Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Divergent climatic, geographic, and taxonomic associations of human gut resistome burden and composition across 34 countries

Maja Mikolás
University of Debrecen
Péter Dávid
University of Debrecen
Anna Szilágyi-Rácz
University of Debrecen
Emese Szilágyi-Tolnai
University of Debrecen
Péter Fauszt
University of Debrecen
Amal Alsayahien
University of Debrecen
Ibrahim Abousharabia
Aalto University
Njomza Gashi
University of Debrecen
Zsombor Szabó
University of Debrecen
Trinh Pham
University of Turku and Åbo Akademi University
Sándor Bíró
University of Debrecen
László Stündl
University of Debrecen
Judit Remenyik
University of Debrecen
Melinda Paholcsek
paholcsek.melinda@agr.unideb.hu

University of Debrecen

Research Article

Keywords: antimicrobial resistance, gut resistome, gut microbiome, environmental exposure, Human Development Index (HDI), cross-country comparison, global health

Posted Date: October 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11305142/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Understanding how antimicrobial resistance (AMR) relates to gut microbial community structure is essential for interpreting the human gut resistome in its ecological context.

Methodology

We investigated the relationships between resistance burden, resistance-class composition, and microbial taxonomic profiles in 98 healthy young adults from 34 countries, recruited among first-year medical students. Country-level pooled shotgun metagenomics was integrated with individual-level 16S rRNA gene sequencing.

Results

At the country-pool level, clustering of resistance-class profiles identified three clusters (RCPC1-RCPC3). RCPC1 combined the highest AMR burden with greater relative abundance of taxa reported as potentially pathogenic, genotoxic, or aerotolerant. This cluster also showed a higher representation of putative short-chain fatty acid-producing taxa, indicating that elevated resistance burden can coexist with taxa associated with beneficial metabolic functions. At the individual level, clustering of 16S-based predicted metabolic pathways identified four clusters (PMC1-PMC4) sharing a broadly conserved functional background.

Country-level analyses further distinguished factors associated with resistance abundance from those associated with resistance-class composition. Mean annual temperature showed an exploratory positive association with AMR abundance, corresponding to approximately 22% higher abundance per 10°C increase after adjustment for continental grouping and pool size. PM2.5 exposure showed a threshold-like pattern across WHO-defined tiers. Neither environmental variable was associated with resistance-class composition, whereas countries assigned to the Asian group showed enrichment of macrolide-lincosamide-streptogramin resistance.

Conclusion

Resistance burden and composition capture complementary features of the gut resistome and should be interpreted alongside microbial community structure. Higher resistance burden can coexist with putatively beneficial taxa, highlighting the limits of single-metric assessments and the potential value of integrated resistome-taxonomic profiling for One Health surveillance.

Background

Antimicrobial resistance has emerged as one of the defining global public health challenges of the 21st century [1]. In 2019, an estimated 1.27 million deaths were directly attributable to bacterial AMR, with an

additional 4.95 million deaths associated with drug-resistant infections [1]. Projections suggest that without concerted global action, AMR-related mortality could rise substantially in the coming decades, with the burden disproportionately concentrated in low- and middle-income countries (LMICs), which currently account for up to 90% of global AMR deaths [2]. Effective mitigation strategies depend on a mechanistic understanding of the factors that drive resistance emergence, accumulation, and spread across diverse global populations.

The human gut microbiome represents one of the largest reservoirs of antimicrobial resistance genes (ARGs) in the human body and serves as both a repository and a potential source of clinically relevant resistance determinants [3]. The composition and diversity of the gut microbiome are shaped by multiple factors including diet, environment, geography, antibiotic exposure, and host genetics [4]. Landmark international sequencing initiatives such as the Human Microbiome Project and the Metagenomes of the Human Intestinal Tract (MetaHIT) project have provided foundational insight into gut microbial composition across populations, revealing that lifestyle and geography are among the most consistent structuring factors [5].

Beyond geographical variation, dietary patterns and lifestyle transitions have been recognised as powerful modulators of gut microbial composition [5]. Over recent decades, gut microbiome composition in westernised populations has changed substantially, driven by shifts including frequent antibiotic and pharmaceutical use, industrialised sanitation practices, and reduced intake of dietary microbiota-accessible carbohydrates as processed foods replace traditional whole-food diets [5]. Traditional and hunter-gatherer populations retain markedly higher microbial diversity than industrialised populations, and transitional societies exhibit intermediate patterns that reflect their nutritional and environmental context [5, 6].

Environmental and climatic factors have increasingly been recognised as important amplifiers of AMR. Temperature affects fundamental bacterial processes, including growth rates and horizontal gene transfer, the primary mechanism by which resistance genes are exchanged between bacterial populations [7, 8]. Local temperature and population density have both been associated with elevated antibiotic resistance rates in clinical pathogens [7], and climate-driven ecological changes, including altered precipitation regimes, flooding, contamination of water systems, and expansion of vector habitats, are projected to further amplify AMR transmission and burden [9]. Ambient air pollution, particularly Particulate Matter 2.5 (PM2.5), has been shown to harbour resistant bacteria and ARGs and to correlate with clinical resistance rates in spatial-temporal analyses [10]. These findings position climate change and environmental exposure as global-scale amplifiers of AMR that operate independently of, and additively to, direct antibiotic use.

Despite growing evidence that both anthropogenic factors and environmental factors contribute to AMR, few studies have integrated these dimensions simultaneously across a globally diverse cohort using human gut metagenomics as a direct biological readout. Most existing global AMR surveillance relies

either on clinical isolate surveys, which capture only symptomatic infections in medical settings, or on sewage-based metagenomics, which integrate exposure at the population level but obscure inter-individual variation. Direct characterisation of the human gut resistome across countries with diverse climatic and socioeconomic remains underexplored, particularly for adult populations and for LMICs, where the AMR burden is greatest.

In the present study, we characterise the human gut resistome and its microbial correlates across 34 countries spanning four continents, using country-level pooled shotgun metagenomic sequencing of stool samples collected from a deliberately homogeneous population of young adult medical students newly enrolled at the University of Debrecen in Hungary, complemented by individual-level 16S rRNA gene sequencing in a subset of participants. This sampling frame is a distinctive feature of our design: recruiting an internationally diverse but socioeconomically and demographically comparable cohort within a short residence window reduces some of the within-country heterogeneity that typically complicates cross-country comparisons, although country-of-origin exposures could not be fully disentangled from more recent, shared exposures following relocation.

By integrating country-level environmental, socioeconomic, and veterinary antimicrobial-use data with shotgun metagenomic profiles, we sought to address three questions: (i) which climatic and environmental factors are associated with total resistance gene abundance; (ii) which factors are associated with the compositional structure of the resistome; and (iii) whether regional resistance-class signatures are consistent with known patterns of veterinary antimicrobial use. We further asked whether these resistance signatures could be interpreted in relation to the surrounding microbial community, rather than as an isolated gene pool, given that resistance determinants are carried within, and potentially shaped by, that community. To this end, we characterised interindividual variation in predicted metabolic function and its taxonomic correlates, together with country-level variation in resistance-class profiles and their associated taxonomic signatures. Through this integrative approach, we aimed to provide a framework for interpreting global heterogeneity in the gut resistome and to identify patterns that may inform future, more targeted One Health surveillance and stewardship research.

Results

Study cohort

Our cohort comprised 98 healthy young adult medical students originating from 34 countries, with a mean age of 22.3 ± 3.3 years. All participants were enrolled at the University of Debrecen, Hungary, voluntarily agreed to participate in the study, and provided informed consent. Eligibility criteria for participants contributing individual faecal samples included no reported antibiotic use within the three months preceding sample collection and no known gastrointestinal infection.

Two complementary sequencing approaches were used: individual faecal samples were analysed using 16S rRNA gene sequencing, whereas samples pooled by country of origin underwent shotgun

metagenomic sequencing. This pooling strategy generated 34 country-level metagenomic datasets, which served as the primary units of analysis for the shotgun metagenomic comparisons. For geographical comparisons, countries were assigned to four continental groups: Asia (17 countries), Africa (9 countries), Europe (5 countries), and the Americas (3 countries) (**Supplementary Table 1**). These categories are hereafter referred to as continental groups for analytical convenience.

Predicted microbial metabolic profiles and associated genus-level variation across individual faecal samples

Microbial functional potential was inferred from 16S rRNA gene sequencing data for 98 individually analysed faecal samples across 27 metabolic pathways selected for their literature-supported relevance to microbial metabolism and host-microbiome interactions in our study population (**Figure 1**).

We first used within-sample Z-score standardisation to describe the relative representation of each pathway within an individual predicted profile (**Figure 1a**). We identified four broad pathway blocks in the hierarchically clustered heatmap and designated them Blocks I-IV.

Block I (pathways 1-5) consistently exhibited the highest within-sample Z-scores and included amino sugar and nucleotide sugar metabolism(mean Z-score: 2.07), starch and sucrose metabolism (1.77), cysteine and methionine metabolism (1.43), pyruvate metabolism (1.63), and glycolysis/gluconeogenesis (1.79).

Block II (pathways 6-14) generally exhibited negative Z-scores and included primary and secondary bile acid biosynthesis (primary: -1.17, secondary: -1.03), xenobiotic metabolism by cytochrome P450 (-1.2), tryptophan metabolism (-1.08), taurine and hypotaurine metabolism (-0.93), lysine degradation (-0.95), benzoate degradation (-0.94), fatty acid degradation (-0.83), and vitamin B6 metabolism (-0.77).

Block III (pathways 15-21) generally exhibited moderately positive or near-zero Z-scores and included fructose and mannose metabolism (0.66), fatty acid biosynthesis (0.41), pantothenate and CoA biosynthesis (0.31), butanoate metabolism (0.33), propanoate metabolism (0.33), histidine metabolism (0.05), and nicotinate and nicotinamide metabolism (-0.06).

Block IV (pathways 22-27) generally exhibited near-zero or slightly negative Z-scores and comprised sulfur metabolism (-0.48), pentose and glucuronate interconversions (-0.28), folate biosynthesis (-0.22), biotin metabolism (-0.26), arginine and proline metabolism (-0.31), and drug metabolism-other enzymes (-0.3).

To characterize interindividual variation, we delineated four predicted metabolic profile clusters (PMC1-PMC4) by hierarchical clustering (**Figure 1b**). Of the 98 samples, 16 (16.3 %) were assigned to PMC1, 27 (27.6 %) to PMC2, 28 (28.6 %) to PMC3, and 27 (27.6 %) to PMC4.

Cluster membership was visualized using principal coordinate analysis (PCoA), with the first two axes accounting for 25.3% and 12.2% of the variation, respectively (**Figure 1c**). Although the groups

overlapped in ordination space, differences among clusters were detected by PERMANOVA (R^2 = 0.18, p = 0.001). All six pairwise comparisons were significant (R^2 = 0.088-0.206; all p = 0.001).

Genus-level comparisons further characterised the taxonomic differences associated with the PMC groups (**Figure 1d**). Kruskal-Wallis testing identified 45 genera that differed significantly across the four clusters (BH-corrected q < 0.05), and Dunn's post-hoc testing of these genera showed that the number of significantly differing genera varied substantially between pairwise comparisons. The largest number of significantly different genera was observed between PMC1 and PMC4 (33), followed by PMC1 versus PMC2 and PMC1 versus PMC3 (22 each).

In comparisons of PMC1 with the remaining clusters, higher relative abundances were observed for genera including *Erysipelatoclostridium* (log2FC: 3.7), *Gemella* (log2FC: 3.3), the *Ruminococcus gnavus* group (log2FC: 3.1), *Flavonifractor* (log2FC: 3.02), and *Escherichia-Shigella* (log2FC: 2.07), whereas several Ruminococcaceae groups, *Coprococcus* 2 (log2FC: -12.6), and *Oxalobacter* (log2FC: -9.7) showed lower relative abundances (**Figure 1e**).

Three country-level resistance-class clusters identify distinct taxonomic signatures

Resistance-class profiles derived from shotgun metagenomic sequencing data yielding 34 country-specific profiles. Hierarchical clustering identified three resistance-class profile clusters (RCPC1-RCPC3), comprising seven (20.6%), nine (26.5%), and 18 (52.9%) countries, respectively (Fig. 2a). The heatmap revealed distinct combinations of standardised resistance-class signals across countries.

Exploratory pairwise PERMANOVA comparisons yielded R^2 values of 0.639 for RCPC1 versus RCPC2, 0.382 for RCPC1 versus RCPC3, and 0.266 for RCPC2 versus RCPC3 (all p = 0.001). The largest pairwise effect size was therefore observed between RCPC1 and RCPC2.

Log_{10}-transformed AMR read counts were subsequently compared across the three clusters (Fig. 2b). RCPC1 showed the highest tendency (4.6 ± 0.18), whereas RCPC2 (4.5 ± 0.2) and RCPC3 (4.5 ± 0.16) exhibited lower, substantially overlapping distributions, although these differences did not reach statistical significance (p > 0.05).

The relative representation of taxa with reported pathogenic or genotoxic potential was highest in RCPC1 (Fig. 2c, d). Putative pathogenic taxa accounted for a relative frequency of 0.120 in RCPC1, compared with 0.026 in RCPC2 and 0.033 in RCPC3. Taxa with reported genotoxic potential followed the same ordering, with relative frequencies of 0.190, 0.035 and 0.046, respectively. Both categories therefore followed the descriptive order RCPC1 > RCPC3 > RCPC2. These differences did not reach statistical significance (p > 0.05).

The combined relative representation of taxa assigned to SCFA-producing groups was also highest in RCPC1, followed by RCPC3 and RCPC2, with approximate total values of 0.47, 0.36 and 0.28, respectively

(Fig. 2e). Putative acetate producers accounted for relative frequencies of 0.27 in RCPC-1, 0.06 in RCPC-2 and 0.20 in RCPC-3, whereas putative butyrate producers were similarly represented across clusters (0.10, 0.11, and 0.11, respectively). Putative propionate producers showed a different ordering, with the highest relative frequency in RCPC2 (0.11), followed by RCPC1 (0.09) and RCPC3 (0.04). The higher overall representation of SCFA-associated taxa in RCPC1 therefore primarily reflected the acetate-associated component.

Finally, aerotolerant taxa represented approximately 0.13 of the classified community in RCPC1, compared with 0.01 in RCPC-2 and 0.02 in RCPC3 (Fig. 2f). The complementary proportions assigned to obligate anaerobes were approximately 0.87, 0.99 and 0.98, respectively. RCPC1 therefore combined a higher relative representation of aerotolerant taxa with higher frequencies of taxa assigned pathogenic, genotoxic, and acetate-producing potential.

b) Boxplots illustrate the log10-transformed read counts of antibiotic resistance genes in the identified three resistance-class profile clusters (RCPC1-3).

Barplots show the relative abundance of taxa with reported pathogenicity (c), genotoxic potential (d), taxa assigned to SCFA-producing groups (e), and the ratio of literature-based aerotolerant and anaerobic taxa (f).

Continental patterns of antibiotic resistance burden and composition

Total log10-transformed antimicrobial resistance read counts varied across the 34 country-level datasets (Africa, 9 countries/26 participants; America, 3/7; Asia, 17/64; Europe, 5/16) but did not differ significantly between continents (Kruskal-Wallis, $p = 0.792$; Fig. 3a). Region alone therefore provided no evidence of scaling total ARG load, motivating a more granular analysis of specific factors, although the small number of countries *per* continent limits the power of this comparison.

Despite the absence of overall burden differences, resistome composition showed continental patterns. We examined the top 10 antibiotic classes, defined as the 10 most abundant classes when aggregated across all 34 country samples (Fig. 3b). When grouped by biological category, these comprised β-lactams (represented by cephamycin), the tetracycline family, the MLSB family (macrolide, lincosamide, streptogramin, and the streptogramin A and B subtypes), aminoglycosides, fluoroquinolones, and rifamycins. Together these top 10 classes accounted for the majority of resistance reads in all four continents. Asia group and Africa group showed the highest cumulative relative frequencies of the top 10 classes (approximately 80.7% and 80.4% respectively), whereas America group and Europe group were lower (approx. 74.3% and 73%), although these aggregated differences did not reach statistical significance ($p > 0.05$).

Tetracycline was the most abundant antibiotic class in all four continents, followed by rifamycin, macrolide, lincosamide, and cephamycin (Fig. 3c). Descriptively, the relative frequencies of these five classes differed between continents. Tetracycline accounted for the largest share of ARG reads in Africa

(43%) and the smallest in America (29%), whereas rifamycin showed the opposite gradient, highest in America (10%) and lowest in Asia (7%). Because multi-class genes (for example, *erm*) contribute reads to each of their classes, the macrolide and lincosamide classes are not independent; their joint enrichment in Asia (macrolide 9%, lincosamide 6%) is therefore treated as a single MLSB-family signal rather than two observations, with the lowest abundance consistently observed in Europe group. Cephamycin followed a similar pattern, peaking in the Asia group (0.06) and lowest in the Europe group (0.04). Due to 15 of the 34 datasets deriving from a single participant and the Europe and America groups comprising only five and three countries, these contrasts are treated as descriptive.

Country-level differences in resistance burden and human development level

Log10-transformed ARG read counts were mapped at the country level to examine global differences in resistome burden (Fig. 4a). The highest values were observed in Sudan (log10AMR burden:4.93; n = 2; Africa group), Mexico (4.89; n = 3; America group), and Iran (4.87; n = 4; Asia group), while the lowest in Ukraine (4.22; n = 1; Europe group) and Uganda (4.24; n = 1; Africa group). Single-participant datasets had lower burden than multi-participant datasets (mean 4.45 *vs* 4.58; $p = 0.051$), so rankings between individual countries should be interpreted with caution. To explore whether national development level could account for these geographical differences, we also mapped the Human Development Index (HDI), a composite indicator integrating health, education, and economic prosperity, for the same set of countries (Fig. 4b). In our cohort Germany had the highest HDI (0.94), followed by Canada (0.93) and Japan (0.92), whereas Sudan (0.50), Nigeria (0.53), and Pakistan (0.53) had the lowest. Visual inspection suggested an inverse spatial pattern between the two maps, with several high-burden countries located in lower-HDI regions; however, the Spearman correlation between HDI and burden was weak and not significant ($\rho = -0.28$, $p = 0.11$), and remained so with pool size as a covariate (HDI coefficient $p = 0.19$). Given the small and unequal datasets, country-level values are regarded as exploratory.

Effects of temperature, air pollution, and precipitation on resistance burden

To explore what factors contribute to the previously described global differences in antibiotic resistance burden, we first investigated the possible role of climatic parameters, namely mean annual temperature (°C), air pollution using PM2.5 (μg/m³), and mean annual precipitation (mm) (Fig. 5).

Mean annual temperature was positively associated with AMR burden (log10 read counts) across our studied countries. Spearman rank correlation indicated a positive association approaching significance ($\rho = 0.34$, $p = 0.054$), and linear regression revealed a significant linear effect ($\beta = 0.0088$ *per* °C, $R^2 = 0.139$, $p = 0.030$). This effect size corresponds to approximately a 22% increase in total AMR reads *per* 10°C increase in mean annual temperature, when back-transformed from the log10 scale (Fig. 5a).

To assess whether this effect was independent of continental geography, we fitted a joint linear model including both predictors. Temperature retained significance in Type II marginal testing (β = 0.011/°C, p = 0.030), while continent showed no significant marginal effect on AMR burden (p = 0.665), indicating that the association was not attributable to regional grouping alone. The temperature association was moderately robust and it remained significant after adjusting for pool size (p = 0.036). Although visual inspection of the temperature-burden scatter suggested an apparent two-cluster pattern, segmented regression with the Davies test for non-linearity detected no significant breakpoint (p = 0.228), but power to detect non-linearity with 34 observations is limited.

Ambient air pollution was assessed using country-level PM2.5 estimates and a three-tier classification according to WHO Air Quality Guidelines thresholds: Low (< 15 μg/m³), Moderate (15–35 μg/m³), and High (> 35 μg/m³). Continuous PM2.5 showed a positive association with AMR burden that approached but did not reach significance in linear regression (p = 0.052), while the three-tier classification revealed significant differences in burden across tiers (Kruskal-Wallis p = 0.020). Post-hoc Dunn testing with Benjamini-Hochberg correction showed that the Low tier carried significantly less AMR than both the Moderate (q = 0.018) and High (q = 0.018) tiers, while Moderate and High did not differ significantly (q > 0.05) (Fig. 5b). Given the small size of the low tier, the tier contrast is regarded as exploratory.

In contrast, annual precipitation showed no significant association with AMR burden (linear regression β = -0.0002 *per* mm, R^2 = 0.01, p = 0.530; Spearman ρ = 0.08, p = 0.66) (Fig. 5c).

1. b) Box plot of log-transformed total AMR read counts across three PM2.5 exposure tiers (Low: < 15 μg/m³; Moderate: 15–35 μg/m³; High: >35μg/m³), defined according to WHO Air Quality Guidelines.
2. c) Scatter plot of log-transformed total AMR read counts as a function of mean annual precipitation (mm), with linear regression line.

Asterisks indicate significant statistical differences (* p < 0.05).

Continental geography shapes the composition of the resistome

We next investigated whether AMR burden and AMR class composition are shaped by the same factors. Notably, none of the climatic and environmental variables previously associated with AMR burden also structured resistome composition: PERMANOVA on Bray-Curtis distances returned no significant compositional effects for mean annual temperature (p = 0.643), annual precipitation (p = 0.562), or PM2.5 tier (p = 0.602). This dissociation indicates that climatic and atmospheric factors scale the quantity of resistance genes without restructuring which specific classes are represented.

To identify factors that do structure resistome composition, we evaluated continental geography and socioeconomic development (Human Development Index, HDI) (Fig. 6). Continental geography significantly structured AMR class composition (PERMANOVA R^2 = 0.16, p = 0.026, Pr = 0.9837). Pairwise testing identified the Asia group as a possible driver of this signal, showing nominal differences from both the Europe group (p = 0.024) and the Africa group (p = 0.035), although these comparisons did not

survive BH correction. No significant pairwise differences were observed among the remaining continental pairs (Fig. 6a).

HDI categorisation (Low, Medium, High and Very High) did not significantly structure resistome composition at the omnibus level (PERMANOVA $p = 0.163$, Pr = 0.31). Pairwise testing revealed a single nominally significant difference between the Medium and High HDI groups ($p = 0.039$) (Fig. 6b).

Identification of per-class differences by continental geography and veterinary antibiotic intensity as a possible contributor

Given that continental geography significantly structured antibiotic class composition, we next identified which specific drug classes drive these compositional differences. Per-class Wilcoxon tests with Benjamini-Hochberg correction were applied to compare the Asia group against the other continents (Fig. 7a).

Comparing Asian group samples to those from other continents, five antibiotic classes showed significant enrichment, but only one after BH correction: macrolide (log2 FC = 0.55, $q = 0.028$), lincosamide (log2 FC = 0.90, raw $p = 0.0045$, $q = 0.057$), streptogramin (log2 FC = 0.79, $p = 0.0054$, $q = 0.057$), streptogramin A (log2 FC = 1.08, $p = 0.0085$, $q = 0.06$), and streptogramin B (log2 FC = 1.10, $p = 0.0095$, $q = 0.06$) (Fig. 7b).

To assess whether anthropogenic factors, such as veterinary antibiotic, could explain the regional compositional differences observed in our cohort, we obtained country-level estimates of veterinary antimicrobial intensity (milligrams per Population Correction Unit, mg/PCU) from Mulchandani et al. (2023), which estimates refer to 2020 and therefore fall within the 2019–2021 period during which the biological samples included in the present study were collected, providing a temporally aligned approximation of veterinary antimicrobial exposure in the countries represented in our cohort [11]. For this analysis, countries were compared according to the four geographical groups defined within the study cohort. The countries included in the African group had significantly lower estimated veterinary antimicrobial intensity per unit of livestock biomass than those included in the American ($q = 0.018$), Asian ($q = 3.8 \times 10^{-6}$), and European ($q = 0.030$) groups (Fig. 7c). Conversely, the countries included in the Asian group showed the highest estimated veterinary antimicrobial intensity in the cohort, with a mean of 90.1 mg/PCU compared with 48.7 mg/PCU among all countries assigned to the remaining geographical groups (Wilcoxon $p = 0.0001$).

Discussion

Antimicrobial resistance is contributing to an estimated 1.14 million deaths directly attributable to bacterial AMR and 4.71 million deaths associated with it in 2021, while being heavily concentrated in low- and middle-income countries (LMICs) [1, 12]. Understanding the drivers of this global variation, both quantitative and qualitative, is essential for designing effective, regionally appropriate interventions.

In this study, we examined the human gut resistome and its associated microbial composition in 98 healthy, young adult medical students from 34 countries, integrating two complementary sequencing approaches: country-level pooled shotgun metagenomics, used to characterise resistome burden and class-level composition, and individual-level 16S rRNA gene sequencing, used to characterise predicted metabolic function and its taxonomic correlates.

The use of a demographically and educationally relatively homogeneous cohort of young medical students was intended to reduce some of the demographic and socioeconomic heterogeneity that typically complicates international comparisons. Nevertheless, the results should be interpreted at the level of the countries and geographical groupings represented in this cohort, rather than as population-representative estimates for the respective countries or continents.

Our analysis identified a broadly shared pattern of predicted metabolic potential alongside interindividual variation associated with differences in microbial composition. Such similarity is not unexpected in a cohort of first-year medical students of comparable age within a shared educational setting, although the contribution of these characteristics was not directly tested.

Functional similarities have also been observed across more heterogeneous populations. The Human Microbiome Project reported comparatively conserved metabolic pathway profiles despite substantial taxonomic variation, suggesting that a shared functional background is a broader feature of human-associated microbial communities rather than a consequence of demographic homogeneity alone [13].

The common pathway pattern nevertheless coexisted with individual differences, including variation within the descriptive pathway blocks. A particularly relevant comparison is provided by Mohr et al. (2024), who studied first-year, dormitory-housed college students and found lower variability in estimated functional and faecal metabolomic features than in genus-level composition within their longitudinal analysis [14]. Their findings support the coexistence of microbial individuality and comparatively conserved functional features during emerging adulthood.

The broadly similar pathway hierarchy observed across taxonomically different communities in our study is consistent with, but does not directly demonstrate, partial functional redundancy, whereby different microbial assemblages retain overlapping functional capacities [15, 16]. Complementing this interpretation, Meawad et al. (2025) described healthy adult gut microbiomes as mixtures of functional archetypes with differing metabolic emphases [17], providing a conceptual parallel to our observation that a shared functional background can coexist with interindividual differences in the relative representation of specific pathways. Communities with overlapping predicted capacities may nonetheless differ in the organisms carrying those functions and in the conditions under which they are expressed [16, 18]; gene-level characterisation and targeted metabolite measurements would be needed to determine the biological consequences of the observed profile differences. For example, butanoate and propanoate metabolism in Block III identify short-chain fatty acid-related functions for further investigation, but their

predicted representation does not establish net butyrate or propionate production or metabolite availability to the host.

Beyond the shared functional background, the four PMC groups differed in their associated genus-level composition, with PMC1 most clearly separated from the remaining clusters. Genera more abundant in PMC1 included *Erysipelatoclostridium*, *Gemella*, the *Ruminococcus gnavus* group, *Flavonifractor*, and *Escherichia-Shigella*, several of which have been reported in association with gut inflammation or reduced microbial diversity in other cohorts [19–22], whereas genera depleted in PMC1 included several Ruminococcaceae groups together with *Coprococcus* 2 and *Oxalobacter*.

This pattern could raise the possibility that PMC1 represents a taxonomically and functionally distinct subgroup within the cohort, rather than simply one end of a continuous gradient shared by all four clusters. However, because PMC assignment was derived from predicted pathway abundances and the associated genus differences were identified in the same 16S dataset, this taxonomic signature should be regarded as a descriptive correlate of the clustering rather than independent confirmation of a biologically distinct community state. The specific genera distinguishing PMC1 would need to be examined in relation to clinical or lifestyle variables, and ideally validated using shotgun sequencing or gene-level functional data, before their relevance to host health can be assessed.

Hierarchical clustering of resistance-class profiles from the 34 country-level shotgun datasets identified three resistance-class profile clusters. These groups differed in their accompanying taxonomic characteristics, with the cluster showing the highest AMR read counts (RCPC1) also exhibiting greater relative abundances of taxa classified as surveillance-relevant, potentially pathogenic, potentially genotoxic, or aerotolerant.

A relevant human comparison is provided by Lee et al. (2023), who identified two resistotypes in human gut metagenomes, one characterised by higher overall ARG abundance and an association with species-specific resistance genes in Proteobacteria [23]. Their analysis also identified mobile ARGs shared between commensals and pathogens, illustrating that the organisation of the gut resistome extends beyond the abundance of conventionally pathogenic organisms.

Based on our data, the greater representation of aerotolerant taxa and taxa with reported pathogenic or genotoxic potential in RCPC1 adds an ecological dimension to its resistance profile. Related observations have been reported in poultry environments: Yasmin et al. (2025) found positive associations between genera including Enterobacter, Acinetobacter and Pseudomonas and predicted AMR-associated functions [24]. However, both the taxonomic profiles and functional predictions originated from 16S rRNA gene data, so these associations provide a methodological and ecological comparison rather than direct evidence of ARG carriage.

Stronger evidence for the contribution of community restructuring comes from shotgun metagenomic studies. Xiong et al. (2018) reported that therapeutic chlortetracycline treatment in broilers reduced the relative abundance of Escherichia/Shigella while enriching tetracycline-resistance genes associated with

other bacterial hosts, including Bifidobacterium, illustrating that resistome changes need not track potentially pathogenic taxa uniformly [25]. Similarly, Plata et al. (2022) reported distinct taxonomic, metabolic, and resistance-associated responses to different antibiotic growth promoters in chickens, illustrating how antibiotic exposure may reshape microbial communities and resistance profiles, but do not establish it as the cause of the differences observed in our human cohort [26].

The coexistence of higher AMR read counts and a greater relative representation of putative SCFA-producing taxa in RCPC1 indicates that resistance-associated features and taxa with potentially beneficial metabolic capabilities are not mutually exclusive at the community level. This observation argues against interpreting the RCPCs along a single favourable-to-unfavourable ecological gradient: the higher proportion of putative SCFA producers in RCPC1 neither counterbalances its resistance-associated signals nor demonstrates improved gut health.

The relative abundance of acetate-associated taxa varied markedly among RCPCs, whereas butyrate-associated taxa remained comparatively similar. Experimental work on human gut isolates provides a biological explanation for why producer categories cannot be interpreted independently: Belenguer et al. (2006) demonstrated that fermentation products generated by Bifidobacterium adolescentis, including acetate and lactate, could support butyrate formation by other bacteria through cross-feeding [27]. Consequently, differences in acetate-associated taxa could be relevant to wider fermentation networks even when butyrate-associated taxa are similarly represented, although such cross-feeding remains a hypothesis in the present dataset, as neither metabolite exchange nor pathway activity was measured. A further limitation specific to this analysis is that each shotgun resistance-class profile represents a single country-level pool; the observed RCPC profiles therefore cannot establish within-country variability or national representativeness, and AMR read counts require consideration of sequencing depth and normalisation before being interpreted as differences in resistance-gene abundance.

The principal conceptual finding from the shotgun-based analyses was that total AMR abundance and resistome class composition were associated with partly distinct sets of country-level variables. Mean annual temperature and veterinary antimicrobial intensity were associated primarily with AMR abundance, whereas continental geography was more strongly associated with the relative composition of the resistome. This distinction suggests that the factors shaping the overall quantity of detected resistance determinants may differ from those shaping the specific antimicrobial classes represented, and emphasises that resistome abundance and composition are complementary, rather than interchangeable, dimensions of AMR ecology.

Mean annual temperature showed the clearest climatic association with total AMR abundance: a 10°C increase was associated with an approximately 22% increase in AMR read abundance, an association that remained significant after adjustment for continental grouping and pool size. The magnitude of this association should not be directly compared with estimates derived from clinical surveillance, because metagenomic AMR abundance and phenotypic resistance in individual pathogens represent different biological endpoints. Nevertheless, the direction of the relationship is consistent with prior observations,

including MacFadden et al. (2018) [7], who reported increasing resistance rates in several bacterial pathogens across warmer regions of the United States. Higher temperatures can plausibly influence bacterial replication, horizontal gene transfer, and the environmental persistence of resistant bacteria and ARGs [7, 28, 29], but these mechanisms were not measured directly here. The observed relationship is therefore best regarded as an ecological association compatible with a potential amplifying role of temperature, rather than as evidence of a direct causal effect. If this association generalises, rising global temperatures could represent an ongoing, climate-linked pressure on AMR burden that would operate alongside, rather than instead of, anthropogenic drivers.

Ambient air pollution, measured as PM2.5 concentration, was also associated with AMR abundance with a threshold-like pattern: countries in the WHO low-exposure tier (< 15 μg/m³) carried significantly lower AMR abundance than the moderate and high tiers, which did not differ from each other.

This observation is consistent with reports that particulate matter can harbour antibiotic-resistant bacteria and resistance genes [30], and with spatiotemporal associations between PM2.5 and prevalence of carbapenem-resistant Enterobacteriaceae in China [10]. However, the mechanisms underlying these associations remain uncertain. PM2.5 may also act as a proxy for correlated conditions such as industrial activity, urbanisation, and sanitation infrastructure, so the association may reflect both direct exposure pathways and broader contextual confounding rather than a specific causal effect of particulate matter itself.

Annual precipitation was not associated with either AMR abundance or resistome composition. A previous study has reported heterogeneous, context-dependent associations between precipitation and resistance [10], and country-level annual precipitation is a coarse measure that does not distinguish between intense rainfall events, seasonal flooding, and evenly distributed rainfall, exposure pathways that may differentially affect wastewater overflow, runoff, and environmental persistence of resistance determinants. The absence of an association in our cohort should therefore not be taken as evidence that hydrological conditions are irrelevant to resistome variation generally.

We also examined whether socioeconomic development, represented by the Human Development Index, was associated with variation in the gut resistome. HDI showed a weak, non-significant negative association with AMR abundance, and was not significantly associated with resistome composition. This is consistent with the view that broad composite indicators have limited ability to capture the specific structural processes relevant to AMR: Dhinsa et al. (2024) [31] similarly found no strong association between Gross National Income *per* capita or HDI and multidrug-resistant tuberculosis incidence, while Satria et al. (2024) [32] identified Universal Health Coverage, rather than broader development indices, as the country-level factor most strongly associated with national AMR capacities. From a structural-determinants perspective, Davis et al. [2] have argued that AMR research has traditionally emphasised biomedical mechanisms over equity-related factors such as water and sanitation, antimicrobial access, and informal pharmaceutical supply chains, which HDI captures only indirectly. Thus, more proximate, operational indicators, healthcare coverage, prescription oversight, and

surveillance capacity may be more informative than aggregate development indices for understanding national AMR patterns, although these structural variables were not directly evaluated here and should be prioritised in future, more comprehensive analyses.

In contrast to the climatic variables, continental geography was significantly associated with resistome class composition, although this signal should be interpreted cautiously: the continental groups were unequal in size (17/9/5/3 countries *per* continental group), and pairwise differences did not survive correction for multiple comparisons. The findings therefore describe compositional patterns among the sampled country groups rather than continent-wide resistance profiles.

Among the countries assigned to the Asian group, macrolide resistance was significantly enriched relative to the remaining countries after BH correction, while lincosamide and streptogramin-related classes showed concordant but FDR-borderline trends. Because these classes share reads from multi-class resistance genes, this pattern is best interpreted as a single MLSB-family signal detected across several overlapping class definitions, and not as five independent findings.

This concordant behaviour is mechanistically plausible: the *erm* gene family mediates methylation of the 23S rRNA target site and can confer cross-resistance across MLSB classes, whereas *msr*-associated ABC-F proteins act predominantly through ribosomal protection [33], so that selection exerted by one class may enrich determinants affecting related classes even without equivalent use of each agent. Gene-level linkage and participant-specific exposures were not assessed here, and this mechanism remains a biologically informed interpretation.

The pattern is also consistent with regional antimicrobial-use context. Asian countries accounted for approximately 59% of global veterinary antimicrobial tonnage in 2020 [11], and macrolides are important in several livestock and aquaculture systems represented across the region [34, 35]. Within our cohort, the countries assigned to the Asian group showed the highest estimated veterinary antimicrobial intensity (mean 90.1 mg/PCU *versus* 48.7 mg/PCU elsewhere), while the African group showed the lowest. These regional statistics do not provide direct evidence of exposure among the specific participants or countries included in our cohort and should be read as contextual information and not as a causal explanation. Human macrolide use may represent an additional contextual factor: elevated macrolide consumption and resistance, including among respiratory pathogens such as *Streptococcus pneumoniae* and *Mycoplasma pneumoniae*, have been reported in parts of East and Southeast Asia, particularly China [36, 37]. Together with intensive veterinary use, these exposure patterns could provide a plausible context for the MLSB-related signal observed among the Asian countries in our cohort, although participant-level antimicrobial histories and country-specific human consumption data were not available, and this interpretation should be regarded as hypothesis-generating.

The relatively low MLSB-family abundance observed in our European samples likely reflects, at least in part, progressive EU restriction of growth-promoter and prophylactic antibiotic use, with declining lincosamide sales reported across 29 EU/EEA countries between 2017 and 2024 [38]; it may also partly

reflect that our European subset did not include Southern European countries where macrolide and lincosamide use remains comparatively high, which may constrain the regional representativeness of this group.

This study combined country-level pooled shotgun metagenomics with individual-level 16S rRNA gene sequencing to characterise the gut resistome and its predicted metabolic and taxonomic correlates. The relative homogeneity of the cohort strengthens internal comparability but weakens population-level representativeness, and several further features of the design constrain interpretation.

Geographical representation was uneven, with the Asian and African groups more extensively represented than the American and European groups; the findings therefore apply to the countries represented in this cohort and should not be read as continent-wide estimates, particularly for the three-country American group. The shotgun analyses were based on 34 country-level pooled datasets of unequal size (1 to 16 participants *per* pool; 15 countries represented by a single participant), which reduced the effective sample size available for country-level comparisons and limited the extent to which covarying country-level exposures (temperature, geography, socioeconomic development, PM2.5, and antimicrobial use) could be individually partitioned; associations that remained significant after adjustment therefore reflect conditional independence within the fitted models rather than demonstrated mechanistic independence.

International participants had resided in Hungary for up to 6–12 months before sampling, and because migration and dietary transition can reshape the gut microbiome, both the shotgun and 16S profiles may reflect a combination of earlier country-of-origin exposures and more recent shared exposures. Most shotgun-based explanatory variables were measured at the country level and represent ecological context rather than individual exposure; the resulting associations are susceptible to ecological fallacy and should not be translated directly into individual-level exposure-response relationships. Country-level covariates also differed in completeness: veterinary antimicrobial-use estimates were model-derived and subject to uncertainty from incomplete reporting and informal distribution, and several structural variables (prescription enforcement, wastewater-treatment coverage, diagnostic access) could not be incorporated owing to insufficiently complete cross-country data.

In conclusion our study identifies a distinction between the quantitative and compositional dimensions of the human gut resistome, complemented by individual-level evidence of predicted metabolic and taxonomic structure within the same cohort. Within the countries represented in this cohort, mean annual temperature was associated primarily with total AMR abundance, whereas continental geography was more strongly associated with resistance-class composition. This dissociation provides a conceptual framework for interpreting global resistome variation: broad environmental conditions may relate to the overall accumulation of resistance determinants, whereas composition may more closely reflect regionally structured antimicrobial practices and microbial community configurations. At the individual level, predicted metabolic profiling identified a largely shared functional background alongside a taxonomically distinct subgroup (PMC1), and resistance-class clustering identified a cluster (RCPC1)

combining higher AMR burden with greater representation of potentially pathogenic and aerotolerant taxa, illustrating that both abundance-based and composition-based groupings can reveal structure not captured by a single summary metric.

Human gut resistome profiling should therefore be viewed as a complementary layer of One Health AMR surveillance rather than a substitute for clinical, veterinary, and environmental monitoring; its principal value lies in detecting integrated resistance signatures that may remain obscured in conventional source-based datasets. Validation will require larger and more geographically balanced cohorts, individual-level exposure histories, longitudinal sampling, gene-resolved and functional data, and parallel characterisation of environmental reservoirs, to determine whether the patterns identified here can support more precise, context-sensitive One Health antimicrobial-stewardship strategies.

Methods

Study design and participants

This study was conducted between June 2019 and April 2021 at the University of Debrecen, Hungary. Stool samples were collected from healthy young adult first-year medical students (aged 22.3 ± 3.3 years) enrolled at the Faculty of Medicine, comprising both Hungarian and international students from a diverse range of countries. Written informed consent was obtained from all participants, and the study protocol was approved by the Regional and Institutional Review Board of the University of Debrecen (ethics approval number: DE RKEB/IKEB 5191 – 2019).

Medical students in an English-language international programme were chosen as a deliberately homogeneous sampling frame, to reduce inter-individual variability in factors that could otherwise confound country-of-origin comparisons of the gut resistome. Such students are expected to share broadly comparable age, educational attainment, and socioeconomic background, which reduces, but does not eliminate, within-country heterogeneity.

Inclusion criteria were no known chronic gastrointestinal disorders and no reported antibiotic use within the three months preceding sample collection. For non-Hungarian participants, an additional criterion required that they had begun their studies at the University within the 6–12 months preceding sampling, on the assumption that their gut microbiome would still predominantly reflect their country of origin rather than the local Hungarian dietary and environmental context. This criterion was intended to limit, but cannot fully remove, the influence of dietary and lifestyle changes associated with residence in Hungary.

Sample numbers

A total of 194 stool samples were collected. After excluding repeat samples from the same individual, 118 individual samples remained, representing 35 countries of origin. For shotgun metagenomics, DNA

was pooled by country of origin. One country (Azerbaijan; 4 individuals) did not yield a successful sequencing library and was excluded, so 114 individual samples (45 female, 67 male) in 34 country-level pools (Asia, 17 countries; Africa, 9; Europe, 5; America, 3) formed the primary unit of analysis for the resistome. Pool sizes ranged from 1 to 114 individuals, and 15 pools derived from a single individual.

For individual-level 16S rRNA gene sequencing, samples were available from 116 individuals in 34 countries (Tunisia was not included in the 16S set); 98 of these (53 men, 45 women) were retained after exclusion of samples to approximate a balanced sex distribution, with countries represented by a single individual retained.

The 16S and shotgun analyses were therefore performed on overlapping but not entirely identical sample sets and countries (Azerbaijan is represented only in the 16S data, and Tunisia only in the shotgun data), and are reported separately. For the predicted metabolic profile analysis, one single-individual sample (Malaysia) was excluded as an extreme outlier, leaving 97 samples. Exact sample numbers per country are provided in Supplementary File 1.

Data sources for environmental and socioeconomic covariates

Environmental variables

Country-level annual mean ambient *PM2.5 concentrations* (μg/m³) were obtained from the World Health Organization Global Health Observatory (WHO GHO Ambient Air Pollution database; https://www.who.int/data/gho/data/themes/air-pollution/ambient-air-pollution) [39]. Countries were subsequently classified into three PM2.5 exposure tiers using thresholds aligned with the WHO 2021 Global Air Quality Guidelines: Low (< 15 μg/m³), Moderate (15–35 μg/m³), and High (> 35 μg/m³) [40].

Temperature (°C) and *precipitation* (mm) data were obtained from the World Bank Climate Change Knowledge Portal (https://climateknowledgeportal.worldbank.org) [41].

Socioeconomic variables

Human Development Index (HDI) values for 2014–2018 were obtained from the United Nations Development Programme Human Development Report (https://hdr.undp.org/data-center/human-development-index#/indicies/HDI) [42].

Veterinary antimicrobial intensity (milligrams per Population Correction Unit, mg/PCU) was obtained from Mulchandani et al. (2023) [11], which provides modelled 2020 estimates of veterinary antimicrobial use across food-producing animals in 190 countries, calibrated against country-level reports from the

World Organisation for Animal Health (WOAH). The dataset was accessed *via* Our World in Data (https://ourworldindata.org/grapher/antibiotic-usage-in-livestock) [43].

DNA extraction and sequencing

Faecal samples were self-collected into sterile, nuclease-free sampling tubes, transported to the laboratory, and processed immediately upon arrival. For country-level pooling, an equal amount of faecal material was combined from each individual sample. DNA was extracted using the QIAamp® PowerFecal® DNA Kit (Qiagen, Germany; Cat. 12830-50) according to the manufacturer's instructions, and concentrations were determined fluorometrically using a Qubit® HS dsDNA Assay Kit (Thermo Fisher Scientific, USA) on a Qubit® 4.0 fluorometer.

16S rRNA sequencing

16S rRNA gene amplicon libraries were prepared following Illumina's 16S Metagenomic Sequencing Library Preparation protocol (15044223 Rev. B). The V3-V4 hypervariable region was amplified using the universal primer pair 341F (5'-CCTACGGGNGGCWGCAG-3') and 785R (5'-GACTACHVGGGTATCTAATCC-3'), each flanked by Illumina overhang adapters (forward: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3';

reverse: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'; Sigma-Aldrich, USA), yielding amplicons of approximately 460 bp. Amplification was performed using 2× KAPA HiFi HotStart ReadyMix, and dual indexing was carried out with the Nextera XT Index Kit (FC-131-1001/2, Illumina). Amplicons were purified and size-selected to 580–630 bp using MagSi-NGSPREP Plus (MDKT00010075) following the Illumina protocol, and their quality and size distribution were verified on an Agilent D1000 ScreenTape system (reagents 5067–5582 and 5067–5583).

Amplicons were normalised to 4 nM and pooled into a single library at equimolar concentration. The pooled library was denatured with 0.2 M NaOH, diluted to 8 pM, and sequenced on an Illumina MiSeq platform using the MiSeq Reagent Kit v3 (618-cycle, MS-102-3003) in paired-end mode (2 × 301 bp), with a 20% PhiX spike-in (PhiX Control Kit v3, FC-110-3001) included as an internal sequencing quality control.

Shotgun sequencing

Shotgun metagenomic sequencing was performed on an Illumina NovaSeq 6000 instrument (Illumina, USA) with a 150-bp paired-end run at Novogene Bioinformatics Technology (Beijing, China), generating a minimum of 20 million reads per sample. To ensure the required read depth, samples were re-isolated until reaching neccessary purity (OD260/280 = 1.8-2.0) and concentration (≥ 10 ng/μL).

Metadata processing

16S rRNA metagenomic data

Paired-end reads were demultiplexed and converted to FASTQ format using Illumina BaseSpace. Subsequent processing was performed in QIIME 2 (version 2026.7; https://library.qiime2.org/). Adapter sequences (CTGTCTCTTATACACATCT) were trimmed from the 3' ends of reads using the Cutadapt plugin (v4.0) [44]. Quality filtering, trimming, and chimera removal were performed with the DADA2 plugin (v1.26) [45], with forward and reverse reads trimmed by one base at the 5' end and truncated at 300 bases; amplicon sequence variants (ASVs) were inferred from the filtered reads. Multiple sequence alignment was performed using MAFFT (v7.526; https://mafft.cbrc.jp/alignment/software/).

Taxonomic classification was performed using a naïve Bayes classifier trained on the SILVA reference database (version 138.2; https://www.arb-silva.de/), tailored to the V3-V4 region defined by the 341F-785R primer pair. Predicted functional profiles were generated from the inferred ASVs using PICRUSt2 (v2.6.3; https://github.com/picrust/picrust2), comprising phylogenetic placement, hidden-state prediction of gene families, and pathway-level inference; gene families and pathways were annotated against Kyoto Encyclopedia of Genes and Genomes (KEGG) identifiers. All functional outputs represent predicted rather than directly measured biological activity; no KEGG pathway maps or other copyrighted KEGG content were reproduced.

Shotgun metagenomic data

The processing of the metada was performed using the SqueezeMeta pipeline (v1.6.3) [46] with the co-assembly option and no binning. Quality filtering was conducted via Trimmomatic (v0.39) [47] with the parameters: LEADING:8, TRAILING:8, SLIDINGWINDOW:10:15, and MINLEN:30. MEGAHIT was used for the assembly of paired-end reads. The KIFÜ Hungarian High-Performance Computing Competence Center (HPC CC) Komondor HPC cluster was used for the execution of the analyses, utilizing 48 CPU cores and 90 GB of RAM per sample. Antibiotic resistance profiling was performed with KneadData for quality control (using Trimmomatic and Bowtie2), and the resistome was predicted using RGI software with the CARD database.

Statistical analysis

All statistical analyses and graphs were performed in RStudio (2024.12.1 + 563). All statistical tests were two-tailed, with a significance level set at $p < 0.05$. Given the high-dimensional structure of metagenomic resistome data, we applied non-parametric methods throughout as the primary analytical framework.

Sensitivity analyses for pool size. Because country-level datasets were derived from pools of unequal size, we assessed whether pool size (number of individual participants *per* country) confounded the

observed associations. The association between pool size and logAMR was tested by Spearman correlation, and differences in pool size among comparison groups (e.g. continents, resistance-class clusters) by the Kruskal-Wallis test. Pool size was additionally included as a covariate in linear models of logAMR, with models fitted using pool-size weighting and restricted to countries with at least three participants as further sensitivity checks. For compositional analyses, PERMANOVA was repeated excluding single-participant datasets. Where associations remained consistent in direction and magnitude across these approaches, the unadjusted result is reported as the primary estimate.

Hierarchical clustering of standardised (Z-score) metabolic pathway profiles (16S-derived, predicted) and resistance-class relative-abundance profiles (shotgun-derived) was performed using Ward's minimum variance method (ward.D2) on Euclidean distance, implemented with the 'pheatmap' R package (version 1.0.12). Differences in Bray-Curtis compositional dissimilarity between clusters were assessed by PERMANOVA, including pairwise comparisons with BH correction. Genera differing significantly in relative abundance across the four PMC groups were identified using the Kruskal-Wallis test (BH-corrected $q < 0.05$), followed by Dunn's post-hoc test with BH correction to identify genera distinguishing specific pairwise cluster comparisons.

Group comparisons of logAMR were performed using the Wilcoxon rank-sum test (Mann-Whitney U). For associations between logAMR and continuous covariates (temperature, precipitation, PM2.5, HDI, and mg/PCU), we used Spearman rank correlation as the primary test, complemented by linear regression where the relationship approximated linearity. Multivariate linear models were used to test the independence of covariate effects (Type II sums of squares).

Country-level resistome composition was analysed at the antibiotic class level, using relative frequency data. Beta diversity was quantified using Bray-Curtis dissimilarity (vegan package v2.6-4). Compositional divergence between groups was tested using permutational multivariate analysis of variance (PERMANOVA; adonis2 function, vegan) with 999 permutations. Homogeneity of multivariate dispersion was assessed using the betadisper function to distinguish genuine centroid shifts from differential within-group variance. PERMANOVA analyses were performed for both categorical and continuous covariates. For categorical variables with more than two levels, pairwise PERMANOVA comparisons were performed using the pairwise.perm.manova function (RVAideMemoire package), with Benjamini-Hochberg correction of pairwise p-values.

For class-level analyses of AMR differential abundance between groups, we used Wilcoxon rank-sum tests on each of the 32 antibiotic classes independently, with log2 fold changes calculated. For continuous covariates, per-class Spearman rank correlations were computed. In both cases, multiple testing correction was performed using the Benjamini-Hochberg (BH) procedure to control the false discovery rate at $q < 0.05$.

The graphs were generated using the 'ggplot2' R package (version 3.5.0) with 'ggrepel' (version 0.9.4) for non-overlapping text labels and 'patchwork' (version 1.2.0) for multi-panel figure composition.

Declarations

Competing interests

The author(s) declare no competing interests.

Funding

Not applicable.

Ethics approval and consent to participate

The study protocol received ethical approval from the Regional and Institutional Review Board of the University of Debrecen (ethics approval number: DE RKEB/IKEB 5191-2019). All procedures were conducted in accordance with the principles of the Declaration of Helsinki and applicable institutional and national regulatory guidelines. Clinical trial number: not applicable.

Consent for publication

All participants received a full explanation of the study by the lead researcher and provided written informed consent before enrolment.

Data availability statement

All sequence data used in the analyses were deposited in the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) under PRJNA1503165 and PRJNA1541683.

Authors contributions

MM, PD, AASZR, ESZT, PF, AA, IA, NG, ZSSZ and TP processed the biological samples, extracted the nucleic acids, conducted the sequencing. and performed the statistical analyses and figure preparation. SB, LS, and JR participated in the conceptualization of the study. MP led the conceptualization, guided the analyses, and was the principal author of the manuscript.

Acknowledgements

MM was supported by the PhD Excellence Scholarship from the Count István Tisza Foundation for the University of Debrecen. Supported by the EKÖP-25-3-II University Research Scholarship Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

References

1. Ranjbar R, Alam M, Antimicrobial Resistance C. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Evid. Based Nurs.* ebnurs-2022-103540 (2023)

10.1136/ebnurs-2022-103540

2. Davis K, et al. Antimicrobial resistance, equity and justice in low- and middle-income countries: an intersectional critical interpretive synthesis. Nat Commun. 2025;16:9078.
3. Wang C, et al. Global landscape of antibiotic resistance genes in the human gut microbiome metagenome-assembled genomes. BMC Microbiol. 2025;26:33.
4. Hazarika P, Chattopadhyay I, Umpo M, Choudhury Y, Sharma I. Elucidating the gut microbiome alterations of tribal community of Arunachal Pradesh: perspectives on their lifestyle or food habits. Sci Rep. 2022;12:18296.
5. Kumar G, Bhadury P. Exploring the influences of geographical variation on sequence signatures in the human gut microbiome. J Genet. 2023;102:51.
6. Daunizeau C, et al. The gut microbiota of Indigenous populations in the context of dietary westernization: a systematic review and meta-analysis. Front Nutr. 2025;12:1652598.
7. MacFadden DR, McGough SF, Fisman D, Santillana M, Brownstein JS. Antibiotic Resistance Increases with Local Temperature. Nat Clim Change. 2018;8:510–4.
8. Rodríguez-Verdugo A, Lozano-Huntelman N, Cruz-Loya M, Savage V, Yeh P. Compounding Effects of Climate Warming and Antibiotic Resistance. *iScience* 23, 101024 (2020).
9. Burnham JP. Climate change and antibiotic resistance: a deadly combination. Ther Adv Infect Dis. 2021;8:2049936121991374.
10. Li B, Zhen X, Ouyang J, Lundborg CS. Association between antibiotic resistance and air pollution and climate factors in China: A multivariable spatial-temporal analysis. One Health. 2025;21:101204.
11. Mulchandani R, Wang Y, Gilbert M, Van Boeckel TP. Global trends in antimicrobial use in food-producing animals: 2020 to 2030. PLOS Glob Public Health. 2023;3:e0001305.
12. Jamil S, et al. The Global Impact of Bacterial Antimicrobial Resistance: Addressing Gaps and Future Strategies. Public Health Chall. 2026;5:e70229.
13. Gevers D, et al. The Human Microbiome Project: A Community Resource for the Healthy Human Microbiome. PLoS Biol. 2012;10:e1001377.
14. Mohr AE et al. Microbial Ecology and Metabolism of Emerging Adulthood: Gut Microbiome Insights from a College Freshman Cohort. Gut Microbes Rep 1, 2387936.
15. Jiang Y, Che L, Li SC. Deciphering the personalized functional redundancy hierarchy in the gut microbiome. Microbiome. 2025;14:17.
16. Louca S, et al. Function and functional redundancy in microbial systems. Nat Ecol Evol. 2018;2:936–43.
17. Meawad M, et al. Functional archetypes in the human gut microbiome reveal metabolic diversity, stability, and influence disease-associated signatures. Microbiome. 2025;13:241.
18. Lewandowski R. Emergent function, not microbial conformity: functional redundancy and the limits of taxonomic inference in microbiome genomics. Microb Genomics. 2026;12:001831.

19. Chen G, et al. Fluoride induced leaky gut and bloom of Erysipelatoclostridium ramosum mediate the exacerbation of obesity in high-fat-diet fed mice. J Adv Res. 2022;50:35–54.
20. Orejudo M, et al. Exploration of fecal microbiota in newly diagnosed patients with inflammatory bowel disease using shotgun metagenomics. Front Cell Infect Microbiol. 2025;15:1595884.
21. Henke MT et al. Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc. Natl. Acad. Sci. U. S. A.* 116, 12672–12677 (2019).
22. Hu J et al. Correlation between altered gut microbiota and elevated inflammation markers in patients with Crohn's disease. Front Immunol 13, (2022).
23. Lee K, et al. Population-level impacts of antibiotic usage on the human gut microbiome. Nat Commun. 2023;14:1191.
24. Yasmin A, et al. Metagenomic insights into microbial diversity and potential pathogenic transmission in poultry farm environments of Bangladesh. BMC Microbiol. 2025;25:318.
25. Xiong W, et al. Antibiotic-mediated changes in the fecal microbiome of broiler chickens define the incidence of antibiotic resistance genes. Microbiome. 2018;6:34.
26. Plata G, et al. Growth promotion and antibiotic induced metabolic shifts in the chicken gut microbiome. Commun Biol. 2022;5:293.
27. Belenguer A, et al. Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. Appl Environ Microbiol. 2006;72:3593–9.
28. Fernández Salgueiro M, Martínez C, Gan JA. Arcos González, P. Climate change and antibiotic resistance: A scoping review. Environ Microbiol Rep. 2024;16:e70008.
29. Li Z, et al. Climate warming increases the proportions of specific antibiotic resistance genes in natural soil ecosystems. J Hazard Mater. 2022;430:128442.
30. Zhu G, et al. Air pollution could drive global dissemination of antibiotic resistance genes. ISME J. 2021;15:270–81.
31. Dhinsa G, Seo B. Multi-Drug Resistant Tuberculosis in Developing and Developed Countries: A Study of GNI per capita, Human Development Index and the incidence of Multi-Drug Resistant Tuberculosis. J Econ Dev Stud 7, (2019).
32. Satria FB, Tsai F-J. Awareness and capacities of 103 countries to address antimicrobial resistance from 2017 to 2020. Sci Rep. 2024;14:30184.
33. Miklasińska-Majdanik M. Mechanisms of Resistance to Macrolide Antibiotics among Staphylococcus aureus. Antibiotics. 2021;10:1406.
34. Pham DK et al. Monitoring Antibiotic Use and Residue in Freshwater Aquaculture for Domestic Use in Vietnam. *EcoHealth* 12, 480–489 (2015).
35. Lulijwa R, Rupia EJ, Alfaro AC. Antibiotic use in aquaculture, policies and regulation, health and environmental risks: a review of the top 15 major producers. Rev Aquac. 2020;12:640–63.

36. Li J, et al. Severe problem of macrolides resistance to common pathogens in China. Front Cell Infect Microbiol. 2023;13:1181633.
37. Darazam IA, et al. Recent Macrolide Resistance Pattern of Mycoplasma Pneumonia in the World: A Systematic Review and Meta-Analysis. Iran J Public Health. 2025;54:530–41.
38. European Medicines Agency. Reflection paper on the use of macrolides, lincosamides and streptogramins (MLS) in food-producing animals in the European Union: development of resistance and impact on human and animal health. EMA/CVMP/AWP/244404/2004 (European Medicines Agency, 2011). https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-macrolides-lincosamides-streptogramins-mls-food-producing-animals-european-union-development-resistance-impact-human-animal-health_en.pdf. Last accessed on 27 July 2026.
39. Ambient air pollution data. https://www.who.int/data/gho/data/themes/air-pollution/ambient-air-pollution. Last accessed on 27 July 2026.
40. Ambient (outdoor) air pollution, editor. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health. Last accessed on 27 July 2026.
41. Climate Change Knowledge Portal |. Climate Change Knowledge Portal. https://climateknowledgeportal.worldbank.org/. Last accessed on 27 July 2026.
42. Nations U. *Human Development Index*. *Human Development Reports* https://hdr.undp.org/data-center/human-development-index. Last accessed on 27 July 2026.
43. Antibiotic usage in livestock per kilogram of meat. *Our World in Data* https://ourworldindata.org/grapher/antibiotic-usage-in-livestock. Last accessed on 27 July 2026.
44. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.
45. Callahan BJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
46. Tamames J, Puente-Sánchez F, SqueezeMeta. A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front Microbiol. 2019;9:3349.
47. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014;30:2114–20.

Figures

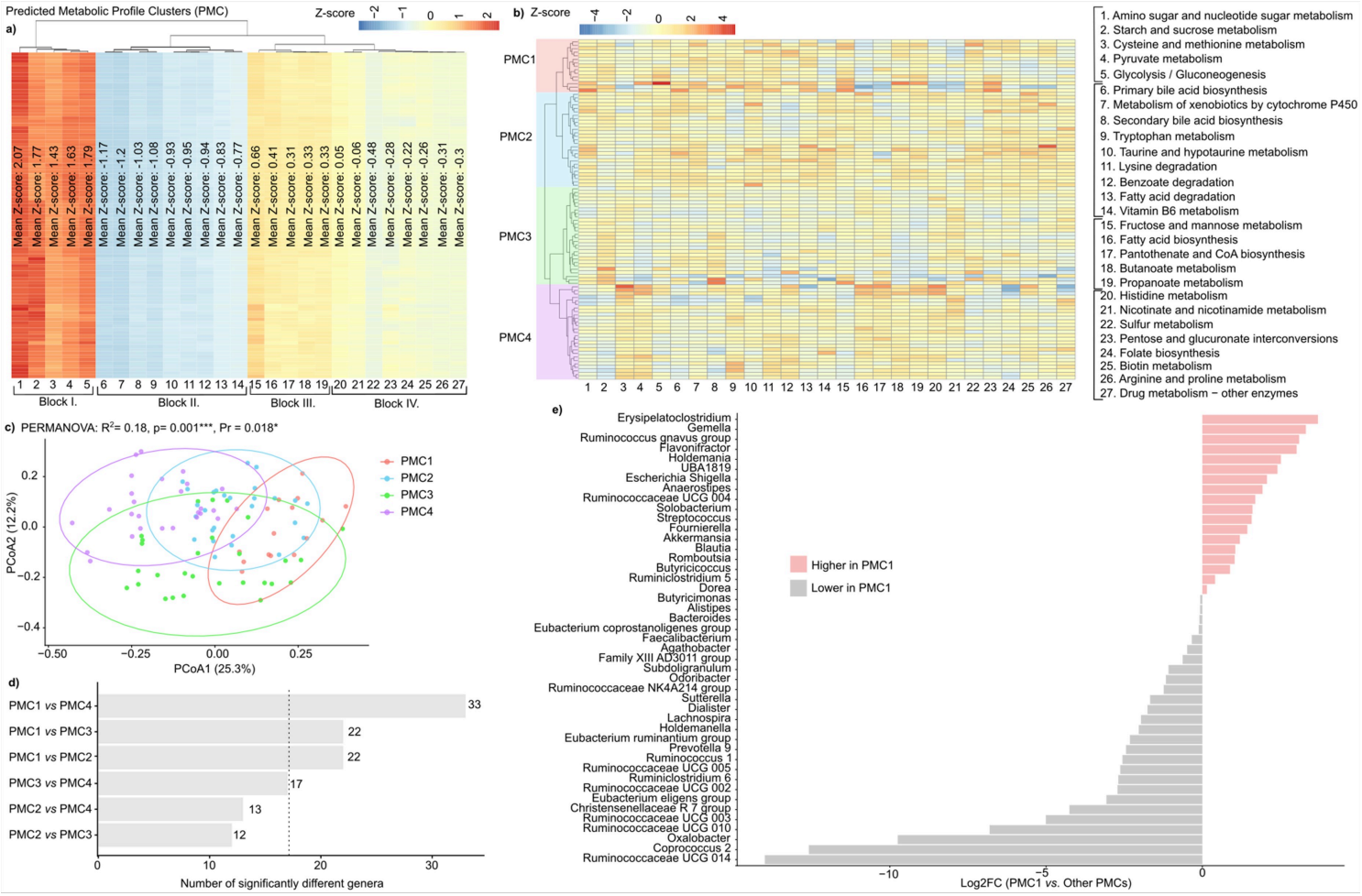


Figure 1

Predicted microbial metabolic profiles, sample clustering, and associated genus-level differences in individual faecal samples.

a) Heatmap of 27 selected metabolic pathways predicted from 16S rRNA gene sequencing data. Rows represent individual samples, and columns represent pathways numbered according to their names. Values were Z-score-standardised within each sample. Hierarchical clustering grouped pathways with similar profiles into four descriptive blocks: I (1-5), II (6-14), III (15-21), and IV (22-27).

b) Sample clustering based on predicted metabolic profiles, delineating four predicted metabolic profile clusters (PMC1-PMC4). Pathway numbering follows panel a. Colour scales are shown separately for panels a and b.

c) Principal coordinate analysis (PCoA), with samples coloured by PMC membership. Each point represents one sample, and ellipses summarise the distribution of samples within each cluster. Axis labels indicate the percentage of variation represented by each coordinate.

d) Number of genera meeting the statistical significance criterion in each pairwise comparison of PMC groups. Genera were first screened by Kruskal-Wallis test across all four PMC groups (BH-corrected q <

0.05, n = 45 genera); pairwise differences among these genera were then assessed by Dunn's post-hoc test with Benjamini-Hochberg correction.

e) Genus-level $\log_2$ fold changes for PMC1 relative to the remaining PMC groups. Positive values (red) indicate higher relative abundance in PMC1, whereas negative values (grey) indicate lower relative abundance.

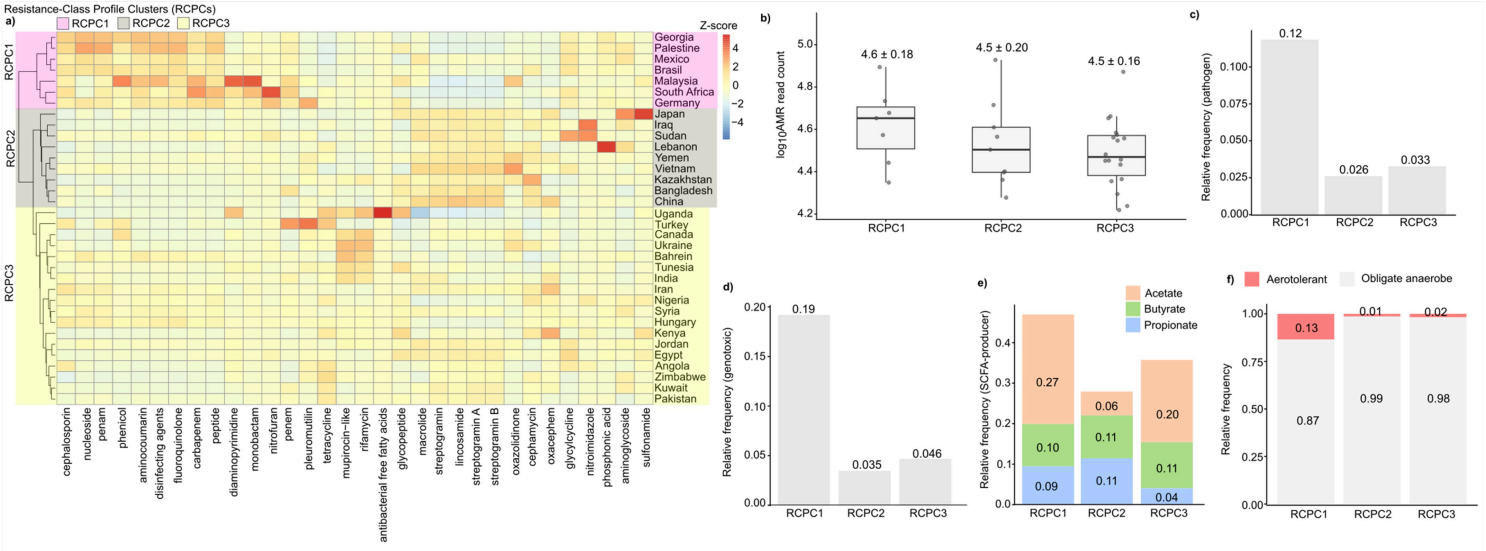


Figure 2

Country-level resistance-class profile clusters and associated microbial taxonomic characteristics. a) Country-level sample clustering based on shotgun resistance-class data, delineating three resistance-class profile clusters (RCPC1-RCPC3). Rows represent country-level shotgun samples, with columns as resistance classes. The colour scale is expressed as Z-score, standardised in each column.

b) Boxplots illustrate the log10-transformed read counts of antibiotic resistance genes in the identified three resistance-class profile clusters (RCPC1-3).

Barplots show the relative abundance of taxa with reported pathogenicity (c), genotoxic potential (d), taxa assigned to SCFA-producing groups (e), and the ratio of literature-based aerotolerant and anaerobic taxa (f).

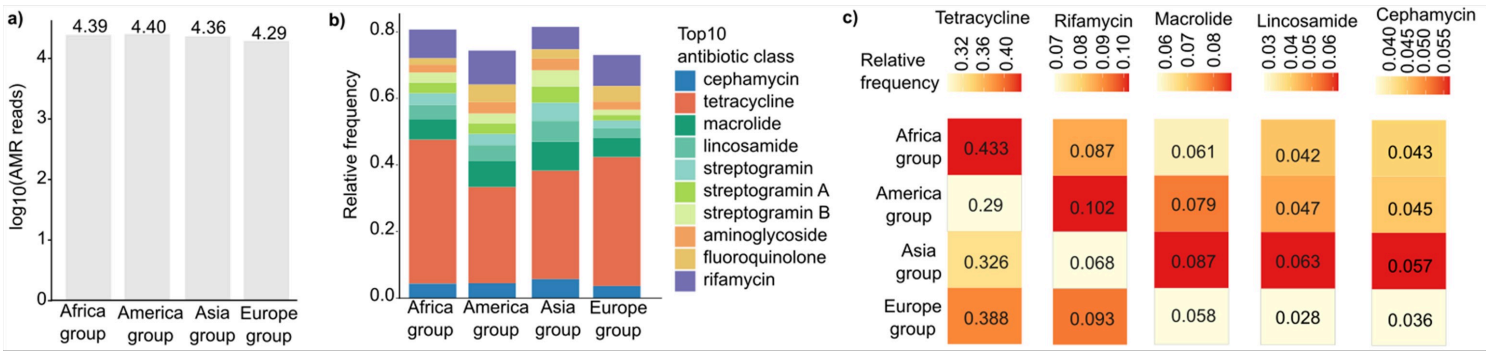


Figure 3

Continental differences in antimicrobial resistance genes. a) Bar plots illustrate the log-transformed read counts of AMRs, aggregated at continental level. b) The relative frequencies of the top 10 antibiotic classes in each continent. c) Heatmap showing cumulative relative frequencies of the top 5 antibiotic classes by continent. Colour scales are adjusted individually per class to highlight relative within-class patterns; ranges are annotated above each panel.

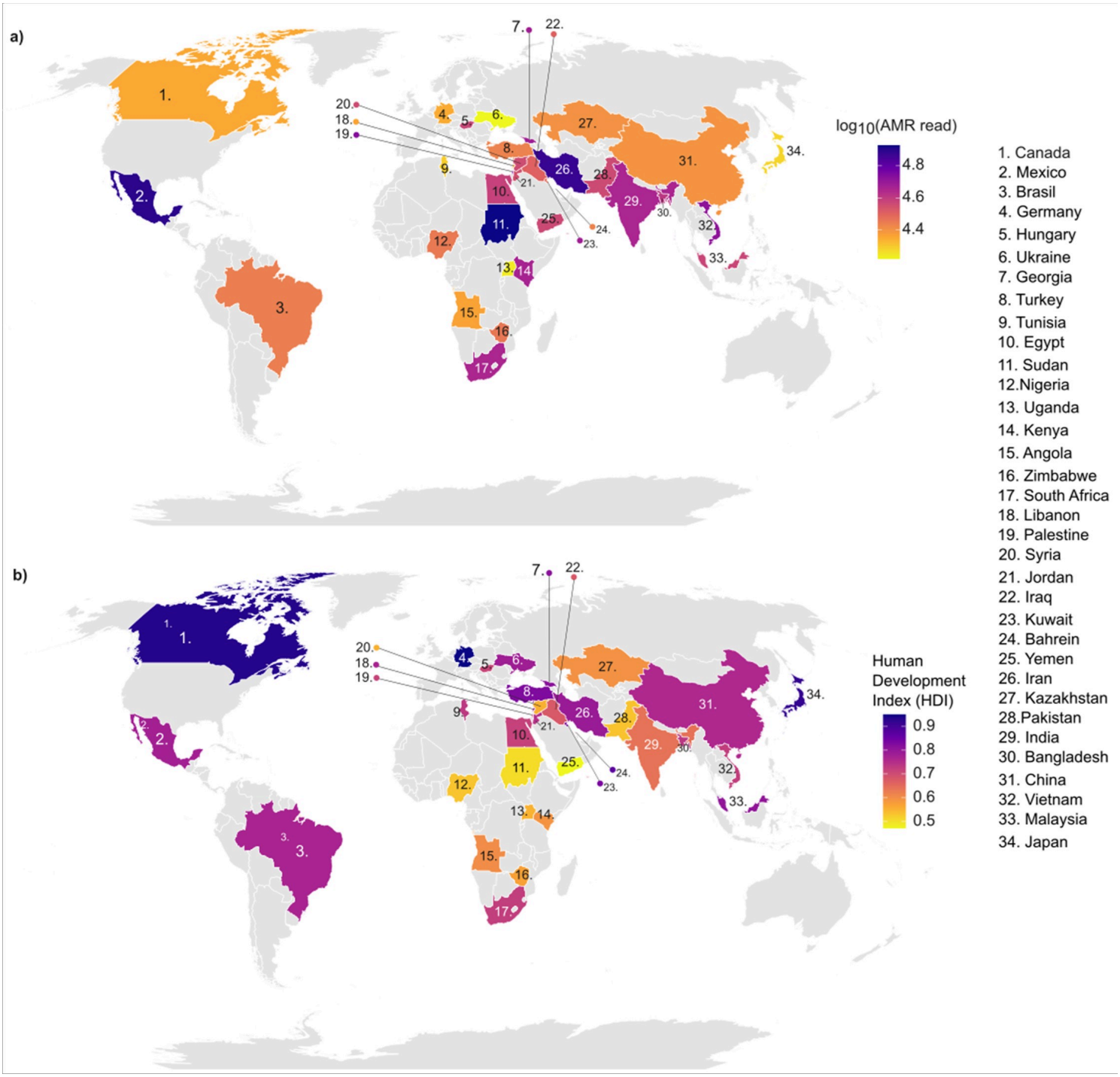


Figure 4

Global differences between resistance burden and development levels across 34 countries. World maps illustrating the a) log10-transformed AMR read counts, and b) Human Development Index (HDI) in our study population.

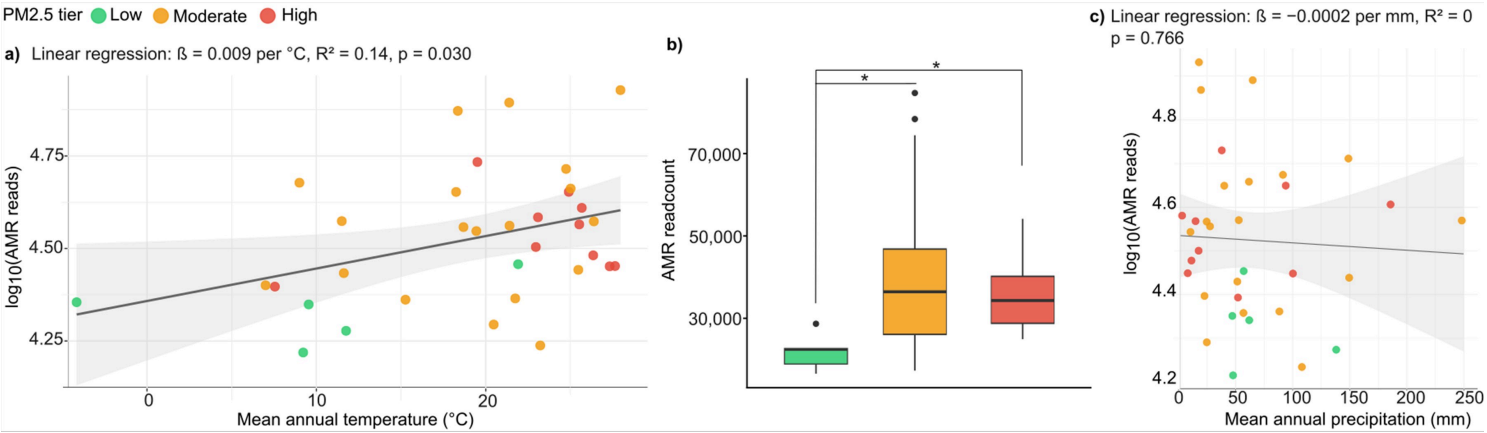


Figure 5

Environmental and climatic factors shaping the resistome. a) Scatter plot of log10-transformed total AMR read counts as a function of mean annual temperature (°C), with linear regression line and 95% confidence interval.

b) Box plot of log-transformed total AMR read counts across three PM2.5 exposure tiers (Low: < 15 µg/m³; Moderate: 15-35 µg/m³; High: >35µg/m³), defined according to WHO Air Quality Guidelines.
c) Scatter plot of log-transformed total AMR read counts as a function of mean annual precipitation (mm), with linear regression line.

Asterisks indicate significant statistical differences (* $p < 0.05$).

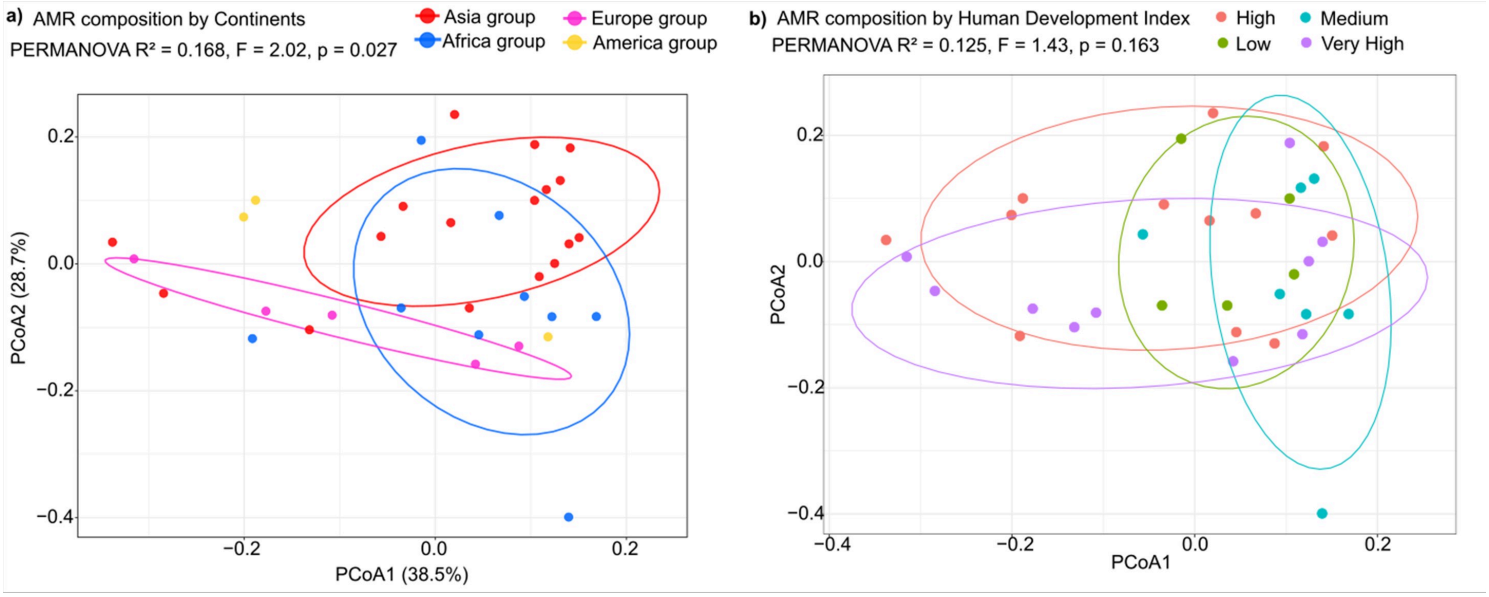


Figure 6

Factors shaping the composition of antibiotic resistance. PCoA plots illustrating the effect of a) continental geography, and b) socioeconomic development, expressed as the Human Development Index, on the composite of the resistome

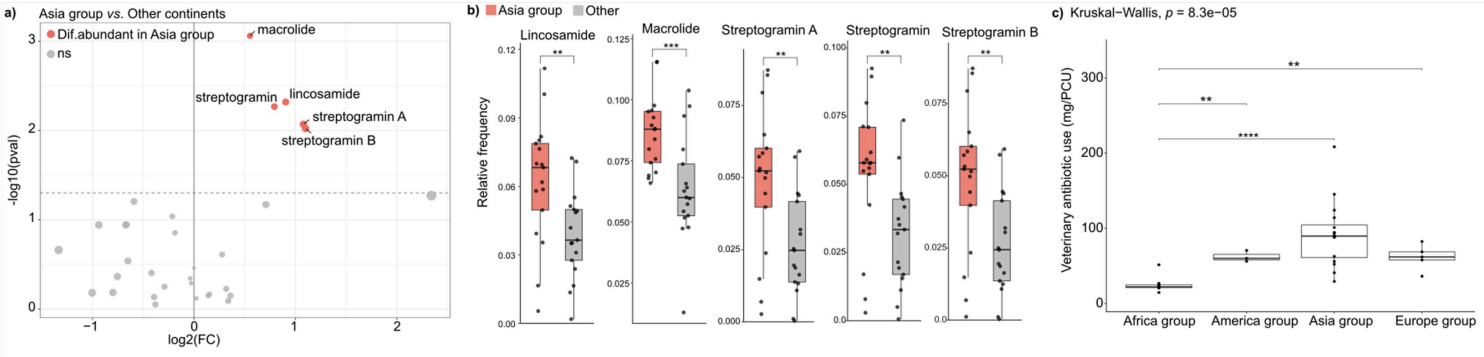


Figure 7

Differentially abundant antibiotic classes between continents and dietary habits. a) Volcano plot comparing Asian to non-Asian samples (per-class Wilcoxon test). Significant MLSB-family classes are highlighted. b) Box plots of MLSB-family relative abundances comparing the Asia group with the other continental groups. c) Regional distribution of veterinary antibiotic use (mg/PCU). Asterisks indicate significant statistical differences (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx