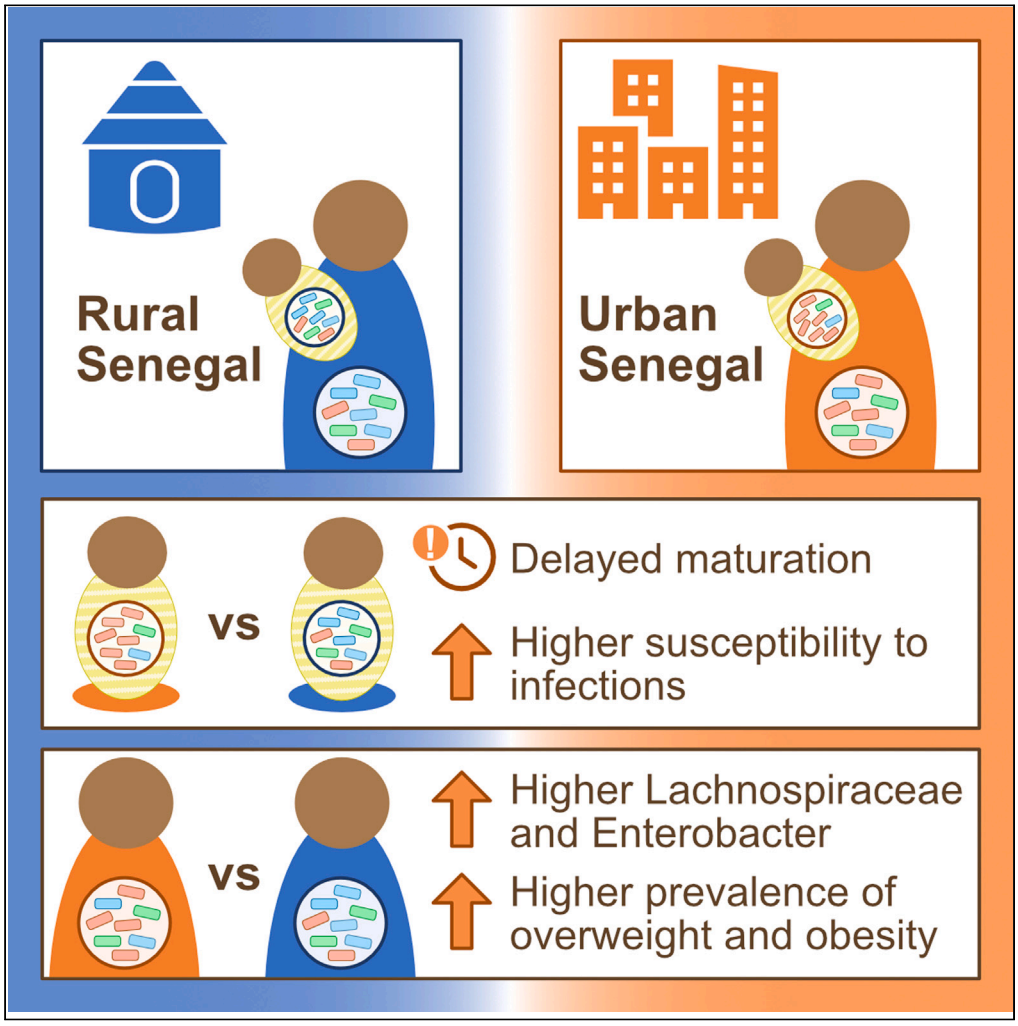
iScience

CellPress
OPEN ACCESS

Article

# Urbanization associates with restricted gut microbiome diversity and delayed maturation in infants




Francesco Morandini, Kevin Perez, Loic Brot, ..., Jean-Pierre Grill, Enguerran Macia, Philippe Seksik

philippe.seksik@aphp.fr

## Highlights

We compare gut microbiomes of mothers and infants from rural and urban Senegal

Urban infants show delayed microbiome maturation compared to rural infants

Urban infants were more susceptible to infectious diseases

Urban mothers had higher abundance of *Lachnospiraceae* and *Enterobacter*

Morandini et al., iScience *26*, 108136
November 17, 2023 
https://doi.org/10.1016/j.isci.2023.108136

iScience

CellPress
OPEN ACCESS

Article

# Urbanization associates with restricted gut microbiome diversity and delayed maturation in infants

Francesco Morandini,[1] Kevin Perez,[2] Loic Brot,[3,4] Sidy Mohammed Seck,[5] Laurence Tibère,[6] Jean-Pierre Grill,[4] Enguerran Macia,[7,8] and Philippe Seksik[3,4,9,10,*]

## SUMMARY

**Alterations of the microbiome are linked to increasingly common diseases such as obesity, allergy, and inflammatory bowel disease. Post-industrial lifestyles are thought to contribute to the gut microbiome alterations that cause or aggravate these diseases. Comparing communities across the industrialization spectrum can reveal associations between gut microbiome alterations and lifestyle and health, and help pinpoint which specific aspect of the post-industrial lifestyle is linked to microbiome alterations. Here, we compare the gut microbiomes of 60 mother and infant pairs from rural and urban areas of Senegal over two time points. We find that urban mothers, who were more frequently overweight, had different gut microbiome compositions than rural mothers, showing an expansion of *Lachnospiraceae* and *Enterobacter*. Urban infants, on the other hand, showed a delayed gut microbiome maturation and a higher susceptibility to infectious diseases. Thus, we identify new microbiome features associated with industrialization, whose association with disease may be further investigated.**

## INTRODUCTION

Gut microbiome composition is known to affect a vast number of biological processes, including metabolism, immunity, and behavior.[1–4] At the same time, gut microbiome composition is highly plastic, and responds to lifestyle and environmental factors.[5] Such factors change greatly as societies become industrialized, acquiring access to antibiotics and better sanitization, as well as transitioning to a westernized diet, typically characterized by increased consumption of refined carbohydrates, saturated fats, and reduced consumption of fruit, vegetables, and whole grains.[6] The post-industrial microbiome has drifted significantly from its pre-industrial state and several health conditions that are more common in industrialized societies such as obesity and inflammatory bowel disease are known to be linked with microbiome composition. As the microbiota and their human hosts have coevolved to form a symbiotic relationship, it is theorized that the perturbation of this equilibrium caused by the post-industrial lifestyle may be partially to blame for these health conditions. Comparison of post-industrial populations with communities that still practice pre-industrial subsistence strategies could allow us to identify which microbiome alterations associate with industrialization and may offer clues as to which lifestyle and environmental factors are responsible for these changes.[7]

In the last decades, the African continent has been making rapid advances toward industrialization; nonetheless, it is home to many communities whose lifestyle is still pre-industrial, as well as other communities that are in the midst of the transition. With this in mind, the societal and environmental diversity of Africa offers the possibility to study the microbiota of communities that share similar environments and genetics but occupy different points along the industrialization spectrum. Nonetheless, to this day, the number of microbiome studies carried out in Africa is much smaller than those carried out in Europe or Northern America.[7,8] Therefore, it is still unclear which microbial features can be considered universal aspects of the transition to a post-industrial lifestyle and which are specific to the regions and subsistence strategies of the studied populations. Moreover, especially when dealing with populations in remote areas, microbiome data are often scarce, comprising few samples; single time points and detailed information regarding overall health and lifestyle are not always available. The lack of multiple time points is particularly disadvantageous when studying the development of the infant gut microbiome as its composition

[1]Biology department, University of Rochester, Rochester, NY 14620, USA
[2]Biomedical Sciences department, University of Lausanne, 1005 Lausanne, Vaud, Switzerland
[3]Centre de Recherche Saint-Antoine, Assistance Publique-Hôpitaux de Paris, Hôpital Saint Antoine, Service de Gastroentérologie, Inserm, Sorbonne Université, 75012 Paris, France
[4]Sorbonne Université, INSERM UMRS-938, Centre de Recherche Saint-Antoine, CRSA, AP-HP, 75012 Paris, France
[5]Internal medicine/Nephrology department, Faculty of Health Sciences and IRL-3189 "Environnement, Santé, Sociétés", University Gaston Berger, Saint-Louis 00234, Senegal
[6]Centre d'Études et de Recherche: Travail, Organisation, Pouvoir (CERTOP) UMR CNRS 5044, Université de Toulouse, 31013 Toulouse, France
[7]International Research Laboratory "Environnement, Sant, Socits" (CNRS / UCAD / UGB / USTTB / CNRST), Dakar, Senegal
[8]Unit Mixte de Recherche 7268 Anthropologie bio-culturelle, Droit, Ethique et Sant (CNRS / AMU / EFS), Aix-Marseille, France
[9]Paris Center for Microbiome Medicine (PaCeMM) FHU, 75571 Paris, France
[10]Lead contact
*Correspondence: philippe.seksik@aphp.fr
https://doi.org/10.1016/j.isci.2023.108136

is highly dynamic during the first years of life. Finally, not all studies compared populations of the same ethnicity, confounding the effects of environmental and lifestyle factors with those determined by genetic differences.

To bridge these gaps, we have profiled the gut microbiomes of 30 mothers and infants from a pastoralist community in Senegal and compared them to 30 mothers and infants of the same ethnicity from an urban area, at two time points. Moreover, we have collected detailed information regarding the health and lifestyle of both populations. To our knowledge, this study is the first to analyze the microbiome of a pastoralist community in Africa, and to compare rural and urban African populations over multiple time points.

## RESULTS

### Population description

The Fula are an ethnic group spread throughout western Africa, northern central Africa, and northeastern Africa and comprise communities with vastly different lifestyles. For this study, we have recruited, within this ethnic group, two groups of mothers and infants from rural and urban areas of Senegal.

The individuals enrolled in the rural group reside in Widou Thiengoli in the Ferlo region in the north of Senegal. This region is located at the edge of the Sahara Desert and is thus extremely arid. The locals practice a nomadic pastoralist lifestyle, as the dry weather conditions are prohibitive to most food crops, except for the cowpea. Their diet is largely dependent on milk and other dairy products, cowpea, rice, and millet. Some of the rarer ingredients found in their cuisine include beef, dried fish, and a few herbs and spices used for seasoning. Notably, we previously observed some western influences in their diets, such as the occasional use of oil for cooking, bouillon, and sugar, typically added to tea or coffee.[9] The infrastructure in the region is minimal: there is no electricity and drinking water is sourced from ponds or boreholes. Moreover, the local medical centers often lack even basic supplies, including antibiotics. In summary, the lifestyle of the Fula living in Widou Thiengoli is almost completely pre-industrial.

The individuals enrolled in the urban group reside in Dakar, where living conditions contrast strongly with those of Widou Thiengoli. Dakar residents have access to electricity, running water, healthcare, and education. Their diet consists of a mixture of traditional Fulani dishes but also includes dishes of other local traditions as well as "globalized" ingredients and recipes, leading to much more nutritional variety than we observed in the diet of the rural group. Fish and rice were among the most common ingredients, accompanied by a large variety of vegetables and seasonings. Oil and salt or bouillon was often used for cooking. In conclusion, the lifestyle of the individuals in the urban group and the environment in which they live is decidedly post-industrial.

### Study design and sampling

The cohort examined in this study consists of 60 Fula mother and infant pairs, 30 of which formed the rural group from Widou Thiengoli area, and the remaining 30 formed the urban group from Dakar (Figure 1A). We collected stool samples at two time points, the first (T1) within 6 months of delivery and the second (T2) one year after T1 when all infants had been introduced to a (partially) solid diet.

We detected a total of 4901 amplicon sequence variants which we mapped to 21 phyla and 385 genera by referencing to the SILVA 13.8 database.[10] Analysis of rarefaction curves revealed that the number of detected taxa plateaued before 14000 read pairs (Figure S1), meaning that all samples but one had sufficient sequencing depth to capture the full richness of the microbial communities and were thus included in the analysis. Importantly, stool sample collection at both time points was accompanied by a medical examination including personal and family history of diseases, body mass index measurements, current treatments (all medication including antibiotics and NSAIDs), as well as an interview on lifestyle and dietary habits. We note that antibiotic use was overall low in both rural and urban groups during the study: only two infants belonging to the urban group had recently been administered amoxicillin, shortly before T1. Additionally, all infants except two (1 rural, 1 urban) had been delivered naturally. Two urban infants were partially formula-fed, whereas the remainder of the infants were fed with fresh milk. The key descriptors of the studied cohort are summarized in Tables 1 and 2, while the full metadata is available in Table S1.

### The gut microbiome of mothers is shaped by urbanization status, time since pregnancy, and colonization by *Entamoeba coli*

First, we explored broad differences between the gut microbiomes of rural and urban mothers (Figure 1B). At T1, the most prevalent phyla were Firmicutes (79%), Bacteroidota (11%), Proteobacteria (5%), and Actinobacteriota (4%), with no significant differences between rural and urban groups (Mann-Whitney U test, $p = 0.542, 0.390, 0.889, 0.752$). At T2, the most prevalent phyla were Firmicutes (73%), Actinobacteriota (10%), Bacteroidota (9%), and Proteobacteria (4%) with lower abundance of Firmicutes in rural individuals compared to urban individuals, and no other significant differences ($p = 0.036, 0.607, 0.503, 0.221$). Between T1 and T2, we observed an increase in the abundance of Actinobacteriota in both rural and urban mothers (Wilcoxon signed-rank test, rural: $p = 0.020$, urban: $p = 0.039$) which seemed to correspond to a decrease in Firmicutes in rural mothers but not in urban ones (rural: $p = 0.046$, urban: $p = 0.324$).

Alpha diversity, measured by multiple metrics, was higher in urban mothers, compared to rural, although not significantly (Figures 2A and S2). Curiously, we observed an increase in alpha diversity between T1 and T2 for both rural and urban mothers. As the two time points were processed in separate batches, this difference could be due to a technical batch effect. On the other hand, previous studies have shown that alpha diversity decreases during pregnancy,[11,12] thus the increase in alpha diversity between T1 and T2 might reflect a recovery from the state of low diversity at the end of pregnancy. Indeed, at T1, the Chao1 index of mothers was positively correlated with time since delivery (Figure 2B). We looked for further associations between alpha diversity and lifestyle/environmental factors by constructing linear regression

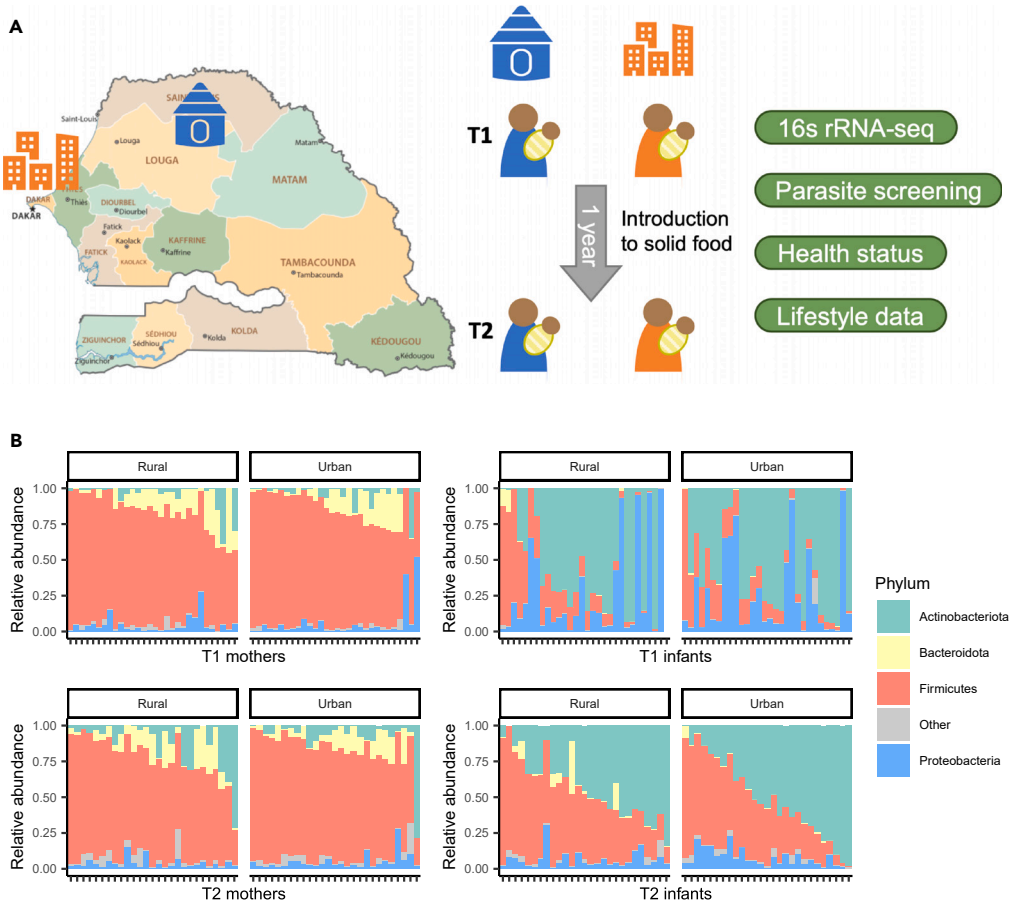



**Figure 1. Design of the study and phylum-level gut microbiome composition of the study groups**

(A) Design of the study.

(B) Proportions of most common phyla for mothers and infants at both time points. One bar corresponds to one individual. Individuals have been ordered by decreasing proportion of Firmicutes.

models with the Chao1 index as dependent variable and selecting explanatory variables by stepwise feature selection. The resulting model selected the covariates for time since pregnancy, rural/urban environment, and colonization by *Entamoeba coli* (t since preg.: $p < 0.001$, environment: $p = 0.038$, *Entamoeba coli*: $p = 0.028$). *Entamoeba coli* is a non-pathogenic amoeba which was previously found to strongly affect gut microbiome composition in several Cameroonian communities.[13] In our cohort, colonization by *Entamoeba coli* was associated with increased alpha diversity (Figure 2C) and was more common in the rural group than the urban one, albeit not significantly (Fisher's exact test: odds ratio = 0.42, $p = 0.107$).

Beta diversity analysis using permanova[14] produced different results based on the metric of beta diversity employed: differences between rural and urban environments had much stronger effect and significance on Bray-Curtis, Aitchison, and unweighted UniFrac distances than on

**Table 1. Summary of the mother's cohort characteristics**

| | Rural mothers T1 | Urban mothers T1 | Rural mothers T2 | Urban mothers T2 |
|---|---|---|---|---|
| Number | 30 | 30 | 27 | 27 |
| Age (years) | 26 ± 8.6 | 26.5 ± 6.1 | 27.9 ± 8.4 | 27.7 ± 6.4 |
| Height (cm) | 164.2 ± 6.2 | 161.7 ± 4.9 | 164.1 ± 6.7 | 161.9 ± 5 |
| Weight (kg) | 57.9 ± 9.2 | 62.7 ± 13.3 | 57 ± 10.1 | 64.1 ± 18.6 |
| BMI | 21.4 ± 2.4 | 24 ± 5.1 | 21.1 ± 2.8 | 24.4 ± 7.1 |
| WHR | 0.78 ± 0.06 | 0.81 ± 0.06 | 0.8 ± 0.04 | 0.83 ± 0.08 |
| *Entamoeba coli* + | 10 (33.3%) | 5 (16.7%) | 6 (22.2%) | 3 (11.1%) |

**Table 2. Summary of the infants' cohort characteristics**

| | Rural infants T1 | Urban infants T1 | Rural infants T2 | Urban infants T2 |
|---|---|---|---|---|
| Number | 30 | 30 | 27 | 27 |
| Age (months) | 3 ± 1.7 | 2.6 ± 1.6 | 14.8 ± 1.7 | 14.5 ± 1.9 |
| Male \| Female | 19 \| 10 | 17 \| 13 | 17 \| 9 | 16 \| 11 |
| Height (cm) | 60.6 ± 6.3 | 58.6 ± 5.6 | 78.3 ± 4.4 | 74.2 ± 3.4 |
| Weight (kg) | 5.8 ± 1.7 | 5.6 ± 1.7 | 9.3 ± 1.4 | 9.4 ± 1.3 |
| BMI | 15.4 ± 2 | 16 ± 2.5 | 15.2 ± 1.8 | 17 ± 1.7 |
| Head circumference (cm) | 39.9 ± 3.2 | 39.5 ± 2.6 | 45.6 ± 1.7 | 45.9 ± 1.2 |
| C section delivery | 1 (3.3%) | 1 (3.3%) | 1 (3.7%) | 0 (0%) |
| Time since solid diet introduced (months) | 0 ± 0 | 0 ± 0 | 9 ± 2.6 | 7.7 ± 2.3 |
| Weaned | 0 (0%) | 0 (0%) | 1 (3.7%) | 2 (7.4%) |

the weighted UniFrac distance (Bray-Curtis: R2 = 0.019, p = 0.001, Aitchison: R2 = 0.015, p = < 0.001, UniFracU: R2 = 0.016, p = 0.027, UniFracW: R2 = 0.010, p = 0.348). As the type of environment had a stronger effect on the unweighted UniFrac distance than on its weighted counterpart, this could mean that the microbiome of mothers residing in rural areas differed from the microbiome of urban dwellers mostly by the presence or absence of rare, phylogenetically diverse taxa. We note that, despite significance, the variance explained by the environment was relatively low (R2 ~ = 1%–2%). This is, however, in line with similar studies which investigated associations between lifestyle, environment, and gut microbiome composition in relatively homogeneous populations.[13,15,25] Next, we repeated permanova including terms for potential confounders (Figures 2D and S3). As for alpha diversity, time since pregnancy and *Entamoeba coli* parasitism had a significant effect. The effect of time since pregnancy was stronger on unweighted rather than weighted UniFrac distances, whereas the opposite was true for *Entamoeba coli* parasitism. Notably, *Entamoeba coli*-positive individuals formed a clear cluster in the first two components of principal coordinate analysis (PCoA) on Bray-Curtis dissimilarities (Figure 2E). Finally, BMI and age had generally weak effects among the explored variables, but they were nonetheless significant according to most metrics.

Lastly, we used Maaslin2[16] to look for microbial genera whose abundance differed between rural and urban environments. We included covariates for time since pregnancy and *Entamoeba coli* parasitism due to the strong effects that they showed on alpha and beta diversity, as well as random effect terms for individuals and time points to account for repeated measurements and batch effects (Figure 2F). Focusing on differences related to urbanization, we observed a loss of *Enterococcus* and *Lactococcus* in urban mothers. These genera are closely related and found in fermented dairy foods.[17,18] Thus, their difference in abundance between the rural and urban group is probably due to higher consumption of dairy products in the pastoralist rural population. *Butyrivibrio* and *Lachnospira* are two genera within *Lachnospiraceae*, a family with conflicting associations with human health.[19] While several studies found that higher overall levels of *Lachnospiraceae* and *Lachnospira* associated with obesity and metabolic syndrome, not all genera within *Lachnospiraceae* associated with negative health conditions.[20,21] Nonetheless, the group of urban mothers showed increased abundance of *Butyrivibrio* and *Lachnospira* and higher BMI (Table 1) compared to the rural group, possibly confirming the association between *Lachnospiraceae* and obesity in this population. Lastly, *Enterobacter*, another genus with association to obesity, was more abundant in urban mothers.[22,23]

## The gut microbiome of urban infants shows delayed maturation

Next, we investigated the effects of urban conditions on the composition and development of the infant microbiome. At T1, when infants were exclusively milk-fed, the most abundant phyla were Actinobacteriota (55%), Proteobacteria (26%), Firmicutes (19%), and Bacteroidota (<1%), with no significant differences between the rural and urban group (Figure 1B, Mann-Whitney U test, p = 0.572, 0.958, 0.898, 0.388). Interestingly, we observed a sub-group (n = 12) of individuals whose microbiome was almost entirely populated by Proteobacteria. At T2, after the introduction of solid foods, the infant's microbiome mainly comprised Actinobacteriota (46%), Firmicutes (43%), Proteobacteria (7%), and Bacteroidota (2%) with a higher abundance of Bacteroidota in rural infants and no other significant differences (Mann-Whitney U test, p = 0.242, 0.335, 0.400, 0.037). None of the infants at T2 exhibited the high abundance of Proteobacteria observed at T1. Between T1 and T2, both rural and urban groups saw an increase in abundance of Firmicutes (rural: p = 0.002, urban: p = 0.001) and a decrease in Proteobacteria (rural: p = 0.019, urban: p = 0.003). The increase in abundance of Bacteroidota was more pronounced in the urban group than in the rural one, even though Bacteroidota were more abundant in the microbiome or rural infants (rural: p = 0.095, urban: p = 0.033).

Before continuing our analysis, we briefly investigated the Proteobacteria-high sub-group observed at T1. Individuals with high Proteobacteria abundance (>50%) were equally common in the rural and urban groups (Fisher's exact test: odds ratio = 0.98, p = 1) and were neither younger nor older than the rest of the infants (Mann-Whitney U test, p = 0.503). In the PCoA of weighted UniFrac distances computed on all samples, the Proteobacteria-high group formed a cluster that was equidistant from both the mothers and the remaining infants (Figure S1E). A longitudinal study profiling the gut microbiome of Nigerian infants similarly found high levels of Proteobacteria in young infants.[24] As the high-Proteobacteria group appeared to have a completely distinct phenotype from the rest of the infants and was a minority in our dataset, we opted to remove them from the analysis.

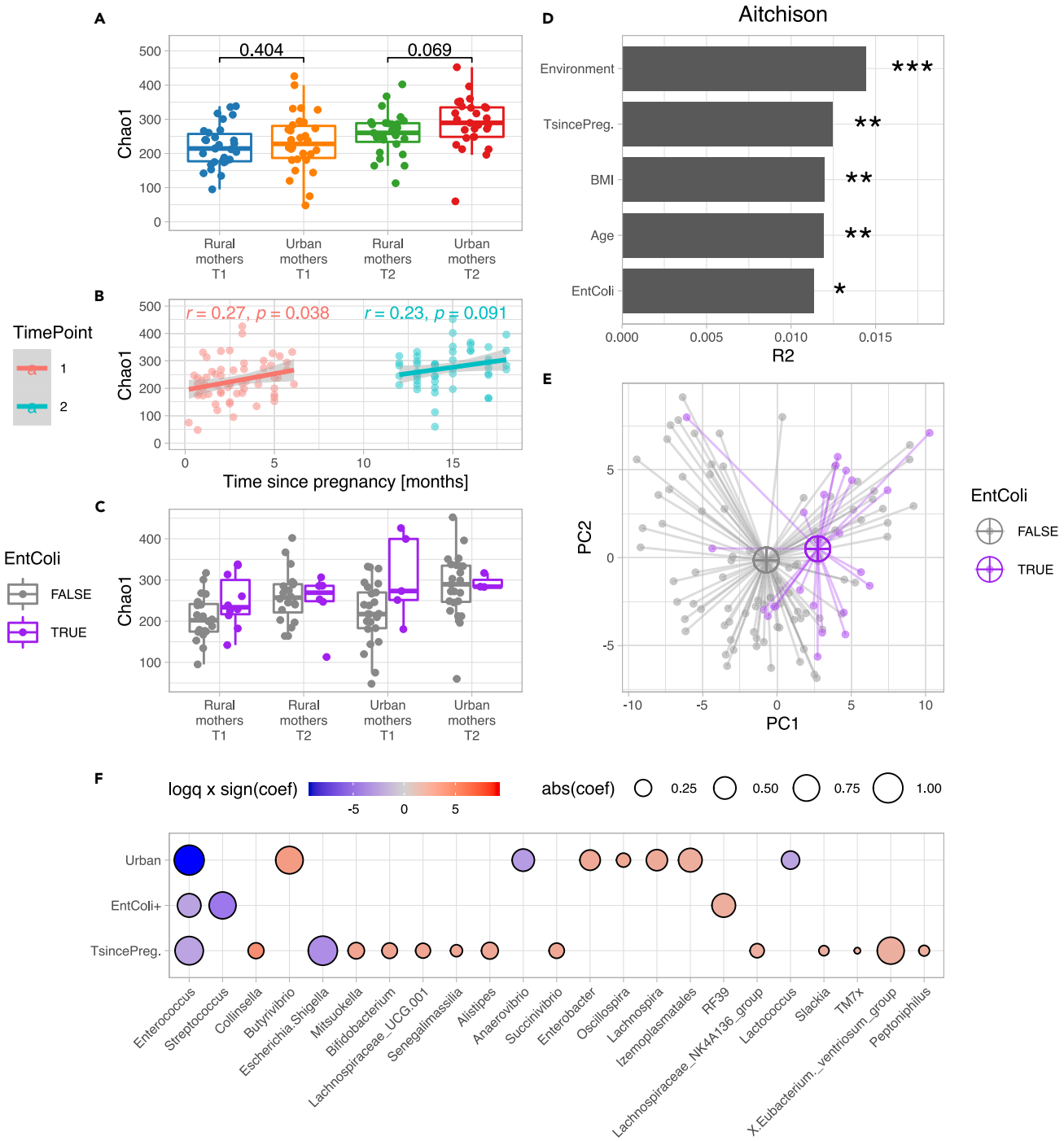



**Figure 2. Factors affecting gut microbiome diversity and composition in mothers**

(A) Comparison of alpha diversity (Chao1) between rural and urban mothers at multiple time points.

(B) Correlation between time since delivery and alpha diversity in mothers (Chao1).

(C) Effect of *Entamoeba coli* parasitism in mothers, separated by rural/urban and time point.

(D) Variables explaining beta diversity (Aitchison) between mothers ranked by effect size. (. = p < 0.1, * = p < 0.05, ** = p < 0.01, *** = p < 0.001).

(E) PCoA of Aitchison beta diversity, showing clustering of *Entamoeba coli* + mothers. Group centroids are included as a visual aid.

(F) Associations between genera abundance in mothers and rural/urban environment, *Entamoeba coli* parasitism, and time since delivery.

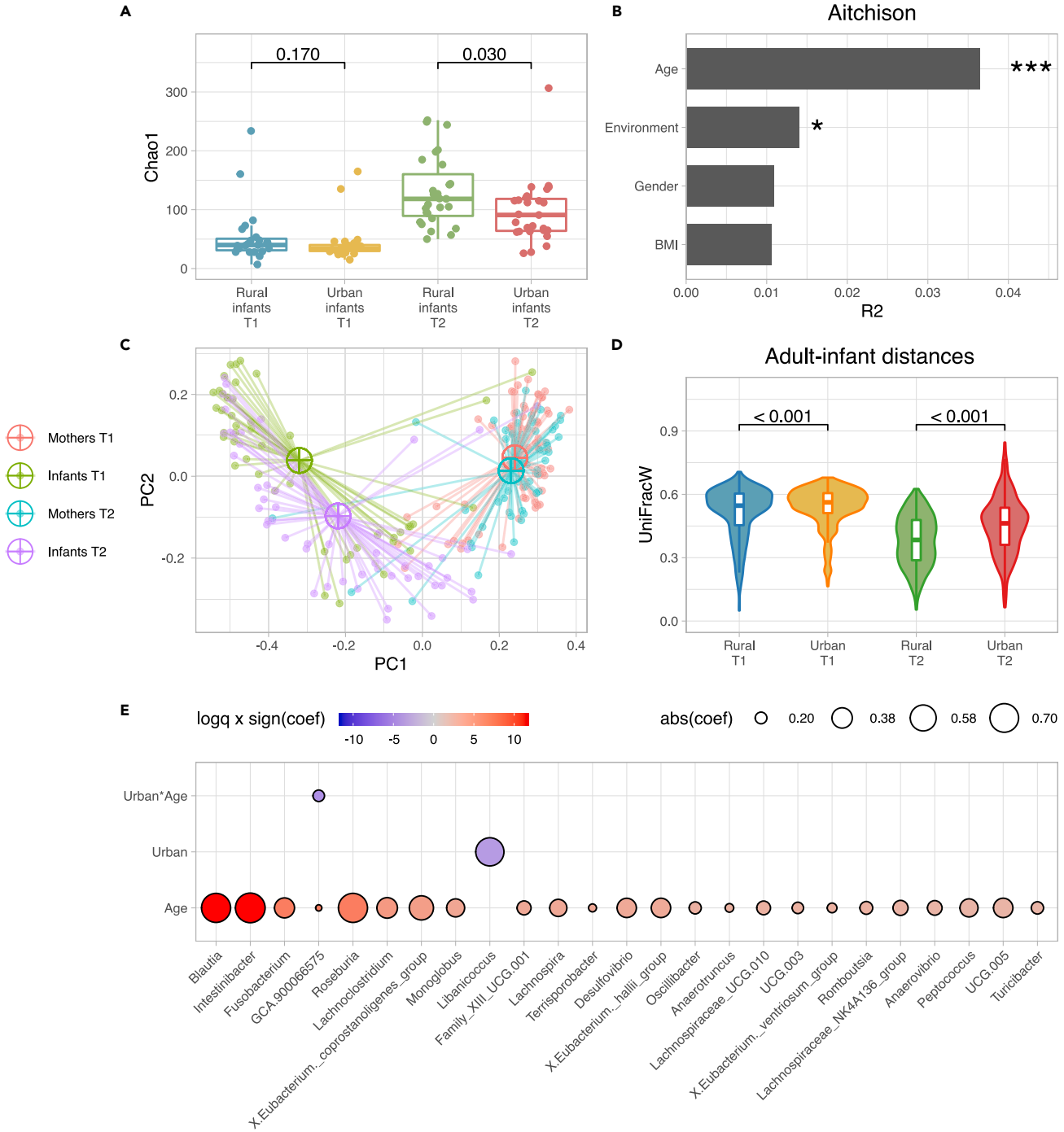



**Figure 3. Factors affecting gut microbiome diversity and composition in infants**

(A) Comparison of alpha diversity (Chao1) between rural and urban infants at multiple time points.

(B) Variables explaining beta diversity (Aitchison) between infants ranked by effect size. (. = p < 0.1, * = p < 0.05, ** = p < 0.01, *** = p < 0.001) (C) PCoA of Aitchison beta diversity, showing a maturation trajectory from newborn (T1) to adult microbiome. Group centroids are included as a visual aid.

(D) UnifracW beta diversity between adults and infants by group.

(E) Associations between genera abundance in infants and rural/urban environment, infant age, and the interaction of environment and age.

Alpha diversity did not differ between the rural and urban groups at T1; however, at T2, all examined metrics indicated higher alpha diversity in rural infants (Figures 3A and S2). We once again looked for lifestyle/environmental variables associated with alpha diversity in infants by constructing linear regression models. The selected model included terms for infant age, rural/urban environment, and their interaction but no additional lifestyle-related variables (Age: $p < 0.001$, environment: $p = 0.821$, age*environment: $p = 0.081$).

Permanova analysis of beta diversity revealed that in infants, like in mothers, the type of environment had a significant effect on Bray-Curtis, Aitchison, and unweighted UniFrac distances but not on the weighted UniFrac distance (Bray-Curtis: $R2 = 0.021$, $p = 0.027$, Aitchison: $R2 = 0.015$, $p = 0.008$, UniFracU: $R2 = 0.021$, $p = 0.026$, UniFracW: $R2 = 0.015$, $p = 0.180$). The type of environment had a stronger effect when performing permanova on infants at T2 separately (Bray-Curtis: $R2 = 0.039$, $p = 0.030$, Aitchison: $R2 = 0.028$, $p = 0.006$, UniFracU: $R2 = 0.039$, $p = 0.014$, UniFracW: $R2 = 0.038$, $p = 0.103$). When more terms were included, we observed a strong effect related to infant age while BMI and gender of the infant did not appear to have any influence on beta diversity (Figures 3B and S3). PCoA of weighted UniFrac distances computed on all samples placed the infants on a trajectory of maturation from newborn to adult (Figure 3C). Because urban infants showed delayed maturation of alpha diversity, we asked if beta diversity would indicate that the gut microbiomes of urban infants were further separated from adult microbiomes, compared to rural infants. Indeed, we found that at both time points, beta diversity between urban mothers and infants was higher than between rural mothers and infants according to all metrics (Figures 3D, S4A, S4C, S4E, and S4G) despite rural infants being almost exactly the same age as urban ones (9 days difference in means, t test $p = 0.809$). This finding is in agreement with a prior study which found a higher similarity between the gut microbiome of rural infants and adults than between the microbiome of urban infants and adults.[25] Lastly, we wondered if the infant microbiome could be more similar to the microbiome of the infants' own mothers as compared to the microbiome of unrelated adults due to genetics, exposure to common environmental sources, or a shared lifestyle; however, this did not appear to be the case at either of the time points (Figures S4B, S4D, S4F, and S4H).

The gut microbiome affects the development of the immune system in infants, influencing the response to infection and the development of allergies.[26] Thus, we wondered if the delayed maturation of the gut microbiome in urban infants (lower alpha diversity, higher beta diversity from mothers compared to rural) associated with weakened immune defenses or a higher prevalence of allergy. Indeed, our medical examinations reported a higher prevalence of diseases (mostly infections, respiratory and dermatologic) in urban infants compared to rural infants at T2 (Experiencing symptoms/total group size: urban 16/27, rural 6/27; Fisher's exact test: odds ratio = 4.93, $p = 0.012$). Notably, 5 infants from the urban group experienced some symptoms of allergy during or prior to the T2 medical examination, while only one child in the rural group experienced such symptoms (odds ratio = 5.74, $p = 0.192$). Because urban environments correspond to increased human density, the higher prevalence of infectious diseases in the urban group at T2 could simply be explained by facilitated infection spreading in densely populated areas.[27] However, we note that disease prevalence was actually higher in the rural group at T1, albeit not significantly (Experiencing symptoms/total group size: urban 5/24, rural 11/23, odds ratio = 0.30, $p = 0.069$).

Finally, we again used Maaslin2 to look for associations between the abundance of specific genera and type of environment, infant age, and the interaction of the former two variables (Figure 3E). To our surprise, only 2 of 27 significant associations were related to the type of environment, whereas the remaining 25 genera were positively associated with infant age. In particular, urban infants showed lower abundance of the genus *Libanicoccus* which is closely related to *Olsenella*, and within the *Coriobacteriales* order.[28] This order is common in human microbiomes and carries out important functions such as the conversion of bile salts and the activation of dietary polyphenols.[29] Not surprisingly, the remaining genera such as *Blautia*, *Roseburia*, *Lachnospira*, *Lachnoclostridium*, *Eubacterium*, and *Clostridium*, which increased in relative abundance with infant age, are found in infant gut microbiota and usually associate with mature adult microbiota and aging.[30] In conclusion, urbanization showed only minor associations with abundance of microbiota at the genus level. However, before running Maaslin2, we filtered out rare genera as abundance data for rare genera are highly sparse and sensitive to outliers. It is possible that the main differences between the microbiome of rural and urban infants consist in the presence or absence of rare genera, rather than the abundance of common ones.

## DISCUSSION

Our study investigates how urbanization, lifestyle, and environment associate with gut microbiome composition in adulthood and during development. To approach this question, we performed a longitudinal study on a genetically homogeneous population, living in drastically opposite conditions such as urban and rural environments. Our results clearly show that urbanization associates with changes in gut microbiome composition in both adults and infants, although seemingly, in very different ways between the two groups. While urbanization associated with slowed growth of alpha diversity in infants, it seemed to slightly boost it in mothers. A possible explanation is that the better sanitary conditions of urban environments reduce the exposure of infants to microbial sources, while in adults, alpha diversity has had time to "saturate". Beta diversity revealed urbanization-related compositional changes in the microbiome of both adults and infants, but this effect was stronger in adults. This could once again be explained by the difference in timescales relevant for adults and infants: by year 1, infants have had limited exposure to factors which differ between rural and urban environments. Diet, for example, only starts diversifying after the infants are introduced to solid food. Conversely, lifestyle and environment have ample time to affect the microbiome of adults. This is also evidenced by the larger number of factors which affected beta diversity in mothers (urbanization, *Entamoeba coli* parasitism, time elapsed since pregnancy, BMI, and age) as compared to infants (age and urbanization). Similarly, multivariable association discovery using Maaslin2 showed that the abundance of several genera in the microbiome of mothers was associated with urbanization, whereas only two genera were associated with urbanization in infants. Taken together, these pieces of evidence suggest that urbanization alters the composition of the adult gut microbiome by boosting or inhibiting the growth of specific commensals by

unknown mechanisms (e.g., diet, sanitization, pollution …). In contrast, urbanization may act on the gut microbiome of infants by limiting the exposure to diverse genera, delaying the growth of alpha diversity but only driving minor compositional changes with respect to rural gut microbiomes, at least in this population.

Our findings point to interesting links between urbanization, gut microbiome, and human health. Gut microbiota interact with the immune system, and alterations to the microbiome can disrupt immune functions.[4,5,26] In this study, we observed that urban infants, who showed delayed maturation of the gut microbiome at T2, were more susceptible to infectious diseases at T2 compared to rural infants but not at T1. Furthermore, symptoms of allergy tended to be more common in urban infants compared to rural. These results would need to be confirmed in a larger cohort and with additional time points but are nonetheless intriguing, given the intimate relationship between the gut microbiome and maturation of the immune system during infancy.[26] As for the mothers in our study: the urban group had higher BMIs than the rural group and showed increased abundance of microbial taxa previously found to associate with obesity such as *Lachnospiraceae* and *Enterobacter*.[20–23] It would be worth investigating the link between these taxa and weight gain both in a general setting and in recent mothers, who may gain weight as a consequence of pregnancy.

Although our study is mainly descriptive, we believe that our results allow us to draw strong conclusions due to several strengths of the study design. First and foremost, we chose to study rural and urban communities of matching (Fula) ethnicity. This was critical to study the effects of lifestyle and environment without the genetic confounder, which often limits the interpretability of studies contrasting urban and rural societies across the world. As confirmation of genetic homogeneity, we did not find the gut microbiomes of related mother-infant pairs to be more similar to each other than those of unrelated pairs. Second, the inclusion of two time points in the study design allowed us to investigate the gut microbiome development in infants over the transition to solid foods. Thanks to repeated measurements, we were also able to identify and correct for the confounding effect of pregnancy on the microbiome composition in mothers. Lastly, we were able to obtain highly detailed information on the health and dietary habits of the studied populations. This level of detail is often unavailable for communities living in remote areas and allowed us to test for associations between many lifestyle variables and gut microbiome composition.

Despite the rich health and lifestyle data, we were not able to pinpoint the sources of microbiome alterations associated with urbanization. The transition to an urban lifestyle is characterized by many changes in habits (nutrition, work, social structure …) and in the environment (sanitization, pollution, human density …). Separating the individual effects of these factors on the microbiome and health is indeed challenging. It is not clear why the microbiome of rural infants showed a faster development than that of urban infants. We hypothesize that pollution and water sanitation, two factors which we did not measure in this study, may have played a role. Additionally, we did not have access to quantitative information regarding caloric intake and macronutrients. Finally, we may have missed associations present in our data due to insufficient statistical power. Future studies interested in determining how exactly urbanization exerts a pressure on the gut microbiome should strive to collect detailed information on nutrition, water sanitation, and parasite colonization.

In conclusion, our study shows several ways in which urbanization can influence the gut microbiome in adulthood and infancy. These findings contribute to our understanding on how industrialization has altered gut microbiomes in developed and developing countries, leading to the rise of several non-communicable diseases. We hope that continued efforts in studying the transition from pre- to post-industrial microbiomes will help produce treatments and health recommendations that prevent the emergence of non-communicable diseases as more of the world becomes industrialized.

### Limitations of the study

Although we collected information on antibiotic usage, and are confident that data regarding recent antibiotic use are complete, we did not have access to information on lifetime antibiotics usage. Additionally, the nutrition data we collected included lists of ingredients and meals, but it did not include calorie or macronutrient intake quantifications.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Subject enrollment and questionnaires
  - Stool sample collection
  - 16s rRNA sequencing
  - ASV calling, taxonomic assignment
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quality control and exploration

- Alpha diversity
- Beta diversity
- Differential abundance

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108136.

## ACKNOWLEDGMENTS

This research was funded by the French National Center of Scientific Research (CNRS, MITI, PEPS MUTALIM 2016, 2017). All the authors thank Enterome for their logistical support. F.M. thanks Cheyenne Rechsteiner for her help in polishing the figures for the final version of the manuscript.

## AUTHOR CONTRIBUTIONS

Conceptualization: E.M., J.-P.G., and P.S.; methodology: E.M., F.M., K.P., P.S., and S.M.S.; formal analysis: F.M. and K.P.; resources: L.B.; data curation: E.M., F.M., K.P., and P.S.; writing – original draft: F.M.; writing – review & editing: E.M., F.M., K.P., and P.S.; supervision: E.M. and P.S.; project administration: E.M., P.S., and L.T.; funding acquisition: E.M. and P.S.

## DECLARATION OF INTERESTS

P.S. declares financial support for scientific works from Biocodex, MSD, Takeda, Janssen, and Sandoz, and consultant fees from Abbvie, Merk, MSD, Gilead, Pfizer, Sandoz, Janssen, and Fresenius Kabi.

## INCLUSION AND DIVERSITY

We worked to ensure that the study questionnaires were prepared in an inclusive way. We avoided "helicopter science" practices by including the participating local contributors from the region where we conducted the research as authors on the paper.

Received: March 26, 2023
Revised: July 16, 2023
Accepted: October 2, 2023
Published: October 6, 2023

## REFERENCES

1. Gupta, H., Min, B.-H., Ganesan, R., Gebru, Y.A., Sharma, S.P., Park, E., Won, S.-M., Jeong, J.-J., Lee, S.-B., Cha, M.-G., et al. (2022). Gut Microbiome in Non-Alcoholic Fatty Liver Disease: From Mechanisms to Therapeutic Role. Biomedicines *10*, 550. https://doi.org/10.3390/biomedicines10030550.
2. Lavelle, A., and Sokol, H. (2020). Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. Nat. Rev. Gastroenterol. Hepatol. *17*, 223–237. https://doi.org/10.1038/s41575-019-0258-z.
3. Morais, L.H., Schreiber, H.L., and Mazmanian, S.K. (2021). The gut microbiota–brain axis in behaviour and brain disorders. Nat. Rev. Microbiol. *19*, 241–255. https://doi.org/10.1038/s41579-020-00460-0.
4. Pickard, J.M., Zeng, M.Y., Caruso, R., and Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunol. Rev. *279*, 70–89. https://doi.org/10.1111/imr.12567.
5. Conlon, M.A., and Bird, A.R. (2014). The Impact of Diet and Lifestyle on Gut Microbiota and Human Health. Nutrients *7*, 17–44. https://doi.org/10.3390/nu7010017.
6. Statovci, D., Aguilera, M., MacSharry, J., and Melgar, S. (2017). The Impact of Western Diet and Nutrients on the Microbiota and Immune Response at Mucosal Interfaces. Front. Immunol. *8*, 838.
7. Brewster, R., Tamburini, F.B., Asiimwe, E., Oduaran, O., Hazelhurst, S., and Bhatt, A.S. (2019). Surveying Gut Microbiome Research in Africans: Toward Improved Diversity and Representation. Trends Microbiol. *27*, 824–835. https://doi.org/10.1016/j.tim.2019.05.006.
8. Allali, I., Abotsi, R.E., Tow, L.A., Thabane, L., Zar, H.J., Mulder, N.M., and Nicol, M.P. (2021). Human microbiota research in Africa: a systematic review reveals gaps and priorities for future research. Microbiome *9*, 241. https://doi.org/10.1186/s40168-021-01195-7.
9. Macia, E., Tibère, L., Ka, A., Seksik, P., Faye, B., Boëtsch, G., and Duboz, P. (2021). L'alimentation des Peuls du Sénégal. Anthropol. Food. https://doi.org/10.4000/aof.11708.
10. Glöckner, F.O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., and Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. J. Biotechnol. *261*, 169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198.
11. Koren, O., Goodrich, J.K., Cullender, T.C., Spor, A., Laitinen, K., Bäckhed, H.K., Gonzalez, A., Werner, J.J., Angenent, L.T., Knight, R., et al. (2012). Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy. Cell *150*, 470–480. https://doi.org/10.1016/j.cell.2012.07.008.
12. Mesa, M.D., Loureiro, B., Iglesia, I., Fernandez Gonzalez, S., Llurba Olivé, E., García Algar, O., Solana, M.J., Cabero Perez, M.J., Sainz, T., Martinez, L., et al. (2020). The Evolving Microbiome from Pregnancy to Early Infancy: A Comprehensive Review. Nutrients *12*, 133. https://doi.org/10.3390/nu12010133.
13. Morton, E.R., Lynch, J., Froment, A., Lafosse, S., Heyer, E., Przeworski, M., Blekhman, R., and Ségurel, L. (2015). Variation in Rural African Gut Microbiota Is Strongly Correlated with Colonization by Entamoeba and Subsistence. PLoS Genet. *11*, e1005658. https://doi.org/10.1371/journal.pgen.1005658.
14. Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecol. *26*, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.
15. He, Y., Wu, W., Zheng, H.-M., Li, P., McDonald, D., Sheng, H.-F., Chen, M.-X., Chen, Z.-H., Ji, G.-Y., Zheng, Z.-D.-X., et al. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat. Med. *24*, 1532–1535. https://doi.org/10.1038/s41591-018-0164-x.

16. Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren, B., Schwager, E.H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. PLoS Comput. Biol. *17*, e1009442. https://doi.org/10.1371/journal.pcbi.1009442.
17. Ogier, J.-C., and Serror, P. (2008). Safety assessment of dairy microorganisms: The Enterococcus genus. Int. J. Food Microbiol. *126*, 291–301. https://doi.org/10.1016/j.ijfoodmicro.2007.08.017.
18. Song, A.A.-L., In, L.L.A., Lim, S.H.E., and Rahim, R.A. (2017). A review on Lactococcus lactis: from food to factory. Microb. Cell Factories *16*, 55. https://doi.org/10.1186/s12934-017-0669-x.
19. Vacca, M., Celano, G., Calabrese, F.M., Portincasa, P., Gobbetti, M., and De Angelis, M. (2020). The Controversial Role of Human Gut Lachnospiraceae. Microorganisms *8*, 573. https://doi.org/10.3390/microorganisms8040573.
20. Chávez-Carbajal, A., Nirmalkar, K., Pérez-Lizaur, A., Hernández-Quiroz, F., Ramírez-Del-Alto, S., García-Mena, J., and Hernández-Guerrero, C. (2019). Gut Microbiota and Predicted Metabolic Pathways in a Sample of Mexican Women Affected by Obesity and Obesity Plus Metabolic Syndrome. Int. J. Mol. Sci. *20*, 438. https://doi.org/10.3390/ijms20020438.
21. Lippert, K., Kedenko, L., Antonielli, L., Kedenko, I., Gemeier, C., Leitner, M., Kautzky-Willer, A., Paulweber, B., and Hackl, E. (2017). Gut microbiota dysbiosis associated with glucose metabolism disorders and the metabolic syndrome in older adults. Benef. Microbes *8*, 545–556. https://doi.org/10.3920/BM2016.0184.
22. Fei, N., and Zhao, L. (2013). An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. ISME J. *7*, 880–884. https://doi.org/10.1038/ismej.2012.153.
23. Keskitalo, A., Munukka, E., Toivonen, R., Hollmén, M., Kainulainen, H., Huovinen, P., Jalkanen, S., and Pekkala, S. (2018). Enterobacter cloacae administration induces hepatic damage and subcutaneous fat accumulation in high-fat diet fed mice. PLoS One *13*, e0198262. https://doi.org/10.1371/journal.pone.0198262.
24. Oyedemi, O.T., Shaw, S., Martin, J.C., Ayeni, F.A., and Scott, K.P. (2022). Changes in the gut microbiota of Nigerian infants within the first year of life. PLoS One *17*, e0265123. https://doi.org/10.1371/journal.pone.0265123.
25. Ayeni, F.A., Biagi, E., Rampelli, S., Fiori, J., Soverini, M., Audu, H.J., Cristino, S., Caporali, L., Schnorr, S.L., Carelli, V., et al. (2018). Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria. Cell Rep. *23*, 3056–3067. https://doi.org/10.1016/j.celrep.2018.05.018.
26. Zhang, H., Zhang, Z., Liao, Y., Zhang, W., and Tang, D. (2022). The Complex Link and Disease Between the Gut Microbiome and the Immune System in Infants. Front. Cell. Infect. Microbiol. *12*, 924119.
27. Neiderud, C.-J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. Infect. Ecol. Epidemiol. *5*, 27060. https://doi.org/10.3402/iee.v5.27060.
28. Bilen, M., Cadoret, F., Richez, M., Tomei, E., Daoud, Z., Raoult, D., and Fournier, P.-E. (2018). Libanicoccus massiliensis gen. nov., sp. nov., a new bacterium isolated from human stool. New Microbes New Infect. *21*, 63–71. https://doi.org/10.1016/j.nmni.2017.11.001.
29. Clavel, T., Lepage, P., and Charrier, C. (2014). The Family Coriobacteriaceae. In The Prokaryotes: Actinobacteria, E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Springer), pp. 201–238. https://doi.org/10.1007/978-3-642-30138-4_343.
30. Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J.Z., Abe, F., and Osawa, R. (2016). Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. BMC Microbiol. *16*, 90. https://doi.org/10.1186/s12866-016-0708-5.
31. Duboz, P., Macia, E., Diallo, A.H., Cohen, E., Bergouignan, A., and Seck, S.M. (2021). The good life in rural and urban Senegal: A qualitative and quantitative study. PLoS One *16*, e0252134. https://doi.org/10.1371/journal.pone.0252134.
32. Duboz, P., Boëtsch, G., Guisse, A., and Macia, E. (2020). Assessing health impacts of an environmental pan-African development project: A migration perspective. SSM Popul. Health *11*, 100633. https://doi.org/10.1016/j.ssmph.2020.100633.
33. Sokol, H., Leducq, V., Aschard, H., Pham, H.-P., Jegou, S., Landman, C., Cohen, D., Liguori, G., Bourrier, A., Nion-Larmurier, I., et al. (2017). Fungal microbiota dysbiosis in IBD. Gut *66*, 1039–1048. https://doi.org/10.1136/gutjnl-2015-310746.
34. Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864. https://doi.org/10.1093/bioinformatics/btr026.
35. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
36. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9.
37. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583. https://doi.org/10.1038/nmeth.3869.
38. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.
39. Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., and Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome *6*, 90. https://doi.org/10.1186/s40168-018-0470-z.
40. Robeson, M.S., 2nd, O'Rourke, D.R., Kaehler, B.D., Ziemski, M., Dillon, M.R., Foster, J.T., and Bokulich, N.A. (2021). RESCRIPt: Reproducible sequence taxonomy reference database management. PLoS Comput. Biol. *17*, e1009581. https://doi.org/10.1371/journal.pcbi.1009581.
41. Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. Bioinformatics *26*, 1463–1464. https://doi.org/10.1093/bioinformatics/btq166.
42. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., et al. (2022). Vegan: Community Ecology Package.
43. Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S (Springer). https://doi.org/10.1007/978-0-387-21706-2.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| Raw 16S rRNA data | This paper | SRA: PRJNA899851 |
| Oligonucleotides | | |
| 16S rRNA V3-V4 region primers, forward: TACGGRAGGCAGCAG reverse: CTACCNGGGTATCTAAT | This paper | |
| Software and algorithms | | |
| Analysis code | This paper | https://github.com/SunScript0/Fula-microbiome |
| QIIME2 version 2021.11 | Bolyen et al.[36] | https://qiime2.org/ |
| R version 4.0.4 | | https://www.r-project.org/ |
| Maaslin2 | Mallick et al.[16] | https://huttenhower.sph.harvard.edu/maaslin/ |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Philippe Seksik (philippe.seksik@aphp.fr).

#### Materials availability

This study did not generate any new unique materials.

#### Data and code availability

- The raw 16s RNA-seq data generated in this study is available on SRA under the accession code: PRJNA899851.
- The code used to call ASVs, assign taxonomic annotations, perform the analysis, and create figures is available at https://github.com/SunScript0/Fula-microbiome.
- Instructions to recreate the entirety of the analysis and figures are included with the code.

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study involved 120 human participants of Fula ethnicity. Deidentified demographic, nutritional and health data is included in Table S1 and summarized in Tables 1 and 2. The study was approved by Ethical committee No. 272/MSAS/DPRS/CNRS 28 May 2014 and informed consent was ensured.

### METHOD DETAILS

#### Subject enrollment and questionnaires

The two sampling campaigns (T1 and T2) were conducted in July 2017 and July 2018 in two regions (urban and rural). Mother-child pairs were deemed eligible for the study if the child had been born less than six months prior to T1. For each region, the first 30 mothers who accepted and their newborns were included in the study after informed consent (Ethical committee No. 272 /MSAS/DPRS/CNERS 28 May 2014). Thus, 60 mother and infant pairs were part of the study at T1, however, six of them dropped out of the study before T2. At each sampling campaign, general anthro-biometric, health, and dietary data questionnaires were filled. The questionnaire was written in French and the questions were asked in Wolof by the interviewers. Briefly, questionnaires were developed to collect data on the socio-demographic characteristics of the households, the health status, and feeding practices of the mother and child as previously described.[31,32] The feeding practice information included a list of ingredients used in any meal throughout the day for both mothers and infants, and the number of individuals present at meals. Additional information regarding the infants included the first meal that infants received after birth, the time (months) since introduction to solid food at T2, and whether the infants had been weaned. Infants received fresh milk unless otherwise reported.

### Stool sample collection

A fecal sample from the mother and child was collected and immediately sampled in three aliquots: two tubes of 200 mL in RNA-later tubes® were frozen at -20°C on site within 12 h for gut microbiota analysis and the remainder of the sample (> 40 g) was stored at 4°C for parasite screening performed in Dakar (UMI 3189, UCAD, CNRS). At the end of each campaign, which lasted less than a week, all samples were shipped frozen to Dakar (UMI 3189, UCAD, CNRS) and then transferred on dry ice from Senegal to the MI2 lab (microbiota, intestine & inflammation - CRSA UMRS_938, Inserm, Sorbonne université) and frozen on arrival at -80°C.

### 16s rRNA sequencing

DNA was extracted from fecal samples by both mechanical and chemical methods, as previously described.[33] Microbial lysis was performed by both mechanical and chemical methods. Briefly, mechanical lysis was performed with glass beads and following isopropanol precipitation of nucleic acids for 10 min at room temperature, samples were incubated on ice for 15 min and then centrifuged for 30 min at 20 000 g and 4°C. The resulting pellets were suspended in phosphate buffer (450 µL) and potassium acetate (50 µL). Following RNase treatment and DNA precipitation, recovery of nucleic acids was performed via centrifugation at 20 000 g and 4°C for 30 min. The DNA pellet was suspended in 80 µL of trypsin-EDTA buffer. Amplicon sequencing of the V3-V4 region of the 16S ribosomal RNA gene was employed for microbiota analysis. The primers used for this analysis were – 16S sense 5′-TACGGRAGGCAGCAG-3′ and anti-sense 5′-CTACCNGGGTATCTAAT-3′. This was performed using an optimized and standardized 16S amplicon library preparation protocol (Metabiote, GenoScreen, Lille, France). 16S DNA PCR was performed with 5 ng of genomic DNA with barcoded primers (Metabiote MiSeq Primers) according to the manufacturer's protocol (Metabiote) at a final concentration of 0.2 µmol/L, with an annealing temperature of 50°C for 30 cycles. PCR product purification was performed with Agencourt AMPure XP-PCR purification system (Beckman Coulter, Brea, CA, USA) and was quantified according to the manufacturer's protocol with samples multiplexed at equal concentrations. An Illumina MiSeq platform (Illumina, San Diego, CA, USA) was used for sequencing both time points in two different sequencing runs: a 250 bp paired-end sequencing protocol and a 300 bp paired-end sequencing protocol, at GenoScreen. Raw paired-end sequencing reads were subjected to the following initial procedures at GenoScreen: (1) quality filtering with the PRINSEQ-lite PERL script,[34] truncating bases from the 3′ end with a quality <30 (based on the Phred algorithm) and (2) using CutAdapt to remove primers, with no mismatches allowed in the primer sequences.[35] Only sequences with perfectly matching forward and reverse primers were retained for further analysis.

### ASV calling, taxonomic assignment

Pre-processed sequences were imported in qiime2[36] version 2021.11 and denoised using the dada2 qiime plugin[37] with –p-trim-left-f 10, –p-trim-left-r 0, –p-trunc-len-f 0 and –p-trunc-len-r 0. A phylogenetic tree of ASV representative sequences was generated using align-to-tree-mafft-fasttree with default settings.[38] ASVs were assigned to a taxonomic annotation using a naïve bayes classifier pre-trained on the SILVA 16s database clustered to 99% similarity.[10,39,40] Abundance tables at genus and phylum levels were also generated by collapsing ASVs using the qiime taxa collapse tool.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Quality control and exploration

Following processing with qiime2, the data was imported in R version 4.0.4 for analysis. We considered samples with more than 14000 read pairs to be of good quality as rarefaction curved had reached a plateau at that sequencing depth, meaning that all samples but one had sufficient depth to capture the full richness of the microbial communities. After removal of the sample with insufficient sequencing depth, the relative abundance of the most common phyla (mean relative abundance across all samples > 1%) was summarized and compared across groups. Having noted a subpopulation of infants with an abnormally high abundance of proteobacteria (> 50%) we opted to remove them from the remainder of the analysis.

### Alpha diversity

Alpha diversity was evaluated with three metrics (Shannon diversity, Chao1 index and Faith's phylogenetic diversity (PD), in the Vegan and Picante R packages[41,42]) applied to unrarefied ASV data and compared across groups. To explore variables associated with alpha diversity we constructed linear regression models with alpha diversity (Chao1) as the dependent variable and selected independent variables by performing stepwise feature selection using the stepAIC function in the MASS package.[43] Feature selection for the mothers' model started from a null model and potential independent variables included rural/urban environment, time since pregnancy and their interaction, age, BMI, Entamoeba coli colonization, number of unique ingredients used in food preparation, and average number of individuals present at meals. Feature selection for the infants' model started from a null model and potential independent variables included rural/urban environment, age and their interaction, maternal age, BMI, sex, and the first meal the infants had after birth (to investigate potential effects on initial gut colonization by microbiota). Seven samples from infants with abnormally high alpha diversity within their group were excluded from regression analysis (Chao1 < Q1 – 1.5*IQR or Chao1 > Q3 + 1.5*IQR).

### Beta diversity

Beta diversity was evaluated with four metrics (Bray-Curtis dissimilarity, Aitchison distance, weighted and unweighted UniFrac) applied to unrarefied ASV data. The effects of explanatory variables on beta diversity were evaluated using Permanova implemented by the adonis2 function in the Vegan package with 2000 permutations. When multiple variables were included, effect significance was evaluated by margin. For mothers and infants both, the effect of environment was evaluated individually first and later accounting for potential confounders. Variables were considered as potential confounders either based on prior reports showing they may have an effect on beta diversity (e.g. BMI, age) or based on having an effect on alpha diversity within this study (e.g. Entamoeba coli colonization, time since pregnancy). PCoAs were computed with the cmdscale function in the stats package on all samples (Figures 3C and S1E) or on mothers only (Figure 2E).

### Differential abundance

Multivariate association between lifestyle/environmental factors and abundance of microbial genera was investigated using Maaslin2.[16] To increase the interpretability of the results we removed taxa with undefined terms in their taxonomic annotation (genus unknown or containing the term "uncultured" at any taxonomic level at or above genus). Additionally, to decrease the effect of outliers, we ignored genera that were not detected in at least five samples in at least one group (e.g. urban infants at T2). This reduced the number of genera from 385 to 142, however, the removed taxa accounted for only 3.5% of microbial abundance on average. Maaslin2 was run on unrarefied counts with the following arguments: normalization = "TSS", transform = "LOG", analysis_method = "LM", max_significance = 0.1, min_prevalence = 0. Fixed effects were chosen based on the results of alpha and beta diversity analyses. Random effects for individuals and time points were included to account for the longitudinal nature of the measurements and potential batch effects. The full Maaslin2 results for mothers and infants can be found in Data S2 and S3 respectively. Additionally, Data S4 and S5 contain Masslin2 results for mothers and infants without prior filtering to genera detected in more than 5 samples.