



OPEN

A plasmid network from the gut microbiome of semi-isolated human groups reveals unique and shared metabolic and virulence traits

Liliane Costa Conteville[✉] & Ana Carolina Paulo Vicente

The plasmids in gut microbiomes have the potential to contribute to the microbiome community, as well as human health and physiology. Nevertheless, this niche remains poorly explored. In general, most microbiome studies focus on urban-industrialized groups, but here, we studied semi-isolated groups from South America and Africa, which would represent a link between ancestral and modern human groups. Based on open metagenomic data, we characterized the set of plasmids, including their genes and functions, from the gut microbiome of the Hadza, Matses, Tunapuco, and Yanomami, semi-isolated groups with a hunter, gather or subsistence lifestyle. Unique plasmid clusters and gene functions for each human group were identified. Moreover, a dozen plasmid clusters circulating in other niches worldwide are shared by these distinct groups. In addition, novel and unique plasmids harboring resistance (encompassing six antibiotic classes and multiple metals) and virulence (as type VI secretion systems) genes were identified. Functional analysis revealed pathways commonly associated with urban-industrialized groups, such as lipopolysaccharide biosynthesis that was characterized in the Hadza gut plasmids. These results demonstrate the richness of plasmids in semi-isolated human groups' gut microbiome, which represents an important source of information with biotechnological/pharmaceutical potential, but also on the spread of resistance/virulence genes to semi-isolated groups.

The human gut microbiome is a complex and dynamic ecosystem that performs several essential functions for human metabolism, nutrition, physiology, and immunity. Due to its high density and diversity of microorganisms, this environment is prone to constant genetic exchange between resident and transient bacteria¹⁻⁴. Genetic transfers by plasmids in the gut have the potential to impact the ecology and evolution of the gut microbes. The acquisition of a plasmid and its cargo genes by gut bacteria or archaea may assist the microorganism's survival, fitness, and adaptation to the environment⁴⁻⁷. Some traits provided by plasmids in the gut have been associated with antibiotic (AR) and metal resistance (MR)⁷⁻¹⁰, salt tolerance^{11,12}, and interbacterial competition¹³. Therefore, the set of plasmids in the human gut has the potential to modulate the gut microbiome community and introduce new traits, such as antibiotic resistance and virulence, to both resident and transient gut microbes⁴⁻¹⁴.

Many studies have assessed the microbial profile present in human gut microbiomes¹⁵⁻¹⁹. Although these studies have improved our knowledge on the subject, the set of plasmids in the gut microbiome is still poorly investigated despite its potential to modulate the microbiome physiology and, consequently, human health. Many human life aspects modulate the gut microbiome and its plasmids, including diet, lifestyle, therapy, consumption of pre- and probiotics, and diseases^{7,15-21}. These aspects have been dramatically transformed during human history, which is evidenced by the current plethora of diets and lifestyles around the globe. The gut microbiome composition and functionality are products of a long coevolutionary host-microbe relationship, and the human gut plasmids also suggestively reflect this coevolution. As an example of the diet impacting mobile elements from the human gut, the seaweed in the Japanese daily diet was proposed to have transferred algal polysaccharide degradation genes to the Japanese microbiome²⁰. The different diets of Fijians and Americans have been linked to the different glycoside hydrolase families encoded by mobile genes in the gut microbiome of these groups⁶.

Laboratório de Genética Molecular de Microrganismos, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brazil. ✉email: lilianeconteville@gmail.com

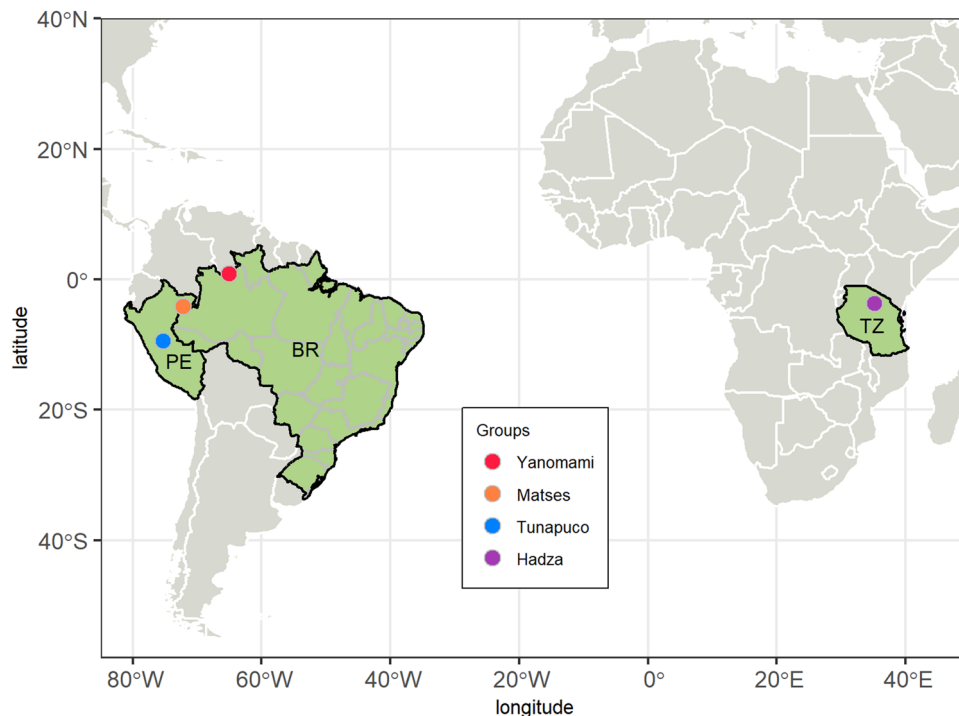


Figure 1. Geographic locations of the four semi-isolated groups. Map generated in R, version 4.0.3. (<https://www.R-project.org/>) with the packages `rnatrualearth`, `brazilmaps`, `sf` and `ggplot2`.

Moreover, the use of antimicrobials has been associated with the increase in the prevalence of mobile elements carrying functional resistance genes associated with synthetic antibiotics^{7–10,22–26}.

Most of the microbiome studies have been focusing on urban-industrialized groups, and few have considered semi-isolated human groups^{15–19,27}. However, the study of gut plasmids from semi-isolated groups has the potential to unravel genetic elements related to their unique lifestyle and environment. These groups retain valuable information about the microbiome before urbanization and industrialization impacted human diet and lifestyle, and also clues that can expand the field of prebiotics and probiotics for modern disorders prevention and treatment, as well as biomarkers that can form the basis for health and prognostic disease tests. So here, based on the gut microbiome survey from semi-isolated human groups, we investigated their plasmid content, including the set of accessory genes associated with these mobile elements that can modulate human health and physiology. We analyzed the set of plasmids and its accessory genes from the gut microbiome of four semi-isolated human groups: the Yanomami, the largest indigenous group from the Brazilian Amazon Forest¹⁹, the Matses, a remote hunter-gatherer group from the Peruvian Amazon Forest, the Tunapuco, a traditional agricultural community from the Peruvian Andean highlands¹⁶, and the Hadza, an indigenous ethnic group from northwestern Tanzania¹⁷. The geographical locations of these groups are presented in Fig. 1 and a summary of their diets and lifestyles is presented in Supplementary Note S1.

Results

We identified a total of 290 putative plasmids in the gut microbiome of the Yanomami, Matses, Tunapuco, and Hadza. These putative plasmids were identified based on the nucleotide similarity with sequences of replication initiation protein (*rep*), mobilization protein (relaxase), mate-pair formation (MPF), and origin of transfer (*oriT*). The Yanomami was the group with significantly more plasmids identified ($n = 184$, FDR-corrected Kruskal–Wallis multiple comparisons: $P < 0.05$), followed by the Hadza ($n = 52$), the Matses ($n = 36$), and the Tunapuco ($n = 18$). The number of plasmids per metagenome varied from zero to 17 in the Yanomami, Matses, and Hadza, but in the Tunapuco it varied from zero to seven plasmids. However, one Yanomami metagenome stands out for harboring 90 plasmids (STable 1).

The plasmids were classified as conjugative, mobilizable, and non-mobilizable based on three plasmid markers: relaxase, MPF, and *oriT*. Thus, among the 290 plasmids identified in this study, 8 (~2.7%) were classified as conjugative, for harboring both relaxase and MPF; 89 (~30.6%) as mobilizable, for harboring relaxase and/or *oriT* without MPF; and the majority ($n = 193$, ~66.5%) as non-mobilizable, for not presenting relaxase and *oriT*. Considering the plasmid length, most of them (79.6%) is less than 10 Kb long, with 20% having between 10 and 90 Kb, and the longest plasmid has 115 Kb. These are the lengths of the contigs characterized as plasmids and they may not represent the length of the original plasmid. Information regarding each plasmid identified can be found in STable 1.

Moreover, mobilization (MOB) typing was used to classify the relaxase and *oriT* sequences of the conjugative and mobilizable plasmids into six MOB types. MOBP was the most prevalent type, followed by MOBQ, and

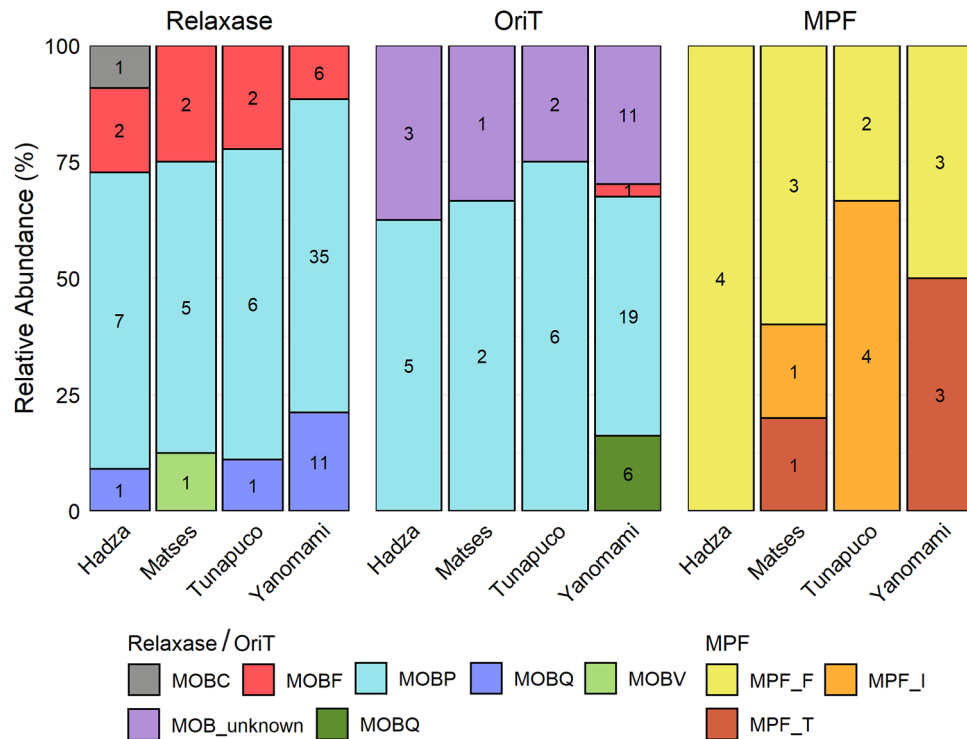


Figure 2. Stacked bar chart representing the relative abundance (y-axis) of the MOB families and MPF types of the relaxase, *oriT*, and MPF identified in the gut plasmidomes of each group. Raw counts are also shown.

MOBF (Fig. 2). Two other MOB families (MOBC and MOBV) were assigned to relaxases from the Hadza and Matses, respectively. MOBH was the only MOB family not identified. Besides, 16 *oriT* sequences did not match any of the six known MOB families and were classified as unknown. The MPF was also classified into classes and the most prevalent was MPFF, followed by MPFI and MPFT (Fig. 2).

We explored the protein's functions encoded by the gut plasmids of the groups. The most abundant protein categories were Signaling/Cellular Processes (28–53%) and Genetic Information Processing (13–23%) (Stables 2, 3). The majority of the proteins associated with Signaling/Cellular Processes are part of Secretion (IV and VI) and Prokaryotic Defense Systems. The majority of proteins categorized as associated with Genetic Information Processing are related to Replication and Repair (Stables 2, 3). Some relevant gene functions identified in the gut plasmids differed by group. Considering the four human groups, functions related to lipopolysaccharide biosynthesis were exclusive to the Hadza, plant-pathogen interaction, immune system, and siderophore biosynthesis were exclusive to the Matses, flagellar assembly and drug and xenobiotics metabolism were exclusive to the Yanomami. Moreover, genes associated with Carbohydrate Metabolism and Other Amino Acids were only identified in the Matses and Yanomami. We also observed that the genes that generate energy metabolize different nutrients according to each group: Sulfur, Nitrogen and Methane in the Hadza, Matses and Yanomami, respectively.

Considering Prokaryotic Defense Systems, we performed focused analyses on identifying antimicrobial and metal resistance, and virulence, which in some cases, can directly impact the host's health. AR and MR genes and virulence factors (VFs) were found in 27 plasmids identified in 19/72 metagenomes (Stable 1). The Hadza was the group in which the gut set of plasmids harbored the highest number of AR, MR, or VFs ($n = 25$), followed by the Matses ($n = 18$), the Yanomami ($n = 10$), and the Tunapuco ($n = 4$). Considering all AR genes identified, they are associated with six antibiotic classes: aminoglycosides, β -lactams, fluoroquinolones, fosfomicin, tetracyclines, and trimethoprim. The most abundant AR genes found were *aph* (aminoglycoside resistance), *tet* (tetracycline resistance), and *tem* (beta-lactamase). The MR genes are associated with copper, iron, mercury, and multiple metals resistance. Concerning the VFs, one plasmid presents a heat-stable enterotoxin (*astA*), but the other VFs are mainly yersiniabactin, which besides being a VF, is also associated with iron resistance. The gene *ybt*, which is a siderophore yersiniabactin, was identified in a cluster of eight *ybt* genes in one plasmid recovered from one metagenome from the Matses (Stable 1).

Moreover, we also identified 119 transposase sequences in 49 plasmids. Therefore, some plasmids have more than one transposase. The vast majority have one to three transposases, but a plasmid recovered from Tunapuco had seven. The transposases were classified into families and the most prevalent was IS3 ($n = 61$).

Based on the clustering classification, the plasmids were split into three groups: 1, 2, and 3. Group 1 comprises plasmids that were assigned to plasmid clusters of the MOB-suite program database, therefore known plasmids. While plasmids that exceeded the distance threshold with all references in the database were classified in Groups 2 and 3, and therefore novel plasmids could be part of these groups. The former comprises plasmids

ID	Mobility	Groups	ARGs	Locality/Continent	Samples/Hosts
A	Non-mobilizable	H/M/T	<i>dfjf</i> (Hadza)	Asia, Europe, N./S. America	Avian, Food, Human, Housefly, Cattle, Rat
B	Mobilizable	H/M	–	Eastern Asia, Europe, New Zealand, USA	Avian, Environment, Food, Human, Cattle
C	Non-mobilizable	H/M	–	China, Europe, Islas Malvinas, USA	Avian, Human, Housefly, Livestock, Rat
D	Non-mobilizable	H/M	–	Denmark, Islas Malvinas	Rat
E	Non-mobilizable	H/Y	–	Africa, Asia, Australia, Europe, USA	Environment, Food, Human, Cattle, Rat, Sewage
F	Non-mobilizable	H/Y	<i>tem</i> , <i>aph</i> (H) <i>tem</i> (Y)	China, France	Environment, Food
G	Mobilizable	H/Y	–	Africa, Asia, Australia, Europe, N. America	Avian, Environment, Human, Cattle
H	Mobilizable	H/Y	–	Africa, Australia, China, Europe, USA	Environment, Human, Swine
I	Mobilizable	H/Y	–	Europe, India, N./S. America	Food, Human
J	Mobilizable	H/Y	–	Australia, Asia, Europe, N/S. America	Avian, Canine, Environment, Human, Cattle, Sewage
K	Mobilizable	T/Y	–	Asia, Australia, Europe, USA	Avian, Environment, Human, Livestock, Mollusk
L	Mobilizable	T/Y	–	Africa, Asia, Australia, Europe, N. America	Avian, Canine, Environment, Food, Human, Livestock
M	Conjugative	M	<i>fosA</i>	Africa, Asia, Australia, Europe, USA	Avian, Canine, Human, Livestock, Panda
N	Mobilizable	M	<i>ctx</i>	Asia, Australia, Europe, N./S. America	Human
O	Conjugative	H	<i>aph</i> , <i>tet</i>	Asia, Europe, N. America	Avian, Canine, Human, Livestock
P	Conjugative	T	<i>ctx</i>	Asia, Europe, USA	Avian, Human
Q	Mobilizable	T	<i>tetB</i>	Asia, Australia, Europe, N. America	Human, Livestock
R	Non-mobilizable	Y	<i>tetU</i>	–	–
S	Mobilizable	Y	<i>tem</i>	Asia, Europe, Oceania, N./S. America	Avian, Canine, Human, Livestock
T	Mobilizable	Y	<i>qnrB</i>	Asia, Europe, N./S. America	Avian, Human

Table 1. Mobility prediction and antibiotic resistance genes of the plasmid clusters A–T from the network. The hosts and niches shown refer to the plasmids that had high similarity to this cluster by blast analysis. Groups: H—Hadza; M—Matses, T—Tunapuco, Y—Yanomami.

shared among the metagenomes and were clustered, and the latter comprises unique plasmids. The plasmids of Group 1 (n = 96) were grouped into 63 clusters, Group 2 (n = 65) into 22 clusters, while the remaining 129 were singletons (Group 3).

To infer the plasmids shared among the studied human groups, we generated a network based on the plasmid clusters characterized by MOB-suite (Fig. 2). The network revealed that 12/85 clusters (Group 1, n = 9; and Group 2, n = 3) are shared among the human groups, mostly between the Hadza and the Yanomami. Moreover, to expand our scenario, we also performed BLAST analysis on the plasmids shared among the groups. In this way, we found that the set of plasmids in the semi-isolated groups' gut microbiome has a high similarity with plasmids spread worldwide in bacteria from humans, animals, environments, and foods (Table 1, Stable 4). Even Group 2 clusters, potentially comprising novel plasmids, presented similarities with known plasmids. However, the BLAST analysis of the plasmid cluster "R" returned no matches, indicating the presence of novel plasmids in this cluster.

Surprisingly, the plasmids from the clusters "A", "E", and "C" (Fig. 3), previously characterized as non-mobilizable (Stable 1), were the clusters identified in a higher number of metagenomes. The cluster "A" was the most prevalent in the metagenomes analyzed, present in the Hadza (n = 3), Matses (n = 3), and Tunapuco (n = 1). The same *rep* and *TolA* sequences characterized the plasmids from cluster "A", including the Hadza plasmids (n = 3), which also harbor a *dfjf* (dihydrofolate reductase) trimethoprim resistance gene. Besides that, this cluster was inferred by the MOB-suite as potentially hosted by bacteria of the genus *Campylobacter*. The clusters "E" and "C" harbor known plasmids, with no relevant functions so far assigned. The mobilizable cluster "F" shared between the Hadza and Yanomami carries beta-lactamase resistance genes, besides some Hadza's plasmids also carry *aph* (aminoglycoside resistance). Conjugative plasmids carrying AR genes were identified in cluster "M" (*fosA* gene), cluster "O" (*aph* and *tet* genes), and cluster "P" (*ctx* gene) (Fig. 3).

Besides the plasmids shared among the human groups, we also observed plasmids with the potential to be human group-specific as some from Group 2 clusters (n = 8) identified only in metagenomes from Hadza (n = 2) and Yanomami (n = 6). The Yanomami-specific clusters were classified as mobilizable (n = 4), but the others and Hadza's plasmids were non-mobilizable. These "human group-specific" plasmid clusters carry genes associated with Metabolism, mostly related to Glycolysis/Gluconeogenesis, and also genes that encode Transport Proteins and Transposases.

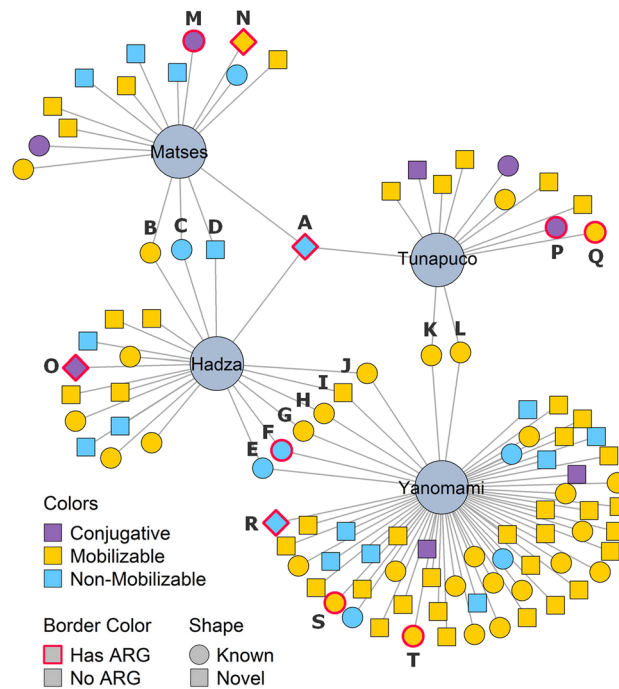


Figure 3. Network based on the plasmid clusters (smaller-sized nodes) harbored and shared between the human groups (bigger-sized nodes). Non-mobilizable plasmids that were identified in only one metagenome were removed from this analysis. Identification of the plasmid clusters A-T is also shown. Network generated in R, version 4.0.3. (<https://www.R-project.org/>) with the packages igraph and ggplot2.

Discussion

Diet and lifestyle have changed dramatically throughout human evolution, first with the introduction of farming (~ 10,000 years ago), and more recently, with the introduction of industrially processed foods, hygienic conditions, antimicrobials, pollution, and other factors^{27,28}. These transformations in human lifestyle have been widely associated with the selection and modulation of the gut microbes and with a substantial loss of gut bacterial diversity^{15–19,27}. However, this microbial profile restructuring not only involves the microbial taxa but, responding to selective pressures, also the mobilome⁴. Moreover, modern diseases, such as diabetes, obesity, and allergies, partially driven by the lifestyle, have been associated with the loss and variation of bacterial taxa and mobile elements diversity in the gut microbiome^{28,29}.

In the present study, we analyzed the set of plasmids and their accessory genes from the gut microbiome of four semi-isolated human groups: the Yanomami from the Brazilian Amazon Forest¹⁹, the Matses from the Peruvian Amazon Forest, the Tunapuco from the Peruvian Andean highlands¹⁶, and the Hadza from northwestern Tanzania¹⁷. These datasets are some of the few high-quality shotgun metagenomes available from hunter-gatherers, human groups whose gut microbiomes represent a link between the ancestral and modern human groups. The bacterial composition and functionality of these four human groups have been previously analyzed, and although they have a similar contrasting composition compared to urban-industrialized groups, each of them has particular characteristics related to their different ecological niches, genetics, and diets^{15–19}. The gut plasmids analysis showed a similar pattern, in which specific plasmid clusters and gene functions are unique for each human group, yet they also share significant parts of their set of plasmids. Moreover, a dozen of plasmid clusters are shared among these semi-isolated human groups living in distant and distinct geographical niches. In addition, some of these shared plasmid clusters have also been detected worldwide in humans, animals, environments, and foods. This could indicate that these plasmids are part of our ancestral microbiome and that they remain in the human population to this day, but could also reveal that the global inter-niche exchange of plasmids eventually carried out by modern practices could be reaching distinct semi-isolated human groups and their environments.

In general, plasmids can be classified by their relaxase and *oriT*, which comprise six MOB families, and MPF, which comprises four families^{23,30}. Based on a database of plasmid markers, the gut plasmids of our study were mostly characterized by the families MOBP, MOBQ, and MOBF. The MOBP, a cluster of actively evolving relaxases, is the most prevalent family in plasmids databases, featuring a considerable proportion of short plasmids^{30,31}. Indeed, most of the plasmids that we identified are short plasmids (< 10 Kb). Moreover, the MOBC and MOBV families, low prevalent families in plasmid databases, were identified in two novel plasmids of the Hadza and Matses. This highlights the yet underexplored scenario that modulates the plasmids and diversity in the gut microbiome^{30,31}. Concerning the quantitative aspect, the Yanomami presented the largest set of plasmids in their gut and the Tunapuco the smallest, likely reflecting their remarkably different diet, lifestyle, and environment. Although both groups inhabit the Amazon Region, the Yanomami's subsistence is based on gathering and hunting

wild plants and animals in the Amazon Forest, the world's largest and most biodiverse tropical rainforest^{19,32,33}. In contrast, Tunapuco's subsistence is based on local agricultural produce and homegrown small animals located in the Peruvian Andean highlands, at an elevation between 2500 and 3100 m above sea level^{16,32,33}.

The gut microbes usually process undigested carbohydrates, proteins, and fat, producing additional energy from the diet and also metabolites that can impact human physiology³⁴. These metabolic functions, encoded by chromosomes and plasmids of the gut microbes, are selected/modulated according to the components of the human's host diet^{7,20,34,35}. In these groups, the food seasonality and their eventual nomadic behavior are selection pressures that can drive a variety of changes in the gut throughout the year, including in the set of plasmids. So, to rapidly adapt to the new/seasonal dietary components, specific enzymes must be selected and spread among the gut microbes^{34,35}. This may explain why only the groups that inhabit the Amazonian Forest (Yanomami and Matses) have plasmid accessory genes associated with Carbohydrate and Other Amino Acids Metabolism. These groups live in a hotspot of diversity, therefore they are always coming into contact with various dietary components, which must be metabolized^{32–35}. Moreover, their niches and consequently, the distinct selective pressures, contribute to the selection of plasmids' content that participates in the metabolism of molecules from their diet. The functions identified as group-exclusive follow this context. For example, only in the Hadza set of plasmids was observed the lipopolysaccharide (LPS) biosynthesis pathway. This pathway is known for inducing strong inflammatory responses, but contrastingly, it also has anti-inflammatory actions when produced by certain gut bacteria, such as *Bacteroidetes*. Indeed, the Hadza harbor this phylum in their gut microbiomes^{15,17}. LPS has been proposed to be used in the prevention or treatment of chronic inflammatory diseases that are induced by pro-inflammatory LPS³⁶. The gut accessory genome, represented by the gut plasmids, constitutes an underexplored plethora of genes and functions and an important source of information with potential biotechnological or pharmaceutical value^{37,38}.

The prevalent gene functions that were identified in the gut plasmids of all groups, such as Prokaryotic Defense Systems, Secretion Systems, and Replication and Repair, are associated with core plasmid genes^{39,40}. However, the type VI Secretion Systems (T6SS) is not plasmid ubiquitous, moreover, it is associated with inter-bacterial competition, bacteria-host interactions, and virulence^{41–43}. Therefore, if the T6SS is harbored by a commensal microbe, it can prevent the microbiome disassembly by an invading bacteria. In contrast, if T6SS is carried by an invading bacteria, the community assembly can be disestablished, resulting in dysbiosis^{41–43}.

Interestingly, we observed a resistome in the gut plasmids of these human groups that have no or limited exposure to synthetic antibiotics, as well as reported in other isolated or semi-isolated groups/locations^{7,24,44}. Most of these AR genes are associated with β -lactam and tetracycline resistance, which are commonly seen in soil and water environments⁴⁵. In fact, Hadza and Yanomami gut plasmids harbor accessory genes with the function of biosynthesis of Penicillin and Cephalosporin. Therefore, the AR genes in these semi-isolated groups may have originated from an environment where antibiotic-producing microbes naturally occur, and also may have reached these groups via contact with antibiotic-exposed populations^{22,24,46,47}. Since the Hadza and Matses presented the highest numbers of resistance and virulence genes in their gut plasmids, we speculate that these groups have been more impacted by urban-industrialized lifestyles, in a way that may be favoring the selection and expansion of these genes. A previous study with the Hadza, in fact, showed the existence of antibiotic resistance genes, in this population that has little to no antibiotic exposure, even though they have a lower amount of AR genes in comparison with an urban-industrialized population¹⁷.

In the same way, metal resistance genes present in the gut plasmids can reflect bacterial adaptations to environmental conditions. Indeed, the Yanomami niche is likely to be contaminated with cadmium due to the continuous discharge of batteries by them over decades¹⁹, however, the natural or artificial presence of metal in the other group's environment is unknown. Moreover, MR genes can impact the community assembly of the gut microbiomes and consequently affect host health⁴⁸. Concerning the virulome, there is no prevalence of bacterial virulence-associated factors in the gut plasmids, with the expectation of T6SS. In fact, these factors are not prevalent within environmental bacteria but occur more often in pathogens in clinical settings⁴⁹. In addition, the plasmids recovered also harbor transposases, mainly the IS3 family, which is widely distributed and one of the largest and most abundant Insertion Sequence families among plasmids⁵⁰.

The gut plasmids recovered and characterized here represent an achievement in the study of the gut microbiome of semi-isolated groups, providing a snapshot of the forces that modulate the gut microbiome given the unique as well as global scenarios to which these human groups are exposed. Although we have focused only on plasmids from the human gut, other mobile elements, such as bacteriophages, must also be relevant in this context. Beyond that, we are aware that our study has biases and limitations, for example, the plasmid identification that is based on the current plasmid markers in the databases. However, different studies have successfully carried out bioinformatic analysis of sequences and addressed them to plasmids, mainly when the plasmids were in vitro isolated from the strains^{51,52}. The semi-isolated groups may harbor plasmid markers that have not yet been characterized and this may have resulted in the general lack of relaxase, MPF, and oriT sequences identified. As a result, most plasmids were classified as novel and non-mobilizable. Indeed, new MOB families have often been reported^{30,53–57}. Besides, the complexity of metagenomic assembly tends to produce shorter contigs and may have hampered the identification of the whole genetic content of the plasmids⁵⁸. In addition, the largest number of genes found in the plasmids have no known functionality. Indeed, the set of plasmids from the gut microbiome is a recognized source of novel genes and gene products, so the presence of unknown or novel elements is expected³⁸. However, these results highlight the knowledge gap in this field and in databases, which is often attributed to the difficulties in studying plasmids from complex environments, such as microbiomes^{40,59}.

Methods

For this study, we analyzed shotgun metagenomic data from previously published gut microbiome studies of semi-isolated human groups: semi-isolated hunter-gatherer communities from South America and Africa (Yanomami from Brazilian Amazon¹⁹, n = 15; Matses from Peruvian Amazon¹⁶, n = 24; and Hadza from Tanzania¹⁷, n = 27), and a rural agriculturalist community from the Andean highlands in Peru (Tunapuco¹⁶, n = 12). These datasets comprise paired-end sequences generated on Illumina sequencers, and bioinformatic processing was performed in parallel. All methods were performed in accordance with the relevant guidelines and regulations. The diet of the hunter-gatherer communities is mainly based on the consumption of highly fibrous tubers and vegetal foods, while the diet of the rural agriculturalist community is based on the local agricultural produce of stem and root tubers, and homegrown small animals.

Each metagenome was independently subjected to de novo assembly through metaSPAdes v.3.13⁶⁰ using default parameters. For plasmid identification, MOB-recon and MOB-typer modules from the MOB-suite program were used⁶¹. These modules identified putative plasmids from the contigs of each metagenome and searched for replication gene (*rep*), mobilization protein (relaxase), mate-pair formation (MPF), and the origin of transfer (*oriT*). Based on the presence/absence of these plasmid markers, the plasmids' mobility was predicted⁶¹. Optional tools for the classification of potential circular or linear plasmids were not performed and it is reasonable to expect the recovery of both types of plasmids. To avoid false positives we removed from the analysis the plasmids with no markers identified by the MOB-suite. For characterization, pairwise genomic distances between each plasmid identified and each reference plasmid cluster was calculated. Thus, each plasmid identified was assigned to their closest reference plasmid cluster or as a novel plasmid depending on the predicted genomic distance, using the MOB-suite database as reference⁶¹. Following the MOB-suite analysis, the function "dist" of the Mash program⁶² was used to calculate the pairwise genomic distances between all novel plasmids identified in the groups. If the distance between the novel plasmids was lower than 0.05, they were assigned to the same novel cluster. This distance metric allows the clustering of plasmids with considerable differences in length, which is useful in this case, since plasmids can undergo diverse changes in their sequence content⁶¹. Based on the clustering classification, the plasmids were split into three groups. Group 1 comprises plasmids that were assigned to plasmid clusters from the reference database. Group 2 comprises plasmids assigned as novel and clustered into novel clusters. Group 3 comprises unique plasmids, with no significant similarity/coverage to the plasmids in the database or to other plasmids found in the metagenomes. Moreover, only plasmid clusters identified in at least three metagenomes from one same human group were considered with the potential to be group-specific.

We also used FragGeneScan v.1.31⁶³ to predict Open Reading Frames (ORFs) in the contigs assigned as plasmids. Thereafter, Abricate (<https://github.com/tseemann/abricate>) was used to screen the ORFs against the MegaRes v.2⁶⁴ and VFDB⁶⁵ databases. Abricate results were filtered for a minimum DNA identity of 75% and minimum coverage of 50%⁶⁶. Functional annotation and assignment of KEGG functional categories to the ORFs were performed using the KOfamKOALA tool⁶⁷, considering a threshold of 0.01. The plasmids were also annotated using Prokka⁶⁸ for transposase identification.

Network analysis of the plasmids identified in the four groups was performed based on the plasmid clusters characterized by MOB-suite. The network was constructed with the igraph R package (<http://igraph.org/r/>). In this network, the smaller-sized nodes represent the plasmids, which are linked to the bigger-sized nodes that represent the groups that harbor these plasmids. We removed from this analysis non-mobilizable plasmids that were identified in only one metagenome.

In order to investigate which other hosts and niches these plasmids have been identified, we performed a search analysis on each plasmid using the online BLAST tool⁶⁹ with the nt/nr database. We removed results with similarity and coverage less than 70% and e-value less than 1e-5. To recover metadata information from the accession numbers of the plasmids that matched our putative plasmids, we used a custom python script based on the Entrez module of the Biopython package.

Data availability

The metagenomic data of the gut microbiomes used in this study were downloaded from the National Center for Biotechnology Information Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) BioProject PRJNA268964 for Reference 16, BioProject PRJNA278393 for Reference 17, and BioProject PRJNA527208 for Reference 19. The putative plasmids generated and analyzed during the current study are available from the corresponding author on reasonable request.

Received: 20 January 2022; Accepted: 8 July 2022

Published online: 15 July 2022

References

- Dionisio, F., Matic, I., Radman, M., Rodrigues, O. R. & Taddei, F. Plasmids spread very fast in heterogeneous bacterial communities. *Genetics* **162**(4), 1525–1532 (2002).
- Liu, L. *et al.* The human microbiome: A hot spot of microbial horizontal gene transfer. *Genomics* **100**(5), 265–270 (2012).
- Coyne, M. J., Zitomersky, N. L., McGuire, A. M., Earl, A. M. & Comstock, L. E. Evidence of extensive DNA transfer between bacteroidales species within the human gut. *MBio* **5**, e01305-e1314 (2014).
- Lerner, A., Matthias, T. & Aminov, R. Potential effects of horizontal gene exchange in the human gut. *Front. Immunol.* **8**, 1630 (2017).
- Licht, T. R. & Wilcks, A. Conjugative gene transfer in the gastrointestinal environment. *Adv. Appl. Microbiol.* **58C**, 77–95 (2005).
- Broaders, E., Gahan, C. G. & Marchesi, J. R. Mobile genetic elements of the human gastrointestinal tract: potential for spread of antibiotic resistance genes. *Gut Microbes* **4**(4), 271–280 (2013).
- Brito, I. L. *et al.* Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**(7612), 435–439 (2016).

8. Gumpert, H. *et al.* Transfer and persistence of a multi-drug resistance plasmid in situ of the infant gut microbiota in the absence of antibiotic treatment. *Front. Microbiol.* **26**(8), 1852 (2017).
9. Wein, T., Hulter, N. F., Mizrahi, I. & Dagan, T. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nat. Commun.* **10**, 2595 (2019).
10. Tepekule, B., Abel Zur Wiesch, P., Kouyos, R. D. & Bonhoeffer, S. Quantifying the impact of treatment history on plasmid-mediated resistance evolution in human gut microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **116**(46), 23106–23116 (2019).
11. Culligan, E. P., Sleator, R. D., Marchesi, J. R. & Hill, C. Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome. *ISME J.* **6**(10), 1916–1925 (2012).
12. Broaders, E., O'Brien, C., Gahan, C. G. & Marchesi, J. R. Evidence for plasmid-mediated salt tolerance in the human gut microbiome and potential mechanisms. *FEMS Microbiol. Ecol.* **92**(3), 19 (2016).
13. Hols, P., Ledesma-García, L., Gabant, P. & Mignolet, J. Mobilization of microbiota commensals and their bacteriocins for therapeutics. *Trends Microbiol.* **27**(8), 690–702 (2019).
14. Sarowska, J. *et al.* Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: Recent reports. *Gut Pathog.* **21**(11), 10 (2019).
15. Schnorr, S. L. *et al.* Gut microbiome of the Hadza hunter-gatherers. *Nat. Commun.* **15**(5), 3654 (2014).
16. Obregon-Tito, A. J. *et al.* Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.* **25**(6), 6505 (2015).
17. Rampelli, S. *et al.* Metagenome sequencing of the hadza hunter-gatherer gut microbiota. *Curr. Biol.* **25**(13), 1682–1693 (2015).
18. Mancabelli, L. *et al.* Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ. Microbiol.* **19**(4), 1379–1390 (2017).
19. Contevelle, L. C., Oliveira-Ferreira, J. & Vicente, A. C. P. Gut Microbiome biomarkers and functional diversity within an amazonian semi-nomadic hunter-gatherer group. *Front. Microbiol.* **30**(10), 1743 (2019).
20. Hehemann, J. H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**(7290), 908–912 (2010).
21. Ravi, A. *et al.* Association of the gut microbiota mobilome with hospital location and birth weight in preterm infants. *Pediatr. Res.* **82**(5), 829–838 (2017).
22. D'costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**(7365), 457 (2011).
23. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. & de la Cruz, F. Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* **74**(3), 434–452 (2010).
24. Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Sci Adv.* **1**(3), e1500183 (2015).
25. Pehrsson, E. C. *et al.* Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**(7602), 212–216 (2016).
26. Smits, S. A. *et al.* Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* **357**(6353), 802–806 (2017).
27. Gupta, V. K., Paul, S. & Dutta, C. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front. Microbiol.* **23**(8), 1162 (2017).
28. Zuo, T., Kamm, M. A., Colombel, J. F. & Ng, S. C. Urbanization and the gut microbiota in health and inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **15**(7), 440–452 (2018).
29. Levy, M., Kolodziejczyk, A. A., Thaiss, C. A. & Elinav, E. Dysbiosis and the immune system. *Nat. Rev. Immunol.* **17**(4), 219–232 (2017).
30. Garcillán-Barcia, M. P., Francia, M. V. & de la Cruz, F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **33**(3), 657–687 (2009).
31. Douarre, P. E., Mallet, L., Radomski, N., Felten, A. & Mistou, M. Y. Analysis of COMPASS, a new comprehensive plasmid database revealed prevalence of multireplicon and extensive diversity of IncF plasmids. *Front. Microbiol.* **24**(11), 483 (2020).
32. Sioli, H. *The Amazon and its Main Affluents: Hydrography, Morphology of the River Courses, and River Types* 127–165 (Springer, New York, 1984).
33. Verissimo A, Capobianco JPR. Biodiversidade na Amazônia Brasileira: Avaliação E Ações Prioritárias Para a Conservação, Uso Sustentável e Repartição de Benefícios. Instituto Socioambiental. (2001). Available at: https://books.google.com.br/books/about/Biodiversidade_na_amazonia_brasileira.html?id=VrhcAAAAMAAJ&redir_esc=y Accessed from 4 May 2021.
34. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their impact on host health. *Microbiome* **7**(1), 91 (2019).
35. Fragiadakis, G. K. *et al.* Links between environment, diet, and the hunter-gatherer microbiome. *Gut Microbes* **10**(2), 216–227 (2019).
36. Lin, T. L. *et al.* Like cures like: Pharmacological activity of anti-inflammatory lipopolysaccharides from gut microbiome. *Front. Pharmacol.* **30**(11), 554 (2020).
37. Ogilvie, L. A., Firouzmand, S. & Jones, B. V. Evolutionary, ecological and biotechnological perspectives on plasmids resident in the human gut mobile metagenome. *Bioeng. Bugs* **3**(1), 13–31 (2012).
38. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**(1), 105–114 (2021).
39. Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**(4), 169–181 (2007).
40. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**(7376), 241–244 (2011).
41. Logan, S. L. *et al.* The *Vibrio cholerae* type VI secretion system can modulate host intestinal mechanics to displace gut bacterial symbionts. *Proc. Natl. Acad. Sci. U. S. A.* **115**(16), E3779–E3787 (2018).
42. Peñil-Celis, A. & Garcillán-Barcia, M. P. Crosstalk between type VI secretion system and mobile genetic elements. *Front. Mol. Biosci.* **13**(6), 126 (2019).
43. Custodio, R. *et al.* Type VI secretion system killing by commensal *Neisseria* is influenced by expression of type four pili. *Elife* **7**(10), e63755 (2021).
44. Perez, M. F. *et al.* First report on the plasmidome from a high-altitude lake of the andean puna. *Front. Microbiol.* **11**, 1343 (2020).
45. Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**(1), 207–216 (2015).
46. Benveniste, R. & Davies, J. Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **70**(8), 2276–2280 (1973).
47. Forsberg, K. J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**(6098), 1107–1111 (2012).
48. Chiu, K., Warner, G., Nowak, R. A., Flaws, J. A. & Mei, W. The impact of environmental chemicals on the gut microbiome. *Toxicol. Sci.* **176**(2), 253–284 (2020).
49. Leitão, J. H. Microbial virulence factors. *Int. J. Mol. Sci.* **21**(15), 5320 (2020).
50. Chandler, M., Fayet, O., Rousseau, P., Ton Hoang, B. & Duval-Valentin, G. Copy-out–paste-in transposition of IS 911: A major transposition pathway. *Microbiol. Spectr.* **3**(4), 3–4 (2015).

51. Dumas, E. *et al.* Mycobacterial pan-genome analysis suggests important role of Plasmids in the radiation of type VII secretion systems. *Genome Biol. Evol.* **8**(2), 387–402 (2016).
52. Morgado, S. M. & Paulo Vicente, A. C. Genomics of Atlantic Forest Mycobacteriaceae strains unravels a mobilome diversity with a novel integrative conjugative element and plasmids harbouring T7SS. *Microb. Genom.* **6**(7), mgen000382 (2020).
53. Wisniewski, J. A. *et al.* TcpM: A novel relaxase that mediates transfer of large conjugative plasmids from *Clostridium perfringens*. *Mol. Microbiol.* **99**(5), 884–896 (2016).
54. Coluzzi, C. *et al.* A glimpse into the world of integrative and mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. *Front. Microbiol.* **20**(8), 443 (2017).
55. Ramachandran, G. *et al.* Discovery of a new family of relaxases in Firmicutes bacteria. *PLoS Genet.* **13**(2), e1006586 (2017).
56. Guzmán-Herrador, D. L. & Llosa, M. The secret life of conjugative relaxases. *Plasmid* **104**, 102415 (2019).
57. Soler, N. *et al.* Characterization of a relaxase belonging to the MOB family, a widespread family in Firmicutes mediating the transfer of ICEs. *Mob. DNA* **10**, 18 (2019).
58. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**(7285), 59–65 (2010).
59. Acman, M., van Dorp, L., Santini, J. M. & Balloux, F. Large-scale network analysis captures biological features of bacterial plasmids. *Nat. Commun.* **11**(1), 2452 (2020).
60. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27**(5), 824–834 (2017).
61. Robertson, J. & Nash, J. H. E. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.* **4**(8), e000206 (2018).
62. Ondov, B. D. *et al.* Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**(1), 132 (2016).
63. Rho, M., Tang, H. & Ye, Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**(20), e191 (2010).
64. Doster, E. *et al.* MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res.* **48**(D1), D561–D569 (2020).
65. Chen, L. *et al.* VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* **33**, D325–D328 (2005).
66. Carroll, L. M. *et al.* Whole-genome sequencing of drug-resistant salmonella enterica isolates from dairy cattle and humans in New York and Washington States reveals source and geographic associations. *Appl. Environ. Microbiol.* **83**(12), e00140–e217 (2017).
67. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**(7), 2251–2252 (2020).
68. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**(14), 2068–2069 (2014).
69. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinf.* **15**(10), 421 (2009).

Acknowledgements

We thank the support given by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [Finance Code 001]; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ); and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Author contributions

Conceptualization: L.C. and A.V.; Methodology, Formal analysis and Investigation: L.C.; Supervision: A.V.; Writing: L.C. and A.V.

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/s41598-022-16392-z>.

Correspondence and requests for materials should be addressed to L.C.C.

Reprints and permissions information is available at 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) 2022