Full length article

# Characterising global antimicrobial resistance research explains why One Health solutions are slow in development: An application of AI-based gap analysis

Cai Chen [a,b,1], Shu-Le Li [a,b,1], Yao-Yang Xu [a,c], Jue Liu [d,e], David W. Graham [f,*], Yong-Guan Zhu [a,c,g,*]

[a] Key Laboratory of Urban Environment and Health, Ningbo Observation and Research Station, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Zhejiang Key Laboratory of Urban Environmental Processes and Pollution Control, CAS Haixi Industrial Technology Innovation Center in Beilun, Ningbo 315830, China
[d] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
[e] Institute for Global Health and Development, Peking University, Beijing 100191, China
[f] School of Engineering, Newcastle University, Newcastle, UK
[g] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

## ARTICLE INFO

## ABSTRACT

The global health crisis posed by increasing antimicrobial resistance (AMR) implicitly requires solutions based a One Health approach, yet multisectoral, multidisciplinary research on AMR is rare and huge knowledge gaps exist to guide integrated action. This is partly because a comprehensive survey of past research activity has never performed due to the massive scale and diversity of published information. Here we compiled 254,738 articles on AMR using Artificial Intelligence (AI; i.e., Natural Language Processing, NLP) methods to create a database and information retrieval system for knowledge extraction on research perfomed over the last 20 years. Global maps were created that describe regional, methodological, and sectoral AMR research activities that confirm limited intersectoral research has been performed, which is key to guiding science-informed policy solutions to AMR, especially in low-income countries (LICs). Further, we show greater harmonisation in research methods across sectors and regions is urgently needed. For example, differences in analytical methods used among sectors in AMR research, such as employing culture-based versus genomic methods, results in poor communication between sectors and partially explains why One Health-based solutions are not ensuing. Therefore, our analysis suggest that performing culture-based and genomic AMR analysis in tandem in all sectors is crucial for data integration and holistic One Health solutions. Finally, increased investment in capacity development in LICs should be prioritised as they are places where the AMR burden is often greatest. Our open-access database and AI methodology can be used to further develop, disseminate, and create new tools and practices for AMR knowledge and information sharing.

## 1. Introduction

The threat of antimicrobial resistance (AMR) to global health is exemplified by progressively increasing infection rates by AMR-carrying bacterial pathogens, creating substantial human, animal, food supply, environmental, and socioeconomic burdens (World Health Organization, 2016). An estimated 1.27 million deaths directly attributed to AMR were seen in 2019 (Murray et al., 2022) and total annual deaths may exceed 10 million by 2050 (O'Neill, 2014). Furthermore, AMR impacts worktime and productivity (Alan Krupnick, 2020) with an estimated 3.5 million additional sick days per year due to waterborne AMR alone, costing $300 million yearly (World Economic Forum, 2021). Growing AMR encompasses the dissemination of antibiotic resistance bacteria (ARB) and antibiotic resistance genes (ARGs) as well as resistance to

fungal, viral and other microorganisms (Larsson and Flach, 2022). Bacterial pathogens, which is the focus here, can develop AMR through mutations or acquisition of ARGs from other bacteria (Collignon et al., 2018). ARB are natural, existing widely across nature in both pathogenic and commensal strains, but readily cross boundaries in host organisms (human beings or animals) and between environments (Berendonk et al., 2015). However, new ARB and ARGs are increasing, largely due to anthopogenic activity that has accelerated microbial evolution, and ARB and ARGs are spreading across sectors and over regions, being transmitted via different vectors (Kumarasamy et al., 2010; Liu et al., 2016). This is why solutions to AMR using One Health approach are critical.

The concept of One Health is holistic and interdisciplinary, grounded in the idea that the well-being of humans, domestic and wild animals, plants, and the broader environment (including ecosystems) is intricately interconnected and interdependent (World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, United Nations Environment Programme, 2023). In accordance with the One Health concept, addressing the challenge of AMR necessitates assessments of AMR "hotspots" and "background" conditions to determine true burdens across regions and sectors (Hernando-Amado et al., 2019). Originating in the early 2000s, the One Health approach recognises the interconnectedness of humans, animals, and the environment (Lee and Brumme, 2012). In 2010, the World Health Organization (WHO), the World Organization for Animal Health (WOAH), and the Food and Agricultural Organization of the United Nations (FAO) started an alliance to address the close cooperation to mitigate the global health challenges of One Health, known as the Tripartite (Food and Agriculture Organization of the United Nations, World Organization for Animal Health, World Health Organization, 2010). However, in 2022, the United Nations Environment Programme (UNEP) was included in the collaboration and formalized as the new Quadripartite group, enabling the new One Health Joint Plan of Action (2022–2026) for addressing global health threats across sectors encompassing human health, animal health, plant health, food systems, and the environment (World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, United Nations Environment Programme, 2022). The new joint action plan for One Health places AMR within wider considerations between human health and ecosystems, requiring integrating more empirical evidence to guide future AMR monitoring and surveillance.

Direct or potential pathways that disseminate AMR have been widely reported in studies (Chen et al., 2016; Qiao et al., 2018; Manaia, 2017). ARB that can colonize and proliferate in humans and ARGs can directly transmit from the environment (e.g., air, water, or soil), animals (e.g., companion animals), or agricultural products to humans (Manaia, 2017; Allel et al., 2023; Zhou et al., 2023). Also, healthcare, community, and animal waste effluents and other pollutant pathways (e.g., microplastics, solid waste leachates, urban runoff etc.) can mediate ARB transfer and retention in the environment, increasing the possibility of evolution and horizontal gene transfer, and the selection of new AMR genes and strains in any location (Berendonk et al., 2015; Zhou et al., 2023; Liguori et al., 2022; Hu et al., 2022).

In parallel, complex but interlinked socioeconomic factors also impact the emergence, transmission, and spread of AMR (e.g., the quality of healthcare systems, the quality of water, sanitation, and hygiene (WASH) infrastructure), particularly in LICs and lower-middle-income countries (LMICs) – places where population sizes and the demand for antibiotics is increasing (Allel et al., 2023); but where intrastructure for environmental protection is lagging behind (Graham et al., 2019). It is estimated that LMICs have the highest AMR infection burden (Murray et al., 2022) and some ARGs, such as $bla_{NDM-1}$ and $mcr-1$, appear to have emerged and spread from LMICs (Kumarasamy et al., 2010; Liu et al., 2016). Within this context, it is urgent that human, animal, plant, agri-food, and ecosystem conditions be investigated relative to AMR in a more holistic manner, including defining the current status and knowledge gaps, essential for future AMR surveillance across sectors and

regions. However, previous studies have often focused on sector or region-specific work, not work aimed at developing solutions to AMR at global scales.

One of the main problems is that the breadth of previous studies and data are massive, and create solutions we need better methods for integrating knowledge, increasing the value of past AMR research and prioritizing new activities. AI technologies (including NLP, Large Language Models (LLM), etc.) are now needed to effectively synthesize evidence, gather information, and mine data and text, enabling a semi-automated way to comprehend the rapidly growing research, which is reflected by peer-reviewed publications over time (Berrang-Ford et al., 2021; Nunez-Mir et al., 2016). NLP represents an interdisciplinary domain positioned at the convergence of AI and linguistics, aiming to empower computational systems to comprehend and analyze statements or lexical units articulated in human languages, enabling the execution of functional operations (Jurafsky and Martin, 2000). This multifaceted field spans a broad spectrum of methodologies within the realm of scientific literature analysis. These methodologies encompass various techniques, including topic modeling (Berrang-Ford et al., 2021; Sietsma et al., 2024), text summarization (Awasthi et al., 2021); and information extraction (Dagdelen et al., 2024; Tshitoyan et al., 2019; Krallinger et al., 2017), contributing to the comprehensive exploration and interpretation of scientific literature (Hirschberg and Manning, 2015).

Here we provide a powerful new tool and database for examining research over time, which considers 254,738 AMR publications from 178 sampling countries for between 2003 and 2023. Based on the database, we used NLP to systematically harvest and sort global AMR research and distill knowledge and information. Comprehensive global maps of AMR research are provided, including extensive detail on previous work performed on AMR pathogens, ARGs, and existing AMR surveillance methodologies around world. This is essential information needed to guide solutions to AMR based on a One Health approach.

## 2. Materials and methods

### 2.1. Search strategy and literature download

The overall search and screening strategy for our NLP-based analytical framework is presented in Fig. 1. Existing peer-reviewed manuscripts in Scopus were first archived and searched using a comprehensive strategy detailed in the Supporting Information (SI). Briefly, inclusion of documents in our "AMR database" was based on titles, keywords, or abstracts that contained "antimicrobial resistance" and related alliance or abbreviations (Chau et al., 2022). Our searches were restricted to the 20 years from 2003 to 2023 and published in the English language. Books (whole books or chapters), letters, notes, conference papers, editorials, and errata were not included (i.e., sources with unknown levels of peer review). After deduplication, information from 254,738 publications was downloaded and archived based on the title, keywords, abstract, DOI, etc. A complete downloaded list is provided in additional file. All the HTML files were fetched and parsed into sub-components (e.g., title, authors, abstracts, and keywords) by analyzing ElementTree with package lxml in Python (Behnel and Martijn, 2005).

### 2.2. NLP and the optimized search system

NLP was utilized to convert the texts into numeric information according to the frequency distribution of words. The NLP processing pipeline in this study contains four steps: tokenization, part-of-speech (POS) tagging, lemmatization, and noun phrase extraction. Tokenization splits the text into meaningful segments non-destructively. After tokenization, POS tagging was compiled to analyze the content in the abstracts by a pre-trained transformer-based artificial intelligence model (Matsuo, 2023). Verbs, nouns, and adjectives were analysed by parsing and making predictions based on both definitions and contexts.
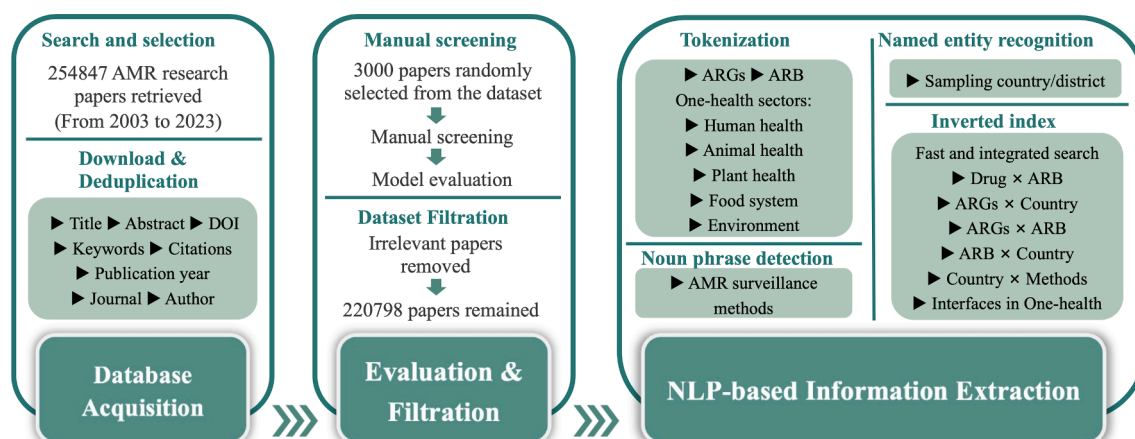
**Fig. 1.** NLP-based framework. The framework contains three parts: dataset acquisition, database evaluation and filtration, and NLP-based information extraction.

Lemmatization converted each word into its root form, including tense (e.g., converted "found"/"finds" into "find"), complex number (e.g., converted "genes" into "gene"), and other grammatical factors. Noun phrases, alias and abbreviation from abstracts was identified and extracted by R package Spacyr (version 1.2.1) (Matsuo, 2023). Detailed extraction procedure was based on the retrieval system and search strategies, which were provided on the following contents.

### 2.3. Construction of information retrieval system and complex search approach

After normalizing the text data in the database by NLP, inverted indexes were built for each noun phrase to construct an efficient information retrieval system. The inverted index data structure allows searching for large volumes of text in the origin documents, and it works by mapping each individual information point (i.e., term, noun phrases or named entity) to the publications in which it occurs. We used Hash-Map to store inverted index, the detailed guidance for users is provided in the user manual.

When profiling AMR research, search "restricts" are sometimes needed to avoid the time delay of simple grid search. We designed the complex search strategies in this study, which could be divided into the following three simple Boolean queries:

Conjunction query: the intersection operation to the two index lists in inverted index. (e.g., searching results for all documents containing both *mecA* and *Escherichia coli* equal to the intersection of indexes of *mecA* and *E. coli*);
Disconjunction query: the union operation to the two index lists in inverted index. (e.g., searching results for all documents containing either *Escherichia coli* or *E. coli*, equal to the union of indexes of *Escherichia coli* and *E. coli*);
Negation query: the complement operation to the index list in the inverted index. (e.g., searching results for all documents excluding *E. coli* equal to the complement of index of *E. coli*).

### 2.4. Text mining and literature categorizing

Text mining was used to extract the information of research country, ARB, and ARGs, and the literature was categorized into different One Health sectors. Additionally, the combination of AMR research terms was extracted by the complex search as above mentioned.

The process was as follows. First, a pre-trained Named-Entity Recognition (NER) model (Matsuo, 2023) was constructed to identify research countries – the named entities – in the abstracts of database. We constructed a country list containing 211 countries with their corresponding synonyms and abbreviations, and 172 research countries (i.e.,

possible sampling countries or surveillance countries in AMR research) were recognized. Income level and region of each country was downloaded from the world bank website (The World Bank Group, 2022).

Grey literature that reports recent guidance on AMR priorities then was reviewed to identify the primary pathogen lists of different countries from websites of WHO, the European Centre for Disease Prevention and Control, the US Centers for Disease Control and Prevention, the Australian Commission on Safety and Quality in Health Care, and the Japan Nosocomial Infections Surveillance, etc. We used The Comprehensive Antibiotic Resistance Database (CARD; card.mcmaster.ca) to download the list of ARGs (Alcock et al., 2023). Based on the new One Health Joint Plan of Action (2022–2026), we established a broad list related to human health, plant health, animal health, environment, and food systems to comprehensively understand the research status of ARB and ARGs. The definitions and Semantic "term" lists for One Health sectors were provided in Supplementary Materials (User Manual).

### 2.5. Screening and validation

Screening and validation of the data from our 254,738 selected publications was performed manually, underpinned by statistical analysis. This was needed because we recognise that the title, keywords and-or abstract of a given manuscript may reflect its true content (results shown in Fig. 2A). Therefore, 3,000 papers were randomly sub-selected from the database and manually verified (World Health Organization, 2016) that an article actually was AMR research and (Murray et al., 2022) how the article should be categorized into the One Health sectors. Through this step, each article got an identification score (0 for false and 1 for true) about their relevance and correctly belonging in a category. The manual scores then were compared with the predicted score to validate the accuracy of AI-based identification.

Two reviewers (Cai Chen and Shu-Le Li) separately screened all titles, abstracts, and full texts. Any disagreement on a specific manuscript for inclusion or exclusion was resolved by a senior researcher (Yong-Guan Zhu). As for AI-based identification, all predicted scores of whether the article was related to AMR research is defined as 1; and all predicted scores of which category or categories the article belongs to are calculated by the presence of AMR words in the abstracts. The quantification methods (Sokolova and Lapalme, 2009; Powers, 2011) and results of bias level in the AI classification results are provided in Tables S1–S3. All subsequent data analysis below is based upon the literature database.

### 3. Results

The NLP AI methodology presented here efficiently and accurately distilled knowledge from the massive available literature (Tables S2 and
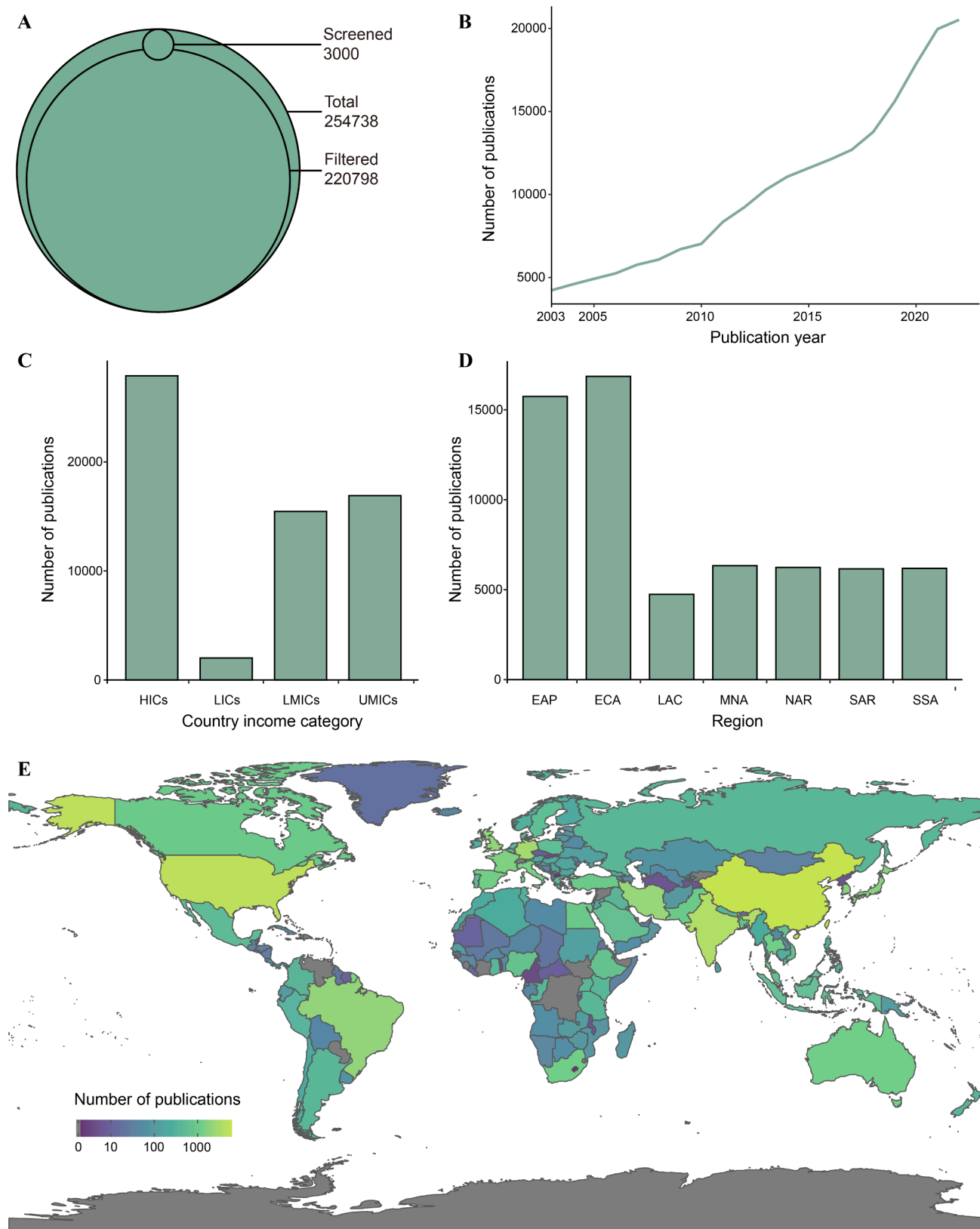
**Fig. 2.** Descriptive summary of publications on AMR. (A) The initial number of publications in the database (254,738), the number of publications by manually screened (3,000), and the number of publications after filtration (220,798). (B) The trend of publication numbers from 2003 to 2022. The publications number of surveillance countries detected in abstracts, categorized by income level (C) and region (D). (E) The global mapping of publication numbers. The region and income level lists are referred to WorldBank categories: HICs, high-income countries; LICs, low-income countries; LMICs, lower middle-income countries; UMICs, upper middle-income countries; EAP, East Asia and the Pacific; ECA, Europe and Central Asia; LAC, Latin America and the Caribbean; MNA, Middle East and North Africa; NAR, North America; SAR, South Asia; SSA, Sub-Saharan Africa.

S3). The searching system took about 0.6 microseconds to extract all related literature for a specific topic, regardless of whether the topic contained a single word or complex combinations of words. Based on our manual assessment of the 3,000 trial papers, the NLP-based model correctly categorized 97.4 % of the papers for AMR and 98.5 % according to One Health sector.

Preliminary examination of the database show that AMR research, as reflected by AMR publications, is growing quickly (Fig. 2B). In 2022, there were 20,519 relevant manuscripts included in the database – 4.86 times more than in 2003 – with the publications growing on average by 8.68 % per year. However, rates of publication from 2018 to 2022 have dramatically increased, growing annualy by 10.49 %.

Using NER methods, we detected identified country names mentioned in abstracts, providing a comprehensive survey of AMR publications and associated research across the globe (Fig. 2C and Fig. 2D). Highlights are as follows. 27,897 (44.8 %) of 62,270 studies including sampling for AMR in high-income countries (HICs), which as almost double that of LMICs and nearly 14 times greater than in LICs. The number of publications from LMICs was close to those from upper-middle-income countries (UMICs).

Publications on AMR displayed large regional differences. The number of publications in East Asia and Pacific (15,746) was close to Europe and Central Asia (16,865), with the Asian and Pacific manuscripts being largely driven by publications from China (close to 40 %). Latin America & Caribbean had lowest number of publications, with only 4,744 in total. In North America, the number of publications was

6,235, which was dominated by the United States – more than 84 % of the regional publcations (5,243).

### 3.1. Pathogens studied across regions and income levels

The resulting pathogens ranked by occurrence frequency in different regions (categorized by region, and income level) are shown in Table 1. Information categorized by country and year can be found in Tables S4–S11. *E. coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa, Salmonella* spp*., and Klebsiella pneumoniae* were the top five reported pathogens across all publications. The Europe and Central Asia region had the highest overall number of publications, i.e., more than half of the pathogens reported in Table 1 are from Europe or Central Asia.

Conversely, the Latin America and the Caribbean region had far fewer publications related to pathogens than other regions – almost all types of pathogens were studied much less than other countries. Publications related to pathogens in North America were about 38.3 % of publications in Europe and Central Asia region (shown in Table 1). However, each region had specific pathogens more commonly studied. *Salmonella* spp., *Acinetobacter* spp., *Shigella* spp., and *Neisseria gonorrhoeae* were more often studied in East Asia and the Pacific compared with other regions. South Asia had the highest number of publications related to *Salmonella typhi*. *Shigella* spp. also was predominant in publications from South Asia.

The countries were categorized by income level and significant income-related effect was observed – the publications related to

**Table 1**
Occurrence of human bacterial pathogens in different countries categorized by region and income level, ordered in terms of frequency of occurrence.

| | East Asia & the Pacific | Europe & Central Asia | Latin America & the Caribbean | Middle East & North Africa | North America | South Asia | Sub-Saharan Africa | HIC | LIC | LMIC | UMIC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Escherichia coli* | 1122 | 1321 | 336 | 783 | 456 | 666 | 564 | 2102 | 176 | 1666 | 1304 |
| *Staphylococcus aureus* | 790 | 1068 | 276 | 723 | 455 | 514 | 666 | 1764 | 215 | 1504 | 1009 |
| *Salmonella* spp. | 1015 | 951 | 372 | 424 | 527 | 600 | 518 | 1737 | 170 | 1338 | 1162 |
| *Klebsiella pneumoniae* | 812 | 817 | 243 | 563 | 229 | 401 | 298 | 1268 | 109 | 1006 | 980 |
| *Acinetobacter* spp. | 755 | 734 | 228 | 601 | 255 | 350 | 171 | 1216 | 76 | 867 | 935 |
| *Pseudomonas aeruginosa* | 533 | 797 | 193 | 592 | 327 | 341 | 261 | 1290 | 64 | 1000 | 690 |
| *Enterobacter* spp. | 363 | 418 | 113 | 240 | 145 | 199 | 156 | 667 | 48 | 498 | 421 |
| *Acinetobacter baumannii* | 413 | 380 | 111 | 369 | 129 | 130 | 70 | 632 | 29 | 436 | 505 |
| *Mycobacterium tuberculosis* | 370 | 307 | 185 | 125 | 63 | 247 | 297 | 345 | 117 | 516 | 616 |
| *Streptococcus pneumoniae* | 318 | 488 | 101 | 147 | 225 | 81 | 150 | 840 | 60 | 293 | 317 |
| *Proteus* spp. | 153 | 282 | 58 | 175 | 70 | 143 | 163 | 379 | 52 | 371 | 242 |
| *Campylobacter* spp. | 201 | 301 | 71 | 80 | 149 | 51 | 80 | 527 | 19 | 177 | 210 |
| *Shigella* spp. | 209 | 106 | 60 | 119 | 55 | 183 | 125 | 221 | 50 | 390 | 196 |
| *Enterococcus faecium* | 158 | 266 | 40 | 90 | 116 | 40 | 35 | 438 | 6 | 129 | 172 |
| *Enterococcus faecalis* | 159 | 220 | 44 | 91 | 87 | 52 | 43 | 357 | 5 | 149 | 185 |
| *Haemophilus influenzae* | 155 | 224 | 41 | 45 | 102 | 32 | 66 | 402 | 23 | 108 | 132 |
| *Neisseria gonorrhoeae* | 178 | 152 | 26 | 16 | 97 | 28 | 41 | 335 | 11 | 69 | 123 |
| *Staphylococcus epidermidis* | 97 | 142 | 42 | 83 | 52 | 30 | 79 | 215 | 19 | 154 | 137 |
| *Proteus mirabilis* | 96 | 147 | 37 | 82 | 40 | 38 | 47 | 212 | 12 | 134 | 129 |
| *Enterobacter cloacae* | 122 | 120 | 40 | 69 | 37 | 42 | 35 | 198 | 11 | 122 | 134 |
| *Helicobacter pylori* | 166 | 128 | 33 | 78 | 17 | 27 | 16 | 226 | 4 | 112 | 123 |
| *Serratia* spp. | 80 | 141 | 51 | 60 | 39 | 38 | 50 | 207 | 13 | 111 | 128 |
| *Salmonella typhi* | 54 | 34 | 8 | 25 | 5 | 211 | 77 | 48 | 16 | 298 | 52 |
| *Streptococcus pyogenes* | 50 | 121 | 21 | 31 | 41 | 22 | 25 | 190 | 10 | 50 | 61 |
| *Streptococcus agalactiae* | 80 | 73 | 35 | 31 | 24 | 9 | 24 | 121 | 8 | 45 | 102 |
| *Serratia marcescens* | 41 | 70 | 23 | 31 | 19 | 15 | 21 | 105 | 4 | 51 | 60 |
| *Klebsiella oxytoca* | 46 | 71 | 17 | 27 | 22 | 14 | 22 | 104 | 9 | 50 | 56 |
| *Stenotrophomonas maltophilia* | 61 | 60 | 13 | 23 | 31 | 17 | 9 | 110 | 7 | 35 | 62 |
| *Vibrio cholerae* | 43 | 12 | 22 | 13 | 4 | 85 | 25 | 21 | 10 | 129 | 44 |
| *Providencia* spp. | 33 | 42 | 19 | 32 | 17 | 20 | 18 | 69 | 4 | 56 | 52 |
| *Morganella* spp. | 31 | 48 | 11 | 26 | 9 | 14 | 12 | 73 | 2 | 41 | 35 |
| *Enterobacter aerogenes* | 29 | 43 | 13 | 22 | 14 | 10 | 13 | 64 | 2 | 35 | 43 |
| *Clostridioides difficile* | 23 | 25 | 6 | 13 | 17 | 3 | 6 | 62 | 0 | 13 | 18 |
| *Salmonella paratyphi* | 13 | 7 | 1 | 6 | 0 | 55 | 5 | 13 | 1 | 64 | 9 |

pathogens in LICs were only 10.8 % of other regions on average. LICs had fewer publications on all types of pathogens with some pathogens not appearing in LIC publications, such as *Clostridioides difficile*.

Conversely, almost all types of pathogens more common in publications from HICs compared with LMICs and UMICs. The exceptions are for *Mycobacterium tuberculosis, Shigella* spp., *S. typhi, Vibrio cholerae*, and *Salmonella paratyphi*. For these pathogens, *M. tuberculosis* occurred most in publications from UMICs, while others occurred more often in publications from LMICs. Specifically, the numbers of publications related to *Shigella* spp., *S. typhi, V. cholerae*, and *S. paratyphi* in LMICs were 1.76, 6.21, 6.14, and 4.92 times higher than those of HICs, respectively. The number of publications related to *M. tuberculosis* of UMICs was 1.79 and 1.19 times higher than those from HICs and LMICs, respectively.

### 3.2. Pathogen-drug studied across regions and income levels

Data for publications related to pathogen-drug combinations from the WHO Global priority list of ARB are provided in Table 2 (additional pathogen-drug combinations can be found in Tables S13–S20). Search results indicate that methicillin-resistant *S. aureus* (MRSA) was the most studied pathogen-drug topic (1,763) across the globe, followed by ESBL-producing *E. coli* (1,193) and carbapenem-resistant Enterobacteriaceae (952). The world map of publications related to MRSA is shown in Fig. 3, with related research expanding across all regions, especially in China, the United States, Iran, and India. The region with the largest amount of MRSA research is Europe and Central Asia (3,740), 1.52 times more than East Asia and the Pacific, the second productive region. As before, the

**Table 2**
Occurrenc of drug-pathogen combinations in different countries categorized by region and income level, ordered in terms of frequency of occurrence.

| | East Asia & the Pacific | Europe & Central Asia | Latin America & the Caribbean | Middle East & North Africa | North America | South Asia | Sub-Saharan Africa | HIC | LIC | LMIC | UMIC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Methicillin-resistant *Staphylococcus aureus* | 303 | 513 | 92 | 306 | 177 | 176 | 196 | 778 | 53 | 568 | 364 |
| ESBL-producing *Escherichia coli* | 224 | 369 | 75 | 184 | 73 | 149 | 119 | 505 | 35 | 388 | 265 |
| Carbapenem-resistant Enterobacteriaceae | 242 | 291 | 66 | 115 | 116 | 86 | 36 | 478 | 13 | 194 | 267 |
| ESBL-producing Enterobacteriaceae | 148 | 304 | 50 | 119 | 60 | 65 | 93 | 404 | 47 | 218 | 170 |
| Vancomycin-resistant *Staphylococcus aureus* | 105 | 187 | 39 | 157 | 87 | 100 | 83 | 290 | 34 | 271 | 163 |
| Penicillin-non-susceptible *Streptococcus pneumoniae* | 149 | 244 | 44 | 87 | 83 | 33 | 66 | 375 | 23 | 140 | 168 |
| Carbapenem-resistant *Pseudomonas aeruginosa* | 134 | 228 | 67 | 98 | 76 | 42 | 30 | 321 | 11 | 143 | 200 |
| Carbapenem-resistant *Acinetobacter baumannii* | 139 | 205 | 52 | 107 | 32 | 45 | 25 | 261 | 9 | 147 | 188 |
| Carbapenem-resistant *Klebsiella pneumonia* | 146 | 166 | 39 | 87 | 39 | 58 | 34 | 211 | 11 | 157 | 190 |
| ESBL-producing *Klebsiella pneumonia* | 95 | 146 | 30 | 120 | 26 | 57 | 55 | 190 | 21 | 194 | 124 |
| Fluoroquinolone-resistant *Salmonella* spp. | 106 | 73 | 27 | 20 | 33 | 102 | 33 | 141 | 6 | 161 | 86 |
| Vancomycin-resistant *Enterococcus faecium* | 68 | 155 | 16 | 50 | 42 | 25 | 16 | 215 | 2 | 76 | 79 |
| Carbapenem-resistant *Enterobacter* spp. | 87 | 128 | 27 | 42 | 37 | 33 | 13 | 173 | 5 | 80 | 109 |
| Carbapenem-resistant *Escherichia coli* | 65 | 86 | 26 | 43 | 46 | 51 | 33 | 150 | 10 | 96 | 94 |
| ESBL-producing *Enterobacter* spp. | 54 | 112 | 16 | 58 | 32 | 36 | 30 | 156 | 12 | 116 | 54 |
| Clarithromycin-resistant *Helicobacter pylori* | 107 | 100 | 20 | 52 | 12 | 13 | 9 | 166 | 2 | 65 | 80 |
| Ampicillin-resistant *Haemophilus influenzae* | 60 | 84 | 12 | 22 | 19 | 8 | 14 | 133 | 4 | 38 | 44 |
| ESBL-producing *Proteus* spp. | 21 | 60 | 11 | 36 | 13 | 18 | 16 | 84 | 3 | 57 | 31 |
| Fluoroquinolone-resistant Campylobacter | 37 | 77 | 11 | 4 | 26 | 6 | 7 | 120 | 0 | 17 | 31 |
| Cephalosporin-resistant *Neisseria gonorrhoeae* | 48 | 37 | 7 | 0 | 24 | 8 | 10 | 78 | 2 | 18 | 36 |
| Carbapenem-resistant *Proteus* spp. | 19 | 52 | 6 | 17 | 12 | 20 | 6 | 56 | 1 | 40 | 35 |
| Carbapenem-resistant *Serratia* spp. | 20 | 30 | 8 | 12 | 5 | 4 | 4 | 43 | 0 | 14 | 26 |
| ESBL-producing *Serratia* spp. | 9 | 37 | 5 | 11 | 6 | 3 | 9 | 46 | 1 | 20 | 13 |
| Fluoroquinolone-resistant *Shigella* spp. | 28 | 7 | 0 | 7 | 6 | 16 | 4 | 19 | 1 | 31 | 17 |
| Carbapenem-resistant *Providencia* spp. | 6 | 8 | 7 | 9 | 6 | 9 | 2 | 21 | 1 | 14 | 11 |
| Fluoroquinolone-resistant *Neisseria gonorrhoeae* | 17 | 11 | 4 | 0 | 5 | 3 | 7 | 21 | 2 | 7 | 17 |
| ESBL-producing *Morganella* spp. | 8 | 14 | 3 | 12 | 2 | 4 | 3 | 24 | 2 | 13 | 7 |
| Carbapenem-resistant *Morganella* spp. | 8 | 10 | 1 | 10 | 1 | 3 | 2 | 16 | 0 | 9 | 10 |
| ESBL-producing *Providencia* spp. | 6 | 6 | 1 | 9 | 0 | 2 | 1 | 11 | 1 | 8 | 5 |
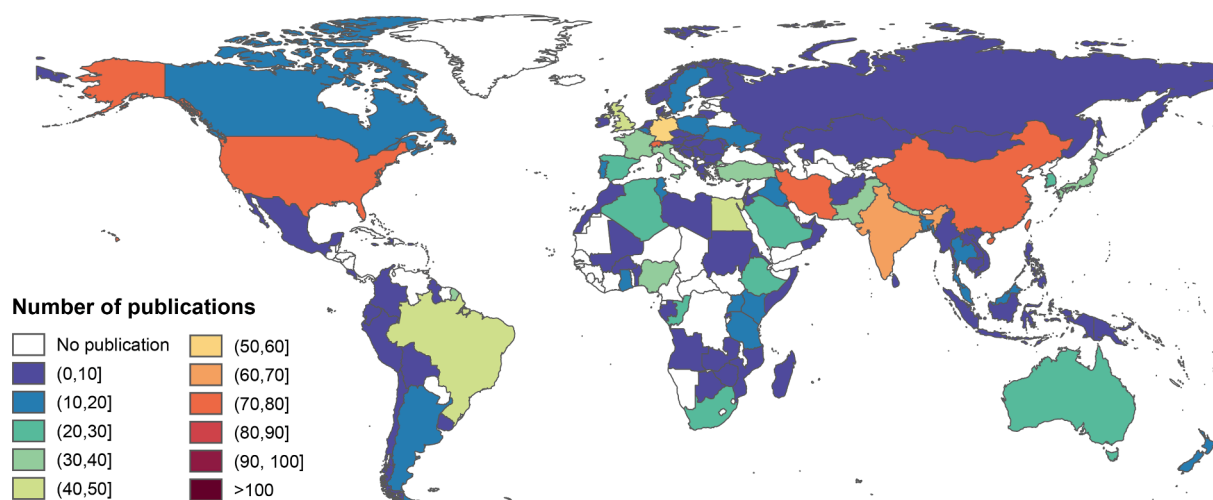
**Fig. 3.** Global mapping of publication numbers related to MRSA.

least number of related publications are from the Latin American and the Caribbean region.

In Europe and Central Asia, studies performed on carbapenem-resistant and ESBL-producing Enterobacteriaceae (including *K. pneumonia, E. coli, Enterobacter* spp., *Serratia* spp., *Proteus* spp., and *Providencia* spp., *Morganella* spp.) were more common than other regions. In East Asia and the Pacific, studies related to drug-resistant *N. gonorrhoeae* (including Cephalosporin-resistant *N. gonorrhoeae* and Fluoroquinolone-resistant *N. gonorrhoeae*) accounted for more than one-third of the total related publications. Furthermore, the proportion of fluoroquinolone-resistant *Salmonella* spp. and fluoroquinolone-resistant *Shigella* spp. related publications in the Asian region (including East Asia and the Pacific and South Asia) was 52.8 % and 64.7 %, respectively, greatly exceeding other regions. Finally, studies on ESBL-producing *K. pneumonia* made up a large proportion of manuscripts from the Middle East and North Africa. In North America, no particular pathogen-drug combination was dominant.

Similar to AMR pathogens, clear differences in predominant pathogen-drug research was apparent with income status, with the number of studies from HICs being much greater than other regions, the same as number as all publications from LMICs and UMICs combined. Publications on pathogen-drug were limited in LICs, which was less than one-tenth of those from other regions. However, fluoroquinolone-resistant *Salmonella* spp. and fluoroquinolone-resistant *Shigella* spp. were particularly prevalent in publications from LMICs. It should be noted that studies on carbapenem-resistant *P. aeruginosa* and carbapenem-resistant *Acinetobacter baumannii*, which are critical ARB on the 2017 WHO Global priority list, were not among the top five pathogen-drug combinations when it comes to published research, suggesting a critical research gap.

### 3.3. Antibiotic resistance genes (ARGs) research across regions and income levels

Global research publications on ARGs, including the survey of ARG prevalence by country and co-occurrence of ARG-pathogen combinations are provided is shown in Fig. 4. The number of studies related to ARGs is dominated by China, the United States, India, Iran, and the Republic of Korea. Among all AMR publications, *mecA* is the most frequent ARG studied across all countries. The greatest number of *mecA* publications are from Iran (95), followed by Brazil (75), the United States (63), Egypt (60), and China (54). However, when one compares *mecA* publications per country based on the total number of AMR publications from that country, Saudi Arabia had the highest local level (9.66 %) followed by Egypt (9.26 %) then Iran (8.36 %), indicating these

are countries where *mecA* research is a high priority. However, this also suggest these countries are centers of experience on *mecA* issues, potentially making them valuable partners in future cooperative research on the topic.

ARG foci vary between countries. For example, China researchers published more on *gyrA*, *sul1*, *sul2*, and *parC* compared with other countries. Further, AMR research focused on *mcr-1* had the largest proportion in China (4.99 %) and also ranked first in Thailand (5.30 %) and Vietnam (7.27 %). In India, the ARG with greatest focus is $bla_{NDM-1}$, more than 6.00 % of AMR research. Conversely, AMR research in Australia has most focused on *vanA* (7.82 %), followed by research on *parC* (4.47 %), *mecA* (3.91 %), and *gyrA* (3.63 %). Details on ARG-related research publications per country and their proportion are provided in Tables S21 and S22.

Research on the co-occurrence of ARGs and pathogens in publications was examined and is summarised in Fig. 4B, which relates to pathogens currently included in the Global Antimicrobial Resistance and Use Surveillance System (GLASS) (World Health Organization, 2023). Co-occurrence publications on *E. coli* and various ARGs is common. As examples, *ampC* was the most frequently reported ARG in studies on *E. coli* with 424, whereas *mcr-1* has been published upon 263 times. Similarly, publications related to *Salmonella* spp. frequently co-occurred with *gyrA* (278), *parC* (186), *sul1* (205), *sul2* (152), and *tetA* (183), and publications related to *Streptococcus pneumoniae* most associated with *gyrA* (62) and *parC* (85). Within the research on *S. aureus*; *mecA* was the most common topic with a frequency of 1250, whereas *Enterococcus faecium* research most frequently co-occurred with *vanA* (shown in Table S23).

### 3.4. AMR research methods used across regions, income levels, and One health sectors

The types of AMR surveillance and analytical methods reported in publications are summarised in Fig. 5 (details provided in the Tables S24–S29). Most research activity over the past 20 years has included some form of antimicrobial susceptibility testing (AST) and PCR-based methods (see Fig. 5A). Whereas, publications including whole genomic sequencing (WGS) methods on ARB isolates have dramatically increased over the last decade, with the number of publications in 2022 being over 25 times greater than one decade ago. Further, the use of metagenomics methods on microbial communities also has emerged over the last decade, especially since 2018 – the number of manuscripts in 2022 being over triple that of 2018, which is probably due to reducing costs and greater accessibility to sequencing methods.
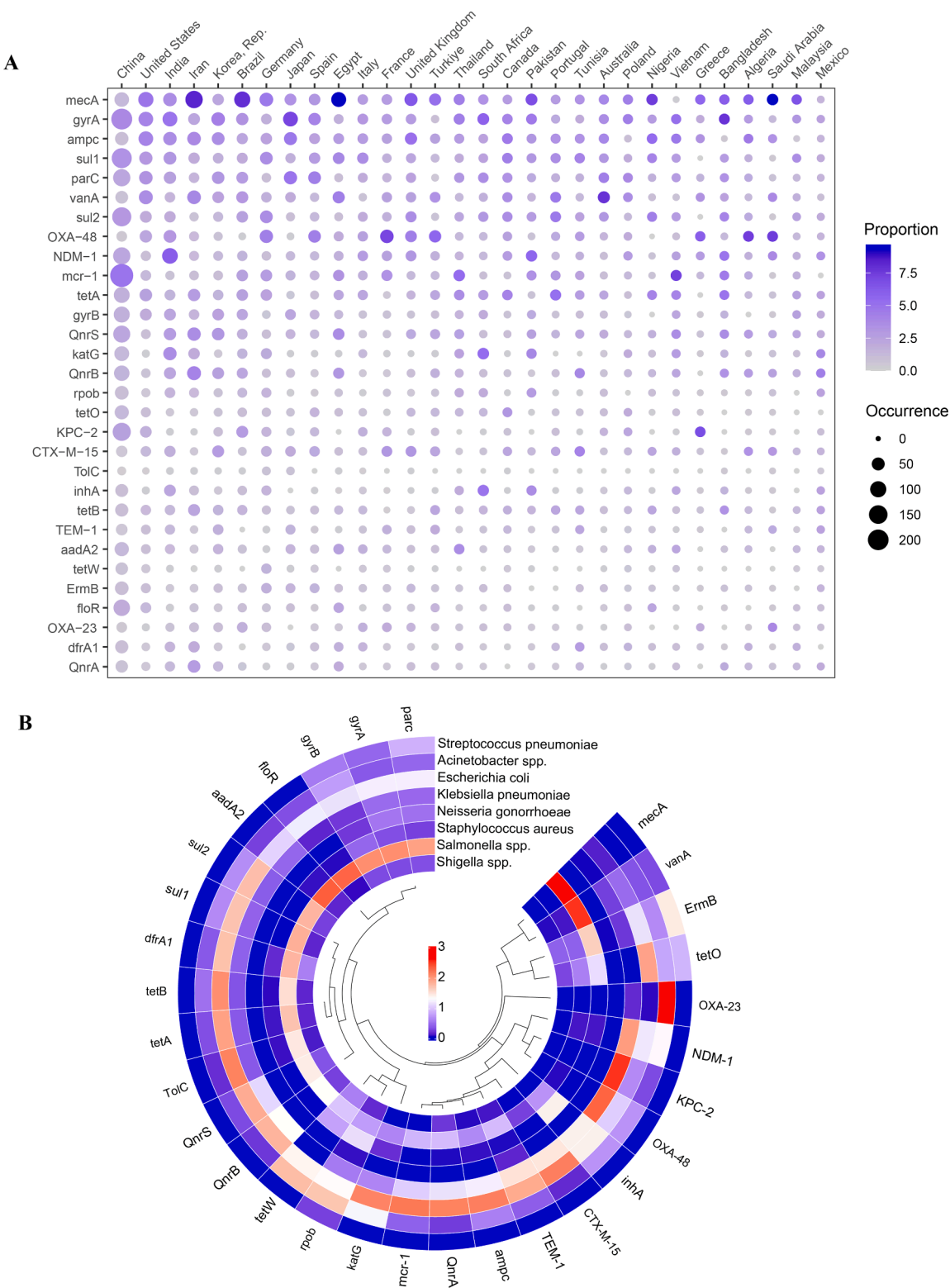
**Fig. 4.** Descriptive summary of ARGs in publications. (A) The bubble plot shows the frequency of ARGs detected in the publications of different countries. The size of the bubble represents the number of publications reporting the ARG subtype in the country. The colour of the bubble represents the proportion that the ARG accounts for all ARGs in the country. (B) The co-occurrence of pathogens and ARGs in publications. The pathogens are included in GLASS. The prevalence of specific ARGs in research publications across countries provides a snapshot of global AMR "hotspots" according to (A). Here are two examples: First, publication on *mcr-1* are very common in China, which is consistent with the fact that *mcr-1*, a plasmid-mediated colistin resistance gene, was firstly isolated from food-producing animals and subsequently in humans in China (Liu et al., 2016; Wang et al., 2017). Alternately, *vanA* research was common in Australian publications, which is consistent with previous reporting (Coombs et al., 2018; The Australian Group on Antimicrobial Resistance, 2021). From 2017 to 2019, the rates of vancomycin-resistant *E. faecium* in Australia ranked first, second, and fourth, compared with the data from European Antimicrobial Resistance Surveillance Network (EARS-Net) countries, and the *vanA*-harbouring isolates were broadly detected.
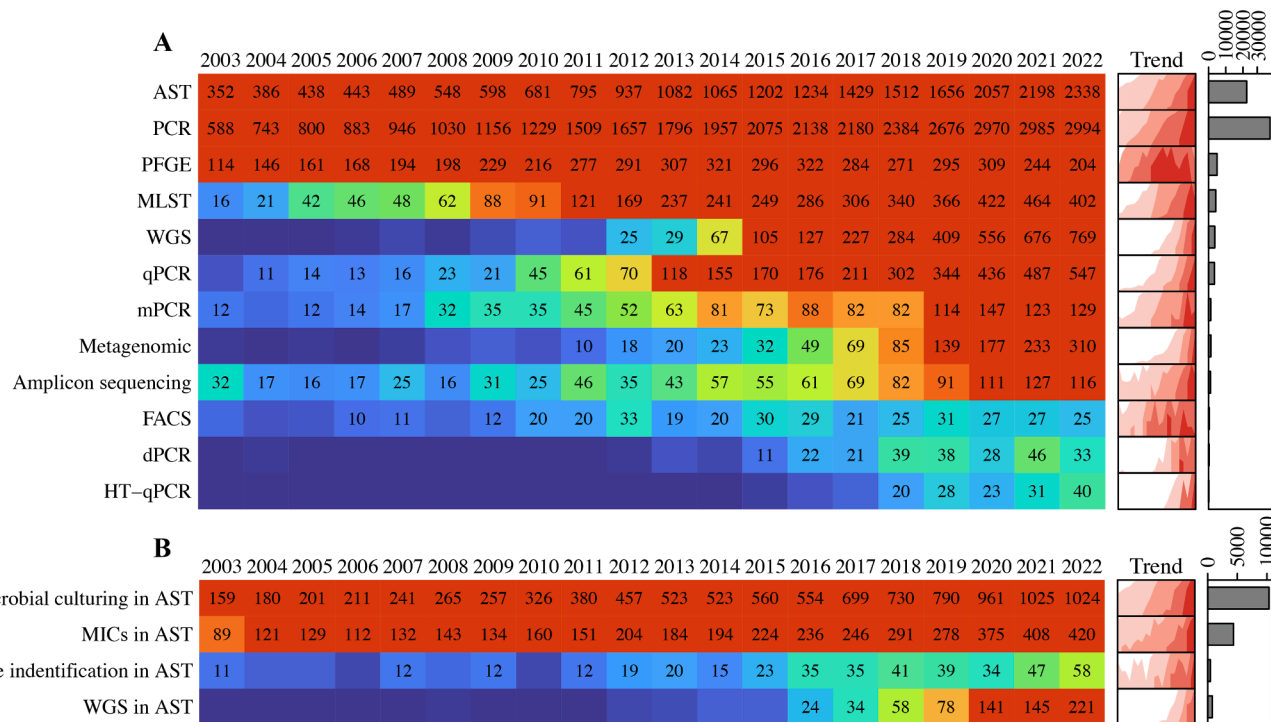
**Fig. 5.** Annual trends of (A) existing AMR surveillance methods and (B) AST-based methods in publications. Due to the fact there are numerous different AST options, the searches were divided four categories to get a better sense of the actual methods employed around the world over time. The AST methods divided into 1) microbial culturing, 2) determination of minimum inhibitory concentrations (MICs), 3) identification of specific isolates (including biochemical testing, mass spectrometry, nucleic acid amplification tests, MALDI etc.) and 4) WGS (WGS was not included in pathogen identification because it has different purposes in AST, such determining resistances genotypes and their genetic context within host isolates).

Most AST-related studies center on culture-based methods (Fig. 5B), with the number of AMR publications with microbial culturing rapidly increasing over the last 20 years. Conversely, although research that includes MIC determination has increased over time, the rate of increase is much lower than studies that employ culturing without MIC determination. However, AMR methods used in published research has changed dramatically over the last five years, with WGS being much more frequently combined with AST testing. However, specific isolate identification has not increased in tandem, with publications using WGS being four times greater than manuscripts reporting isolate identification, which is a critical research gap, especially in the environmental sector.

AMR surveillance methods were examined in publications from specific countries and national income status dramatically impacted methods employed (shown in Table S27). Publications employing "traditional AST-based methods" and PCR were common across income levels. However, the number of publications including WGS were higher in HICs (925) – more than the sum of LMICs (302), UMICs (527), and LICs (28) combined. Similarly, metagenomics-based research has expanded in HICs, with 1.46 times more publications than the sum of LMICs, UMICs, and LICs. The publications using high throughput qPCR (HT-qPCR) has recently and dramatically increased in UMICs, particularly driven by research from China.

Results of existing AMR research methods across One Health sectors (i.e., human health, plant health, animal health, environment, and food systems) was provided in Tables S28 and S29. The AST-based methods were most frequently used in human health sector, with a proportion of exceeding one third of total research. Metagenomics methods are most frequently used in environmental sector research, close to four times greater than in human health sector. Additionally, more than a half of research on plant health sector was conducted with PCR-based methods, sometimes combined with AST-based methods and qPCR methods. Meanwhile, the amplicon sequencing methods were more frequently

used in plant health sector than other sectors.

### 3.5. Stages of AMR research over the last 20 years

Using hierarchical clustering, we found that the development of AMR research can be divided into three stages: 2003–2011, 2012–2018, and 2019–2022 (Fig. 6A). The total number of publications for all five major sectors increased across the stages. However, the methods employed in each sector differed.

To understand underpinning patterns across the three stages, co-occurrence network patterns for the intra-sectors and inter-sectors were characterised (Fig. 6B-G). From stage 1 to stage 3, the total number of edges grew by more than four times, which reflects the sharp rise in the AMR research. Specifically, inter-sector edges increased by 3.5 times, reflecting broad and growing associations in AMR publications between sectors. Additionally, the weighted edges for inter-sector and intra-sector in stage 1 were 3,751 and 4,853, respectively, while the numbers in stage 3 were 29,516 and 24,715, suggesting much higher levels of cross-sectoral activity in the last five years. In fact, the number of publications crossing inter-sectors grew 1.3 times more from stage 1 to 3 than intra-sectoral publications, confirming slowly increasing levels of multi-sectoral research.

Relative to sector-specific activity, publications related to human health predominated AMR research from stage 1 to stage 3, with publication numbers of 3,451, 7,658, and 13,323, respectively. Other sectors also had increasing levels of AMR research, e.g., manuscripts related to animal health and environment increased in stage 3 vs stage 1 by seven and 14 times, respectively. When it comes to the overlap of research between the two sectors, publications between plant health and other sectors increased sharply for stage 1 to stage 3, especially overlapping research between the environment and plant health sectors, growing by over 21 times. Over this period, the number of publications related to plant health itself grew by 5.67 times. Further, research
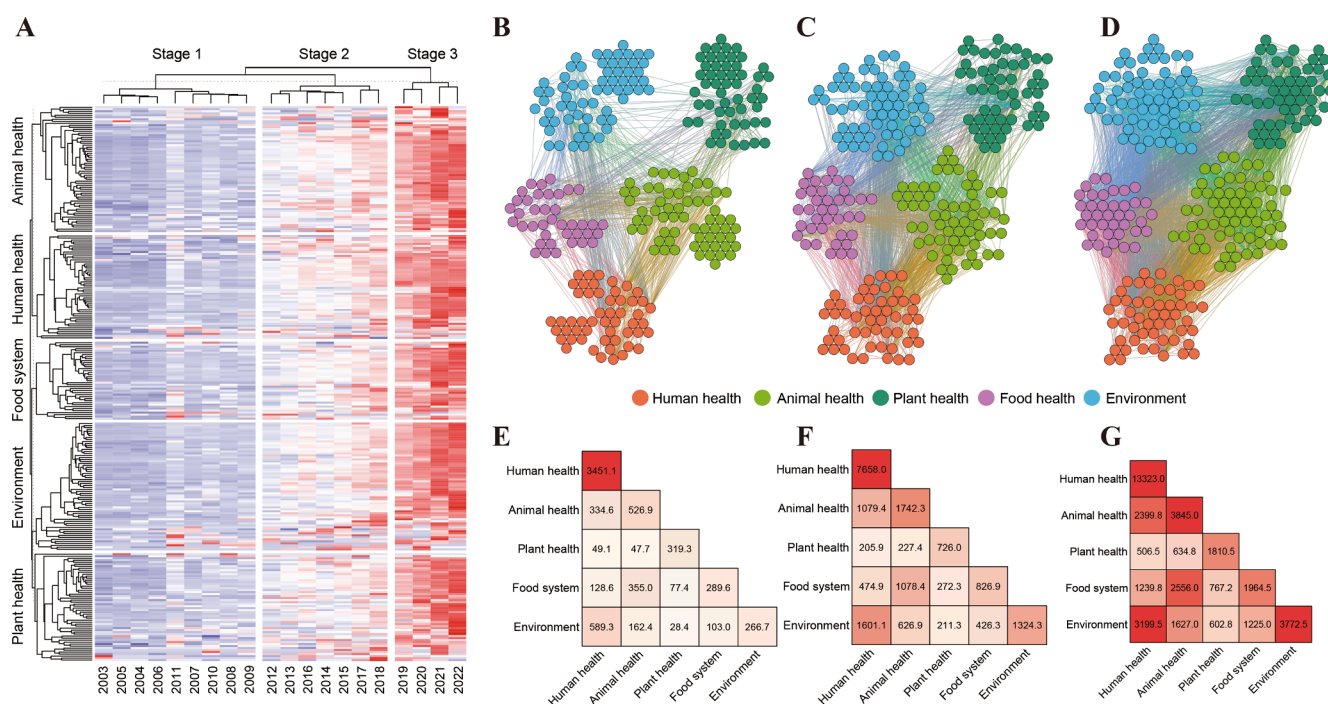
**Fig. 6.** AMR research under One Health sectors The One Health sectors contain human health, plant health, animal health, environment, and food systems. (A) Annual publication number heatmap of One Health terms, normalized by row. Three stages are determined by hierarchical clustering: stage 1, 2003–2011; stage 2, 2012–2018; stage 3, 2019–2022. (B-D) Co-occurrence networks of publications in One Health sectors in stage 1 (B), stage 2 (C), and stage 3 (D). In the co-occurrence network, each node represents a distinct term, with the color of the node indicating the One Health sector to which the term belongs. When terms co-occur within a single publication, an edge is formed between the corresponding nodes. Specifically, an edge within the same sector signifies that the term is exclusively associated with that sector within the publication. Conversely, edges between different sectors indicate terms that are simultaneously associated with both sectors within the publication. Additionally, the weight of an edge represents the average number of publications involving the respective terms within or between sectors at a given stage. (E-G) The heat matrix of the average number of intra-sectoral and inter-sectoral publications per year. Detailed information is provided in Table S30.

conducted on the interface of the environment sector and food system sector increased by 4.14 and 11.9 times in stage 1 to stage 2 and stage 1 to stage 3, respectively.

In stage 3, the most common inter-sectoral research was at the interface of human health and environment (3,199), followed by research at the interface of the animal health and food system sectors (2,556), and the interface of the animal health and human health sectors (2,399). An interesting finding was that the publication numbers at the interface of the food system and animal health sectors exceeded the number of publications on food system themselves, which was not seen in any other combinations between sectors. Finally, although the interface of plant health research with other sectors grew sharply, the total amount of publications was low, e.g., publications at the plant and human health interface was one-sixth of those at the environment and human health interface.

Numbers of mono-sector and multi-sector publications were compared in stage 3 to assess the extent of current One Health research (Fig. 7). We found 33 publications contained all five sectors and there were 426 publications containing four sectors (i.e., mostly crossing the human health, animal health, food system, and environment sectors; other combinations of four sectors were >100). Publications straddling animal health, food system, and the environment were greatest among three-sector publications.

Among all the publications that contained more than two sectors (2,239 in total), more than 40 % were conducted in HICs. The number of publications contained more than two sectors from LMICs (601) was close to those from UMICs (660), while the number of LICs was 59. Interestingly, no publications in LICs including all five sectors. More than 60 % of the top 100 most cited publications were related to HICs; in contrast, only one publication was related to LICs.

## 4. Discussion

AI-based methodologies were employed to interrogate the huge bank of literature, using NLP to map global patterns in AMR research. The work demonstrates how AI can be used consolidate and help interpret large-scale scholarly activity, in this case identifying knowledge gaps in past AMR research and motivating greater One Health focused work in the future. Sifting and consolidating large literature databases is a huge challenge for humans because of hundreds of thousands publications. However, our innovative AI approach rapidly, accurately, and efficiently categorized information from diverse sources, systematically characterizing AMR research around the world over time, across income statuses, and One Health sectors. Our goal was to understand what and how AMR research was being performed now to identify knowledge gaps and guide future integrated One Health research activity.

The AI work provided many valuable results. First, among the top five pathogens studied in publications, four were ESKAPE pathogens (i. e., *S. aureus*, *P. aeruginosa, Salmonella* spp*.,* and *K. pneumoniae*) (Rice, 2008); which are common but concerning strains that can become multidrug-resistant and virulent (Semenec et al., 2023). These pathogens and *E. coli*, the most widely studied bacterial genus, are widely studied by HIC researchers, reflecting the interest these bacteria from a research perspective, where greater resources are available for research. However, *Salmonella* spp. was the most studied pathogen in LMICs, especially in Asia (including East Asia, the Pacific, and South Asian regions), and the most studied pathogen in South Asia was *S. typhi*. Such foci are consistent with regional health issues, especially where resources are more limited for research. Such regional focus is logical and justified.

Typhoid and paratyphoid infections (i.e., human diseases resulting from *Salmonella* spp. infections) pose a great burden of illness and death
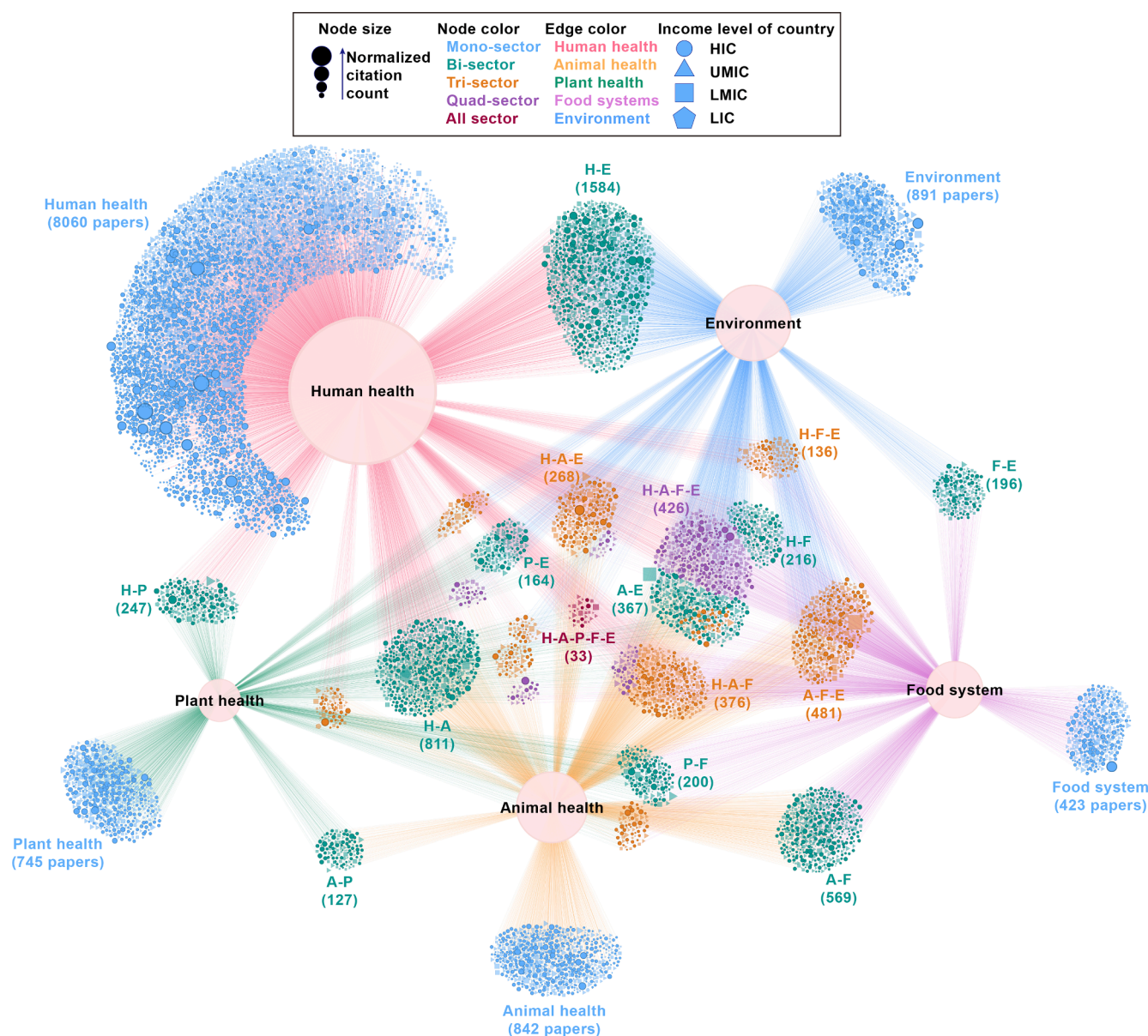
**Fig. 7.** Membership-based network of publications in stage 3 (2019–2022) distributed in the sector-based groups in One Health. The size of the big pink circles reflects the number of associated papers from the five sectors. The publications are grouped by interface of One Health sectors: H, human health; A, animal health; P, plant health; F, food system; E, environment). The colour of the node represents the number of sectors that were conducted in the publication. The size of the node represents the citation number of the publication. The shape of the node represents the different income levels of the surveillance country in the publication. The colour of the edge represents the One Health sector that the publication belongs to. Groups with more than 100 papers are labeled, except for publications conducting all One Health sectors. Detailed information is provided in Table S31. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in LMICs, largely due to poor quality water, inadequate sanitation, and hygiene infrastructure, especially impacting children (Crump et al., 2004). Typhoid fever is estimated to cause over one hundred thousand deaths per year in LMICs (Stanaway et al., 2019); and is particularly problematic in South and Southeast Asia (Cao et al., 2021; Walker et al., 2023). However, when fluoroquinolone was recommended for typhoid treatment in the 1990s, fluoroquinolone nonsusceptibility (FQNS) resulted in mutations in quinolone resistance determining regions (QRDR) in *S. typhi*, which are now ubiquitous across South Asia (da Silva et al., 2022; Mogasale et al., 2014).

Similarly, our results showed that *Salmonella* spp. frequently co-occurred with fluoroquinolone-resistant genes (*gyrA*, *parC*). In general terms, the AI work on global research patterns verified that South Asia is a critical hub for the generation and spread of AMR in general terms (da

Silva et al., 2022). Therefore, more resources are urgently needed for this region to control typhoid and its consequences, which our results on local research foci confirm. However, research experience from this region, similar to other regions with dominant local problems (e.g., *mecA* in Middle Eastern countries), a valuable resource to guide future activity within the region and elsewhere in the world on the topic.

Other patterns are apparent. For example, sulfonamide-resistant genes (*sul1*, *sul2*), and tetracycline-resistant genes (*tetA*), which are often result from the use of broad-spectrum antimicrobial agents in veterinary treatments, were frequently detected with *Salmonella* spp.. Although speculative, extensively drug-resistant (XDR) *Salmonella* spp. appeared to emerge from Pakistan and spread to other countries (Nizamuddin et al., 2021; Bharathan and Kurian, 2020; López-Segura et al., 2019); reflecting the need for controlling the antibiotic use in

livestock and poultry breeding to prevent and mitigate the overbreak of *Salmonella* spp. in animal-based food with more newly emerging drug resistance caused by selection pressure.

Results found publications related to fluoroquinolone-resistant *Shigella* spp. were often from in the Asia region; however, one must recognise that Africa also has a great shigellosis disease burden and risks exist of intercontinental spread of fluoroquinolone-resistant *Shigella* spp. from South Asia (Stenhouse et al., 2023; Kahsay and Muthupandian, 2016; Chung The et al., 2019). This means that our AI approach based on literature alone cannot learn everything, i.e., in places with limited research capacity, observing limited disease burden may be due to a lack of resources for research and nothing is reported. In this case, greater investment in research and surveillance is needed to determine the epidemiology of *Shigella* spp. in Africa. The broad lack of publications from Africa suggest a major research gap, suggesting it should be the focus for new funding and work, especially cross-sectoral integrated research. This gap is confirmed by the limited number of integrated AMR studies in African LICs.

Overall, patterns of AMR research in different low-income settings tend to be region-specific, closely related to local healthcare priorities. Conversely, research in HICs tends to be broader. Despite region-specific activity, the highest AMR burden exists in LICs and LMICs (Murray et al., 2022; Kotloff et al., 2013; Fuhrmeister et al., 2023); i.e., despite known regional needs, problems like limited laboratory infrastructure, incomplete NAPs, and-or inadequate financial support restrict locally important research.

This leads to another important finding of this study, which is the dramatic differences in AMR quantification and characterization methods used across country groups. For example, Iran has the largest number of studies on *mecA*, with most of the publications related to *S. aureus*. However, more than 82 % of the publications in Iran were conducted using only PCR-based methods, indicating that the detection of *mecA* gene or antibiotic-resistant strains was not performed using more advanced WGS of isolates or HT-qPCR methods, both of which provide richer context of associated AMR. The burden of MRSA is very high in Iran (Murray et al., 2022; Antimicrobial Resistance Collaborators, 2022); suggesting that advanced testing methods would be invaluable, although there are socioeconomic and political explanations for this gap (World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, United Nations Environment Programme, 2022). In contrast, in China, AMR characterization that employed only PCR-based methods was in around 43 % of the studies. More often, other AMR research methods, such as WGS and metagenomics, were also included in assessments. Further, HT-qPCR methods for the detection of ARGs, which is among the newer and most promising qPCR technologies (Zhu et al., 2013); were mostly from China.

There are innovations in genomic sequencing and analysis, including metagenomics of clinical samples and environmental samples, tracking of ARGs and plasmids, and machine learning for predicting resistance mechanisms and resistance rate changes (Wheeler et al., 2023). However, these new technologies for AMR research seldom occur or are used in LICs. Therefore, reducing imbalances in capacity between LICs and HICs, e.g., through initiatives such as increasing co-developed research outputs, advancing technological capacity, and fostering more holistic AMR surveillance planning, is crucial for addressing AMR on a global scale. This is due to the interconnectedness of sectors and countries, which is a primary driver behind the ongoing increase in AMR worldwide, along with the urgent necessity to promote local research ownership within each country. Capacity building, such as investing in new laboratories, instruments, and equipment, providing guidance and training, and improving integrated monitoring, epidemiology, and surveillance is critical. The goal is to make all countries and regions sustainable within their own domains (Laxminarayan et al., 2013).

On a positive note, our findings show a clear trend in AMR research towards One Health approaches, although this trend is most evident in progressively higher income settings. Increasing One Health thinking initiated in 2011, which potentially explains this international response, highlighted by Tripartite (the alliance of WHO, WOAH, and FAO) (Food and Agriculture Organization of the United Nations, World Organization for Animal Health, World Health Organization, 2010). Although One Health approaches were slow to be applied, it is clear that by 2019 research that includes human health and other sectors were taking hold, especially in the interface between human health and environment. This is possibly a result of increased awareness of the impact of environmental factors on human health due to the COVID-19 pandemic (Lefrançois et al., 2023). However, there are clear gaps in cross-sectoral research, especially linking plant health and others. Specifically, cross-sectoral research including plant health rarely including more than one other sector. This is concerning because human health, animal health, and security of food systems are almost certainly directly or indirectly impacted by plant health, although research at related interfaces has been limited. Food and phytomedicine provision for human and animals includes niches for microorganisms, managing ecosystems and sustainable agriculture (Destoumieux-Garzón et al., 2018; Rizzo et al., 2021; One Health High-Level Expert Panel (OHHLEP) et al., 2022).

It should also be noted that the culture-based AST methods mostly conducted in human health sector, while metagenomics methods dominate the environment sector. These trends are not surprising since culture-based methods provide essential phenotype information needed for clinical decision-making, such as MIC data on specific isolates, the gold standard for quantifying the degree of resistance in an isolated pathogen. However, combining culture-based methods with WGS provides a richer narrative on resistance mechanisms, which may become central to AMR surveillance in future (Grundmann et al., 2017). The culture-based methods may be generally time-consuming (World Health Organization, 2019); however, most of national and regional surveillance databases center on phenotypic AST data.

In contrast, metagenomics methods have been used in environmental samples to provide deeper insights into the genetic character of environmental resistance without bias caused by culturing (over 99 % percent of microbes in the natural environment are not culturable (Amann et al., 1995), but such methods are rarely combined with WGS of environmental isolates that allow genetic comparisons between clinical and environmental isolates of specific medical concern (Bengtsson-Palme et al., 2017). Validation of new or novel resistance mechanisms rely on the analysis of specific AMR isolates because molecular or metagenomics methods still say little about actual phenotypic resistance (Larsson and Flach, 2022; World Health Organization, 2019). As such, there is an urgent need to combine culture-based methods with non-culture-based molecular AMR methods across all sectors to provide consistent and more comprehensive data that is essential for One Health solutions. We feel one of the main reasons why environmental AMR surveillance data has not been recognised as important to healthcare researchers, who primarily rely on culture-based methods, is that the methods between the two sectors do not intersect. Harmonization of methods between sectors, therefore, is critical.

Although our work provides critical information needed for policy- and decision-makers at regional and international levels, the study has some limitaitons. For example, due to the limitations in data availability, it was not possible to perform full-text interrogation of all the articles assessed here. Further, one bibliographic database (Scopus) was employed to extract titles, abstracts, and keywords, which may cause some bias by decreasing the number of documents identified through the NER process. An example is due to sometimes imprecise terms relative to sampline in manauscript abstracts. Also, searches strategy was restricted to English manuscripts, which may result in a language bias. Finally, definitions of One Health are often not very specific. Therefore, terms used in each sector and complex relationships in definitions across sectors need refinement in future version of our analysis.

However, this study provides consequential evidence of global AMR

patterns based on research trends, and importantly, a powerful way of combining and integrating data from thousands of disciplinary, cross-disciplinarity, and cross-sectoral studies. Our analysis is the first to collate and analyse global AMR research using dispassionate NLP methods and provides a platform for equitable data sharing now and in the future. This means our databases can updated and reinterrogated over time as AMR research proceeds, providing a method of quantifying trends in AMR transmission and spread across sectors and countries. This clearly will be invaluable in policy guidance and creating governance frameworks for global collaboration on the AMR threat. Finally, our study shows what is possible when AI-based methods are used in a beneficial and socially constructive manner.

## CRediT authorship contribution statement

**Cai Chen:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Shu-Le Li:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Yao-Yang Xu:** Supervision, Project administration, Funding acquisition, Conceptualization. **Jue Liu:** Writing – review & editing, Supervision. **David W. Graham:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Yong-Guan Zhu:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2024.108680.

## References

One Health High-Level Expert Panel (OHHLEP), Adisasmito, W.B., Almuhairi, S., Behravesh, C.B., Bilivogui, P., Bukachi, S.A., et al., 2022. One Health: a new definition for a sustainable and healthy future. PLoS Pathog. 18 (6), e1010537.

Alan Krupnick, M.L.C., 2020. Primer on Costs of Action/Inaction and Communication to Policymakers.

Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya Amogelang, R., Wlodarski, M. A., et al., 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 51 (D1), D690–D699.

Allel, K., Day, L., Hamilton, A., Lin, L., Furuya-Kanamori, L., Moore, C.E., et al., 2023. Global antimicrobial-resistance drivers: an ecological country-level study at the human–animal interface. Lancet Planet. Health. 7 (4), e291–e303.

Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59 (1), 143–169.

Antimicrobial Resistance Collaborators, 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet (London, England). 399 (10325), 629–655.

Awasthi, I., Gupta, K., Bhogal, P.S., Anand, S.S., Soni, P.K. (Eds.), 2021. Natural language processing (NLP) based text summarization - a survey. In: 2021 6th International Conference on Inventive Computation Technologies (ICICT); 2021 20-22 Jan.

Behnel, S.F., Martijn, Bicking, Ian, 2005. lxml: XML and HTML with Python.

Bengtsson-Palme, J., Larsson, D.G.J., Kristiansson, E., 2017. Using metagenomics to investigate human and environmental resistomes. J. Antimicrob. Chemother. 72 (10), 2690–2703.

Berendonk, T.U., Manaia, C.M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., et al., 2015. Tackling antibiotic resistance: the environmental framework. Nat. Rev. Microbiol. 13 (5), 310–317.

Berrang-Ford, L., Sietsma, A.J., Callaghan, M., Minx, J.C., Scheelbeek, P.F.D., Haddaway, N.R., et al., 2021. Systematic mapping of global research on climate and health: a machine learning review. Lancet Planet. Health 5 (8), e514–e525.

Bharathan, G., Kurian, B., 2020. First reported case of ceftriaxone-resistant typhoid fever in the middle-east. Int. J. Travel Med. Glob. Health 9, 39–41.

Cao, Y., Karthikeyan, A.S., Ramanujam, K., Raju, R., Krishna, S., Kumar, D., et al., 2021. Geographic pattern of typhoid fever in India: a model-based estimate of cohort and surveillance data. J. Infect. Dis. 224 (Supplement_5), S475–S483.

Chau, K.K., Barker, L., Budgell, E.P., Vihta, K.D., Sims, N., Kasprzyk-Hordern, B., et al., 2022. Systematic review of wastewater surveillance of antimicrobial resistance in human populations. Environ. Int. 162, 107171.

Chen, Q., An, X., Li, H., Su, J., Ma, Y., Zhu, Y.-G., 2016. Long-term field application of sewage sludge increases the abundance of antibiotic resistance genes in soil. Environ. Int. 92–93, 1–10.

Chung The, H., Boinett, C., Pham Thanh, D., Jenkins, C., Weill, F.-X., Howden, B.P., et al., 2019. Dissecting the molecular evolution of fluoroquinolone-resistant Shigella sonnei. Nat. Commun. 10 (1), 4828.

Collignon, P., Beggs, J.J., Walsh, T.R., Gandra, S., Laxminarayan, R., 2018. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. Lancet Planet. Health 2 (9), e398–e405.

Coombs, G.W., Daley, D.A., Shoby, P., Mowlaboccus, S., 2018. Australian Group on Antimicrobial Resistance (AGAR) Australian Enterococcal Surveillance Outcome Program (AESOP). Commun. Dis. Intell. 2022, 46.

Crump, J.A., Luby, S.P., Mintz, E.D., 2004. The global burden of typhoid fever. Bull. World Health Organ. 82 (5), 346–353.

da Silva, K.E., Tanmoy, A.M., Pragasam, A.K., Iqbal, J., Sajib, M.S.I., Mutreja, A., et al., 2022. The international and intercontinental spread and expansion of antimicrobial-resistant Salmonella Typhi: a genomic epidemiology study. Lancet Microbe 3 (8), e567–e577.

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A.S., Ceder, G., et al., 2024. Structured information extraction from scientific text with large language models. Nat. Commun. 15 (1), 1418.

Destoumieux-Garzón, D., Mavingui, P., Boetsch, G., Boissier, J., Darriet, F., Duboz, P., et al., 2018. The One Health concept: 10 years old and a long road ahead. Front. Vet. Sci. 5.

Food and Agriculture Organization of the United Nations, World Organization for Animal Health, World Health Organization, 2010. The FAO–OIE–WHO collaboration: a Tripartite concept note. Available from: https://www.who.int/publications/m/item/the-fao-oie-who-collaboration.

Fuhrmeister, E.R., Harvey, A.P., Nadimpalli, M.L., Gallandat, K., Ambelu, A., Arnold, B. F., et al., 2023. Evaluating the relationship between community water and sanitation access and the global burden of antibiotic resistance: an ecological study. Lancet Microbe 4 (8), e591–e600.

Graham, D.W., Giesen, M.J., Bunce, J.T., 2019. Strategic approach for prioritising local and regional sanitation interventions for reducing global antibiotic resistance. Water 11 (1), 27.

Grundmann, H., Glasner, C., Albiger, B., Aanensen, D.M., Tomlinson, C.T., Andrasević, A. T., et al., 2017. Occurrence of carbapenemase-producing Klebsiella pneumoniae and Escherichia coli in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. Lancet Infect. Dis. 17 (2), 153–163.

Hernando-Amado, S., Coque, T.M., Baquero, F., Martínez, J.L., 2019. Defining and combating antibiotic resistance from One Health and Global Health perspectives. Nat. Microbiol. 4 (9), 1432–1442.

Hirschberg, J., Manning, C.D., 2015. Advances in natural language processing. Science 349 (6245), 261–266.

Hu, X., Waigi, M.G., Yang, B., Gao, Y., 2022. Impact of plastic particles on the horizontal transfer of antibiotic resistance genes to bacterium: dependent on particle sizes and antibiotic resistance gene vector replication capacities. Environ. Sci. Technol. 56 (21), 14948–14959.

Jurafsky, D., Martin, J.H., 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall PTR.

Kahsay, A.G., Muthupandian, S., 2016. A review on Sero diversity and antimicrobial resistance patterns of Shigella species in Africa, Asia and South America, 2001–2014. BMC. Res. Notes 9 (1), 422.

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., et al., 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet (London, England) 382 (9888), 209–222.

Krallinger, M., Rabal, O., Lourenço, A., Oyarzabal, J., Valencia, A., 2017. Information retrieval and text mining technologies for chemistry. Chem. Rev. 117 (12), 7673–7761.

Kumarasamy, K.K., Toleman, M.A., Walsh, T.R., Bagaria, J., Butt, F., Balakrishnan, R., et al., 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect. Dis. 10 (9), 597–602.

Larsson, D.G.J., Flach, C.F., 2022. Antibiotic resistance in the environment. Nat. Rev. Microbiol. 20 (5), 257–269.

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A.K.M., Wertheim, H.F.L., Sumpradit, N., et al., 2013. Antibiotic resistance—the need for global solutions. Lancet Infect. Dis. 13 (12), 1057–1098.

Lee, K., Brumme, Z.L., 2012. Operationalizing the One Health approach: the global governance challenges. Health Policy Plan. 28 (7), 778–785.

Lefrançois, T., Malvy, D., Atlani-Duault, L., Benamouzig, D., Druais, P.-L., Yazdanpanah, Y., et al., 2023. After 2 years of the COVID-19 pandemic, translating One Health into action is urgent. Lancet 401 (10378), 789–794.

Liguori, K., Keenum, I., Davis, B.C., Calarco, J., Milligan, E., Harwood, V.J., et al., 2022. Antimicrobial resistance monitoring of water environments: a framework for standardized methods and quality control. Environ. Sci. Technol. 56 (13), 9149–9160.

Liu, Y.-Y., Wang, Y., Walsh, T.R., Yi, L.-X., Zhang, R., Spencer, J., et al., 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect. Dis. 16 (2), 161–168.

López-Segura, N., Corberó-Rivali, C., Maldonado-Fernández, M.C., Calpe-Fraile, S., Peyra-Ros, J., Martínez-Roig, A., 2019. Imported extensively drug resistant typhoid fever in a child travelling to Spain from Pakistan. J. Travel Med. 26 (8).

Manaia, C.M., 2017. Assessing the risk of antibiotic resistance transmission from the environment to humans: non-direct proportionality between abundance and risk. Trends Microbiol. 25 (3), 173–181.

Matsuo, K.B.A., 2023. spacyr: Wrapper to the 'spaCy' 'NLP' Library.

Mogasale, V., Maskery, B., Ochiai, R.L., Lee, J.S., Mogasale, V.V., Ramani, E., et al., 2014. Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. Lancet Glob. Health 2 (10), e570–e580.

Murray, C.J.L., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., et al., 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399 (10325), 629–655.

Nizamuddin, S., Ching, C., Kamal, R., Zaman, M.H., Sultan, F., 2021. Continued outbreak of ceftriaxone-resistant salmonella enterica Serotype Typhi across Pakistan and assessment of knowledge and practices among healthcare workers. Am. J. Trop. Med. Hyg. 104 (4), 1265–1270.

Nunez-Mir, G.C., Iannone, B.V., Pijanowski, B.C., Kong, N., Fei, S., 2016. Automated content analysis: addressing the big literature challenge in ecology and evolution. Methods Ecol. Evol. 7 (11), 1262–1272.

O'Neill, 2014. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations.

Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. ArXiv. abs/2010.16061.

Qiao, M., Ying, G.-G., Singer, A.C., Zhu, Y.-G., 2018. Review of antibiotic resistance in China and its environment. Environ. Int. 110, 160–172.

Rice, L.B., 2008. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE. J. Infect. Dis. 197 (8), 1079–1081.

Rizzo, D., Lichtveld, M., Mazet, J., Togami, E., Miller, S., 2021. Plant health and its effects on food safety and security in a One Health framework: four case studies. One Health Outlook 3.

Semenec, L., Cain, A.K., Dawson, C.J., Liu, Q., Dinh, H., Lott, H., et al., 2023. Cross-protection and cross-feeding between Klebsiella pneumoniae and Acinetobacter baumannii promotes their co-existence. Nat. Commun. 14 (1), 702.

Sietsma, A.J., Theokritoff, E., Biesbroek, R., Canosa, I.V., Thomas, A., Callaghan, M., et al., 2024. Machine learning evidence map reveals global differences in adaptation action. One Earth 7 (2), 280–292.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45 (4), 427–437.

Stanaway, J.D., Reiner, R.C., Blacker, B.F., Goldberg, E.M., Khalil, I.A., Troeger, C.E., et al., 2019. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect. Dis. 19 (4), 369–381.

Stenhouse, G.E., Keddy, K.H., Bengtsson, R.J., Hall, N., Smith, A.M., Thomas, J., et al., 2023. The genomic epidemiology of shigellosis in South Africa. Nat. Commun. 14 (1), 7715.

The Australian Group on Antimicrobial Resistance. AGAR Surveillance Outcome Programs 2021 Report 2022.

The World Bank Group, 2022. The World by Income and Region.

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., et al., 2019. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature 571 (7763), 95–98.

Walker, J., Chaguza, C., Grubaugh, N.D., Carey, M., Baker, S., Khan, K., et al., 2023. Assessing the global risk of typhoid outbreaks caused by extensively drug resistant Salmonella Typhi. Nat. Commun. 14 (1), 6502.

Wang, Y., Tian, G.-B., Zhang, R., Shen, Y., Tyrrell, J.M., Huang, X., et al., 2017. Prevalence, risk factors, outcomes, and molecular epidemiology of mcr-1-positive Enterobacteriaceae in patients and healthy adults from China: an epidemiological and clinical study. Lancet Infect. Dis. 17 (4), 390–399.

Wheeler, N.E., Price, V., Cunningham-Oakes, E., Tsang, K.K., Nunn, J.G., Midega, J.T., et al., 2023. Innovations in genomic antimicrobial resistance surveillance. Lancet Microbe 4 (12), e1063–e1070.

World Economic Forum, 2021. Antimicrobial Resistance and Water: The risks and costs for economies and societies.

World Health Organization, 2016. Global action plan on antimicrobial resistance. Available from: https://www.who.int/publications/i/item/9789241509763.

World Health Organization, 2019. Global Antimicrobial Resistance Surveillance System (GLASS): molecular methods for antimicrobial resistance (AMR) diagnostics to enhance the Global Antimicrobial Resistance Surveillance System.

World Health Organization, 2023. Available from: https://www.who.int/initiatives/glass/glass-routine-data-surveillance.

World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, United Nations Environment Programme, 2022. Antimicrobial resistance TrACSS Iran (Islamic Republic of) 2022 country profile: National Action Plans and Monitoring and Evaluation (NPM). Available from: https://www.who.int/publications/m/item/Antimicrobial-resistance-tracss-irn-2022-country-profile.

World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, United Nations Environment Programme, 2022. One health joint plan of action (2022–2026): working together for the health of humans, animals, plants and the environment. Available from: https://www.who.int/publications/i/item/9789240059139.

World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, United Nations Environment Programme, 2023. One health and the United Nations Sustainable Development Cooperation Framework: guidance for United Nations country teams. Available from: https://www.who.int/publications/i/item/WHO-OHI-2023.2.

Zhou, Z., Shuai, X., Lin, Z., Yu, X., Ba, X., Holmes, M.A., et al., 2023. Association between particulate matter (PM)2·5 air pollution and clinical antibiotic resistance: a global analysis. Lancet Planet. Health 7 (8), e649–e659.

Zhu, Y.G., Johnson, T.A., Su, J.Q., Qiao, M., Guo, G.X., Stedtfeld, R.D., et al., 2013. Diverse and abundant antibiotic resistance genes in Chinese swine farms. PNAS 110 (9), 3435–3440.