

# Diet changes due to urbanization in South Africa are linked to microbiome and metabolome signatures of Westernization and colorectal cancer

Received: 13 July 2023

Accepted: 15 February 2024

Published online: 20 April 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Transition from traditional high-fiber to Western diets in urbanizing communities of Sub-Saharan Africa is associated with increased risk of non-communicable diseases (NCD), exemplified by colorectal cancer (CRC) risk. To investigate how urbanization gives rise to microbial patterns that may be amenable by dietary intervention, we analyzed diet intake, fecal 16 S bacteriome, virome, and metabolome in a cross-sectional study in healthy rural and urban Xhosa people (South Africa). Urban Xhosa individuals had higher intakes of energy (urban:  $3,578 \pm 455$ ; rural:  $2,185 \pm 179$  kcal/d), fat and animal protein. This was associated with lower fecal bacteriome diversity and a shift from genera favoring degradation of complex carbohydrates (e.g., *Prevotella*) to taxa previously shown to be associated with bile acid metabolism and CRC. Urban Xhosa individuals had higher fecal levels of deoxycholic acid, shown to be associated with higher CRC risk, but similar short-chain fatty acid concentrations compared with rural individuals. Fecal virome composition was associated with distinct gut bacterial communities across urbanization, characterized by different dominant host bacteria (urban: Bacteriodota; rural: unassigned taxa) and variable correlation with fecal metabolites and dietary nutrients. Food and skin microbiota samples showed compositional differences along the urbanization gradient. Rural-urban dietary transition in South Africa is linked to major changes in the gut microbiome and metabolome. Further studies are needed to prove cause and identify whether restoration of specific components of the traditional diet will arrest the accelerating rise in NCDs in Sub-Saharan Africa.

Worldwide, colorectal cancer (CRC) is the second most common cause of cancer in women and the third in men<sup>1</sup>. There is a remarkable geographical variation in CRC incidence, with the highest rates in North America, Australia/New Zealand, and Europe (all  $>40:100,000$ ) and the lowest in rural Africa ( $<5:100,000$ ). Convincing evidence supports the now classic observations of Burkitt, working in East

Africa, and Walker in South Africa in the 60s–80s, which demonstrated an inverse association between CRC risk and traditional African diets rich in fiber<sup>2</sup>. Recent large-scale prospective studies and meta-analyses confirmed that a high intake of dietary fiber correlates with low CRC risk<sup>3,4</sup>, whereas consumption of ultra-processed foods low in fiber is linked to higher CRC rates<sup>5</sup>. Dietary fiber is fermented by gut bacteria

 e-mail: [sjokeefe@pitt.edu](mailto:sjokeefe@pitt.edu)

to short-chain fatty acids (SCFA), including butyrate, which serves as the chief source of energy for the colonic epithelium and has anti-inflammatory and anti-neoplastic effects<sup>6,7</sup>. Fiber also reduces intestinal exposure to dietary carcinogens by lowering the colonic transit time and binding tumor-promoting secondary bile acids<sup>8</sup>, collectively promoting gut microbiota homeostasis. In addition to the critical importance of dietary fiber in maintaining colonic health, recent studies highlighted the implication of a high-fiber diet to reduce disease risk beyond CRC: the intake of fiber is associated with decreased all-cause mortality and lower risk of non-communicable diseases (NCD) such as diabetes, cardiovascular diseases and obesity<sup>4,9</sup>. Being an upper-middle-income country, South Africa faces high rates of NCDs that pose a massive threat with increasing urbanization and started to extend even to the rural segments of the population<sup>10–12</sup>. About 63% of South Africans are already living in urban areas, and the number is expected to rise to 71% by 2030<sup>13</sup>. As part of this ongoing urbanization process, the adoption of Western dietary patterns will promote the rapid increase of NCDs throughout Sub-Saharan Africa. Thus, NCDs are set to become the leading cause of disability and mortality in this region and must be recognized as major barriers to attaining the African Sustainable Development Goals<sup>14,15</sup>. Here, we investigate how the ongoing transition from traditional to urban lifestyle affects the gut bacteriome, virome and its metabolome in samples of the second largest section of the black population (amaXhosa) in South Africa to identify signatures that might be amenable to change through dietary intervention.

## Results

### Urban individuals have a Western dietary pattern associated with lower diversity and distinct composition of the gut microbiota

We performed a cross-sectional, observational pilot study in two healthy middle-aged rural or urban South African amaXhosa (Xhosa) cohorts. The urban cohort ( $n = 20$ ) was located in Cape Town, Western Cape Province, while the rural group ( $n = 24$ ) was recruited in the rural Eastern Cape Province of South Africa (Table 1). Food consumption was higher in the urban Xhosa cohort, with a mean energy intake of 3,578 kcal/d in urban compared to 2185 kcal/d in rural study participants (Table 2). Analysis of the contribution of carbohydrate, fat, and protein to total energy indicated that the composition was within normal limits corresponding to Western diets for the urban Xhosa cohort, whereas rural individuals consumed a high carbohydrate diet providing 72% of calories, whilst fat only contributed 16.9% of total calories. This translated into a higher absolute intake of dietary fiber in urban individuals but normalized to 1000 kcal, the rural Xhosa cohort had a slightly higher fiber intake (Table 2). Consistent with their less Westernized dietary pattern, the rural Xhosa diet was characterized by a lower intake of cholesterol and animal protein, as well as higher amounts of plant protein when normalized to kcal/d intake (Table 2). Similarly, dietary diversity and polyphenol consumption were higher in the rural Xhosa cohort because of a higher amount, variety, and frequency of plant foods eaten (e.g., 'imifino', onion, cabbage, and pumpkin leaves).

For fecal microbiota analysis by 16S rRNA gene sequencing, we noted lower  $\alpha$ -diversity indices in urban samples (Fig. 1A) and a distinct compositional clustering based on generalized UniFrac distances according to rural–urban status (Fig. 1B). Despite no clear separation of the two cohorts at phylum level, regional sub-clusters can be identified by high abundance of Bacteroidota (rural) or Proteobacteria (urban) (Fig. 1C). Urban fecal samples were characterized by high numbers of operational taxonomic units (OTU) related to *Bacteroides*, and other genera such as *Lachnoclostridium*, *Fusobacterium*, *Alistipes*, and *Bifidobacterium* (Fig. 1D). Despite the similar fiber intake, a higher abundance of OTUs assigned to *Prevotella*, *Faecalibacterium*, *Dialister* that are involved in saccharolytic fermentation and SCFA production was

**Table 1 | Demographic data of study participants**

Variable	Rural	Urban	p Value
Geographic area	Zithulele, Eastern Cape	Khayelitsha, Cape Town	
Number of participants (n)	24	20	
Age (mean $\pm$ SD)	56.7 $\pm$ 11.2	50.4 $\pm$ 6.7	0.0313*
Sex: Female	11	10	
Male	13	10	
Height (mean $\pm$ SD), m	1.6 $\pm$ 0.1	1.6 $\pm$ 0.1	0.3629
Weight (mean $\pm$ SD), kg	68.0 $\pm$ 10.8	68.9 $\pm$ 17.7	0.5163
BMI (mean $\pm$ SD), kg/m <sup>2</sup>	26.3 $\pm$ 5.0	25.8 $\pm$ 7.0	0.4032
HIV positive	5 (21%)	11 (55%)	
Antiretroviral therapy	5 (21%)	11 (55%)	
Diabetes mellitus	0	0	
Other medication <sup>a</sup>	11 (46%)	3 (15%)	
Smoker	0	8 (40%)	
Alcohol consumption	3 (13%)	16 (80%)	

SD standard deviation.

Statistical analysis was performed by two-sided unpaired t-test (data normally distributed) or non-parametric Mann–Whitney-U-test (data not normally distributed), statistical significance is highlighted as \* $p < 0.05$ .

<sup>a</sup>Other medication includes anti-hypertensive, anti-depressant, and pain relief medication.

One of the 20 urban participants was underweight (BMI  $< 18$  kg/m<sup>2</sup>).

observed in the rural Xhosa fecal microbiota (Fig. 1D). Rural fecal samples also showed high levels of OTUs related to *Treponema* (Fig. 1D), an organism that seems to be associated with a traditional rural lifestyle<sup>16</sup>.

### Urban diet is linked to high levels of deoxycholic acid and a distinct bacteria-metabolite pattern

Global profiling of urine and fecal water from rural and urban Xhosa people was acquired and analyzed using PCA and OPLS-DA. Scores plots of PCA did not show clear clustering between the two groups and no statistically significant OPLS-DA were obtained from the two groups (Suppl. Fig. 1A, B). Targeted quantification of gut microbial metabolites revealed similar levels of the main SCFA, including butyrate, in the feces of rural and urban Xhosa people (Fig. 2A), while numbers of bacteria expressing the butyryl coenzyme A (CoA):acetate-CoA transferase (*bcoA*, involved in butyrate production) were higher in urban Xhosa fecal samples (Suppl. Fig. 1C). In contrast, levels of cholic acid (CA), deoxycholic acid (DCA) and ursodeoxycholic acid (UDCA) were higher in feces of urban Xhosa people, whereas chenodeoxycholic acid (CDCA) and lithocholic acid (LCA) were similar to the rural Xhosa cohort (Fig. 2B, C). The levels of the essential amino acids lysine and threonine, as well as serine and glutathione, were lower in fecal samples of urban Xhosa individuals (Suppl. Table 1). In contrast, the concentration of the non-dietary amino acid homocysteine was higher in urban Xhosa individuals (Suppl. Table 1), possibly reflecting their higher animal protein intake<sup>17</sup>.

We identified a strong positive correlation between *Bacteroides*, *Dialister*, *Megasphaera*, and *Muribaculacea* with total energy intake as well as total consumption of carbohydrates, fiber, fat, protein, and several micronutrients (Fig. 2D). Conversely, *Catenibacterium*, *Christensenellaceae* and, to a minor extent, *Prevotella* showed a uniform inverse association with these dietary factors. *Prevotella*, which was enriched in the fecal microbiota of rural Xhosa individuals, had the strongest positive association with fecal SCFA levels (Fig. 2D). While this may indicate a fundamental role of *Prevotella* in saccharolytic

**Table 2 | Daily dietary intake of macro- and micro-nutrients in rural and urban Xhosa cohorts<sup>a</sup>**

Daily nutrient intake	Rural	Urban	P value	Rural	Urban	p Value
Energy (kcal)	2,185 ± 179	3,578 ± 455	0.0041**	Normalized to per 1000 kcal		
Total carbohydrates (g)	395 ± 32	511 ± 60	0.0812	184 ± 6	150 ± 7	0.0003***
Total dietary fiber (g)	28 ± 2	40 ± 5	0.034*	14 ± 1	12 ± 1	0.0391*
Total fat (g)	41 ± 6	107 ± 17	0.0004***	19 ± 2	28 ± 2	0.0004*
Dietary saturated fat (g)	12 ± 2	31 ± 5	0.0013**	5 ± 1	8 ± 1	0.005**
Cholesterol (mg)	172 ± 34	468 ± 108	0.0073**	75 ± 10	122 ± 20	0.029*
Total protein (g)	63 ± 7	109 ± 17	0.0128*	28 ± 2	29 ± 2	0.63
Animal protein (g)	25 ± 6	60 ± 13	0.0129*	11 ± 2	15 ± 2	0.0926
Plant protein (g)	37 ± 3	48 ± 6	0.0964	17 ± 1	14 ± 1	0.0009***
Folate (µg)	532 ± 43	665 ± 95	0.1837	254 ± 15	197 ± 13	0.0086**
Calcium (mg)	642 ± 75	809 ± 165	0.3348	299 ± 27	208 ± 22	0.0144*
Iron (mg)	18 ± 2	27 ± 5	0.0582	8 ± 1	7 ± 1	0.0192*
Zinc (mg)	11 ± 1	18 ± 3	0.0181*	5 ± 1	5 ± 1	0.6172
Vitamin A (µg)	183 ± 67	625 ± 197	0.0273*	69 ± 19	150 ± 36	0.0433*
Vitamin C (mg)	61 ± 10	161 ± 69	0.1235	32 ± 6	37 ± 9	0.6269
Total polyphenols (mg)	1400 ± 212	994 ± 156	0.1503	830 ± 178	344 ± 86	0.026*

<sup>a</sup>Values are shown as mean ± SEM (standard error of the mean). Statistical analysis for rural  $n=24$  and urban  $n=20$  samples using t-tests (two-sided) corrected for false discovery rate (FDR, 1%) according to the Benjamini–Hochberg method with statistical significance highlighted as \* $p < 0.05$ , \*\* $p < 0.01$  and \*\*\* $p < 0.001$ . Source data are provided in Supplementary Data File 1.

fermentation, bacterial abundance and metabolite levels may not show linear associations. This is also exemplified by other complex polysaccharide-degrading and SCFA-producing genera: *Faecalibacterium*, *Holdemanella* and *Roseburia* were not strongly associated with fecal SCFA, whereas *Butyrivibrio* and the *Eubacterium coprostanoligenes* group were inversely associated (Fig. 2D). The fecal bile acids CA and DCA were positively associated with *Bacteroides*, *Butyrivibrio* and *Ruminococcus torques* group, whereas *Dialister* and *Succinivibrio*, both notable SCFA producers, showed an inverse correlation (Fig. 2D). A similar pattern was detected for CDCA, but not LCA, suggesting differential responses of bacteria to different classes of bile acids.

### Differences in composition but not diversity of fecal virome are linked to separate phage-bacteria-metabolite dynamics in rural and urban cohorts

In total, 900 viral contigs (VC = viral OTUs) with lengths longer than 3 kb were identified, of which 176 were shared between the two cohorts (Fig. 3A), suggesting that a high portion remained unique to each group. The vast majority (>85%) of VCs were double-stranded DNA, and more than 4% were single-stranded DNA. The virome of rural Xhosa individuals had higher proportions of Podoviruses (3.3% vs. 0.3%) and lower levels of Siphoviruses (1.7% vs. 4.6%) and Microviruses (8.3% vs 10.5%) compared to the urban Xhosa cohort (Fig. 3B), yet these differences were not statistically significant. About 82% of the viral sequences were not taxonomically assigned. Of all VCs, 124 were found to be differentially abundant in rural and urban Xhosa individuals: phages predicted to infect *Oscillospiraceae*, *Parabacteroides*, and *Lactococcus* were more often detected in fecal samples of rural Xhosa individuals, whereas urban fecal samples contained more phages directed at *Faecalibacterium* and *Muribaculaceae* (Fig. 3C). Phages targeting *Bacteroides* were enriched in both, rural and urban, samples (Fig. 3C). With regard to viral replication cycle, more temperate phages (8.6% vs. 5%) were identified in urban fecal samples, but more crAsphages (0.2% vs. 0.02%) were found in rural Xhosa individuals. However, these differences were not statistically significant. Although there was no significant difference in  $\alpha$ -diversity for VCs across urbanization (Suppl. Fig. 2A), these demonstrated separate clustering on Bray–Curtis dissimilarity (Fig. 3D).

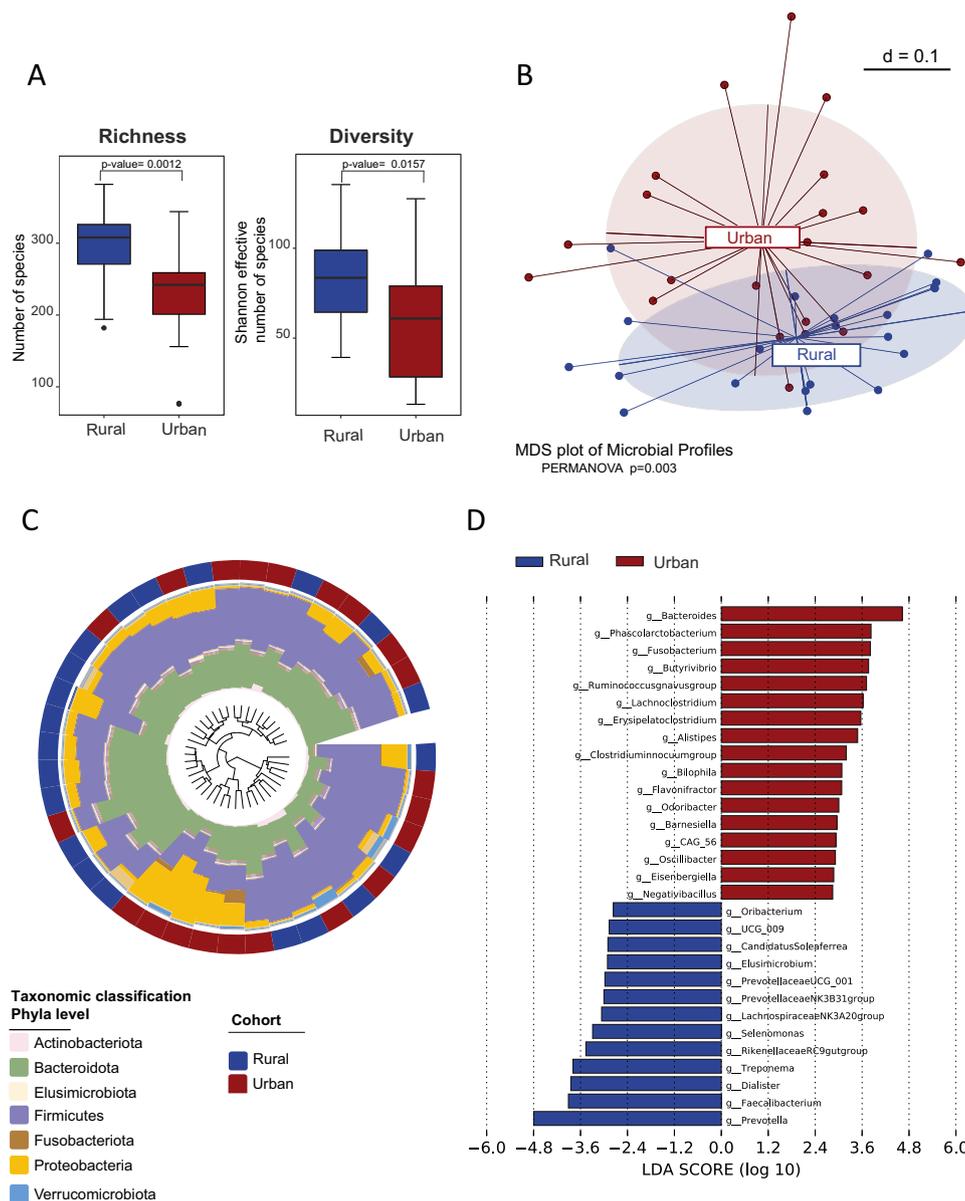
Pearson's correlation of VCs that are differently abundant across urbanization and fecal metabolites or dietary nutrients revealed

multiple novel associations ( $r > 0.6$ ): VC58, a virulent phage (= a phage that is not able to lysogenize) predicted to infect *Muribaculaceae*, and the temperate phage VC80 were strongly associated with fecal butyric acid and vitamin C (Fig. 3E). VC79 was strongly linked to vitamin A levels. VC680, a virulent phage enriched in urban samples, was positively associated with cholesterol, animal protein, and zinc (Fig. 3E). Two virulent phages highly abundant in rural samples, VC139 and VC2, the latter is predicted to infect *Parabacteroides*, showed weak inverse correlation ( $r = 0.2$ ) with dietary fiber intake, but positive correlation with the primary bile acid CA (Fig. 3E). Several VCs were positively associated ( $r > 0.3$ ) with fecal DCA levels (VC6, VC14, VC47, VC65, VC158, VC161), suggesting a link between this secondary bile acid and phages in the gut, which requires further experimental validation.

Urban Xhosa fecal samples harbored more phages predicted to infect the phylum Bacteroidota (Fig. 3F). We were unable to assign bacterial hosts to most of the VCs in the study, with a higher abundance of unassigned phages in rural Xhosa fecal samples (Fig. 3F), highlighting the limited understanding of the gut virome in Sub-Saharan populations. At the level of VCs and bacterial genera that are both differentially abundant in rural or urban Xhosa people, a clear clustering according to dominant bacteria was detected: VC58, VC80, and VC269, all highly abundant in urban fecal samples, were positively associated with *Bacteroides* and *Fusobacterium*, but inversely linked to *Prevotella* (Suppl. Fig. 2B). VC58 and VC269 were predicted to infect Bacteroidota, suggesting active interactions between these phages and their bacterial host. VC269 was positively associated with *Bilophila*, which includes several sulfite reducers<sup>18</sup>, and was also more abundant in urban Xhosa microbiota (Suppl. Fig. 2B).

### The composition of food and skin microbiota is different between rural and urban cohorts

We performed exploratory microbiota analysis of the 24 h food samples collected from representative households of both Xhosa cohorts to characterize how respective food and processing conditions may contribute to the differential microbiota patterns. The microbiota of food samples collected from rural Xhosa individuals showed higher  $\alpha$ -diversity, separate compositional clustering on Bray–Curtis similarity, a higher absolute number of bacteria counts, and higher abundances of OTUs linked to *Rosenbergiella* and *Weissella* (Fig. 4A–D), shown to be present in fruit nectar<sup>19</sup> or fermented foods<sup>20</sup>, respectively.



**Fig. 1 | The fecal microbiota of urban Xhosa individuals is less diverse and shows a distinct clustering and composition.** Analysis of fecal microbiota by 16 S rRNA gene sequencing from healthy rural and urban Xhosa individuals showing **A** bacterial community richness or Shannon Effective as indices for  $\alpha$ -diversity, **B** compositional clustering at the multi-dimensional scaling (MDS) plot (PERMANOVA test  $p = 0.001$ ), **C** cumulative relative abundance at phylum level in a phylogenetic tree and circular plot, where the outer ring indicates cohort distribution with urban (red) and rural (blue) individuals and the inner ring indicates taxonomic

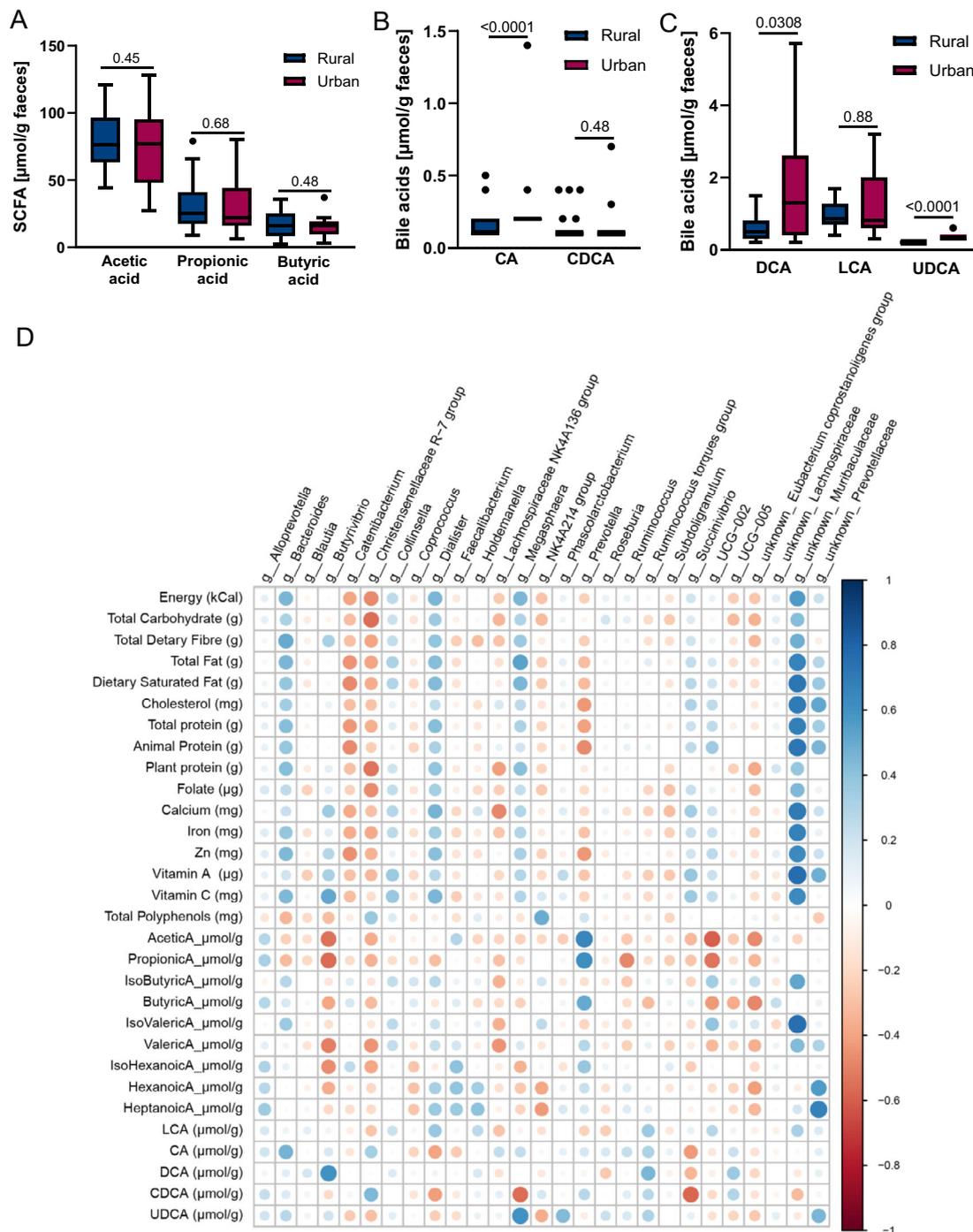
classification covering 100% relative abundance at phylum level for the respective sample, **D** significantly differential bacterial genera identified by linear discriminant analysis (LDA) effect size (LEfSe). Box-and-whisker plots with center as median, box from 25th to 75th percentile, and whiskers showing minima and maxima according to Tukey. Statistical analysis using two-sided Mann–Whitney–U-test or multivariate comparison corrected for multiple testing by the Benjamini–Hochberg method with rural  $n = 21$ , urban  $n = 20$  samples, respectively. A value of  $p < 0.05$  was considered to be statistically significant.

Regarding metabolites, the 24 h collection of the food consumed in both groups revealed higher levels of butyric acid and several other carboxylic acids such as aconitate,  $\alpha$ -ketoglutarate, succinate and fumarate (Suppl. Table 2) in the rural Xhosa cohort.

In another exploratory microbiota analysis, hand-finger swabs of rural and urban Xhosa study participants revealed no difference in  $\alpha$ -diversity, but distinct compositional clustering, characterized by higher levels of Cyanobacteria for rural and higher abundance of Proteobacteria in urban Xhosa individuals (Fig. 5A–C). Several low-abundance genera were more often detected in rural (e.g., *Actinomyces*, *Bacillus*, *Sphingomonas*, *Weissella*) or urban (e.g., *Acinetobacter*, *Jeotgalicoccus*, *Psychrobacter*, *Succinivibrio*) Xhosa individuals (Fig. 5D) (Suppl. Table 3).

## Discussion

Here, we investigate potential biome explanations for the alarming increase in NCDs such as CRC in urban South Africa by comparison to rural Xhosa individuals who continue a traditional lifestyle. We present new evidence on how diet transition across urbanization may be linked to distinct gut microbiota signatures in healthy individuals. The diet of healthy urban Xhosa individuals was characterized by higher energy, fat, cholesterol, and animal protein consumption as well as a relatively high but still  $<50$  g/d intake of dietary fiber that was previously associated with a minimal CRC risk<sup>2</sup>. This was associated with a less diverse gut microbiota with bacterial compositions likely closer to Western cohorts and functionally adapted to Western diet stimuli (e.g., secondary bile acids). Despite the relatively high



**Fig. 2 | Urban Xhosa individuals have similar SCFA but higher DCA levels compared with rural individuals, correlating with distinct microbial genera.**

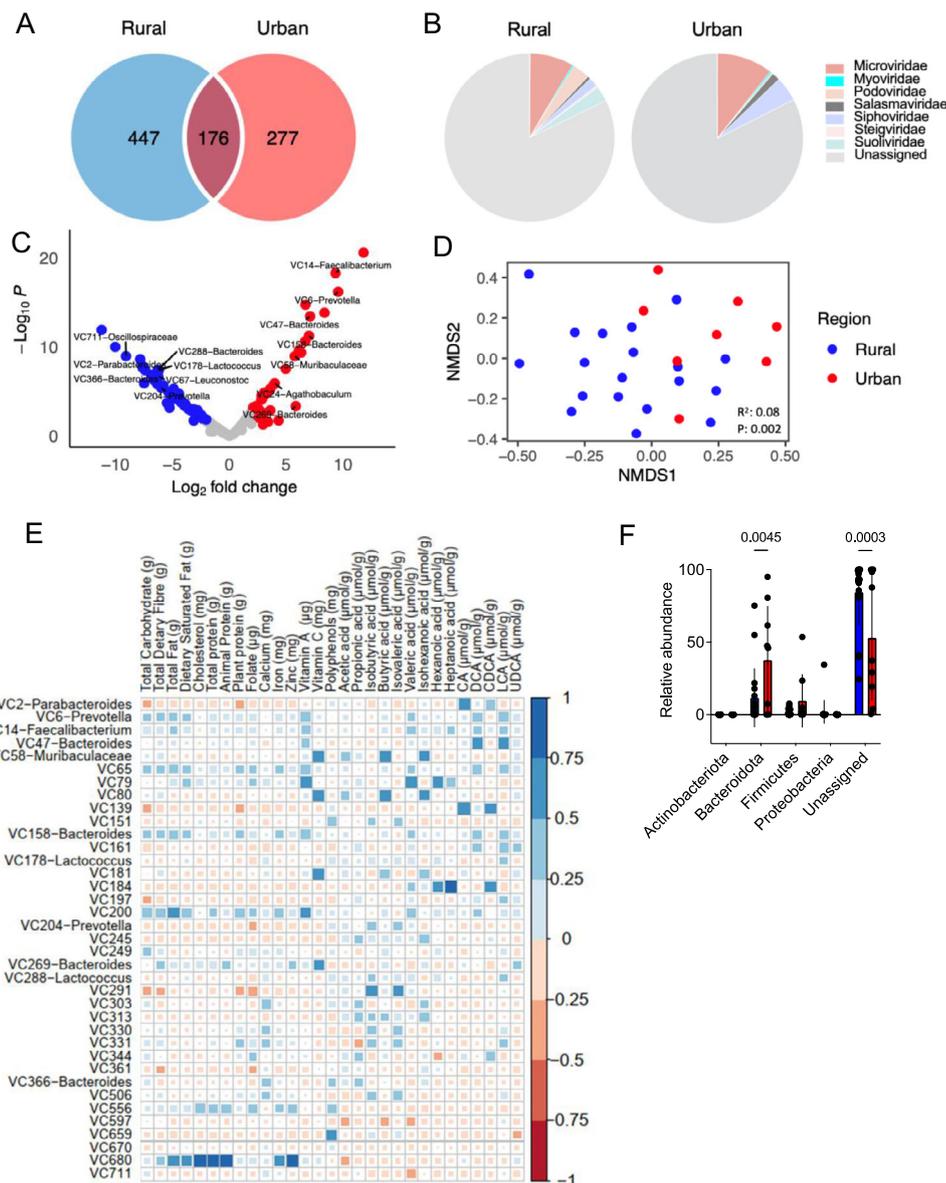
**A** Levels of main short-chain fatty acids detected in feces of rural and urban Xhosa individuals. Levels of **B** primary and **C** secondary bile acids detected in feces of rural and urban Xhosa individuals (CA cholic acid, CDCA chenodeoxycholic acid, DCA deoxycholic acid, LCA lithocholic acid, UDCA = ursodeoxycholic acid). Statistical analysis was performed by two-sided unpaired t-test (data normally distributed) or non-parametric Mann–Whitney-U-test (data not normally distributed) with rural  $n = 24$  and urban  $n = 19$  samples. A value of  $p < 0.05$  was considered to be statistically significant. Box-and-whisker plots with whiskers minima and maxima

according to Tukey, center as median, box from 25th to 75th percentile. Source data are provided in Supplementary Data file 1. **D** Heatmap of Pearson's correlations between dietary information, fecal metabolites and microbial genera detected to be significantly different between rural and urban Xhosa cohorts using 16S rRNA gene sequencing data. Correlations were adjusted for multiple testing using the Benjamini-Hochberg FDR. The color of the circles indicates the type of correlation (positive/negative), and the radius of the circles is proportional to the correlation. Statistical analysis was performed with sample numbers as listed for previous analyses in Table 2, Figs. 1 and 2.

amounts of dietary fiber and fecal SCFA levels detected in urban individuals, their diet resembles the pattern of Western diets and is associated with a higher number of obesity cases ( $n = 6$  in urban vs.  $n = 1$  in rural) (Table 1) and NCD risk-associated fecal metabolites

such as homocysteine that was previously linked to vascular diseases<sup>21</sup>.

The fecal microbiota of the urban Xhosa cohort was enriched for bacteria and related phages that have been associated with a



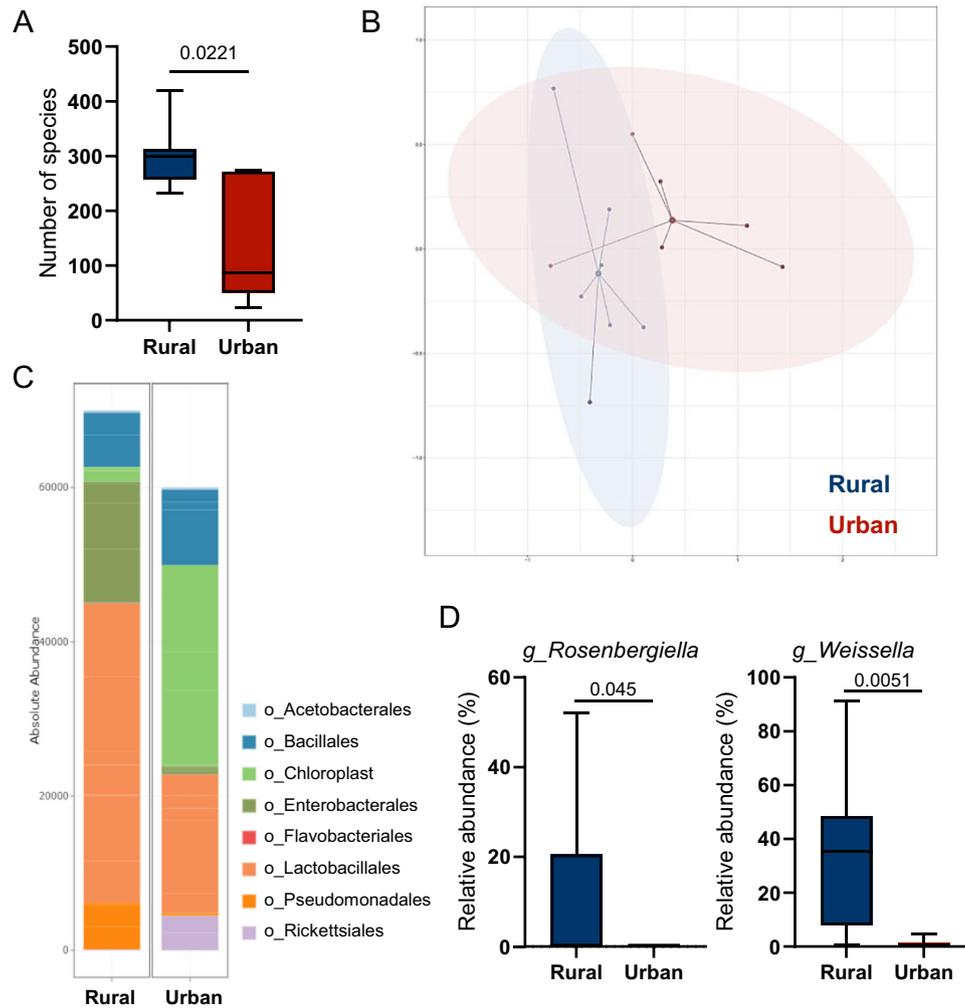
**Fig. 3 | The fecal virome of urban Xhosa individuals shows distinct composition and correlation with fecal metabolites and dietary nutrients.** **A** Shared and unique viral contigs (VCs) between rural ( $n = 19$ ) and urban ( $n = 9$ ) Xhosa individuals. **B** Proportion of VCs based on predictions of their taxa. **C** Differentially abundant VCs in rural ( $n = 19$ ) compared to urban ( $n = 9$ ) Xhosa fecal samples. **D** NMDS plot of Bray–Curtis dissimilarity of VCs between rural ( $n = 19$ ) and urban ( $n = 8$ ) Xhosa fecal samples. **E** Pearson's correlation of VCs differentially abundant

in rural ( $n = 19$ ) or urban ( $n = 9$ ) Xhosa fecal samples and dietary nutrients or fecal metabolites, respectively. **F** Proportion of relative abundance of VCs based on predicted bacterial hosts from rural ( $n = 19$ ) and urban ( $n = 9$ ) Xhosa fecal samples. Statistical analysis was performed by two-way ANOVA corrected for multiple comparisons with Sidak's test. A value of  $p < 0.05$  was considered to be statistically significant. When present, error bars correspond to the standard deviation of host abundances.

Western-style dietary pattern and high CRC risk in previous studies: *Bacteroides* were detected at high levels in diets rich in fat<sup>22–25</sup> and were more often targeted by VCs in the urban Xhosa cohort, suggesting an active interaction between this genus and its associated phages, possibly through fluctuating-selection dynamics observed in these phages<sup>26</sup>. *Fusobacteria* were linked to CRC in previous studies<sup>27,28</sup>, and their phages are potential markers of CRC<sup>29</sup>. Sub-species of *Alistipes* showed experimental pro-inflammatory and tumor-promoting activity<sup>30,31</sup>. *Bilophila*, a group of sulfite-reducing bacteria associated with intestinal inflammatory conditions and adaptation of the gut microbiota to bile acid metabolism<sup>32,33</sup>, was targeted by phages highly abundant in the urban Xhosa cohort. Phages were shown to alter the sulfur metabolism of host cells by encoding auxiliary metabolic genes<sup>34</sup>, suggesting an active role in regulating human health through

their bacterial hosts. Urban fecal samples also contained more *Faecalibacterium* phages that are commonly found in patients with IBD<sup>35</sup> and may contribute to diseases by reducing the abundance of their host through kill-the-winner dynamics.

The higher fat intake of the urban Xhosa cohort is likely to increase the levels and a differential composition of fecal bile acids and to promote the abundance of gut bacteria adapted to bile acid transformation: the genus *Lachnospirillum* includes members involved in or closely linked to bacteria performing bile acid metabolism and was enriched in Alaska Native people who have a very high CRC risk<sup>32</sup>. Higher levels of DCA and 7 $\alpha$ -dehydroxylating bacteria were detected in the feces of CRC patients, in healthy cohorts facing a high CRC risk, and in urban compared with rural Zimbabweans<sup>32,36–39</sup>. Several phages were positively associated with fecal bile acid levels, including CA and



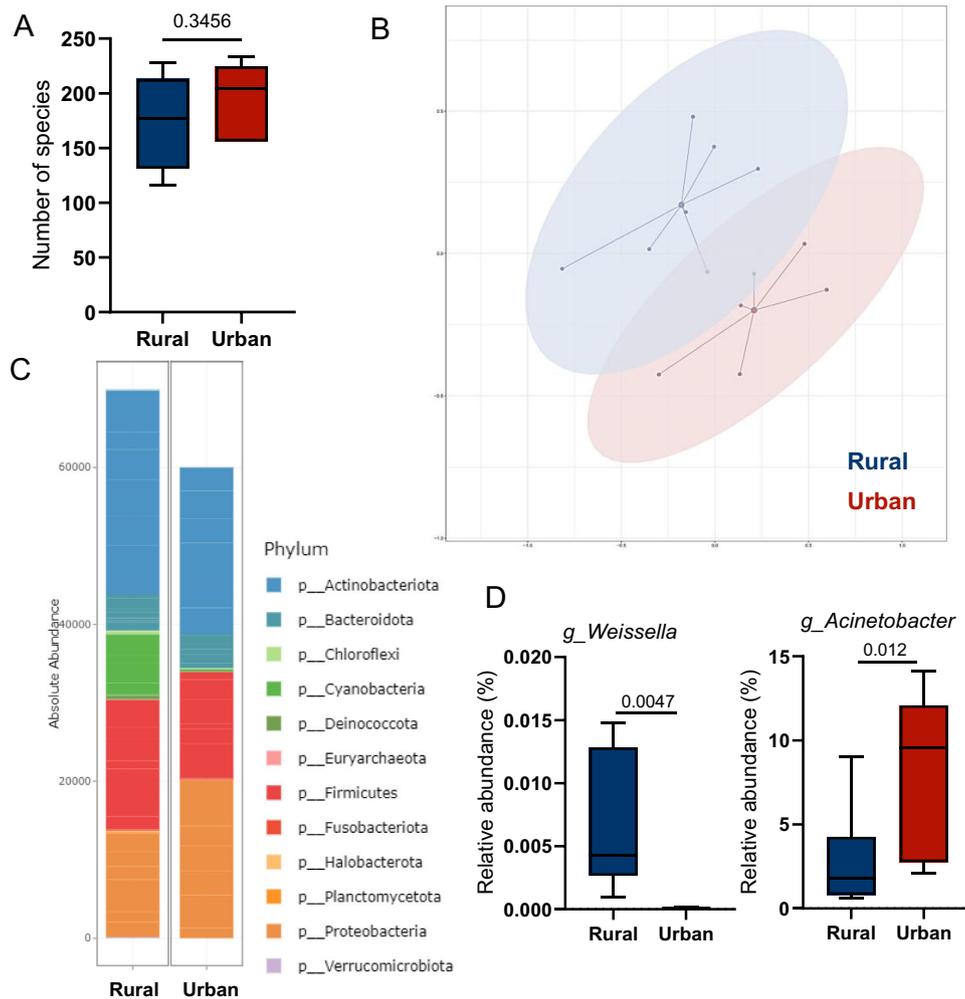
**Fig. 4 | Differences in food microbiota reflect food ingredients of rural and urban Xhosa people.** Analysis of microbiota in frozen food samples by 16S rRNA gene sequencing obtained from rural and urban Xhosa individuals showing **A** richness as an index for  $\alpha$ -diversity, **B** Bray–Curtis dissimilarity at the multi-dimensional scaling (MDS) plot (PERMANOVA test  $p = 0.004$ ), **C** cumulative absolute abundance at the order level, **D** significantly different relative abundance at the

genus level. Statistical analysis using a two-sided Wilcoxon test corrected for multiple testing by the Benjamini–Hochberg method using normalized samples (rural  $n = 7$  and urban  $n = 6$ ). A value of  $p < 0.05$  was considered to be statistically significant. Box-and-whisker plots with whiskers showing minima to maxima, center as median, box from 25th to 75th percentile.

DCA, supporting previous studies that demonstrated phage-mediated alterations of microbial bile acid transformation<sup>40</sup> and enrichment of a phage targeting a DCA-producing bacterium in CRC patients<sup>29</sup>. The higher levels of temperate phages in urban Xhosa fecal samples may be a consequence of diet transition since Western dietary compounds are potent inducers of prophages in bacteria<sup>41</sup> that shape the gut microbiota through host cell lysis and horizontal gene transfer<sup>42</sup>. In addition, the skin microbiota may be a reservoir contributing to changes in microbial setups across urbanization, especially considering different hygiene conditions or food processing approaches: higher abundance of skin-associated Proteobacteria in the urban Xhosa cohort and soil/water-associated genera on the hands of rural individuals indicated different home environments along the urbanization gradient<sup>43</sup>. However, their overall very low abundances limit potent effects on the host, requiring cautious interpretation and further critical analysis (e.g., control for contamination).

Given the Western diet pattern and CRC risk-associated gut microbial composition, concomitant with retained fiber intake and high fecal SCFA levels, the urban Xhosa cohort may show a transitional state between traditional and Western diet, confirming previous results from a study in Zimbabwe<sup>39</sup>. This may indicate that local

traditional and Western diets are interlaced in this study population, which may collectively promote beneficial effects on intestinal health across urbanization. However, functional consequences with regard to long-term CRC or NCD risk remain unclear since urban fecal samples had lower levels of saccharolytic and butyrogenic genera. Both, *Prevotella* and *Faecalibacterium*, were shown to be more abundant in fecal samples from healthy individuals compared to CRC patients in Kenya<sup>44</sup>. Enrichment of *Prevotella* ssp. was demonstrated to be a key characteristic of the gut microbiota in rural Sub-Saharan Africa<sup>24,32,36,45–47</sup>. The relative loss of *Prevotella* members in the urban fecal microbiota may be linked to the higher abundance of *Prevotella* phages in these samples and detrimental to the functional capacity to degrade plant-derived complex carbohydrates despite the high fiber consumption of the host<sup>48,49</sup>. This was further supported by the strong positive association of *Prevotella* with fecal SCFA in this study, but other complex polysaccharide-degrading SCFA-producers (*Holdemanna*, *Roseburia*, *Faecalibacterium*) were not strongly associated with fecal SCFA. The latter may indicate that bacteria abundance and associated metabolite levels do not follow linear associations, which is further supported by the higher *bcoA* copy but similar butyrate levels in urban compared to rural Xhosa fecal samples. Since *bcoA* gene copy



**Fig. 5 | Different bacteria detected on skin swabs of rural and urban Xhosa people.** Analysis of microbiota in skin swab samples (hands) by 16S rRNA gene sequencing obtained from rural and urban Xhosa individuals showing **A** richness as an index for  $\alpha$ -diversity, **B** Bray–Curtis dissimilarity at the multi-dimensional scaling (MDS) plot (PERMANOVA test  $p = 0.003$ ), **C** cumulative absolute abundance at the phylum level, **D** significantly different relative abundance at the genus level.

Statistical analysis using a two-sided Wilcoxon test corrected for multiple testing by the Benjamini–Hochberg method using normalized samples (rural  $n = 7$  and urban  $n = 6$ ). A value of  $p < 0.05$  was considered to be statistically significant. Box-and-whisker plots with whiskers showing minima to maxima, center as median, box from 25th to 75th percentile.

numbers were targeted by degenerate primers designed for Western cohort studies, their higher levels detected in urban fecal samples could also reflect compositional shifts in the gut microbiota due to the transition to urban lifestyle while not covering butyrogenic bacteria that are more prevalent in rural African communities. *Treponema* was more abundant in the fecal microbiota of rural individuals and is another indicative genus for rural gut microbiota in Sub-Saharan Africa<sup>46,50</sup>. The genus *Phascolarctobacterium*, covering bacteria able to produce propionate and acetate, was more abundant in the urban cohort but previously reported to be reduced or lacking in Western populations and urbanized cohorts in South Africa<sup>16,46,51</sup>. Together with the minor differences in global fecal metabolomic profiles, these results suggest that the urban Xhosa cohort analyzed here still retains gut microbial features of rural South African communities, which may be targeted by the restoration of specific components of the traditional diet. This may include access to plant-based local food ingredients (as highlighted by the higher polyphenol intake and abundance of *Rosenbergiella* in rural Xhosa food samples) and usage of traditional food preparation such as fermentation (indicated by higher abundance of *Weissella* in rural Xhosa food samples)<sup>19,20</sup>. Traditional food preparation is also relevant in the context of resistant starch, which is a dietary fiber rich in traditional high carbohydrate diets of rural South

Africa, but not measured in this study as its content is variable and therefore difficult to measure with standard food composition tables<sup>36</sup>. This also highlights the need to characterize dietary fibers in more detail in dietary analyses, accounting for fiber heterogeneity that may affect their functionalization by the gut microbiota<sup>4,52,53</sup>. In this study, other factors (e.g., obesity, alcohol consumption, smoking, medication, different average age between cohorts) may impair the general extrapolation of observed biome signatures to urbanizing communities in low- or middle-income countries. Finally, the cross-sectional study layout, small sample size, and focus on two study sites may not fully reflect the dynamic processes within urbanization, whereas the lack of colonic mucosal biopsies limits the conclusions with regard to the CRC risk of study cohorts. Consistently, a recent systematic review highlighted that changes in the microbiota along the rural-urban gradient in Sub-Saharan Africa are heterogeneous and multifactorial resulting in low universal patterns<sup>54</sup>.

This study demonstrates diet- and microbiota-centered support for the hypothesis that urbanization-associated changes in lifestyle as part of economic development are the main drivers associated with NCD risk in Sub-Saharan Africa<sup>55</sup>. The microbial and metabolic differences of the two geographically distinct South African populations in this study warrant further investigation of the functional

consequences of diet-mediated microbiota modulation with regard to CRC and overall NCD risk. In general, more attention should be given to the quality and quantity of food consumed by inhabitants of Sub-Saharan Africa undergoing Westernization to avoid the otherwise inevitable increase in NCDs in urbanizing populations. In line with this, a focus on nutritional expertise adapted to local requirements, the characterization of undescribed African microbial diversity, including the virome, and the collaboration with local research partners and communities are integral elements to further study Westernization and NCD risk in Sub-Saharan Africa.

## Methods

### Study design and population

A cross-sectional, observational pilot study was performed in two healthy middle-aged rural or urban South African cohorts self-identifying as amaXhosa (Xhosa) with sample collection at the one-time point. The urban cohort was located in Cape Town, Western Cape Province, while the rural group was recruited in the rural Eastern Cape Province of South Africa. Cape Town is the chief city in Western Cape province and has been receiving migrant populations from within and outside the country for centuries. Most of the in-country migrants originate from rural Eastern Cape Province and have settled in the rapidly expanding township of Khayelitsha, the fastest-growing township in South Africa (population about 392,000 in 2011, according to the Department of Statistics South Africa). The study was approved by the Stellenbosch University Health Research Ethics Committee. Informed consent was obtained from all study participants. Inclusion and exclusion criteria were applied as reported previously<sup>36</sup>, and details are provided in the supplementary information. Recruitment of healthy rural participants was based at Zithulele Hospital, Eastern Cape (see additional information in the supplementary material). Recruitment of healthy participants was organized through advertisements and visits to community leaders in public places (e.g., taxi centers and places of community worship) in Khayelitsha. In total,  $n = 24$  healthy individuals were recruited in the rural and  $n = 20$  healthy participants in the urban area of the study (Table 1). Our previous studies in a similar setting demonstrated that samples of this size were sufficient to characterize dietary differences of healthy populations<sup>32,39,47</sup>. Measurements of weights and heights were taken to calculate BMI.

### Dietary assessment

Dietary intake information of rural ( $n = 24$ ) and urban ( $n = 20$ ) study participants was collected at one-time point by an experienced research assistant specifically trained in dietary intake assessment. A South African country-specific, validated semi-quantitative food frequency questionnaire (QFFQ) was employed, developed for the Prospective Urban and Rural Epidemiology study<sup>56</sup> together with a 24-h recall questionnaire. Dietary intake data were quantified in a non-blinded fashion using dietary coding sheets, Food Quantities Manual, and Food Composition Tables developed by the South African Medical Research Council<sup>57,58</sup>. The amounts of nutrients per day and nutrients per specific food items consumed by the two study populations were estimated by computer analysis. Polyphenol values were quantified using the online database Phenol-Explorer (<http://phenol-explorer.eu>). Source data are provided in Supplementary Data File 1.

### Sample collection

Collection tubes with and without DNA stabilization buffer (Invitex Molecular Inc., Germany) for the collection of fecal and urinary (collected in the morning) samples were prepared to facilitate collection in home environments. The sampling took place from November 2019 to February 2020. Duplicates of all meals and drinks consumed over 24 h were collected as “total food intake” from rural ( $n = 7$ ) and urban ( $n = 6$ )

Xhosa individuals, kept cool until complete, and transported to the laboratory for homogenization and sampling. Hand-finger swabs were taken from the same study participants (rural  $n = 7$ , urban  $n = 6$ ) using a DNA stabilization buffer collection tube (DNA Genotek Inc., Canada). Samples were transported on ice to the laboratory, where they were aliquoted and frozen at  $-80^{\circ}\text{C}$  as soon as possible, generally within 6 h. The samples collected in rural Eastern Cape were first frozen at  $-20^{\circ}\text{C}$  at the clinic and then air-couriered on dry ice to the laboratory for further processing.

### Quantification of fecal SCFA by gas chromatography–mass spectrometry (GC–MS)

Fecal SCFA was analyzed using a stringently validated and simplified method for the quantitation of SCFA in human stool using gas chromatography–mass spectrometry<sup>59</sup>. Calibration curves of each SCFA were prepared by extracting aliquots of the Sigma-Aldrich VFAM Certified Reference Material equivalent to 0.05, 12.5, 25.0, 37.5, and 50.0  $\mu\text{mol/g}$  using the internal standard extraction solution following the procedure described.

### Quantification of fecal bile acids by GC–MS

Selected fecal bile acids (cholic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA), lithocholic acid (LCA), ursodeoxycholic acid (UDCA)) were quantified by GC-MS operating in selected ion monitoring mode using D<sub>5</sub>-chenodeoxycholic acid as an internal standard. The LDR was 0.25–5.00  $\mu\text{mol/g}$  feces. The LOQ and detection limit was 0.25  $\mu\text{mol/g}$  feces. Both the coefficients of variation and the relative recoveries were less than 5% and 90–110%, respectively, for most compounds at most levels.

### <sup>1</sup>H NMR spectroscopy-based metabolic profiling of urine and fecal water

Untargeted profiling of urinary and fecal metabolites was performed as described previously<sup>32,36</sup> and is detailed in the supplementary information.

### Quantitation of amino acids by ultra-high-performance liquid chromatography with mass spectrometry (UPLC–MS/MS)

The quantification of amino acids in feces was performed using LC-MS/MS, with details provided in the supplementary information.

### Food metabolomic analysis by LC–MS/MS

Targeted quantitative metabolomics was applied to analyze food samples using a combination of direct injection mass spectrometry with a reverse-phase LC–MS/MS custom assay. This custom assay, in combination with an ABSciex 4000 QTrap (Applied Biosystems/MDS Sciex, MA, USA) mass spectrometer, was used for the targeted identification and quantification of up to 207 different endogenous metabolites including amino acids, acylcarnitines, biogenic amines & derivatives, uremic toxins, glycerophospholipids, sphingolipids and sugars<sup>60</sup>, details are provided in the supplementary information.

### Bacteriome analysis and quantification of microbial *bcoA* copy numbers

For bacteriome analysis of fecal samples stored in DNA stabilization buffer (rural  $n = 24$ , urban = 20) and frozen food (rural  $n = 7$ , urban  $n = 6$ ) samples, 16 S rRNA gene sequencing analysis was conducted as described previously<sup>61</sup> with few changes. Briefly, the DNA of the samples was isolated from 600  $\mu\text{l}$  of the sample-stabilizer mixture after bead beating as described but using the Maxwell (Promega) with the Maxwell<sup>®</sup> RSC Fecal Microbiome DNA Kit. Isolated DNA was used in a 2-step PCR, first using specific primer 341 F (CCT ACG GGN GGC WGC AG) and 785 R (GAC TAC HVG GGT ATC TAA TCC). Both specific primers had adapters attached, which were used in the second PCR to

attach sample-specific barcodes and the P5 and P7 adapters for sequencing. The 16 S rRNA gene sequencing analysis of hand-finger swab samples (rural  $n = 7$ , urban  $n = 6$ ) was performed as previously described<sup>32</sup>. After sequencing, the raw data were processed using the functionality of the IMNGS database<sup>62</sup>, details are provided in the supplementary information. For quantification of the butyryl coenzyme A (CoA):acetate-CoA transferase (*bcoA*) that is involved in butyrate production, extracted microbial DNA was analyzed as described before using the following degenerate primers (*bcoA* F: GCI GAI CAT TTC ACI TGG AAY WSI TGG CAY ATG, *bcoA* R: CCT GCC TTT GCA ATR TCI ACR AAN GC)<sup>32</sup>.

### Fecal virome analysis

Virus-like particles (VLPs) were isolated and processed from 28 fecal samples (rural  $n = 19$ , urban  $n = 9$ ) as described previously<sup>63</sup>. One sample from the urban cohort was excluded as an outlier in the NMDS plot. Viral sequencing data were processed using ViroProfiler<sup>64</sup>. The methods performed for isolating, sequencing, and characterizing the fecal viruses are detailed in the supplementary information.

### Statistical analyses

Statistical analysis of data sets on dietary parameters, fecal SCFA, bile acids, and *bcoA* gene copy numbers was performed by two-tailed unpaired t-test or non-parametric Mann–Whitney-U-test (depending on normal distribution of data) using Prism v.9.0 software (GraphPad Software Inc., CA, USA). Dietary data was FDR (1%) corrected according to the Benjamini–Hochberg method. Statistical analysis of bacteriome data was conducted using the Rhea pipeline and NAMCO microbiome explorer<sup>65,66</sup>. Briefly, adjustment for multiple comparisons was done by the Benjamini–Hochberg correction, and depending on the type of comparison, different statistical tests were applied as described, including Kruskal–Wallis Rank Sum, Fisher’s Exact, and Wilcoxon Rank Sum test. Microbial  $\beta$ -diversity was determined by fitting models with 16 S profiles as distance-based responses using PERMANOVA and visualized by clustering on an MDS plot. For statistical analysis of virome data, Prism v.9.0 software was used to run two-way ANOVA and Sidak’s multiple comparisons test, and the remaining analyses were conducted in R v.4.2.1 (R Core Team, 2022).

### Reporting summary

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

### Data availability

The bacteriome and virome sequencing data generated in this study have been deposited in the NCBI BioProject database under accession code PRJNA1066974 and in the ENA under accession code PRJEB67451, respectively. Dietary and targeted metabolomics data are available in a Supplementary Data File.

### Code availability

Software programs used to analyze the data are either freely or commercially available. All other data relevant to the study are included in the article. Additional data are available on request.

### References

- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- O’Keefe, S. J. The association between dietary fibre deficiency and high-income lifestyle-associated diseases: Burkitt’s hypothesis revisited. *Lancet Gastroenterol. Hepatol.* **4**, 984–996 (2019).
- Aune, D. et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *Br. Med. J.* **343**, d6617 (2011).
- Reynolds, A. et al. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. *Lancet* **393**, 434–445 (2019).
- Wang, L. et al. Association of ultra-processed food consumption with colorectal cancer risk among men and women: results from three prospective US cohort studies. *Br. Med. J.* **378**, e068921 (2022).
- O’Keefe, S. J. D. Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 691–706 (2016).
- Donohoe, D. R. et al. A gnotobiotic mouse model demonstrates that dietary fiber protects against colorectal tumorigenesis in a microbiota- and butyrate-dependent manner. *Cancer Discov.* **4**, 1387–1397 (2014).
- Ocvirk, S. & O’Keefe, S. J. D. Dietary fat, bile acid metabolism and colorectal cancer. *Semin. Cancer Biol.* **73**, 347–355 (2021).
- Waddell, I. S. & Orfila, C. Dietary fiber in the prevention of obesity and obesity-related chronic diseases: from epidemiological evidence to potential molecular mechanisms. *Crit. Rev. Food Sci. Nutr.* **63**, 8752–8767 (2023).
- Roomaney, R. A., van Wyk, B., Cois, A. & Pillay-van Wyk, V. Inequity in the distribution of non-communicable disease multimorbidity in adults in South Africa: an analysis of prevalence and patterns. *Int. J. Public Health* **67**, 1605072 (2022).
- Pillay-van Wyk, V. et al. Mortality trends and differentials in South Africa from 1997 to 2012: second National Burden of Disease Study. *Lancet Glob. Health* **4**, e642–e653 (2016).
- Vorster, H. H., Kruger, A., Wentzel-Viljoen, E., Kruger, H. S. & Margetts, B. M. Added sugar intake in South Africa: findings from the Adult Prospective Urban and Rural Epidemiology cohort study. *Am. J. Clin. Nutr.* **99**, 1479–1486 (2014).
- Rogerson, J. M., Kotze, N. & Rogerson, C. M. Addressing South Africa’s urban challenges. *Urban. Izziv* **25**, S1–S4 (2014).
- Bigna, J. J. & Noubiap, J. J. The rising burden of non-communicable diseases in sub-Saharan Africa. *Lancet Glob. Health* **7**, 1295 (2019).
- Nugent, R. et al. Investing in non-communicable disease prevention and management to advance the sustainable development goals. *Lancet* **391**, 2029–2035 (2018).
- Angelakis, E. et al. Comparison of the gut microbiota of obese individuals from different geographic origins. *N. Microbes N. Infect.* **27**, 40–47 (2019).
- Mann, N. et al. The effect of diet on plasma homocysteine concentrations in healthy male subjects. *Eur. J. Clin. Nutr.* **53**, 895–899 (1999).
- Metwaly, A. et al. Integrated microbiota and metabolite profiles link Crohn’s disease to sulfur metabolism. *Nat. Commun.* **11**, 4322 (2020).
- Halpern, M., Fridman, S., Atamna-Ismaeel, N. & Izhaki, I. *Rosenbergiella nectarea* gen. nov., sp. nov., in the family Enterobacteriaceae, isolated from floral nectar. *Int. J. Syst. Evol. Microbiol.* **63**, 4259–4265 (2013).
- Teixeira, C. G. et al. Weissella: an emerging bacterium with promising health benefits. *Probiotics Antimicrob. Proteins* **13**, 915–925 (2021).
- Misner, B. D. Animal proteins influence blood lipids and homocysteine. *Br. Med. J.* **330**, 111 (2005).
- Chompre, G. et al. A one month high fat diet disrupts the gut microbiome and integrity of the colon inducing adiposity and behavioral despair in male Sprague Dawley rats. *Heliyon* **8**, e11194 (2022).
- Singh, R. P., Halaka, D. A., Hayouka, Z. & Tirosh, O. High-fat diet induced alteration of mice microbiota and the functional ability to utilize fructooligosaccharide for ethanol production. *Front. Cell. Infect. Microbiol.* **10**, 376 (2020).
- Nakayama, J. et al. Impact of Westernized diet on gut microbiota in children on Leyte island. *Front. Microbiol.* **8**, 197 (2017).

25. Martinez-Medina, M. et al. Western diet induces dysbiosis with increased *E coli* in CEABAC10 mice, alters host barrier function favouring AIEC colonisation. *Gut* **63**, 116–124 (2014).
26. Shkoporov, A. N. et al. Long-term persistence of crAss-like phage crAss001 is associated with phase variation in *Bacteroides intestinalis*. *BMC Biol.* **19**, 163 (2021).
27. Castellarin, M. et al. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res.* **22**, 299–306 (2012).
28. Kostic, A. D. et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).
29. Shen, S., Huo, D., Ma, C., Jiang, S. & Zhang, J. Expanding the colorectal cancer biomarkers based on the human gut phageome. *Microbiol. Spectr.* **9**, e00090–21 (2021).
30. Moschen, A. R. et al. Lipocalin 2 protects from inflammation and tumorigenesis associated with gut microbiota alterations. *Cell Host Microbe* **19**, 455–469 (2016).
31. Yang, J. et al. High-fat diet promotes colorectal tumorigenesis through modulating gut microbiota and metabolites. *Gastroenterology* **162**, 135–149.e2 (2022).
32. Ocvirk, S. et al. A prospective cohort analysis of gut microbial co-metabolism in Alaska Native and rural African people at high and low risk of colorectal cancer. *Am. J. Clin. Nutr.* **111**, 406–419 (2020).
33. Devkota, S. et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *IL10<sup>-/-</sup>* mice. *Nature* **487**, 104–108 (2012).
34. Kieft, K. et al. Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat. Commun.* **12**, 3503 (2021).
35. Cornuault, J. K. et al. Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes. *Microbiome* **6**, 65 (2018).
36. O’Keefe, S. J. D. et al. Fat, fibre and cancer risk in African Americans and Rural Africans. *Nat. Commun.* **6**, 6342 (2015).
37. Weir, T. L. et al. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS ONE* **8**, e70803 (2013).
38. Wirbel, J. et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
39. Katsidzira, L. et al. Differences in fecal gut microbiota, short-chain fatty acids and bile acids link colorectal cancer risk to dietary changes associated with urbanization among zimbabweans. *Nutr. Cancer* **71**, 1313–1324 (2019).
40. Campbell, D. E. et al. Infection with bacteroides phage BV01 alters the host transcriptome and bile acid metabolism in a common human gut microbe. *Cell Rep.* **32**, 108142 (2020).
41. Boling, L. et al. Dietary prophage inducers and antimicrobials: toward landscaping the human gut microbiome. *Gut Microbes* **11**, 721–734 (2020).
42. Borodovich, T., Shkoporov, A. N., Ross, R. P. & Hill, C. Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. *Gastroenterol. Rep.* **10**, goac012 (2022).
43. McCall, L.-I. et al. Home chemical and microbial transitions across urbanization. *Nat. Microbiol.* **5**, 108–115 (2020).
44. Obuya, S. et al. A signature of *Prevotella copri* and *Faecalibacterium prausnitzii* depletion, and a link with bacterial glutamate degradation in the Kenyan colorectal cancer patients. *J. Gastrointest. Oncol.* **13**, 2282–2292 (2022).
45. de Goffau, M. C. et al. Gut microbiomes from Gambian infants reveal the development of a non-industrialized *Prevotella*-based trophic network. *Nat. Microbiol.* **7**, 132–144 (2022).
46. Oduaran, O. H. et al. Gut microbiome profiling of a rural and urban South African cohort reveals biomarkers of a population in lifestyle transition. *BMC Microbiol.* **20**, 330 (2020).
47. Ou, J. et al. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *Am. J. Clin. Nutr.* **98**, 111–120 (2013).
48. Sonnenburg, E. D. et al. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016).
49. Gellman, R. H. et al. Hadza *Prevotella* require diet-derived microbiota accessible carbohydrates to persist in mice. *bioRxiv* <https://doi.org/10.1101/2023.03.08.531063> (2023).
50. Tamburini, F. B. et al. Short- and long-read metagenomics of urban and rural South African gut microbiomes reveal a transitional composition and undescribed taxa. *Nat. Commun.* **13**, 926 (2022).
51. Mancabelli, L. et al. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ. Microbiol.* **19**, 1379–1390 (2017).
52. Gurry, T., Nguyen, L. T. T., Yu, X. & Alm, E. J. Functional heterogeneity in the fermentation capabilities of the healthy human gut microbiota. *PLoS ONE* **16**, e0254004 (2021).
53. Veronese, N. et al. Dietary fiber and health outcomes: an umbrella review of systematic reviews and meta-analyses. *Am. J. Clin. Nutr.* **107**, 436–444 (2018).
54. Paulo, L. S. et al. Urbanization gradient, diet, and gut microbiota in Sub-Saharan Africa: a systematic review. *Front. Microbiomes* **2**, 1208166 (2023).
55. Dalal, S. et al. Non-communicable diseases in sub-Saharan Africa: what we know now. *Int. J. Epidemiol.* **40**, 885–901 (2011).
56. Wentzel-Viljoen, E., Laubscher, R. & Kruger, A. Using different approaches to assess the reproducibility of a culturally sensitive quantified food frequency questionnaire. *South Afr. J. Clin. Nutr.* **24**, 143–148 (2011).
57. SAFOODS. *SAMRC Food Quantities Manual for South Africa*. (South African Medical Research Council, 2018).
58. SAFOODS. *SAMRC Food Composition Tables for South Africa*. (South African Medical Research Council, 2017).
59. Eberhart, B. L., Wilson, A. S., O’Keefe, S. J. D., Ramaboli, M. C. & Nesengani, L. T. A simplified method for the quantitation of short-chain fatty acids in human stool. *Anal. Biochem.* **612**, 114016 (2021).
60. Zheng, J., Johnson, M., Mandal, R. & Wishart, D. S. A comprehensive targeted metabolomics assay for crop plant sample analysis. *Metabolites* **11**, 303 (2021).
61. Reitmeier, S., Kiessling, S., Neuhaus, K. & Haller, D. Comparing circadian rhythmicity in the human gut microbiome. *STAR Protoc.* **1**, 100148 (2020).
62. Lagkouvardos, I. et al. IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci. Rep.* **6**, 33721 (2016).
63. Khan Mirzaei, M. et al. Bacteriophages isolated from stunted children can regulate gut bacterial communities in an age-specific manner. *Cell Host Microbe* **27**, 199–212 (2020).
64. Ru, J., Khan Mirzaei, M., Xue, J., Peng, X. & Deng, L. ViroProfiler: a containerized bioinformatics pipeline for viral metagenomic data analysis. *Gut Microbes* **15**, 2192522 (2023).
65. Lagkouvardos, I., Fischer, S., Kumar, N. & Clavel, T. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ* **5**, e2836 (2017).
66. Dietrich, A. et al. Namco: a microbiome explorer. *Microb. Genom.* **8**, mgen000852 (2022).

## Acknowledgements

The authors would like to thank Jinling Xue for her assistance in extracting the viral DNA. S.J.D.O. received funding from the South African Research Foundation (SRG210331591524), the Stellenbosch University Rectors Strategic Research Fund, and the NIH (R01 CA204403).

S.O. received funding from the German Federal Ministry of Education and Research (01KA2103).

## Author contributions

S.J.D.O. designed, acquired funding, conducted, and edited the study. S.O., M.C.R., and M.K.M. prepared the original draft of the paper. M.C.R., L.T.N., A.S.W., S.F. helped with administration and regulatory committees and helped conduct the cross-sector study. S.K.J. and R.S. helped with the dietary analyses. M.K.M. was responsible for the virome analyses and evaluation. B.L.E. performed the targeted metabolite analyses. M.V.-G., A.M., G.B., J.R., D.M.-J., D.C.L., J.Z., Q.C., and M.R.P. helped with the laboratory sample analysis. A.M., S.O., K.N., and B.M. performed the bacteriome analysis. J.V.L., D.W., and R.M. performed the metabolomic analyses. B.G. organized the recruitment of rural volunteers. S.O., D.H., J.V.L., and L.D. supervised and organized the analysis of samples.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-46265-0>.

**Correspondence** and requests for materials should be addressed to S. J. D. O'Keefe.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

**M.C. Ramaboli**<sup>1,15</sup>, **S. Ocvirk**<sup>2,3,4,15</sup>, **M. Khan Mirzaei**<sup>5,6</sup>, **B. L. Eberhart**<sup>2</sup>, **M. Valdivia-Garcia**<sup>7</sup>, **A. Metwaly**<sup>8</sup>, **K. Neuhaus**<sup>9</sup>, **G. Barker**<sup>7</sup>, **J. Ru**<sup>5,6</sup>, **L. T. Nesengani**<sup>10</sup>, **D. Mahdi-Joest**<sup>3</sup>, **A. S. Wilson**<sup>2</sup>, **S. K. Joni**<sup>11</sup>, **D. C. Layman**<sup>11</sup>, **J. Zheng**<sup>12</sup>, **R. Mandal**<sup>12</sup>, **Q. Chen**<sup>7</sup>, **M. R. Perez**<sup>2</sup>, **S. Fortuin**<sup>1</sup>, **B. Gaunt**<sup>13</sup>, **D. Wishart**<sup>12</sup>, **B. Methé**<sup>14</sup>, **D. Haller**<sup>4,8</sup>, **J. V. Li**<sup>7</sup>, **L. Deng**<sup>5,6</sup>, **R. Swart**<sup>11</sup> & **S. J. D. O'Keefe**<sup>1,2</sup>✉

<sup>1</sup>African Microbiome Institute, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. <sup>2</sup>Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. <sup>3</sup>Intestinal Microbiology Research Group, German Institute of Human Nutrition, Potsdam, Germany. <sup>4</sup>ZIEL - Institute for Food and Health, Technical University of Munich, Freising, Germany. <sup>5</sup>Institute of Virology, Helmholtz Centre Munich - German Research Centre for Environmental Health, Neuherberg, Germany. <sup>6</sup>Chair of Microbial Disease Prevention, School of Life Sciences, Technical University of Munich, Freising, Germany. <sup>7</sup>Section of Nutrition, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK. <sup>8</sup>Chair of Nutrition and Immunology, TUM School of Life Sciences, Technical University of Munich, Freising, Germany. <sup>9</sup>Core Facility Microbiome, ZIEL - Institute for Food and Health, Technical University of Munich, Freising, Germany. <sup>10</sup>Department of Agriculture and Animal Health, University of South Africa, Pretoria, South Africa. <sup>11</sup>Department of Nutrition and Dietetics, School of Public Health, University of the Western Cape, Cape Town, South Africa. <sup>12</sup>The Metabolomics Innovation Centre & Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada. <sup>13</sup>Zithulele Hospital, Mqanduli District, Mqanduli, Eastern Cape Province, South Africa. <sup>14</sup>Center for Medicine and the Microbiome, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. <sup>15</sup>These authors contributed equally: M. C. Ramaboli, S. Ocvirk. ✉ e-mail: [sjokeefe@pitt.edu](mailto:sjokeefe@pitt.edu)