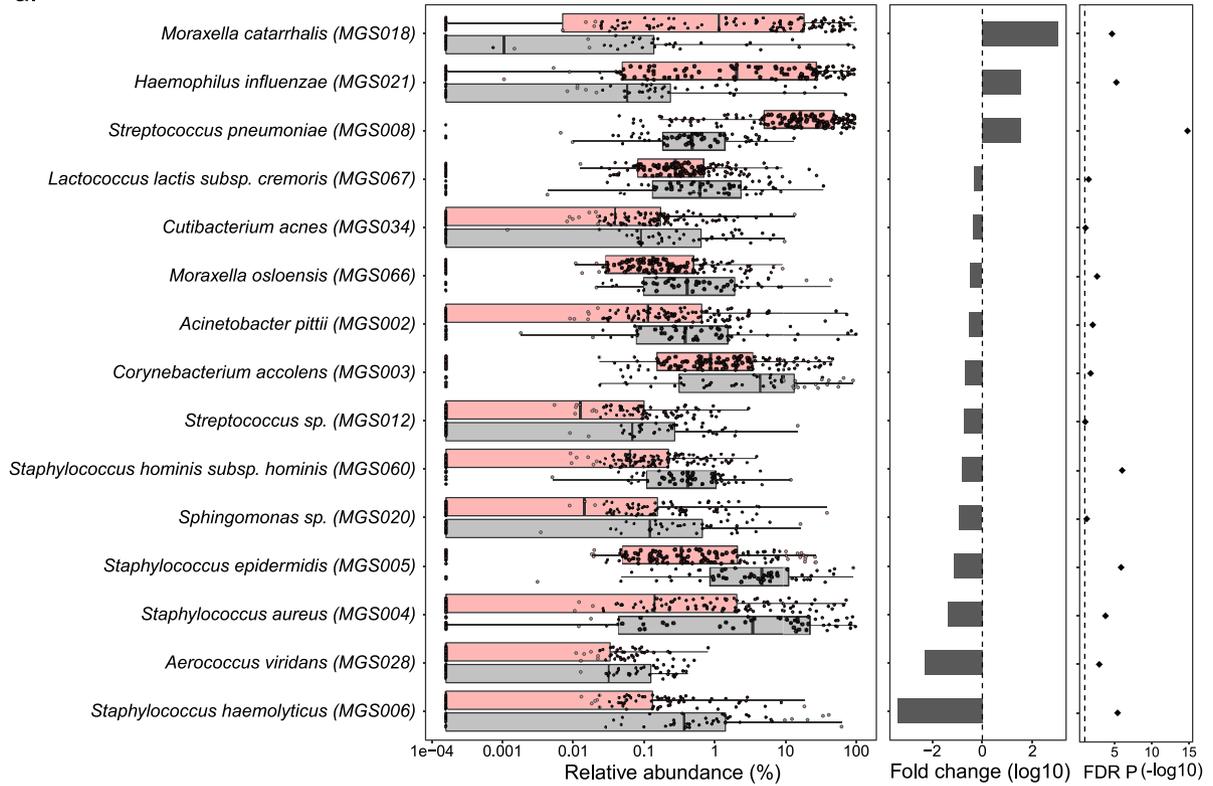
Effect of maternal secretor and Lewis status and extrinsic factors on the nasopharyngeal microbiome. If the protective effect of HMOs were conveyed by modulation of the infant microbiome, we would expect an association between maternal secretor and/or Lewis status and microbiome composition. This was, however, not the case: maternal secretor and Lewis status showed no significant association with the composition of the nasopharyngeal microbiome (Table 2). Other variables previously reported to influence the nasopharyngeal microbiome (such as prior infection, mode of delivery, breastfeeding, and season at sampling) showed no significant association with microbiome except season at sampling, which was more pronounced at 2 months of age.

Nasopharyngeal microbiome composition could not explain subsequent ARI occurrence. To evaluate whether the nasopharyngeal microbiome is linked to subsequent ARIs we employed a wide range of machine learning approaches. However, despite transforming the microbiome data in different ways, taking into account the impact of extrinsic factors on the microbiome, considering different follow-up time periods, and using different machine learning methods (Supplemental Text S1), the 72 models we built exhibited highly unsatisfactory trade-off between specificity and sensitivity, thus precluding linking the nasopharyngeal microbiome with subsequent ARIs (Table S3).

Nasopharyngeal microbiome composition was associated with the presence of bacterial facultative pathogens. The detection by culture and by MGS characterization of the three facultative pathogens (*H. influenzae*, *S. pneumoniae*, and *M. catarrhalis*) was highly consistent; 99% of samples in which a pathogen was detected by culture also showed the corresponding MGS (MGS abundance, >0) (Table S4). Further analysis was conducted to explore microbiome features specifically linked to the presence of pathogens.

While no significant differences were observed in alpha-diversity between colonized (one of the three pathogens was detected by culture) and noncolonized samples, they strongly differed in microbiome composition. Beyond the expected differences in the 3 MGS matching the pathogens detected by culture (MGS008, MGS018, and MGS021), 12 MGS were significantly more abundant among the noncolonized samples at 2 months (two-sided Wilcoxon rank-sum test, FDR < 0.1) (Fig. 6), and this number increased to 16 by

a.



b.

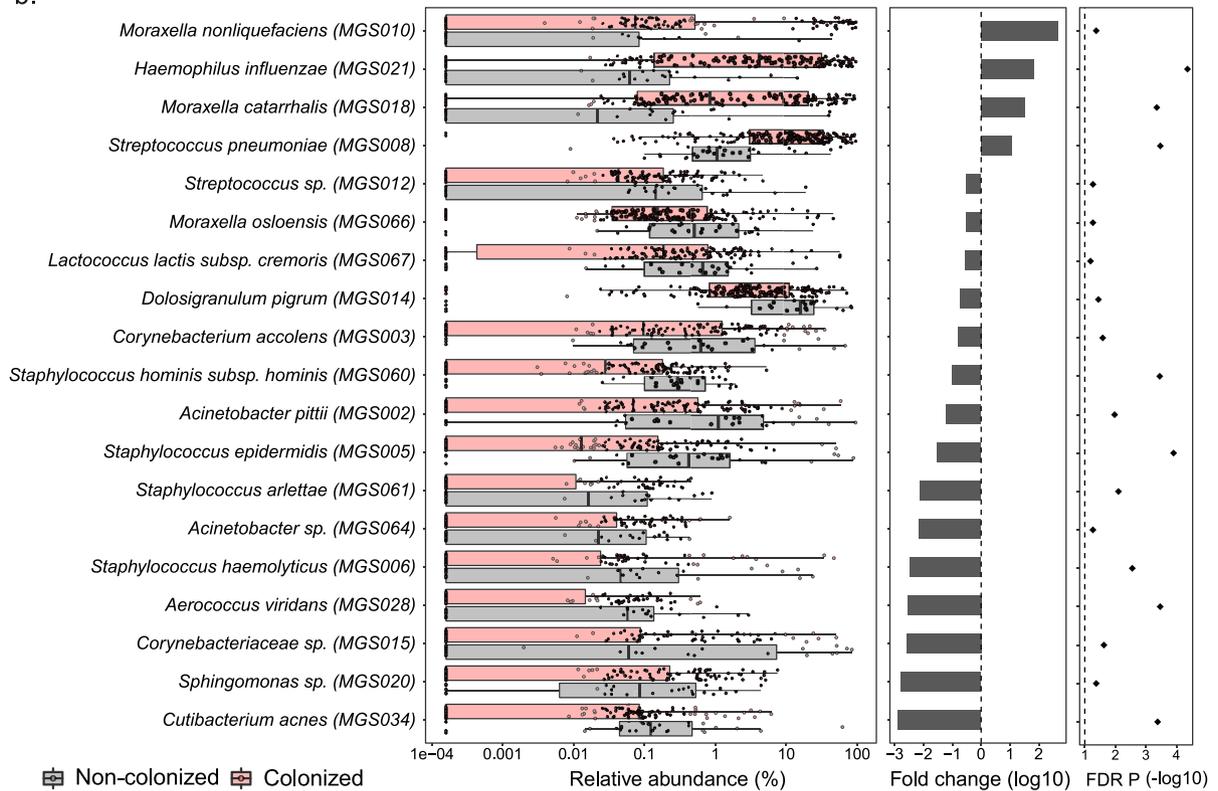


FIG 6 Differentially abundant metagenomic species between colonized and noncolonized samples (FDR $P < 0.1$) at 2 months (a) and at 4 months (b). The left panels display the median and IQR of the relative abundance. For this figure, a small constant (minimum relative abundance observed divided by 2) was added to all abundances to avoid a logarithm of zero. The middle panels show the fold change, and the right panels show the FDR-adjusted P value (FDR P) of a two-sided Wilcoxon rank-sum test for the difference between colonized and noncolonized samples.

4 months of age. All species showing higher abundance in noncolonized samples are considered to possess no or low pathogenic potential, except for *S. aureus*, which showed differential abundance only at 2 months and greatly decreased in overall abundance by 4 months of age. *Moraxella nonliquefaciens* was the only species more abundant in the colonized samples and only at 4 months, an effect seemingly driven by a subset of samples with a high abundance of this species (Figure 6b).

DISCUSSION

Maternal secretor and Lewis status as proxies of HMO composition in maternal milk and infant ARIs. The Microbiota and Health Study provided an opportunity to assess a population of predominantly breastfed infants with high ARI prevalence, peaking between two and 6 months of age. We observed that positive maternal secretor status (defined by the presence of a functional variant of the *FUT2* gene) was associated with a reduced ARI risk. Remarkably, this effect was observed only during the first 6 months of life and thus coincided with the period of exclusive or predominant breastfeeding in the study population. Our study exploited a tight link between genetic variation in *FUT2* and *FUT3* determining secretor and Lewis status and the presence of a distinct group of 1,2- and 1,3/4-fucosylated HMOs in breast milk (10, 11). As maternal secretor status is a reliable marker of the availability of specific HMOs to a breastfed infant (11), our results suggest a protective role of 1,2-fucosylated HMOs against respiratory infections in early infancy. All mutations (in this case in *FUT2*) can have unknown pleiotropic effects; therefore, it cannot be excluded that the effects on the infants are conveyed by yet unknown mechanisms other than the effect on HMO content of milk. For example, strong inhibition of Norwalk virus by secretor milk was conveyed not by free oligosaccharides but, rather, by glycoproteins (29). A mother's secretor status could impact other aspects of her phenotype, including maternal microbiome, with a potential downstream effect on the infant.

Previous studies have suggested a role for 2'FL, which is the most abundant 1,2-fucosylated HMO in secretor milk, in protection against diarrhea in a cohort of Mexican infants (12, 13). However, no association between maternal secretor status and ARI risk was observed in a UK cohort (30). The infants in the latter study were only partially rather than exclusively breastfed, suggesting that exclusive breastfeeding may be necessary to observe a protective effect. Other variables, notably living environment, and ARI etiology could also explain the difference between the studies. Our observations suggesting the protective role of 1,2-fucosylated HMOs are consistent with the results of a randomized clinical trial that observed a reduced risk of LRTI and decreased need for antibiotics and antipyretics in young infants fed formula containing two HMOs, 2'FL and LNnT (15).

Other classes of HMOs might be involved in protection from respiratory infections, as suggested by a pilot study in the United States where the level of lacto-*N*-fucopentaose II (LNFP-II, an 1,3/4-fucosylated HMO) was associated with less respiratory illness in the first 4 months of life (31). In our study we did not identify any significant associations between ARI and maternal Lewis status, a proxy for 1,3/4-fucosylated HMO. However, our results cannot be directly compared to the above-mentioned study where only levels of LNFP-II in milk and no other HMOs were analyzed, in contrast to our approach using genetic markers as a proxy for groups of fucosylated HMO. Nevertheless, results from observational studies propose a protective role of fucosylated HMOs against respiratory infections that needs to be further understood.

Our study also examined the association between infant *FUT2* and *FUT3* polymorphisms and ARIs. Infant secretor but not Lewis status was weakly associated with increased ARI risk during the first 4 months of life. A large cohort study conducted in the United Kingdom found increased susceptibility to respiratory infections in secretor-positive infants occurring between 12 and 24 months, but not in early infancy (30). Earlier studies found that secretors of all ages combined were overrepresented among cases of respiratory infections and individuals diagnosed with respiratory viruses (32). While an individual's secretor status seems to modulate the susceptibility to ARIs, a

better understanding of its interaction with other factors such as age and etiology is needed.

Protective effects of maternal secretor status were not conveyed by effects on respiratory pathogens. We did not find substantial evidence that the protective effect of maternal HMOs is conveyed by modulation of respiratory pathogens. We observed a trend for reduced colonization of infants of secretor-positive mothers with bacterial facultative respiratory pathogens (*S. pneumoniae*, *H. influenzae*, *M. catarrhalis*), and the presence of any of these pathogens was weakly associated with ARI risk in the subsequent 2-month period. Variable results were reported on the link between colonization and risk of subsequent ARIs, in particular, contrasting findings from two studies evaluating the same set of pathogens conducted in relatively similar northern European populations (Denmark and the Netherlands) (21, 33). It appears that the association between pathogen colonization and later ARIs could be driven by more severe infections, LRTIs and bronchitis, where bacterial etiology is more usual, while no such link is observed for milder ARIs. We posit that bacterial pathogens are unlikely etiological ARI agents in our study due to the short duration of the majority of episodes (mean duration of 5 days; Vidal et al., in preparation) and human rhinovirus being a predominant etiological agent. The weak link between secretor-positive milk and reduced colonization on one hand and reduced colonization and subsequent infections on the other appears to more likely be an indicator of generalized immune support conveyed by feeding secretor-positive milk, rather than a direct effect of fucosylated HMO sequestering etiological agents of respiratory infections.

The high prevalence of human rhinovirus in asymptomatic infants (42% and 53% at 2 and 4 months, respectively) may appear surprising, suggesting a state of “colonization” with minimal symptoms and/or persistence of the virus after symptoms have ceased. These results agree with the findings of a large cohort study investigating the etiology of severe infections in South East Asia that also included healthy infants (34), where 34% of infants aged, on average, 13 days had human rhinovirus infection. The high colonization rate by facultative respiratory bacterial pathogens (*S. pneumoniae*, *M. catarrhalis*, *H. influenzae*) is in line with findings from other studies conducted in similar settings (35, 36). Despite variation among studies, higher prevalence appears to be observed in low-resource settings, e.g., 29% at 7 weeks of age in South Africa (37) compared to 8.5% at 6 weeks of age in the Netherlands (33).

Symptomatic infants appeared to be slightly but consistently more likely to be colonized by three main bacterial respiratory pathogens that were detected both by culture throughout the study period as well as by molecular methods at 2 and 4 months. In addition, *K. pneumoniae* (only detected by the molecular method) followed the same pattern. However, as bacterial etiology of ARIs appears unlikely, higher prevalence in symptomatic infants could be an indicator of compromised immunity.

Protective effects of maternal secretor status were not conveyed by modulation of the nasopharyngeal microbiome. Beyond respiratory pathogens, we also hypothesized that HMOs could shape the nasopharyngeal microbiome and thus modulate susceptibility to infections. To address this, the nasopharyngeal microbiome needs to be properly characterized up to the species level, notably to distinguish commensals from pathogens, which are often phylogenetically very close. Most previous studies described the nasopharyngeal microbiome in infancy using 16S rRNA gene sequencing (38–40), while shotgun metagenomics was only utilized in sick infants (41) or as a descriptive proof-of-concept study (42). To our knowledge this is the first study to examine the nasopharyngeal microbiome from a large cohort of healthy infants using shotgun metagenomics. A high-resolution profile of the microbiome composition up to at least the species level was achieved, despite the predominance of human sequence reads in some samples. The high level of consistency between shotgun metagenomics indicating the presence of *H. influenzae*, *S. pneumoniae*, or *M. catarrhalis* and culture data further confirmed the quality of the characterization.

Contrary to our hypothesis, we could not link the nasopharyngeal microbiome at 2 or 4 months in asymptomatic infants to subsequent ARI risk. A recent meta-analysis

demonstrated consistent differences in respiratory microbiota between infants suffering from respiratory illnesses and healthy controls, manifesting mainly as increased presence of *Haemophilus* and *Streptococcus* and reduced diversity during the disease (D. Broderick, D. W. Waite, R. Marsh, C. A. Camargo, P. Cardenas, A. B. Chang, W. O. C. Cookson, L. Cuthbertson, W. Dai, M. L. Everard, A. Gervaix, J. K. Harris, K. Hasegawa, L. R. Hoffman, S. J. Hong, L. Josset, M. S. Kelly, B. S. Kim, K. Yong, S. C. Li, J. Mansbach, A. Mejias, G. A. O'toole, L. Paalanen, M. P. Losada, M. M. Pettigrew, M. Pichon, O. Ramilo, L. Ruokolainen, O. Sakwinska, P. Seed, C. J. Van Der Gast, B. Wagner, H. Yi, E. T. Zemanick, Y. Zheng, N. Pillarisetti, and M. Taylor, submitted for publication). Other studies showed some consistency in associating the lack of taxa such as *Dolosigranulum* and *Corynebacterium* with disease (43–45).

However, cross-sectional design cannot determine whether these changes were a cause or consequence of the disease. Few previous studies employed a longitudinal prospective design and suggested that specific microbiome profiles were associated with higher risk of subsequent ARIs (46–48); the results of these studies were not unequivocal. While an earlier study in the Netherlands reported that the respiratory microbiome dominated by *Moraxella* was associated with a lower incidence of subsequent respiratory infections (47), newer studies conducted in Finland and Australia showed that the *Moraxella*-dominated community type was associated with a higher incidence of ARIs (46, 48, 49). These contradictory findings could be a consequence of the lack of taxonomic resolution based on the sequencing of the 16S rRNA gene leading to variable classification within the *Moraxella* genus. While the pathogenic potential of *M. catarrhalis* is well established, this evidence has been emerging for *M. nonliquefaciens* through genome analysis (50) and associations with sinusitis (44) and viral pneumonia (39). In our data, *M. nonliquefaciens* showed a positive association with the presence of other established respiratory pathogens, while *Moraxella osloensis* displayed an inverse correlation with pathogen presence.

We documented striking antagonistic associations among a wide range of species and the presence of established respiratory pathogens, beyond the inverse association between *S. pneumoniae* and *S. aureus* carriage described before (51). Overall, these data could suggest competition for the same ecological niche. However, the host is not a passive element, and factors such as mucosal immunity likely shape the competitive balance among the nasopharyngeal microbiome community. Although our data did not lend support for the link between the nasopharyngeal microbiome and common uncomplicated ARIs, it is possible that more severe infections such as LRTIs or other respiratory conditions such as asthma later in life could be modulated by the microbiome composition (49). More targeted high-resolution methodology is needed to establish robust associations.

The prevailing environmental factor shaping the nasopharyngeal microbiome in our study was seasonality. Only a few previous studies have observed the same (38, 46). It is plausible that in our study population, the infants were more exposed to the external environment, in contrast to previous studies mostly conducted in industrialized settings. On the other hand, we did not observe the effect of several environmental drivers on the microbiome composition as reported before. Our study population was very homogenous in respect to breastfeeding and did not attend daycare; however, high exposure to other nonsibling children appeared very common, which together with the prevailing effect of the season could have stronger effects than other environmental factors.

Alternative hypotheses. We have not found compelling evidence that protective effects of HMOs were conveyed by modulation of the nasopharyngeal microbiome. This leads us to hypothesize that HMOs could exert an immunomodulatory effect on the host. As a nonexclusive alternative, HMOs shape the gut microbiome composition and function, leading to production of metabolites with an immunomodulatory effect on the host (“gut-lung axis”), and results from animal models lend some support to such a potential mechanism (52). Microbes residing in the gut are increasingly implicated in susceptibility to respiratory infections (53, 54). In the context of HMOs, a shift

in the infant gut microbiome accompanied a reduction in respiratory infections observed in infants consuming two HMOs (2'FL and LNnT) (55).

Limitations. The results of genetic analyses should be interpreted with caution, as the study was of moderate size. The number of nonsecretor individuals was low, especially compared to previous studies in Bangladeshi populations (25, 56); however, this difference was not significant and was most likely due to limited sample size. Namely, in the study by Williams and colleagues (56) there were 20 nonsecretors and 94 secretors, and in ours, 24 nonsecretors and 186 secretors (Chi square test $P = 0.125$). In addition, no data on milk composition was available, and the extrapolation to HMO content relies on genetic markers. However, it is well established that major fucosylated HMOs such as 2'FL and LNFP-I or LNFP-II depend entirely on *FUT2* and *FUT3* genetic factors (10, 11). Still, even though genetic nonsecretors have virtually no fucosylated HMO in milk, the quantities vary substantially among individuals with intact *FUT2*. This typically leads to a higher proportion of nonsecretors when phenotypic criteria (antigen detection in saliva) are used (56) and could lead to the inability to detect true associations, in our case between infant outcomes and HMO content. Ideally, larger studies assessing both HMO milk content, genotyping of both maternal and infant *FUT2* and *FUT3*, and the gut and respiratory microbiome should be performed to provide further insights in the association among HMOs, genetic innate susceptibility, and infection risk.

The in-depth analysis of the nasopharyngeal microbiome and its associations with subsequent ARIs was focused on samples collected from asymptomatic infants at 2 and 4 months of age, when the peak prevalence of ARIs was observed in our study. However, detailed analysis revealed no predictive power regarding future ARI episodes. The presence of respiratory pathogens at birth appeared to be most associated with subsequent ARI episodes. It cannot be excluded that the very early microbiome before 2 months of age could provide predictive power regarding subsequent ARIs.

The etiology of the majority of ARIs could not be established, due to the lack of feasibility to collect samples during the episodes. However, the samples that were collected during scheduled visits provided a good snapshot of the prevailing etiology.

The information on ARI occurrence in our study was collected by active surveillance, but no symptom diaries or parental training was carried out. Our methodology was superior to many studies relying on parental recall only; nevertheless, some episodes were likely missed. Studies with a formal parental diary card listing predefined respiratory symptoms and parental training (57) reported higher incidence (e.g., 58). However, the potential underreporting should not be biased and therefore should not invalidate our conclusions.

Conclusions. In conclusion, our study provided further evidence for an association of fucosylated HMOs with reduced risk of respiratory infections in exclusively or predominantly breastfed infants. In our study, infant secretor status and colonization by respiratory pathogens were at best weak modulators of this risk, while the nasopharyngeal microbiome appeared to have no observed effect, suggesting that any protective effects of HMOs or inherent susceptibility are not conveyed via changes in the nasopharyngeal microbiome. Future larger studies are warranted to provide mechanistic insights on links between secretor status, HMOs, and risk of respiratory infections.

MATERIALS AND METHODS

Study population and sampling. The Microbiota and Health Study was a prospective, community-based, longitudinal study of respiratory and gastrointestinal infections including 267 infants born in Nandipara, a peri-urban area near Dhaka, Bangladesh, between April 2013 and October 2016. The follow-up period started in the third trimester of the pregnancy and continued to 2 years of life, as previously described (24). Ethical approval was obtained from the local independent institutional review board. The trial was registered on Clinicaltrials.gov (registration no. [NCT02361164](https://clinicaltrials.gov/ct2/show/study/NCT02361164)). Written informed consent was obtained from all mothers before enrollment in the study.

For the present analysis, we have selected infants who were either partially or exclusively breastfed until at least 4 months of age, which constituted a large majority of infants in the cohort (Vidal et al., in preparation). An ARI episode was defined as one or more of the following symptoms: cough, runny nose, nasal congestion, ear discharge, and rapid breathing. A new ARI episode was defined as an episode starting after 7 symptom-free days from a previous episode. Incidence rate (IR) of ARIs in 100

infant-years (IY) was determined by dividing the total number of ARI episodes occurring among infants by the total number of days that each infant was followed (IY) multiplied by 100.

To determine *FUT2* and *FUT3* polymorphisms, saliva samples were collected from both mothers and infants using Oragene collection and DNA extraction kits (OG-500 and OG-575, respectively), which were subsequently used for human DNA extraction according to manufacturer recommendations.

Nasopharyngeal samples were collected from the infants at bi-monthly intervals during scheduled visits. The detection of three common facultative bacterial respiratory pathogens (*Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Haemophilus influenzae*) was performed by culture using standard clinical diagnostics methods. The remaining part of the sample was frozen and stored for future molecular analysis of viral and bacterial respiratory pathogens and nasopharyngeal microbiome. Total nucleic acid extraction from nasopharyngeal samples was performed using the QIAamp MinElute virus spin kit (Qiagen, Valencia, CA) according to manufacturer instructions. The starting material was 200 μ l of nasopharyngeal samples (NPS) spiked with 2.5E8 PFU of MS2 bacteriophage (ZeptoMetrix, Buffalo, NY). The resulting nucleic acid extract was subsequently used for the detection of respiratory pathogens with customized TaqMan array cards (TAC) and for the analysis of the nasopharyngeal microbiome using shotgun metagenomics.

All biological samples were initially stored at -20°C and subsequently transferred to -80°C until analysis.

Genetic analyses of *FUT2* and *FUT3* from mothers and infants. The entire *FUT2* and *FUT3* coding sequences and part of the 3' and 5' untranslated regions were PCR amplified for 30 cycles using the Kapa HiFi (Roche), starting from 50 ng DNA. The PCR primer sequences were ACACACCCACACTATGCCTG (*FUT2*-Fw), AAGAGAGATGGGCTCTGCTC (*FUT2*-Re), CCCGGAGCTTTGGTAAGCAG (*FUT3*-Fw), and GAGGGTTGGCCACAAAGGAC (*FUT3*-Re). The same melting temperature of 60°C was used for both amplifications. A positive control (DNA from HapMap NA18523) and a no-template control (water) were included on each PCR plate. The quality and quantity of each *FUT2* and *FUT3* PCR were checked by gel electrophoresis using the LabChip GX Touch (Perkin Elmer).

After purification on Ampure beads (Beckman) at a 1.8 \times ratio, sequencing libraries were prepared from the amplicons using the Nextera XT kit (Illumina) strictly following manufacturer's recommendations. Libraries were quantified with Picogreen (Life Technologies), and their size pattern was validated with a Fragment Analyzer (AATI). Sequencing was performed as a paired-end 250-cycle run with the MiSeq Reagent kits v2 (Illumina).

Variant calling was performed with the software FreeBayes (v1.1.0-3-g961e5f3) (59) using default parameters. The resulting vcf files were postprocessed with the Plink software (v1.9) for quality control (QC) purpose and recoding. The sequence of QC steps was as follows: (i) samples with more than 5% of missing genotypes were filtered out; (ii) variants missing in more than 5% of the samples were filtered out; (iii) variants with a minor allele frequency (MAF) below 1% (computed on the cohort data set) were filtered out.

Individuals were defined as nonsecretors if they were homozygous for the minor allele of any of the following SNPs: rs601338, rs1047781, or rs200157007 (11, 25–27). Individuals were defined as Lewis negative if they were homozygous for the minor allele of any of the following SNPs: rs3745635, rs28362459, rs3894326, or rs812936 (11).

Detection of respiratory pathogens by TaqMan array cards. Nasopharyngeal samples collected at 2 and 4 months of age from infants asymptomatic for ARI were tested for 10 bacterial and 10 viral respiratory pathogens using customized TaqMan array cards (TAC) (Thermo Fisher Scientific, Waltham, MA) according to the methods outlined in Saha et al. 2018 (34). The design included one internal positive control (IPCO1) and two specimen quality control assays (RNaseP and MS2 bacteriophage). A total of 45 μ l of nucleic acid extract was amplified with the qScript XLT 1-step RT-qPCR ToughMix Low-ROX kit (Quanta Biosciences, Gaithersburg, MD) for a total volume of 100 μ l per testing lane. Six specimens were tested per card, along with a positive and negative control on each TAC. The negative control consisted of nuclease-free water, while the positive control was a pool of engineered templates that were designed to amplify at a consistent threshold cycle (C_{t}) range (60). The following cycling conditions were used: 45°C for 10 min, 94°C for 10 min, and 45 cycles of 94°C for 30 s followed by 60°C for 1 min on a QuantStudio 7 real-time PCR system (Thermo Fisher Scientific). Data analysis was performed using the QuantStudio 7 software.

Sequencing and analysis of the nasopharyngeal microbiome. A total of 422 nasopharyngeal samples from 2- and 4-month-old infants asymptomatic for ARI at the time of sampling were available for shotgun metagenomics sequencing on the Illumina HiSeq platform (for details see Supplemental Text S1). Briefly, after quality control and host-read filtering, the remaining reads were *de novo* assembled using MEGAHIT, and genes were predicted on all samples using Prodigal. Gene predictions were merged with genes from respiratory tract microbial species sequenced by the Human Microbiome Project (61), and all genes were clustered at 95% similarity to build a nonredundant nasopharynx gene catalog. Gene abundances for each sample were then computed by mapping filtered reads to the gene catalog. Metagenomic species (MGS) were identified based on the coabundance of genes across the 422 samples, following the method defined of Nielsen and colleagues (62). To annotate the MGS, all the catalog genes were subjected to a BLAST search of the NCBI RefSeq genome database and used different levels of similarity (ranging from 65% for phylum to 95% for species) to annotate at the various taxonomic levels.

Statistical analyses. Clinical data. Statistical analyses of clinical data have been performed using SAS (v9.4). For computation of odds ratios (OR), incidence rate ratios (IRR), confidence intervals (CI) and *P* values, the logistic regression was used. The secretor and/or Lewis status of mothers and/or infants were included as covariates. When the response was measured repeatedly, the repeated statement for subject

and the covariate “visit” were included in the models as well. IRR were used to compare the number of ARIs in given time periods between groups, whereas OR compared the occurrence of at least one ARI between groups. The Wald test was used to compare proportions of infants colonized by respiratory pathogens between groups symptomatic and asymptomatic for ARI at the time of sample collection.

Methods used to associate colonization by respiratory pathogens with ARI. The association between respiratory pathogen colonization at 2 and 4 months and subsequent ARI was evaluated through a Kaplan-Meier analysis, including only pathogens for which at least 20 subjects tested positive.

Machine learning approaches to associate microbiome with explanatory variables and risk of ARI. Machine learning analyses on microbiome composition are detailed in Supplemental Text S1. Briefly, the caret package v6.0-80 (63) in R (v3.5.0) was used to predict whether there was an ARI over the next 1, 2, 6, or 20 months after sampling based on microbiome composition at 2 or 4 months together with selected clinical factors. Models were built using three different approaches—GLMnet, random forest, and LogitBoost—and microbiome abundance data were transformed as either (i) presence/absence (0/1), (ii) raw abundance percentages, or (iii) \log_{10} -transformed. Samples were partitioned into training and testing sets at a ratio of 80/20%, and results were evaluated considering accuracy, sensitivity, and specificity.

Data availability. Genetic data and raw nasopharyngeal microbiome reads depleted of human sequences were deposited in the European Nucleotide Archive (ENA) under the project accession number [PRJEB42539](https://www.ebi.ac.uk/ena/record/PRJEB42539) including descriptive metadata. Additional metadata can be shared by the corresponding author on reasonable request.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.2 MB.

TEXT S1, DOCX file, 0.04 MB.

FIG S1, DOCX file, 0.03 MB.

FIG S2, DOCX file, 0.5 MB.

FIG S3, DOCX file, 0.3 MB.

TABLE S1, DOCX file, 0.03 MB.

TABLE S2, DOCX file, 0.03 MB.

TABLE S3, DOCX file, 0.04 MB.

TABLE S4, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

We are grateful to the participants and their caregivers, as well as the study team members for their contributions. We thank Cheng Chen, Caroline Le Roy, Pavla Kadlecova, Ondrej Slavicek, and Marie Simeckova for support in statistical analysis. We are very grateful to Jonas Winchell for essential input on the design and analysis of TaqMan array cards. Laura M. Gosoni, Ivan Montoliu Roura, and Elisabeth Forbes Blum provided helpful comments on the manuscript. We thank the two anonymous reviewers whose comments helped improve and clarify the manuscript.

The study was funded by Société des Produits Nestlé S A, Vevey, Switzerland.

REFERENCES

- Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, O'Brien KL, Campbell H, Black RE. 2013. Global burden of childhood pneumonia and diarrhoea. *Lancet* 381:1405–1416. [https://doi.org/10.1016/S0140-6736\(13\)60222-6](https://doi.org/10.1016/S0140-6736(13)60222-6).
- Henkle E, Steinhoff MC, Omer SB, Roy E, Arifeen SE, Raqib R, Breiman RF, Caulfield LE, Moss WJ, Zaman K. 2013. The effect of exclusive breast-feeding on respiratory illness in young infants in a maternal immunization trial in Bangladesh. *Pediatr Infect Dis J* 32:431–435. <https://doi.org/10.1097/INF.0b013e318281e34f>.
- Lamberti LM, Zakarija-Grković I, Fischer Walker CL, Theodoratou E, Nair H, Campbell H, Black RE. 2013. Breastfeeding for reducing the risk of pneumonia morbidity and mortality in children under two: a systematic literature review and meta-analysis. *BMC Public Health* 13:S18. <https://doi.org/10.1186/1471-2458-13-S3-S18>.
- Victora CG, Barros AJD. 2000. Effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: a pooled analysis. WHO Collaborative Study Team on the Role of Breastfeeding on the Prevention of Infant Mortality. *Lancet* 355:451–455.
- Kunz C, Rodriguez-Palmero M, Koletzko B, Jensen R. 1999. Nutritional and biochemical properties of human milk. Part I. General aspects, proteins, and carbohydrates. *Clin Perinatol* 26:307–333. [https://doi.org/10.1016/S0095-5108\(18\)30055-1](https://doi.org/10.1016/S0095-5108(18)30055-1).
- Coppa GV, Gabrielli O, Pierani P, Catassi C, Carlucci A, Giorgi PL. 1993. Changes in carbohydrate composition in human milk over 4 months of lactation. *Pediatrics* 91:637–641.
- Newburg DS. 1999. Human milk glycoconjugates that inhibit pathogens. *Curr Med Chem* 6:117–127.
- Doherty AM, Lodge CJ, Dharmage SC, Dai X, Bode L, Lowe AJ. 2018. Human milk oligosaccharides and associations with immune-mediated disease and infection in childhood: a systematic review. *Front Pediatr* 6:91. <https://doi.org/10.3389/fped.2018.00091>.
- Sprenger N, Binia A, Austin S. 2019. Human milk oligosaccharides: factors affecting their composition and their physiological significance. *Nestle Nutr Inst Workshop Ser* 90:43–56. <https://doi.org/10.1159/000490292>.
- Samuel TM, Binia A, de Castro CA, Thakkar SK, Billeaud C, Agosti M, Al-Jashi I, Costeira MJ, Marchini G, Martínez-Costa C, Picaud J-C, Stiris T, Stoicescu S-M, Vanpeé M, Domellöf M, Austin S, Sprenger N. 2019. Impact of maternal characteristics on human milk oligosaccharide composition

- over the first 4 months of lactation in a cohort of healthy European mothers. *Sci Rep* 9:11767. <https://doi.org/10.1038/s41598-019-48337-4>.
11. Lefebvre G, Shevlyakova M, Charpagne A, Marquis J, Vogel M, Kirsten T, Kiess W, Austin S, Sprenger N, Binia A. 2020. Time of lactation and maternal fucosyltransferase genetic polymorphisms determine the variability in human milk oligosaccharides. *Front Nutr* 7. <https://doi.org/10.3389/anut.2020.574459>.
 12. Morrow AL, Ruiz-Palacios GM, Altaye M, Jiang X, Guerrero ML, Meinzen-Derr JK, Farkas T, Chaturvedi P, Pickering LK, Newburg DS. 2004. Human milk oligosaccharide blood group epitopes and innate immune protection against campylobacter and calcivirous diarrhea in breastfed infants. *Adv Exp Med Biol* 554:443–446. https://doi.org/10.1007/978-1-4757-4242-8_61.
 13. Newburg DS, Ruiz-Palacios GM, Altaye M, Chaturvedi P, Meinzen-Derr J, Guerrero M. d L, Morrow AL. 2004. Innate protection conferred by fucosylated oligosaccharides of human milk against diarrhea in breastfed infants. *Glycobiology* 14:253–263. <https://doi.org/10.1093/glycob/cwh020>.
 14. Davis JCC, Lewis ZT, Krishnan S, Bernstein RM, Moore SE, Prentice AM, Mills DA, Lebrilla CB, Zivkovic AM. 2017. Growth and morbidity of Gambian infants are influenced by maternal milk oligosaccharides and infant gut microbiota. *Sci Rep* 7:40466. <https://doi.org/10.1038/srep40466>.
 15. Puccio G, Alliet P, Cajazzo C, Janssens E, Corsello G, Sprenger N, Wernimont S, Egli D, Gosoni L, Steenhout P. 2017. Effects of infant formula with human milk oligosaccharides on growth and morbidity: a randomized multicenter trial. *J Pediatr Gastroenterol Nutr* 64:624–631. <https://doi.org/10.1097/MPG.0000000000001520>.
 16. Morrow AL, Ruiz-Palacios GM, Jiang X, Newburg DS. 2005. Human-milk glycans that inhibit pathogen binding protect breast-feeding infants against infectious diarrhea. *J Nutr* 135:1304–1307. <https://doi.org/10.1093/jn/135.5.1304>.
 17. Yu ZT, Nanthakumar NN, Newburg DS. 2016. The human milk oligosaccharide 2'-fucosyllactose quenches *Campylobacter jejuni*-induced inflammation in human epithelial cells HEP-2 and HT-29 and in mouse intestinal mucosa. *J Nutr* 146:1980–1990. <https://doi.org/10.3945/jn.116.230706>.
 18. Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS. 2003. *Campylobacter jejuni* binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J Biol Chem* 278:14112–14120. <https://doi.org/10.1074/jbc.M207744200>.
 19. Cravioto A, Tello A, Villafan H, Ruiz J, del Vedovo S, Neeser JR. 1991. Inhibition of localized adhesion of enteropathogenic *Escherichia coli* to HEP-2 cells by immunoglobulin and oligosaccharide fractions of human colostrum and breast milk. *J Infect Dis* 163:1247–1255. <https://doi.org/10.1093/infdis/163.6.1247>.
 20. Andersson B, Porras O, Hanson LA, Lagergard T, Svanborg-Eden C. 1986. Inhibition of attachment of *Streptococcus pneumoniae* and *Haemophilus influenzae* by human milk and receptor oligosaccharides. *J Infect Dis* 153:232–237. <https://doi.org/10.1093/infdis/153.2.232>.
 21. Vissing NH, Chawes BL, Bisgaard H. 2013. Increased risk of pneumonia and bronchiolitis after bacterial colonization of the airways as neonates. *Am J Respir Crit Care Med* 188:1246–1252. <https://doi.org/10.1164/rccm.201302-0215OC>.
 22. Sakwinska O, Foata F, Berger B, Brüssow H, Combremont S, Mercenier A, Dogra S, Soh S-E, Yen JCK, Heong GYS, Lee YS, Yap F, Meaney MJ, Chong Y-S, Godfrey KM, Holbrook JD. 2017. Does the maternal vaginal microbiota play a role in seeding the microbiota of neonatal gut and nose? *Benef Microbes* 8:763–778. <https://doi.org/10.3920/BM2017.0064>.
 23. de Steenhuijsen PETERS WAA, Binkowska J, Bogaert D. 2020. Early life microbiota and respiratory tract infections. *Cell Host Microbe* 28:223–232. <https://doi.org/10.1016/j.chom.2020.07.004>.
 24. Vidal K, Sultana S, Patron AP, Binia A, Rahman M, Deeba IM, Brüssow H, Sakwinska O, Sarker SA. 2019. Microbiota and Health Study: a prospective cohort of respiratory and diarrheal infections and associated risk factors in Bangladeshi infants under two years. medRxiv <https://doi.org/10.1101/1900505>.
 25. Mottram L, Wiklund G, Larson G, Qadri F, Svennerholm AM. 2017. FUT2 non-secretor status is associated with altered susceptibility to symptomatic enterotoxigenic *Escherichia coli* infection in Bangladeshis. *Sci Rep* 7:10649. <https://doi.org/10.1038/s41598-017-10854-5>.
 26. Guo M, Luo G, Lu R, Shi W, Cheng H, Lu Y, Jin K, Yang C, Wang Z, Long J, Xu J, Ni Q, Liu C, Yu X. 2017. Distribution of Lewis and Secretor polymorphisms and corresponding CA19-9 antigen expression in a Chinese population. *FEBS Open Bio* 7:1660–1671. <https://doi.org/10.1002/2211-5463.12278>.
 27. de Moraes MTB, Olivares AIO, Fialho AM, Malta FC, da Silva E Mouta Junior S, de Souza Bispo R, Velloso AJ, Alves Leitão GA, Cantelli CP, Nordgren J, Svenson L, Miagostovich MP, Leite JPG. 2019. Phenotyping of Lewis and secretor HBGA from saliva and detection of new FUT2 gene SNPs from young children from the Amazon presenting acute gastroenteritis and respiratory infection. *Infect Genet Evol* 70:61–66. <https://doi.org/10.1016/j.meegid.2019.02.011>.
 28. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn L-J, Knetsch CW, Figueiredo C. 2019. Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front Microbiol* 10:1277. <https://doi.org/10.3389/fmicb.2019.01277>.
 29. Ruvoen-Clouet N, Mas E, Marionneau S, Guillon P, Lombardo D, Le Pendu J. 2006. Bile-salt-stimulated lipase and mucins from milk of 'secretor' mothers inhibit the binding of Norwalk virus capsids to their carbohydrate ligands. *Biochem J* 393:627–634. <https://doi.org/10.1042/BJ20050898>.
 30. Barton SJ, Murray R, Lillycrop KA, Inskip HM, Harvey NC, Cooper C, Karnani N, Zolezzi IS, Sprenger N, Godfrey KM, Binia A. 2019. FUT2 genetic variants and reported respiratory and gastrointestinal illnesses during infancy. *J Infect Dis* 219:836–843. <https://doi.org/10.1093/infdis/jiy582>.
 31. Stepan MBF, Wilhelm SL, Hertzog M, Rodehorst TKC, Blaney S, Clemens B, Polak JJ, Newburg DS. 2006. Early consumption of human milk oligosaccharides is inversely related to subsequent risk of respiratory and enteric disease in infants. *Breastfeed Med* 1:207–215. <https://doi.org/10.1089/bfm.2006.1.207>.
 32. Raza MW, Blackwell CC, Molyneux P, James VS, Ogilvie MM, Inglis JM, Weir DM. 1991. Association between secretor status and respiratory viral illness. *BMJ* 303:815–818. <https://doi.org/10.1136/bmj.303.6806.815>.
 33. van Meel ER, Jaddoe VWW, Looman KIM, de Jongste JC, Moll HA, Duijts L. 2020. Airway bacterial carriage and childhood respiratory health: a population-based prospective cohort study. *Pediatr Allergy Immunol* 31:774–782. <https://doi.org/10.1111/pai.13310>.
 34. Saha SK, Schrag SJ, El Arifeen S, Mullany LC, Shahidul Islam M, Shang N, Qazi SA, Zaidi AKM, Bhutta ZA, Bose A, Panigrahi P, Soofi SB, Connor NE, Mitra DK, Isaac R, Winchell JM, Arvay ML, Islam M, Shafiq Y, Nisar I, Baloch B, Kabir F, Ali M, Diaz MH, Satpathy R, Nanda P, Padhi BK, Parida S, Hotwani A, Hasanuzzaman M, Ahmed S, Belal Hossain M, Ariff S, Ahmed I, Ibne Moin SM, Mahmud A, Waller JL, Rafiqullah I, Quaiyum MA, Begum N, Balaji V, Halen J, Nawshad Uddin Ahmed ASM, Weber MW, Hamer DH, Hibberd PL, Sadeq-Ur Rahman Q, Mogan VR, Hossain T, McGee L, et al. 2018. Causes and incidence of community-acquired serious infections among young children in south Asia (ANISA): an observational cohort study. *Lancet* 392:145–159. [https://doi.org/10.1016/S0140-6736\(18\)31127-9](https://doi.org/10.1016/S0140-6736(18)31127-9).
 35. Darboe MK, Fulford AJ, Secka O, Prentice AM. 2010. The dynamics of nasopharyngeal streptococcus pneumoniae carriage among rural Gambian mother-infant pairs. *BMC Infect Dis* 10:195. <https://doi.org/10.1186/1471-2334-10-195>.
 36. Coles CL, Labrique A, Saha SK, Ali H, Al-Emran H, Rashid M, Christian P, West KP, Klemm R. 2011. Newborn vitamin A supplementation does not affect nasopharyngeal carriage of *Streptococcus pneumoniae* in Bangladeshi infants at age 3 months. *J Nutr* 141:1907–1911. <https://doi.org/10.3945/jn.111.141622>.
 37. Madhi SA, Izu A, Nunes MC, Violari A, Cotton MF, Jean-Philippe P, Klugman KP, von Gottberg A, van Niekerk N, Adrian PV, CIPRA 4 Team. 2015. Longitudinal study on *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus* nasopharyngeal colonization in HIV-infected and -uninfected infants vaccinated with pneumococcal conjugate vaccine. *Vaccine* 33:2662–2669. <https://doi.org/10.1016/j.vaccine.2015.04.024>.
 38. Bogaert D, Keijsers B, Huse S, Rossen J, Veenhoven R, van Gils E, Bruin J, Montijn R, Bonten M, Sanders E. 2011. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One* 6:e17035. <https://doi.org/10.1371/journal.pone.0017035>.
 39. Sakwinska O, Bastic Schmid V, Berger B, Bruttin A, Keitel K, Lepage M, Moine D, Ngom Bru C, Brüssow H, Gervais A. 2014. Nasopharyngeal microbiota in healthy children and pneumonia patients. *J Clin Microbiol* 52:1590–1594. <https://doi.org/10.1128/JCM.03280-13>.
 40. Thorsen J, Rasmussen MA, Waage J, Mortensen M, Brejnrod A, Bonnelykke K, Chawes BL, Brix S, Sørensen SJ, Stokholm J, Bisgaard H. 2019. Infant airway microbiota and topical immune perturbations in the origins of childhood asthma. *Nat Commun* 10:5001. <https://doi.org/10.1038/s41467-019-12989-7>.
 41. Stewart CJ, Mansbach JM, Wong MC, Ajami NJ, Petrosino JF, Camargo CA, Hasegawa K. 2017. Associations of nasopharyngeal metabolome and microbiome with severity among infants with bronchiolitis. A multiomic

- analysis. *Am J Respir Crit Care Med* 196:882–891. <https://doi.org/10.1164/rccm.201701-0071OC>.
42. Manenzhe RI, Dube FS, Wright M, Lennard K, Mounaud S, Lo SW, Zar HJ, Nierman WC, Nicol MP, Moodley C. 2020. Characterization of pneumococcal colonization dynamics and antimicrobial resistance using shotgun metagenomic sequencing in intensively sampled South African infants. *Front Public Health* 8:543898. <https://doi.org/10.3389/fpubh.2020.543898>.
 43. Bosch AATM, de Steenhuijsen Piters WAA, van Houten MA, Chu MLJN, Biesbroek G, Kool J, Pernet P, de Groot P-KCM, Eijkemans MJC, Keijsers B, Sanders EAM, Bogaert D. 2017. Maturation of the infant respiratory microbiota, environmental drivers, and health consequences. A prospective cohort study. *Am J Respir Crit Care Med* 196:1582–1590. <https://doi.org/10.1164/rccm.201703-0554OC>.
 44. Santee CA, Nagalingam NA, Faruqi AA, DeMuri GP, Gern JE, Wald ER, Lynch SV. 2016. Nasopharyngeal microbiota composition of children is related to the frequency of upper respiratory infection and acute sinusitis. *Microbiome* 4:34. <https://doi.org/10.1186/s40168-016-0179-9>.
 45. Kelly MS, Surette MG, Smieja M, Pernica JM, Rossi L, Luinstra K, Steenhoff AP, Feemster KA, Goldfarb DM, Arscott-Mills T, Boiditswe S, Rulaganyang I, Muthoga C, Gaofiwe L, Mazhani T, Rawls JF, Cunningham CK, Shah SS, Seed PC. 2017. The nasopharyngeal microbiota of children with respiratory infections in Botswana. *Pediatr Infect Dis J* 36:e211–e218. <https://doi.org/10.1097/INF.0000000000001607>.
 46. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt BJ, Hales BJ, Walker ML, Hollams E, Bochkov YA, Grindle K, Johnston SL, Gern JE, Sly PD, Holt PG, Holt KE, Inouye M. 2015. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* 17:704–715. <https://doi.org/10.1016/j.chom.2015.03.008>.
 47. Biesbroek G, Tsivtsivadze E, Sanders EAM, Montijn R, Veenhoven RH, Keijsers B, Bogaert D. 2014. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med* 190:1283–1292. <https://doi.org/10.1164/rccm.201407-1240OC>.
 48. Toivonen L, Hasegawa K, Waris M, Ajami NJ, Petrosino JF, Camargo CA, Jr, Peltola V. 2019. Early nasal microbiota and acute respiratory infections during the first years of life. *Thorax* 74:592–599. <https://doi.org/10.1136/thoraxjnl-2018-212629>.
 49. Teo SM, Tang H, Mok D, Judd LM, Watts SC, Pham K, Holt BJ, Kusel M, Serralha M, Troy N, Bochkov YA, Grindle K, Lemanske RF, Johnston SL, Gern JE, Sly PD, Holt PG, Holt KE, Inouye M. 2018. Airway microbiota dynamics uncover a critical window for interplay of pathogenic bacteria and allergy in childhood respiratory disease. *Cell Host Microbe* 24:341–352.e345. <https://doi.org/10.1016/j.chom.2018.08.005>.
 50. Yi H, Yong D, Lee K, Cho YJ, Chun J. 2014. Profiling bacterial community in upper respiratory tracts. *BMC Infect Dis* 14:583. <https://doi.org/10.1186/s12879-014-0583-3>.
 51. Bogaert D, van Belkum A, Sluiter M, Luijendijk A, de Groot R, Rümke HC, Verbrugh HA, Hermans PWM. 2004. Colonisation by *Streptococcus pneumoniae* and *Staphylococcus aureus* in healthy children. *Lancet* 363:1871–1872. [https://doi.org/10.1016/S0140-6736\(04\)16357-5](https://doi.org/10.1016/S0140-6736(04)16357-5).
 52. Antunes KH, Fachi JL, de Paula R, da Silva EF, Pral LP, dos Santos AA, Dias GBM, Vargas JE, Puga R, Mayer FQ, Maito F, Zárata-Bladés CR, Ajami NJ, Sant'Ana MR, Candreva T, Rodrigues HG, Schmiele M, Silva Clerici MTP, Proença-Modena JL, Vieira AT, Mackay CR, Mansur D, Caballero MT, Marzec J, Li J, Wang X, Bell D, Polack FP, Kleeberger SR, Stein RT, Vinolo MAR, de Souza APD. 2019. Microbiota-derived acetate protects against respiratory syncytial virus infection through a GPR43-type 1 interferon response. *Nat Commun* 10:3273. <https://doi.org/10.1038/s41467-019-11152-6>.
 53. Harding JN, Siefker D, Vu L, You D, DeVincenzo J, Pierre JF, Cormier SA. 2020. Altered gut microbiota in infants is associated with respiratory syncytial virus disease severity. *BMC Microbiol* 20:140. <https://doi.org/10.1186/s12866-020-01816-5>.
 54. Budden KF, Gellatly SL, Wood DLA, Cooper MA, Morrison M, Hugenholtz P, Hansbro PM. 2017. Emerging pathogenic links between microbiota and the gut-lung axis. *Nat Rev Microbiol* 15:55–63. <https://doi.org/10.1038/nrmicro.2016.142>.
 55. Berger B, Porta N, Foata F, Grathwohl D, Delley M, Moine D, Chappagne A, Siegwald L, Descombes P, Alliet P, Puccio G, Steenhout P, Mercenier A, Sprenger N. 2020. Linking human milk oligosaccharides, infant fecal community types, and later risk to require antibiotics. *mBio* 11:e03196-19. <https://doi.org/10.1128/mBio.03196-19>.
 56. Williams FB, Kader A, Colgate ER, Dickson DM, Carmolli M, Uddin MI, Sharmin S, Islam S, Bhuiyan TR, Alam A, Nayak U, Mychalekyj JC, Petri WA, Jr, Haque R, Qadri F, Kirkpatrick BD, Lee B. 2020. Maternal secretor status affects oral rotavirus vaccine response in breastfed infants in Bangladesh. *J Infect Dis* 7:1147–1151. <https://doi.org/10.1093/infdis/jiaa101>.
 57. Lambert SB, Ware RS, Cook AL, Maguire FA, Whitley DM, Bialasiewicz S, Mackay IM, Wang D, Sloots TP, Nissen MD, Grimwood K. 2012. Observational Research in Childhood Infectious Diseases (ORChID): a dynamic birth cohort study. *BMJ Open* 2:e002134. <https://doi.org/10.1136/bmjopen-2012-002134>.
 58. Sarna M, Ware RS, Sloots TP, Nissen MD, Grimwood K, Lambert SB. 2016. The burden of community-managed acute respiratory infections in the first 2-years of life. *Pediatr Pulmonol* 51:1336–1346. <https://doi.org/10.1002/ppul.23480>.
 59. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv Preprint* <https://arxiv.org/abs/1207.3907>.
 60. Kodani M, Winchell JM. 2012. Engineered combined-positive-control template for real-time reverse transcription-PCR in multiple-pathogen-detection assays. *J Clin Microbiol* 50:1057–1060. <https://doi.org/10.1128/JCM.05987-11>.
 61. Human Microbiome Project Consortium. 2012. A framework for human microbiome research. *Nature* 486:215–221. <https://doi.org/10.1038/nature11209>.
 62. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto J-M, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezebur F, Casellas F, Doré J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Léonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin J, Raes J, Sørensen S, Tap J, Tims S, Ussery DW, Yamada T, Renault P, Sicheritz-Ponten T, Bork P, Wang J, MetaHIT Consortium, et al. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 32:822–828. <https://doi.org/10.1038/nbt.2939>.
 63. Kuhn M. 2008. Building predictive models in R using the caret package. *J Stat Softw* 28:1–26. <https://doi.org/10.18637/jss.v028.i05>.