Genomic sequencing should extend to diverse priority pathogens for effective study and surveillance of antimicrobial resistance: a systematic review of whole-genome sequencing studies from India

Nazneen Gheewalla[1], Jaisri Jagannadham[1,2], Rintu Kutum[1,3], Shraddha Karve[1*]

[1] Trivedi school of biology, Ashoka University, NH 44, Rajiv Gandhi Education City, Sonipat 131029, India.

[2] Centre for Bioinformatics and Computational Biology, Ashoka University, NH 44, Rajiv Gandhi Education City, Sonipat 131029, India.

[3] Department of computer science, Ashoka University, NH 44, Rajiv Gandhi Education City, Sonipat 131029, India.

*for correspondence: shraddha.karve@ashoka.edu.in

Abstract

Background

Antimicrobial resistance (AMR) is a public health emergency in many low and middle-income countries, including India. To effectively tackle AMR, we need rapid diagnostics, effective surveillance and new antimicrobial drugs. Whole-genome sequencing of pathogens is the first definite step towards achieving these goals.

Methods

In this work, we review all the studies published till date that report whole-genome sequences of select priority AMR pathogens from India. We searched PubMed and Web of Science databases for the studies that involved whole-genome sequencing of AMR priority pathogens from India. For the top two highly sequenced pathogens, *S. typhi* and *K. pneumoniae*, we performed phylogenetic analyses to understand the geo-climatic distribution of genetically diverse strains.

Results

Our search reveals 94 studies that report 2547 unique whole-genome sequences. We find that most sequences are limited to select priority pathogens isolated from a couple of geo-climatic zones of India. Our phylogenetic analyses show that available data does not indicate

systematic differences between the genomes of isolates from different geo-climatic zones. Our search also reveals complete absence of travel-related studies tracking possible movement of AMR pathogens within country. Lastly, we find very few studies that sequence AMR pathogens isolated from food, soil or other environments.

Conclusion

Together, these observations suggest that India should prioritize sequencing of diverse AMR pathogens from clinics as well as from environments and travellers rather than extending the geo-climatic range of already-sequenced pathogens. Our recommendations can be potentially valuable for other low and middle-income countries with limited resources, high prevalence of AMR and diverse geo-climatic conditions.

Introduction

Antimicrobial resistance (AMR) is a global crisis and has been ranked amongst the top ten health concerns by WHO.[1] In the year 2019, 1.27 million people died of infections by drug resistant pathogens.[2] Antibiotic resistance is more prevalent in low and middle-income countries owing to lack of hygiene, resource limitation and other socio-economic factors.[3–5] South Asia has the second largest burden of AMR, surpassed only by the sub-Saharan African region.[2] Within South Asia, India is one of the hotspots for resistance. By the year 2050, global annual deaths due to infections by drug resistant pathogens are predicted to rise to ~10 million with ~2 million deaths in India alone.[4,6] India also has a drug resistance index of 71, highest for any country in the world, indicating very poor efficacy for the existing drugs.[7] AMR is thus a public health emergency for countries like India.

The diverse and dynamic nature of AMR demands continuous surveillance, rapid diagnostics and robust drug development programs, amongst other things. Traditionally, the first step of tackling resistant pathogen is antimicrobial susceptibility testing (AST). AST is a gold standard for determining resistance and is relatively economic to perform. However, AST has large turnaround times, requires customized workflow for every pathogenic group and

cannot reveal the molecular mechanisms of resistance. As a result, the outcomes of ASTs are of little use for drug design or rapid diagnostics or extensive surveillance.

In the recent years, whole-genome sequencing (WGS) has been successfully used for understanding the molecular mechanisms of AMR,[8,9] epidemiology of AMR[10,11] as well as for drug discovery.[12] WGS-based studies of antimicrobial resistant pathogens offer several advantages over other traditional approaches. First, WGS gives us a comprehensive picture of genomic changes in contrast to gene-based studies that are restricted to a few well-characterized genes.[13] Second, WGS data can help establish the trajectories and timelines of resistance evolution.[14–17] Third, WGS based approaches can potentially supplement or replace the traditional methods of AST.[13,18–22] Fourth, current WGS technologies promise shorter turnaround times than traditional approaches.[23] Short turnaround times are especially useful for slow-growing pathogens like *Mycobacterium*[24,25] as well as in the treatment of conditions like septicemia where faster diagnosis can increase the patient survival rates significantly. Lastly, WGS-based approaches can also accurately identify pathogens that may have been misidentified by traditional methods. For instance, it is challenging to distinguish between *E. coli* and *Shigella* by methods other than whole-

genome sequencing.[26] These advantages in terms of diagnostics, therapeutics and surveillance automatically make genomic sequencing the first step in the fight against AMR and countries with high prevalence of AMR need a greater focus on genomic sequencing.[27]

To establish effective genomic sequencing for AMR in India, it is important to have information on the available whole-genome sequences from India. A comprehensive review of whole genome sequences of AMR pathogens from India is lacking. Sequence data available in most databases suggest that genomic sequencing from India is patchy at best. For example, according to the Bacterial and Viral Bioinformatics Resource Centre (BV-BRC),[28] India is the third major contributor towards *S. typhi* sequences worldwide with the contribution of nearly 14%. But similar levels of sequencing efforts are lacking for almost all the other AMR pathogens. For instance, BV-BRC lists 8482 *E. coli* sequences from the USA as opposed to 827 sequences from India, which is four times as populated as the USA. Similarly, BV-BRC lists eight times as many of *K. pneumoniae* sequences from China than India.

Here we systematically review the studies that report whole-genome sequences of AMR pathogens from India. We focus on the pathogens that demand immediate attention according to the Indian Council of Medical Research.[29] We find that the use of whole-genome sequencing in this context is very limited in India. Majority of the data are confined to a couple of geo-climatic zones within the country and is restricted to a few select pathogens. We then perform a phylogenetic analysis for the two pathogenic species that have been sequenced across most geo-climatic zones of India. Our results show that there are no systematic differences between the genomes of pathogens isolated from different geo-climatic zones (but see discussion for the limitations of the available data). This finding suggests that the immediate focus of sequencing efforts should be a number of diverse priority pathogens rather than already-sequenced pathogens from new geo-climatic zones of the country. We also find that the spread of resistance due to human travel remains severely understudied in the country. Lastly, our review shows that genomic studies of pathogens isolated from domestic animals, food and environment are rare in India. Increasing the sequencing efforts in this direction of 'One health'[30] can help gain a holistic understanding of resistance spread and evolution in India.

Methods

Choice of pathogenic species

We selected ten pathogenic species and one pathogenic genus for which immediate action is recommended.[29] Specifically, this included eight species (*Escherichia coli, Enterobacter cloacae, Morganella morganii, Citrobacter koseri, Proteus mirabilis, Providencia rettgeri, Salmonella typhi, Serratia marcescens*) and one genus (*Klebsiella*) from the order *Enterobacterales*, and two other species from diverse taxa (*Acinetobacter baumannii* and *Candida auris)*. For the selected pathogens we included the genomic sequence of every isolate from AMR studies conducted in India, irrespective of the resistance profile of the pathogen. We reasoned that every sequence of a priority pathogen, resistant or sensitive to any specific antibiotic, is important for understanding the molecular mechanisms and evolution of resistance as well as for applications like surveillance and diagnostics.

Literature Search

We searched the databases of PubMed and Web of Science on the 5 May, 2023 for studies that reported whole-genome sequences of selected AMR pathogens. We included only those studies that provided the data for pathogens isolated from Indian patients or from

patients with a history of travel to India or environmental/food/veterinary samples collected in India. We followed PRISMA guidelines[31] while performing the literature search. Our detailed search strategy is outlined in the supplementary material (Supplementary material 1).

Checking for the uniform availability of whole-genome sequences for selected pathogens across different geo-climatic zones

To examine whether the number of available whole-genome sequences are uniform across different geo-climatic zones, we performed the Pearson's chi square test of homogeneity using RStudio (v2022.07.0). To this end, we considered the pooled number of sequences of all the selected pathogens for each geo-climatic zone. We only used data for those isolates for which geo-climatic zone was known. To ascertain the geo-climatic zone of every isolate, we used the information reported in the corresponding study or the NCBI metadata associated with the genomic sequence. In case of a discrepancy in location mentioned in the study and NCBI metadata, we gave preference to the information provided in the relevant study.

Phylogenetic Analysis of *S. typhi* and *K. pneumoniae*.

To determine whether different lineages of a given organism are prevalent in different zones of India, we conducted a phylogenetic analysis. We chose *S. typhi* and *K. pneumoniae* for the phylogenetic analysis, as their genomic sequences were available from most geo-climatic zones of India. We excluded the sequences that lacked the information on the geo-climatic zone (Supplementary material 6), as well as the sequences where the accession numbers were not available.

We performed two separate phylogenetic analyses, one for 503 *S. typhi* genomic sequences and other for 231 *K. pneumoniae* genomic sequences. We first downloaded the genomic sequences from NCBI as complete genomes, raw reads or contigs. Using a custom script, we re-named all the downloaded sequences to denote their geo-climatic zone and the year of the study. We pre-processed all the raw reads, both single and paired-end, and checked the quality with FastQC v0.12.1.[32] We then filtered the low-quality reads and adapters using Trimmomatic, v0.39.[33] We used PhaME (Phylogenetics and Molecular evolution analysis tool, v1.0.2)[34] for phylogenetic tree building. For the phylogenetic analysis of *S. typhi* sequences, we used *S. typhi CT18* as the outgroup reference.[35–38] Similarly, we used *K. pneumoniae*

158 *MGH78578* as an outgroup reference strain for the phylogenetic analysis of *K. pneumoniae*

159 *sequences* and SPAdes genome assembler v3.15.5,[39] for getting the draft genomes. We also

160 performed a genotype analysis for *S. typhi* sequences using Genotyphi v2 scheme,[40] and

161 Kleborate v2.3.2[41] for the genotype analysis of *K. pneumoniae* sequences. We used Iroki[42]

162 for visualization of the phylogenetic trees.

Results

Literature Search

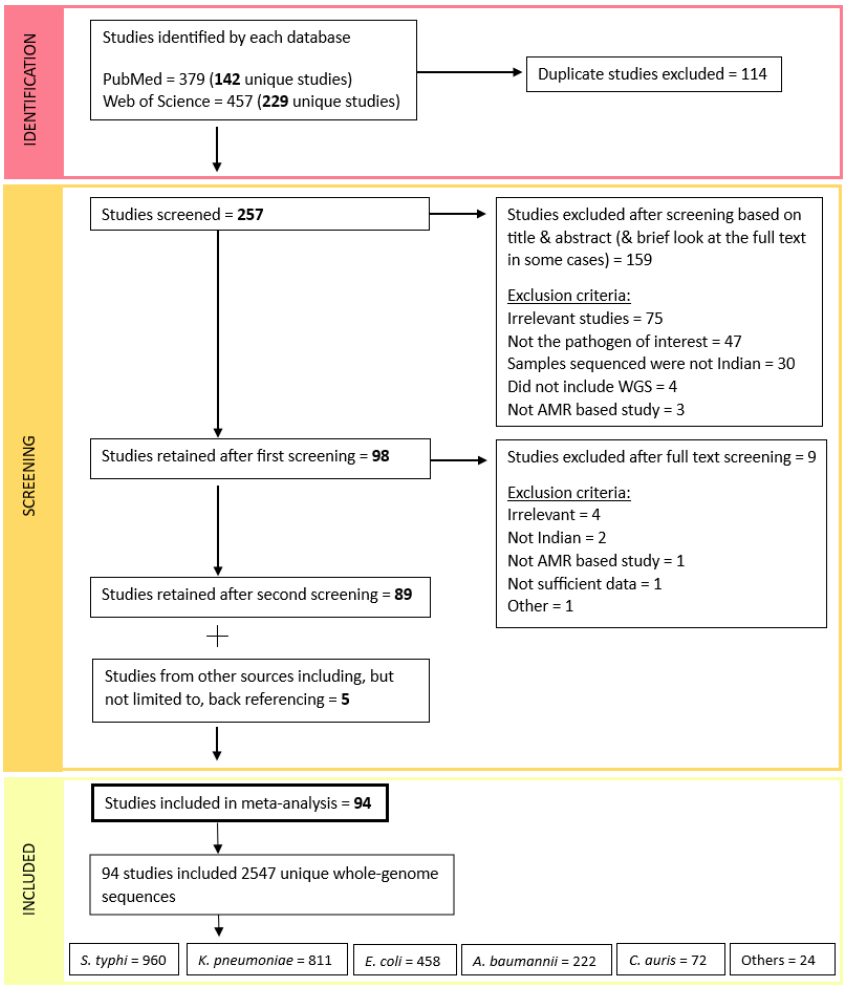

Figure 1. PRISMA flowchart describing the literature search strategy and outcomes for AMR related studies reporting whole–genome sequences of the selected pathogens. We searched PubMed and Web of Science databases for the relevant studies (see Supplementary material 1 for the search phrases). We screened 257 unique studies after removing duplicates within and between the search results from the two databases. We excluded 159 studies after screening the title and abstract. We further excluded 9 studies after the scrutiny of the full text. This resulted in 89 studies matching our inclusion criteria. We added 5 studies found by back-referencing or other sources. Together, we retained 94 studies for this systematic review. After removing the overlapping sequences across studies, the search yielded 2547 unique whole-genome sequences of the selected pathogens. The boxes in the bottom row show the number of genomic sequences for pathogenic species *S. typhi*, *K. pneumoniae*, *E. coli*, *A. baumannii* and *C. auris*. We found very few genomic sequences of *K. quasipneumoniae, S. marcescens, K. aerogenes, E. cloacae, M. morganii* that are clubbed together as 'Others'.

177 The preliminary literature search for the selected pathogens yielded a total of 836 studies.
178 We did not find any studies from India with WGS information on three species from our list,
179 namely *Citrobacter koseri, Proteus mirabilis* and *Providencia rettgeri*. After exclusion of
180 duplicates, we obtained 257 unique studies. From this subset, we further excluded 159
181 studies during the first screening based on the title, abstract and a brief view of full text
182 wherever needed. Nine more studies were excluded during the second screening where we
183 scrutinized the full text for the relevance to our review. This resulted in a total of 89 studies
184 matching our inclusion criteria. We found 5 additional studies via back-referencing. In all we
185 included 94 studies (that spanned over 9 years, from 2014 to 2023) in this systematic review
186 that provided sequences and other data for 2547 isolates.

Organism	Number of studies	Number of genome sequences	Sample categories	References
Salmonella typhi	14	960	Clinical isolates, travel-associated clinical isolates	43–56
Klebsiella pneumoniae	35	811	Clinical isolates, travel-associated clinical isolates	57–91
Escherichia coli	31	458	Clinical isolates, travel-associated clinical isolates, food-associated isolates, environmental (hospital sewage water) isolates, veterinary clinical (milk from cows suffering from mastitis) isolates	59,62,65,77,79,91–116
Acinetobacter baumannii	9	222	Clinical isolates	59,62,65,77,117–121
Candida auris	10	72	Clinical isolates, travel-associated clinical isolates, environmental isolates, food-associated isolates	14,79,122–129
Serratia marcescens	1	17	Clinical and hospital environmental isolates	130
Klebsiella spp (other than *K. pneumoniae*)	4	4	Clinical isolates, environmental isolates	80,81,131,132
Enterobacter cloacae	2	2	Clinical isolates, environmental isolates	133,134
Morganella morganii	1	1	Clinical isolates	135

Table 1: Number of studies and number of unique whole-genome sequences therein for the selected pathogenic groups. The column 'Sample category' indicates whether the isolates were from patients from India (clinical) or from patients with a travel history to India (travel-associated clinical) or from environmental/ food/ veterinary samples collected in India.

Number of whole-genome sequences for any selected pathogen vary widely across different geo-climatic zones of India

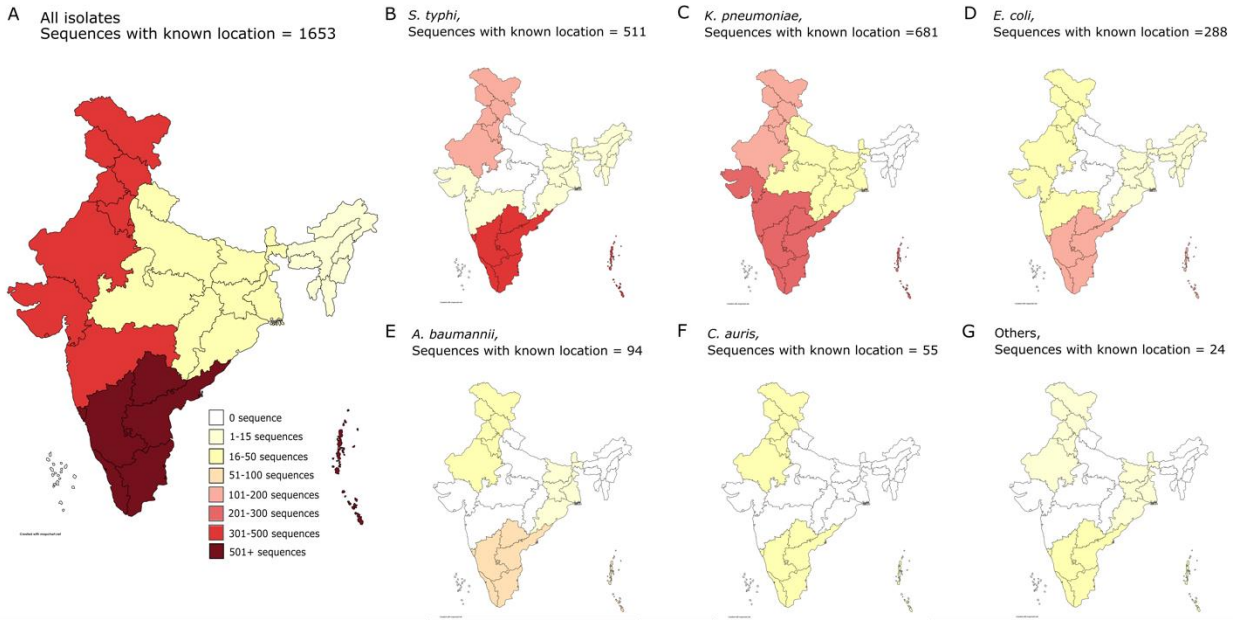

Figure 2: Number of whole-genome sequences of selected pathogens from every geo-climatic zone of India. A. The geo-climatic distribution of the all the 1653 whole-genome sequences with known locations. Panels **B-G** depict the distributions of number of isolates for individual pathogenic groups of *S. typhi, K. pneumoniae, E. coli, A. baumannii, C. auris* and others respectively. The group 'others' includes the species that have very few whole genome sequences, namely, *S. marcescens, K. quasipneumoniae, K. aerogenes, E. cloacae* and *M. morganii*. Only the isolates with known location are included. The colour of each geo-climatic zone in every panel (A-G) represents the range for the number of sequences as per the legend in the panel A. Exact number of sequences and their locations for each pathogenic group are in the supplementary materials 2 and 6 as well as in supplementary dataset. We used mapchart.net for the design and colour of the map, and mapsofindia.com as a reference for determining the geo-climatic zones of India.

The 94 studies that we shortlisted provide WGS data and related information for a total of 2547 distinct isolates. With 960 isolates, *S. typhi* represents a majority (37.69%) of these isolates (Figure 2). This is closely followed by the *Klebsiella* species which together form 31.99% of the isolates for which WGS data are available. For one of the Klebsiella species, *K.*

212 *pneumoniae*, 811 sequences (31.84% of total sequences in our study) are available (Figure
213 2). We also obtained sequence data for 458 *E. coli* isolates, 222 *A. baumannii* isolates and 72
214 *C. auris* isolates. We pooled the sequence data for 17 *S. marcescens* isolates, 2 *E. cloacae*
215 isolates and 1 isolate of *M. morganii*, 3 *K. quasipneumoniae* isolates and 1 *K. aerogenes*
216 isolate during our analysis, (category 'others' in Figure 2) due to very low numbers of
217 isolates from each group. Overall, these numbers show that the extent of whole-genome
218 sequencing is highly variable for different pathogenic groups. *S. typhi* and *K. pneumoniae*
219 are the most sequenced species among the selected pathogens, together comprising
220 69.53% of the total sequences. In contrast, we could not find even a single whole-genome
221 sequence from India for three priority species from *Enterobacterales* - *P. rettgeri, P. mirabilis*
222 and *C. koseri.*
223
224 We found only one study with *S. marcescens* where 17 isolates from patients and hospital
225 environment had been sequenced.[130] Similarly, for *E. cloacae* we found only two studies that
226 reported the total of two whole genome sequences, while for *M. morganii and K. aerogenes*,
227 we found one study with a single sequence each.[132–135]

We also looked at the geographic distribution of the whole-genome sequences across the six different geo-climatic zones of India (Figure 2). Out of the 2547 sequences, we could locate the geo-climatic zones of 1653 sequences (64.90% of the sequences) based on the information in the reported study or from the NCBI metadata associated with the genomic sequence. The number of sequenced genomes were highly variable across different geo-climatic zones for every selected pathogen (Figure 2). From the 1653 sequences with known locations, 893 (i.e. 54.02%) sequences were from the Southern zone, indicating that almost half the genomic sequences were from a single geo-climatic zone of India. Furthermore, out of these 893 genomic sequences from the southern zone, 581 sequences (i.e. more than 65%) belonged to *S. typhi* and *K. pneumoniae* (330 and 251 sequences respectively). These numbers indicate that the majority of the sequenced genomes from India were of two pathogenic species isolated from a single geo-climatic region.

Out of the remaining 760 genomic sequences from other geo-climatic zones, 362 were from the northern, 308 from western, 50 from eastern, 30 from central and only 10 were from the northeastern regions of India (Supplementary materials 2 and 6 and supplementary dataset). Formal statistical analysis validated our observation of large variation in the number of

available whole-genome sequences across different geo-climatic zones for all the selected pathogens combined (Pearson's chi square test for homogeneity, $\chi^2 = 502.25$, $df = 25$, $p < 2.2 \times 10^{-16}$). There is no evidence to indicate that the variation in extent of genomic sequencing reflects the population numbers or the disease burden in the respective geo-climatic zone.

The disparity in number of sequences is most likely the reflection of the disparity in the number of studies that report genomic sequences from each zone. There were only 2 studies which report genomic sequences from the northeastern zone as opposed to 43 studies that report WGS data from the southern zone. Similarly, we found only 2 studies from the central zone of India and both reported genomic sequences of *K. pneumoniae*. These numbers along with the other observations from our literature search suggest that often a particular research group focuses on a specific pathogen. A string of studies reporting WGS data of that pathogen then follows. Sequenced pathogenic isolates then typically belong to the same geo-climatic zone and often collected from the same locality.

261 In sum, genomic sequencing of pathogenic species is patchy in India and the current

262 sequencing effort is focused on only a couple of pathogens mostly isolated from a couple

263 of geo-climatic zones.

Sequences of *S. typhi* from different geo-climatic zones of India do not form phylogenetically distinct clusters

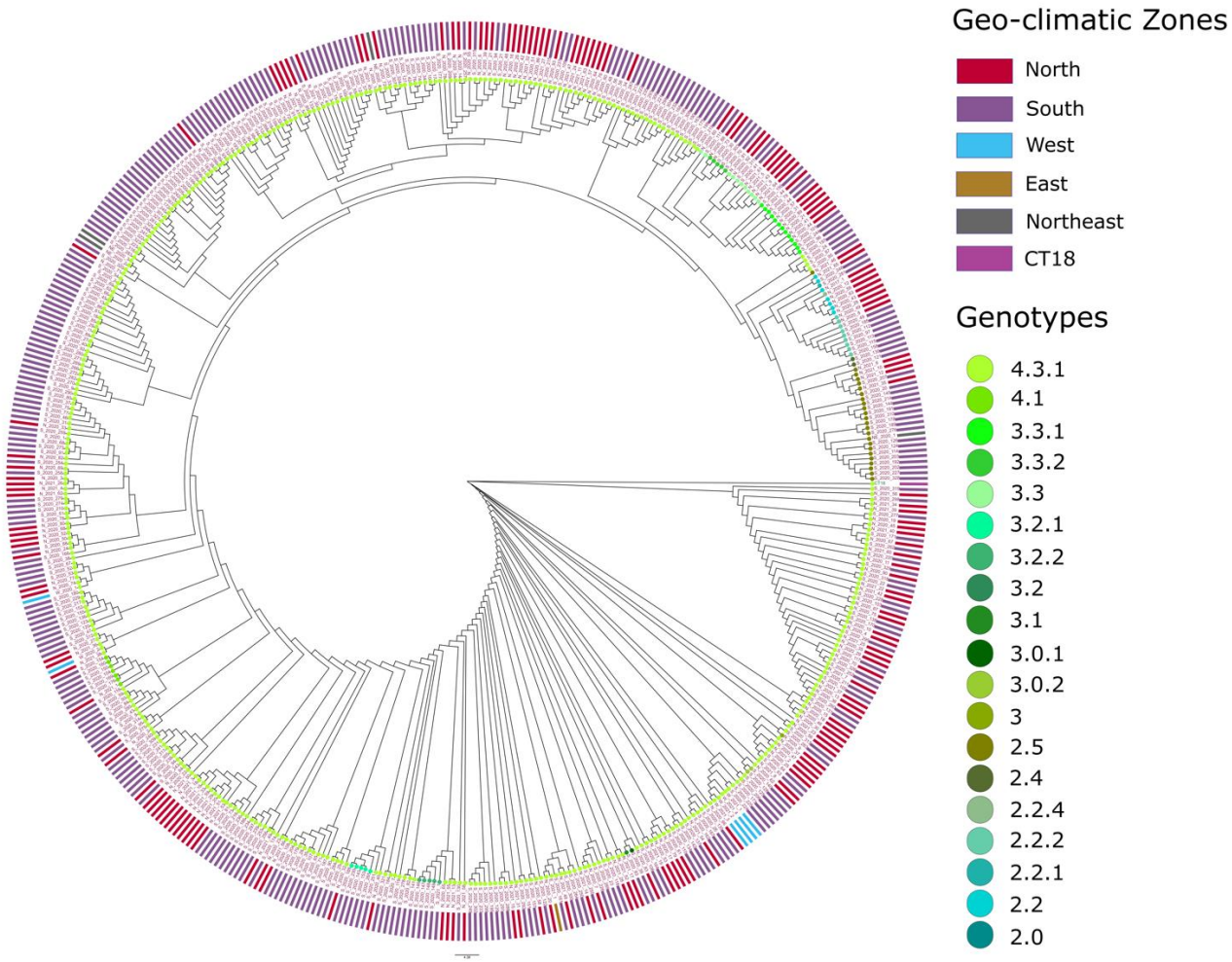

Figure 3: Phylogenetic analysis of *S. typhi* isolates. The geo-climatic zone wise distribution (outer circle) along with the genotype distribution (dots at end of leaf tip) and phylogenetic clustering of the 503 *S. typhi* isolates. The genotyping was done using Genotyphi v2, the phylogenetic tree was built using PhaMe v1.0.2, and visualization was created using Iroki. The name of the isolate indicates its geo-climatic zone and the year of study. Similar genotypes predominantly cluster together but not as per the geo-climatic zone of isolation (outer circle).

Our literature review highlights that whole-genome sequencing of AMR pathogens in India needs to extend to all priority pathogens and cover more geo-climatic zones. But how different are the genomic sequences of isolates from different geo-climatic zones? The literature is equivocal on the sequence-level differences amongst isolates from different geo-climatic conditions with majority of the studies using geographic regions as a proxy for geo-climatic differences.[47,48,136] Moreover, these studies tend to look at larger geo-climatic differences, often considering data across different countries and continents. Some studies suggest that different geographic regions may harbour genetically distant pathogens[47,48,136] while others show genetic homogeneity in isolates from different geographic regions.[136] Understandably, conclusions also vary for different pathogenic species.[136] But little is known about the diversity of whole-genome sequences across different geo-climatic zones within a single country, like India.

To check whether the isolates from different geo-climatic zones of India form distinct sequence clusters we resorted to phylogenetic analysis of the available sequences. We chose the two priority pathogenic species, *S. typhi* and *K. pneumoniae*, which had isolates sequenced from most of the geo-climatic zones. Phylogenetic analysis of the 503 genomic

292 sequences of *S. typhi* showed that the genomic sequences of isolates from the same geo-
293 climatic zone do not cluster together (Figure 3). The conclusion is also supported by the
294 phylogenetic analysis of 231 genomic sequences of *K. pneumoniae* (Supplementary material
295 3).
296
297 One possible reason for such homogeneity could be the spread of pathogens due to human
298 dispersal across geo-climatic zones. The other reason could be the spread through
299 environmental factors such as soil or water as well as food items. Therefore, we next
300 checked for evidence for both these possibilities in the literature.

Genomic sequencing of the pathogens from travel-associated infections is not common in India

Infected individuals traveling from one location to another can spread resistant strains of pathogens and tracking this movement is important to understand the spread of AMR.[137–139] Most of the travel-related studies have a few limitations though. First, it is difficult to track the movement within the country, though some novel approaches may allow this.[140–142] Second, majority of the travel-related studies do not take into account the acquisition of a pathogen prior to or during the travel.[94] This may lead to incorrect identification of the source of infection as the travel destination. Third, most of these studies consider only symptomatic cases of the disease and asymptomatic carriers might be easily overlooked.[47,48,52,79,94,115,123] Despite these limitations however, routine sequencing of isolates from the infected travellers can give a better picture of disease dispersal.[94,115]

Unfortunately, we found only one study that sequenced the genome of a pathogen possibly acquired during the travel between two geo-climatic zones of India.[44] The patient, resident of the eastern zone, had a history of travel to the northern zone, but no clear link between the infection and travel could be established in the study. We found another

318 handful of studies that sequenced isolates from the travellers that were visiting India/South
319 Asia (Table 1 and Supplementary material 4). These studies identified the strain of the
320 pathogen and established an association between the strain and specific geographic region.
321 For example, Ingle et al. observed that *S. typhi* sub-lineage 4.3.1.2. was mainly associated
322 with travellers returning from India and carried mutations in the quinolone resistance-
323 determining genomic regions.[47,48] Similarly, Yaita et al. studied the acquisition of ESBL-
324 producing *E. coli* by travellers returning to Japan. The study found that the majority of the
325 people returning from India (10 out of 14) tested positive for ESBL-producing *E. coli* .[94] Yet
326 another study found that an individual returning to Switzerland from India was colonised
327 with a carbapenamase-producing *E. coli* harbouring a blaOXA-484, a variant of
328 carbapenemase new to Switzerland.[115]
329
330 These examples illustrate the effective tracking of the resistance spread from India to other
331 countries. We did not find similar studies for individuals returning to India from other
332 regions of the world or for individuals traveling across different geo-climatic zones of
333 India. These observations demand increased genomic sequencing of pathogens isolated
334 from travellers to better understand the resistance dynamics within the country.

335 **Whole-genome sequencing of AMR isolates from food and environment is limited in**

336 **India**

337 We next looked for the genomic sequences of priority pathogens isolated from

338 environmental and food samples. AMR studies from all over the world have reported

339 widespread presence of resistance genes in water, soil as well as in food items.[143–145] This

340 'One health' approach of study underlines the importance of tracking resistance across

341 abiotic environments alongside the clinics.[30] We thus looked for studies that reported

342 whole-genome sequences of priority pathogens isolated from environments or food items.

343

344 Our search uncovered 3 such studies of environmental samples from India. It was striking

345 that even such handful of studies (Supplementary material 5) reported isolates that harbour

346 antibiotic resistant markers.[124,132,134] For example, an isolate of *K. aerogenes* from agricultural

347 soil was found to contain ~30 AMR genes.[132] Another sequencing study demonstrated that

348 23 (out of 24) *C. auris* isolates from marshy wetlands and sandy beach areas of Andaman

349 and Nicobar Islands were resistant to a variety of antifungals.[124]

350 We also found 5 studies from India that reported sequences of AMR pathogens isolated

351 from food sources (Supplementary material 5).[93,101,104,109,122] For instance, a study reported

352 resistance patterns in the yeast species found on the surface of apples.[122] Authors found
353 that 73% of the total isolates belonged to *Candida spp.* and 11% of those isolates were of
354 the priority species *C. auris*. Alarmingly, most of these isolates (15 out of 16) had heightened
355 resistance against fluconazole.[122] Similarly, multiple sequencing studies found ESBL
356 producing *E. coli* from poultry samples and one group demonstrated that 78.5% of their
357 poultry isolates were multi-drug resistant.[101,104] Another sequencing study from the
358 Banaskantha district of Gujarat reported that 90% of their samples procured from ~30
359 different farms contained *E. coli*. Nearly 80% of these *E. coli* were ESBL producers.[109] There
360 were also a few studies reporting drug-resistant *E. coli* from bovine mastitis raising a
361 concern of pathogens getting into the milk and milk products.[111,113,114]
362 Overall the results indicate a high prevalence of AMR priority pathogens in the
363 environments and food items but extensive and systematic sequencing studies are needed
364 to obtain a truly comprehensive picture.

Discussion

Our literature search revealed that whole-genome sequencing of AMR priority pathogens from India is patchy, both in terms of species that were sequenced and geo-climatic zones where the pathogens were isolated. For three of the priority pathogens *Citrobacter koseri, Proteus mirabilis, Providencia rettgeri* we did not find even a single genomic sequence reported from India. Recent evidence suggests that all the three species are clinically important.[146–152] For example, multi-drug and pan-drug resistant varieties of *P. rettgeri* caused nosocomial infections in India.[148] Similarly, a longitudinal study from a tertiary healthcare centre in western India reported *C. koseri* with beta-lactam and carbapenem resistance over a period of 3 years.[149] Another study found a nosocomial NICU outbreak of multi-drug resistant *P. mirabilis* that resulted in 80% mortality.[150] Yet another study found ESBL positive *P. mirabilis* in food and livestock samples.[153] Apart from these three pathogens that lack any WGS data from India, we also found other important pathogenic groups that were severely underrepresented. For instance, carbapenem-resistant *Acinetobacter* is a serious threat and leading cause of nosocomial infections in India[29] and yet we found only 222 whole-genome sequences of *Acinetobacter baumannii* from the whole country.

We performed phylogenetic analyses on the sequence data of two highly-sequenced priority pathogens *S. typhi* and *K. pneumoniae*. Results showed that genomic sequences of isolates from different geo-climatic zones do not form distinct clusters (Fig 3, Supplementary material 3). In the absence of evidence for distinct sequence diversity across different geo-climatic zones, it might be beneficial to prioritize the sequencing of diverse pathogenic species over the increased coverage of already-sequenced species across different geo-climatic zones of India. However, there are some limitations of the datasets that we used for our phylogenetic analysis, as discussed below.

Our conclusions are drawn from the limited sequence data that are available from India. The sequences used in the phylogenetic analysis were collected over a period of nine years. Insufficient data for individual years prevented us from performing the analysis separately for each year. Additionally, in many cases there was no clear indication of the month and year of isolation. Adding this information to the phylogenetic analysis may change the inferences drawn. It is reasonable to assume that a particular resistant variant of a pathogen is not likely to arise (or arrive, with an infected individual/food item/environmental agent) in multiple geo-climatic zones at the same time. But how rapidly any resistant variant spreads

399 can be dependent on several factors such as mode of transmission, fitness advantage over
400 the prevalent variants and climatic conditions. Certain resistant variants of any pathogen
401 might already be widespread across all geo-climatic zones. Systematic and continued
402 genome sequencing across all geo-climatic zones is needed to resolve these possibilities for
403 every priority pathogen. It will also help to have a clearer indication of time of isolation and
404 geo-climatic zone for all the reported sequences. For example, from the existing literature,
405 we could not find the geo-climatic zone for 449 sequences of *S. typhi* and adding this
406 information may change the inferences of the phylogenetic analysis.
407
408 We also note that for the phylogenetic analyses we used two gram-negative enteric
409 pathogens, *S. typhi* (Figure 3) and *K. pneumoniae* (Supplementary material 3). Genomic
410 sequences of other non-enteric or gram-positive pathogenic species may cluster according
411 to the geo-climatic zones. It is also known that the ratio of core to accessory (or
412 dispensable) genes varies across different pathogenic species.[154] Such variation can affect
413 the conclusions drawn from phylogenetic and MLST analysis.[155] The predominant nature of
414 the infections, nosocomial vs community-based, may also affect the spread across geo-

415 climatic zones. For instance, *Acinetobacter baumannii* is primarily a nosocomial

416 pathogen[156,157] and specific strains may dominate specifies geo-climatic regions.

417

418 At the time of our literature search, only two geo-climatic zones had reasonable

419 representation for *S. typhi* (330 and 170 sequences each from Southern and Northern zones)

420 while remaining zones of West, Northeast, East had reported 6, 4 and 1 genomic sequence

421 respectively. But genomic sequences from even Southern and Northern zones did not

422 cluster separately. A possible reason for this could be that most of the reported genomic

423 sequences (408 out of 503) belonged to the *S. typhi* lineage 4.3.1, suggesting one dominant

424 infectious strain throughout the country. This may not be the case with other important

425 pathogens and may lead to different observations. However, another study[85], as well as our

426 own analysis with *K. pneumoniae* (Supplementary material 3) genomes, demonstrates lack of

427 distinct clusters for genomic sequences from different geo-climatic regions.

428

429 We reasoned that homogeneity in genomic sequences across different geo-climatic zones

430 might be due to the spread of infectious pathogens/resistance genes through human

431 travellers or food items or other environments. Documentation and sequencing of such

instances is important as it can allow effective mapping of the isolates' ancestry and help understand their dispersal patterns. Sequencing the isolates from infected travellers can also uncover the original incidences of certain infections. For example, a recent study discovered that the first incidence of *C. auris* infection was a 54-year-old male travelling to India in 2007, which was two years before the supposed first case of *C. auris* infection was reported from Japan.[123] This discovery has altered the timeline of emergence of *C. auris* infections. Our literature search revealed very few studies with genomic sequences of pathogens from travel-related infections. Majority of the studies were travellers who acquired the infection in India and travelled abroad. There were almost no studies reporting sequences of pathogens acquired during the incoming travel to India or more importantly, travel within the country. This observation underlines the necessity of tracking travellers for possible infections and extending these investigations to include whole-genome sequencing of the pathogen.

Whole-genome sequences of isolates from soil, water, other environments and foods were also modest in number. Alarmingly, even these few reports uncovered important resistance markers. For example, we found studies that reported colistin-resistant pathogens from food samples. Colistin is one of the final-resort antibiotics and the presence of colistin resistant

449 *Enterobacteriaceae* (group3 pathogens as per ICMR)[29] in food items is concerning. Colistin-

450 resistant *Enterobacteriaceae* (like *Klebsiella, Enterobacter, Citrobacter, E. coli*) and

451 *Pseudomonas* were found in a range of food samples collected from shops and households

452 [93]. Moreover, few of these isolates contained the *mcr-1* gene which is responsible for

453 plasmid-mediated spread of colistin resistance. Extrinsic resistance elements, such as

454 plasmids, can spread the resistance rapidly across pathogens from different groups.[158] Our

455 results underline the need for comprehensive 'One health' approach in WGS studies with

456 extensive sequencing of isolates from environments.

Conclusion

Our review collates the studies that sequence the genomes of priority AMR pathogens from India. We find that many priority pathogens are not routinely sequenced in India while some have not been sequenced at all. Additionally, most genomic sequences are available from only a couple of geo-climatic zones. With the limited sequence data that is available, we infer that genomic sequence diversity is homogeneous across the geo-climatic zones. To assert or refute this conclusion however, we need systematic genomic sequencing for a few more priority pathogens across all geo-climatic zones. This task is resource-intensive and implementation may take a few years at least. While such comprehensive sequencing data is being generated across the country we need to urgently begin sequencing diverse priority pathogens such as *A. baumannii*, *C. koseri, P. mirabilis* and *P. rettgeri*. This should be accompanied by genomic sequencing of isolates from travel-related infections and environment. Our recommendations can be valuable for other low and middle-income countries with diverse geo-climatic conditions, high prevalence of AMR and limited resources.

Acknowledgements

This project has received funding from the Rockefeller Foundation, Mphasis M1 Foundation and Axis Bank. We thank Dr. Srikrishna Subramanian for insightful discussions and Dr. Magdalena San Roman as well as Dr. Abhishek Mishra for comments on the manuscript draft. We also thank Dr. Punit Kaur and Dr. Mohit Bhatia for providing sample accession numbers and metadata on request. We thank Centre for Bioinformatics & Computational Biology supported by Department of Biotechnology, for providing computing resources for running the programs/scripts for phylogenetic analysis.

Author contributions

S.K and N.G. conceptualised the study. N.G. carried out the literature review and data screening. N.G and S.K were involved in the analysis, wrote and edited the paper. J.J performed the phylogenetic analysis and commented on the draft of the manuscript. R.K helped with the statistical analysis and phylogenetic analysis and commented on the draft of the manuscript.

Conflict of interests

The authors declare no conflict of interests.

References

1. World Health Organization. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance. Accessed September 14, 2022.

2. Murray CJ, Ikuta KS, Sharara F *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* 2022; **399**: 629–55.

3. Klein EY, Tseng KK, Pant S *et al*. Tracking global trends in the effectiveness of antibiotic therapy using the Drug Resistance Index. *BMJ Glob Health* 2019; **4**: e001315.

4. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. 2016.

5. Pokharel S, Raut S, Adhikari B. Tackling antimicrobial resistance in low-income and middle-income countries. *BMJ Glob Health* 2019; **4**: e002104.

6. Dixit A, Kumar N, Kumar S *et al*. Antimicrobial Resistance: Progress in the Decade since Emergence of New Delhi Metallo-β-Lactamase in India. *Indian J Community Med* 2019; **44**: 4.

7. The Center for Disease Dynamics, Economics & Policy. ResistanceMap: Drug Resistance Index. 2018. https://resistancemap.onehealthtrust.org/DRI.php. Accessed August 24, 2023.

8. Gomez-Simmonds A, Annavajhala MK, McConville TH *et al.* Carbapenemase-producing Enterobacterales causing secondary infections during the COVID-19 crisis at a New York City hospital. *J Antimicrob Chemother* 2021; **76**: 380–4.

9. Teo JQ-M, Lim JC, Tang CY *et al.* Ceftolozane/Tazobactam Resistance and Mechanisms in Carbapenem-Nonsusceptible Pseudomonas aeruginosa. *mSphere* 2021; **6**: e01026-20.

10. Cuypers WL, Jacobs J, Wong V *et al*. Fluoroquinolone resistance in Salmonella: Insights by whole genome sequencing. *Microb Genom* 2018; **4**: e000195.

511 11. Sabat AJ, Budimir A, Nashev D *et al.* Overview of molecular typing methods for outbreak
512 detection and epidemiological surveillance. *Eurosurveillance* 2013; **18**: 20380.
513 12. Genilloud O. Natural products discovery and potential for new antibiotics. *Curr Opin*
514 *Microbiol* 2019; **51**: 81–7.
515 13. Su M, Satola SW, Read TD. Genome-based prediction of bacterial antibiotic resistance. *J Clin*
516 *Microbiol* 2019; **57**: e01405-18.
517 14. Chow NA, Muñoz JF, Gade L *et al.* Tracing the evolutionary history and global expansion of
518 candida auris using population genomic analyses. *mBio* 2020; **11**: 10.1128/mbio.03364-19.
519 15. Dong Y, Zhang F, Wang B *et al*. Laboratory Evolution Assays and Whole-Genome Sequencing
520 for the Development and Safety Evaluation of Lactobacillus plantarum With Stable Resistance to
521 Gentamicin. *Front Microbiol* 2019; **10**: 1235.
522 16. Miragaia M. Factors Contributing to the Evolution of mecA-Mediated β-lactam Resistance in
523 Staphylococci: Update and New Insights From Whole Genome Sequencing (WGS). *Front*
524 *Microbiol* 2018; **9**: 2723.
525 17. Zhang F, Gao J, Wang B *et al.* Whole-genome sequencing reveals the mechanisms for
526 evolution of streptomycin resistance in Lactobacillus plantarum. *J Dairy Sci* 2018; **101**: 2867–74.
527 18. Aljeldah MM. Antimicrobial Resistance and Its Spread Is a Global Threat. *Antibiotics*, 2022;
528 **11**: 1082.
529 19. Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study
530 antimicrobial resistance. *Nat Rev Genet* 2019; **20**: 356–70.

531 20. Ellington MJ, Ekelund O, Aarestrup FM *et al.* The role of whole genome sequencing in
532 antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clinical*
533 *Microbiology and Infection* 2017; **23**: 2–22.
534 21. Papp M, Solymosi N. Review and Comparison of Antimicrobial Resistance Gene Databases.
535 *Antibiotics* 2; **11**: 339.
536 22. van Camp PJ, Haslam DB, Porollo A. Bioinformatics approaches to the understanding of
537 molecular mechanisms in antimicrobial resistance. *Int J Mol Sci* 2020; **21**: 1363.
538 23. Cao MD, Ganesamoorthy D, Elliott AG *et al*. Streaming algorithms for identification of
539 pathogens and antibiotic resistance potential from real-time MinION$^{(TM)}$ sequencing.
540 *Gigascience* 2016; **5**: 32.
541 24. Chatterjee A, Nilgiriwala K, Saranath D *et al*. Whole genome sequencing of clinical strains of
542 Mycobacterium tuberculosis from Mumbai, India: A potential tool for determining drug-
543 resistance and strain lineage. *Tuberculosis* 2017; **107**: 63–72.
544 25. Soundararajan L, Kambli P, Priyadarshini S *et al.* Whole genome enrichment approach for
545 rapid detection of Mycobacterium tuberculosis and drug resistance-associated mutations from
546 direct sputum sequencing. *Tuberculosis* 2020; **121**: 101915.
547 26. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY *et al*. Accurate
548 differentiation of Escherichia coli and Shigella serogroups: challenges and strategies. *New*
549 *Microbes New Infect* 2018; **21**: 58–62.
550 27. NIHR Global Health Research Unit on Genomic Surveillance of AMR. Whole-genome
551 sequencing as part of national and international surveillance programmes for antimicrobial
552 resistance: a roadmap. *BMJ Glob Health* 2020; **5**: e002244.

553 28. Olson RD, Assaf R, Brettin T *et al.* Introducing the Bacterial and Viral Bioinformatics Resource
554 Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res* 2023; **51**: D678–
555 89.
556 29. ICMR. *Antimicrobial Resistance Research and Surveillance Network Annual Report*. 2020.
557 30. McEwen SA, Collignon PJ. Antimicrobial Resistance: a One Health Perspective. *Microbiol*
558 *Spectr* 2018; **6**: 10.1128/microbiolspec.ARBA-0009-2017.
559 31. Page MJ, McKenzie JE, Bossuyt PM *et al.* The PRISMA 2020 statement: an updated guideline
560 for reporting systematic reviews. *BMJ* 2021; **372**: n71.
561 32. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010.
562 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
563 33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
564 *Bioinformatics* 2014; **30**: 2114–20.
565 34. Shakya M, Ahmed SA, Davenport KW *et al*. Standardized phylogenetic and molecular
566 evolutionary analysis applied to species across the microbial tree of life. *Sci Rep* 2020; **10**: 1723.
567 35. Parkhill J, Dougan G, James KD *et al.* Complete genome sequence of a multiple drug
568 resistant Salmonella enterica serovar Typhi CT18. *Nature* 2001; **413**: 848–52.
569 36. Alokam S, Liu S-L, Said K *et al*. Inversions over the terminus region in Salmonella and
570 Escherichia coli: IS200s as the sites of homologous recombination inverting the chromosome of
571 Salmonella enterica serovar typhi. *J Bacteriol* 2002; **184**: 6190–7.
572 37. Deng W, Liou S-R, Plunkett G 3rd *et al.* Comparative genomics of Salmonella enterica serovar
573 Typhi strains Ty2 and CT18. *J Bacteriol* 2003; **185**: 2330–7.

574 38. Thomson N, Baker S, Pickard D *et al.* The role of prophage-like elements in the diversity of
575 Salmonella enterica serovars. *J Mol Biol* 2004; **339**: 279–300.
576 39. Prjibelski A, Antipov D, Meleshko D *et al*. Using SPAdes De Novo Assembler. *Curr Protoc*
577 *Bioinformatics* 2020; **70**: e102.
578 40. Dyson ZA, Holt KE. Five Years of GenoTyphi: Updates to the Global Salmonella Typhi
579 Genotyping Framework. *J Infect Dis* 2021; **224**: S775–80.
580 41. Lam MMC, Wick RR, Watts SC *et al*. A genomic surveillance framework and genotyping tool
581 for Klebsiella pneumoniae and its related species complex. *Nat Commun* 2021; **12**: 4188.
582 42. Moore RM, Harrison AO, McAllister SM *et al*. Iroki: automatic customization and
583 visualization of phylogenetic trees. *PeerJ* 2020; **8**: e8584.
584 43. Rodrigues C, Kapil A, Sharma A *et al.* Whole-Genome Shotgun Sequencing of Cephalosporin-
585 Resistant Salmonella enterica Serovar Typhi. *Genome Announc* 2017; **5**: e01639-16.
586 44. Samajpati S, Pragasam AK, Mandal S *et al*. Emergence of ceftriaxone resistant Salmonella
587 enterica serovar Typhi in Eastern India. *Infection, Genetics and Evolution* 2021; **96**: 105093.
588 45. Carey ME, Jain R, Yousuf M *et al.* Spontaneous Emergence of Azithromycin Resistance in
589 Independent Lineages of Salmonella Typhi in Northern India. *Clinical Infectious Diseases* 2021;
590 **72**: E120–7.
591 46. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Shankar BA *et al*. Draft genome
592 sequence of blaTEM-1-mediated cephalosporin-resistant Salmonella enterica serovar Typhi
593 from bloodstream infection. *J Glob Antimicrob Resist* 2016; **7**: 11–2.

47. Ingle DJ, Andersson P, Valcanis M *et al.* Genomic epidemiology and antimicrobial resistance mechanisms of imported typhoid in Australia. *Antimicrob Agents Chemother* 2021; **65**: e0120021.

48. Ingle DJ, Nair S, Hartman H *et al.* Informal genomic surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial resistance via returning travellers. 2019; **13**: e0007620.

49. Sah R, Donovan S, Seth-Smith HMB *et al.* A Novel Lineage of Ceftriaxone-resistant Salmonella Typhi From India That Is Closely Related to XDR S. Typhi Found in Pakistan. *Clinical Infectious Diseases* 2020; **71**: 1327–30.

50. Katiyar A, Sharma P, Dahiya S *et al*. Genomic profiling of antimicrobial resistance genes in clinical isolates of Salmonella Typhi from patients infected with Typhoid fever in India. *Sci Rep* 2020; **10**: 8299.

51. Pragasam AK, Pickard D, Wong V *et al.* Phylogenetic Analysis Indicates a Longer Term Presence of the Globally Distributed H58 Haplotype of Salmonella Typhi in Southern India. *Clinical Infectious Diseases* 2020; **71**: 1856–63.

52. Shin E, Park J, Jeong HJ *et al.* Emerging high-level ciprofloxacin-resistant Salmonella enterica serovar typhi haplotype H58 in travelers returning to the Republic of Korea from India. *PLoS Negl Trop Dis* 2021; **15**: e0009170.

53. Britto CD, Dyson ZA, Mathias S *et al.* Persistent circulation of a fluoroquinolone-resistant Salmonella enterica Typhi clone in the Indian subcontinent. *Journal of Antimicrobial Chemotherapy* 2020; **75**: 337–41.

615 54. Tanmoy AM, Saha C, Sajib MSI *et al.* CRISPR-Cas Diversity in Clinical Salmonella enterica
616 Serovar Typhi Isolates from South Asian Countries. *Genes* 2020; **11**: 1365.
617 55. Wong VK, Baker S, Pickard DJ *et al.* Phylogeographical analysis of the dominant multidrug-
618 resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission
619 events. *Nat Genet* 2015; **47**: 632–9.
620 56. Jacob JJ, Pragasam AK, Vasudevan K *et al.* Salmonella Typhi acquires diverse plasmids from
621 other Enterobacteriaceae to develop cephalosporin resistance. *Genomics* 2021; **113**: 2171–6.
622 57. Shankar C, Basu S, Lal B *et al.* Aerobactin Seems To Be a Promising Marker Compared With
623 Unstable RmpA2 for the Identification of Hypervirulent Carbapenem-Resistant Klebsiella
624 pneumoniae: In Silico and In Vitro Evidence. *Front Cell Infect Microbiol* 2021; **11**: 709681.
625 58. Bir R, Gautam H, Arif N *et al.* Analysis of colistin resistance in carbapenem-resistant
626 Enterobacterales and XDR Klebsiella pneumoniae. *Ther Adv Infect Dis* 2022; **9**:
627 20499361221080650.
628 59. Sands K, Carvalho MJ, Portal E *et al.* Characterization of antimicrobial-resistant Gram-
629 negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. *Nat*
630 *Microbiol* 2021; **6**: 512–23.
631 60. Shankar C, Jacob JJ, Sugumar SG *et al.* Distinctive Mobile Genetic Elements Observed in the
632 Clonal Expansion of Carbapenem-Resistant Klebsiella pneumoniae in India. *Microbial Drug*
633 *Resistance* 2021; **27**: 1096–104.
634 61. Shankar C, Kumar S, Venkatesan M *et al*. Emergence of ST147 Klebsiella pneumoniae
635 carrying blaNDM-7 on IncA/C2 with ompK35 and ompK36 mutations in India. *J Infect Public*
636 *Health* 2019; **12**: 741–3.

637 62. Pragasam AK, Jennifer S, Solaimalai D *et al.* Expected plazomicin susceptibility in India based
638 on the prevailing aminoglycoside resistance mechanisms in Gram-negative organisms derived
639 from whole-genome sequencing. *Indian J Med Microbiol* 2020; **38**: 313–8.
640 63. Mathur P, Veeraraghavan B, Devanga Ragupathi NK *et al.* First Report on a Cluster of Colistin-
641 Resistant Klebsiella pneumoniae Strains Isolated from a Tertiary Care Center in India: Whole-
642 Genome Shotgun Sequencing. *Genome Announc* 2017; **5**: e01466-16.
643 64. Paul M, Narendrakumar L, Vasanthakumary AR, Joseph I *et al*. Genome sequence of a
644 multidrug-resistant Klebsiella pneumoniae ST78 with high colistin resistance isolated from a
645 patient in India. *J Glob Antimicrob Resist* 2019; **17**: 187–8.
646 65. Kumari N, Kumar M, Katiyar A *et al.* Genome-wide identification of carbapenem-resistant
647 Gram-negative bacterial (CR-GNB) isolates retrieved from hospitalized patients in Bihar, India.
648 *Sci Rep* 2022; **12**: 8477.
649 66. Dey S, Gaur M, Sahoo RK *et al.* Genomic characterization of XDR Klebsiella pneumoniae
650 ST147 co-resistant to carbapenem and colistin – The first report in India. *J Glob Antimicrob*
651 *Resist* 2020; **22**: 54–6.
652 67. Rodrigues C, Desai S, Passet V *et al*. Genomic evolution of the globally disseminated
653 multidrug-resistant Klebsiella pneumoniae clonal group 147. *Microb Genom* 2022; **8**: 000737.
654 68. Nagaraj G, Shamanna V, Govindan V *et al.* High-Resolution Genomic Profiling of
655 Carbapenem-Resistant Klebsiella pneumoniae Isolates: A Multicentric Retrospective Indian
656 Study. *Clinical Infectious Diseases* 2021; **73**: S300–7.

657 69. Shankar C, Vasudevan K, Jacob JJ *et al.* Hybrid Plasmids Encoding Antimicrobial Resistance
658 and Virulence Traits Among Hypervirulent Klebsiella pneumoniae ST2096 in India. *Front Cell*
659 *Infect Microbiol* 2022; **12**: 875116.
660 70. Shankar C, Shankar BA, Manesh A *et al*. KPC-2 producing ST101 Klebsiella pneumoniae from
661 bloodstream infection in India. *J Med Microbiol* 2018; **67**: 927–30.
662 71. Naha S, Sands K, Mukherjee S *et al.* KPC-2-producing Klebsiella pneumoniae ST147 in a
663 neonatal unit: Clonal isolates with differences in colistin susceptibility attributed to AcrAB-TolC
664 pump. *Int J Antimicrob Agents* 2020; **55**: 105903.
665 72. Bhatia M, Shamanna V, Nagaraj G *et al.* Molecular characterisation of carbapenem-resistant
666 Klebsiella pneumoniae clinical isolates: preliminary experience from a tertiary care teaching
667 hospital in the Himalayas. *Trans R Soc Trop Med Hyg* 2022; **116**: 655–62.
668 73. Pragasam AK, Shankar C, Veeraraghavan B *et al.* Molecular Mechanisms of Colistin
669 Resistance in Klebsiella pneumoniae Causing Bacteremia from India-A First Report. *Front*
670 *Microbiol* 2017; **7**: 2135.
671 74. Mathur P, Khurana S, De Man TJB *et al.* Multiple importations and transmission of colistin-
672 resistant Klebsiella pneumoniae in a hospital in northern India. *Infect Control Hosp Epidemiol*
673 2019; **40**: 1387–93.
674 75. Mathur P, Veeraraghavan B, Devanga Ragupathi NK *et al.* Multiple mutations in lipid-A
675 modification pathway & novel fosA variants in colistin-resistant Klebsiella pneumoniae. *Future*
676 *Sci OA* 2018; **4**: FSO319.

677 76. Naha S, Sands K, Mukherjee S *et al*. OXA-181-Like Carbapenemases in Klebsiella pneumoniae
678 ST14, ST15, ST23, ST48, and ST231 from Septicemic Neonates: Coexistence with NDM-5,
679 Resistome, Transmissibility, and Genome Diversity. *mSphere* 2021; **6**: e01156-20.
680 77. Ragupathi NKD, Bakthavatchalam YD, Mathur P *et al.* Plasmid profiles among some ESKAPE
681 pathogens in a tertiary care centre in south India. *Indian Journal of Medical Research* 2019; **149**:
682 222–31.
683 78. Shankar C, Mathur P, Venkatesan M *et al.* Rapidly disseminating blaOXA-232 carrying
684 Klebsiella pneumoniae belonging to ST231 in India: Multiple and varied mobile genetic
685 elements. *BMC Microbiol* 2019; **19**: 137.
686 79. Khan A, Shropshire WC, Hanson B *et al.* Simultaneous infection with Enterobacteriaceae and
687 Pseudomonas aeruginosa harboring multiple carbapenemases in a returning traveler colonized
688 with candida auris. *Antimicrob Agents Chemother* 2020; **64**: e01466-19.
689 80. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Triplicane Dwarakanathan H *et al.* The
690 Influence of Biofilms on Carbapenem Susceptibility and Patient Outcome in Device Associated K.
691 pneumoniae Infections: Insights Into Phenotype vs Genome-Wide Analysis and Correlation.
692 *Front Microbiol* 2020; **11**: 591679.
693 81. Shankar C, Veeraraghavan B, Nabarro LEB *et al*. Whole genome analysis of hypervirulent
694 Klebsiella pneumoniae isolates from community and hospital acquired bloodstream infection.
695 *BMC Microbiol* 2018; **18**: 1–9.
696 82. Manesh A, Shankar C, George MM *et al.* Clinical and Genomic Evolution of Carbapenem-
697 Resistant Klebsiella pneumoniae Bloodstream Infections over Two Time Periods at a Tertiary
698 Care Hospital in South India: A Prospective Cohort Study. *Infect Dis Ther* 2023; **12**: 1319–35.

699 83. Veeraraghavan B, Perumalla SK, Ragupathi NKD *et al.* Coexistence of Fosfomycin and Colistin
700 Resistance in Klebsiella pneumoniae: Whole-Genome Shotgun Sequencing. *Genome Announc*
701 2016; **4**: e01303-16.
702 84. Raj S, Sharma T, Pradhan D *et al.* Comparative Analysis of Clinical and Genomic
703 Characteristics of Hypervirulent Klebsiella pneumoniae from Hospital and Community Settings:
704 Experience from a Tertiary Healthcare Center in India. *Microbiol Spectr* 2022; **10**: e0037622.
705 85. Shukla S, Desai S, Bagchi A *et al.* Diversity and Distribution of β-Lactamase Genes Circulating
706 in Indian Isolates of Multidrug-Resistant Klebsiella pneumoniae. *Antibiotics* 2023; **12**: 449.
707 86. Bir R, Mohapatra S, Kumar A *et al.* Genomic analysis of Fosfomycin resistance in multi-drug
708 resistant uropathogens and comparison of in-vitro susceptibility methods uropathogens. *Iran J*
709 *Microbiol* 2022; **14**: 636–44.
710 87. Mukherjee S, Bhadury P, Mitra S *et al.* Hypervirulent Klebsiella pneumoniae Causing
711 Neonatal Bloodstream Infections: Emergence of NDM-1-Producing Hypervirulent ST11-K2 and
712 ST15-K54 Strains Possessing pLVPK-Associated Markers. *Microbiol Spectr* 2023; **11**: e0412122.
713 88. Naha S, Sands K, Mukherjee S *et al*. A 12 year experience of colistin resistance in Klebsiella
714 pneumoniae causing neonatal sepsis: two-component systems, efflux pumps,
715 lipopolysaccharide modification and comparative phylogenomics. *J Antimicrob Chemother* 2022;
716 **77**: 1586–91.
717 89. Bean DC, Agarwal A, Cherian BP *et al*. Hypermucoviscous polymyxin-resistant Klebsiella
718 pneumoniae from Kolkata, India: Genomic and phenotypic analysis. *J Glob Antimicrob Resist*
719 2019; **17**: 1–2.

720 90. Rafiq Z, Sam N, Vaidyanathan R. Whole genome sequence of Klebsiella pneumoniae U25, a
721 hypermucoviscous, multidrug resistant, biofilm producing isolate from India. *Mem Inst Oswaldo*
722 *Cruz* 2016; **111**: 144–6.
723 91. Avershina E, Sharma P, Taxt AM *et al.* AMR-Diag: Neural network based genotype-to-
724 phenotype prediction of resistance towards β-lactams in Escherichia coli and Klebsiella
725 pneumoniae. *Comput Struct Biotechnol J* 2021; **19**: 1896–906.
726 92. Ranjan A, Shaik S, Nandanwar N *et al.* Comparative genomics of Escherichia coli isolated
727 from skin and soft tissue and other extraintestinal infections. *mBio* 2017; **8**: e01070-17.
728 93. Ghafur A, Shankar C, GnanaSoundari P *et al.* Detection of chromosomal and plasmid-
729 mediated mechanisms of colistin resistance in Escherichia coli and Klebsiella pneumoniae from
730 Indian food samples. *J Glob Antimicrob Resist* 2019; **16**: 48–52.
731 94. Yaita K, Aoki K, Suzuki T *et al.* Epidemiology of extended-spectrum β-lactamase producing
732 Escherichia coli in the stools of returning Japanese travelers, and the risk factors for
733 colonization. *PLoS One* 2014; **9**: e98000.
734 95. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Gajendiran R *et al*. First Indian report of
735 IncX3 plasmid carrying blaNDM-7 in Escherichia coli from bloodstream infection: potential for
736 rapid dissemination. *New Microbes New Infect* 2017; **17**: 65–8.
737 96. Ragupathi NKD, Veeraraghavan B, Sethuvel DPM *et al.* First Indian report on genome-wide
738 comparison of multidrug-resistant Escherichia coli from blood stream infections. *PLoS One*
739 2020; **15**: e0220428.

740 97. Ranjan A, Shaik S, Hussain A *et al.* Genomic and functional portrait of a highly virulent, CTX-
741 M-15-producing H30-Rx subclone of Escherichia coli sequence type 131. *Antimicrob Agents*
742 *Chemother* 2015; **59**: 6087–95.
743 98. Palani GS, Ghafur A, Krishnan P *et al*. Intestinal carriage of colistin resistant
744 Enterobacteriaceae in hospitalized patients from an Indian center. *Diagn Microbiol Infect Dis*
745 2020; **97**: 114998.
746 99. Ranjith K, SaiAbhilash CR, Sai Prashanthi G *et al*. Phylogenetic Grouping of Human Ocular
747 Escherichia coli Based on Whole-Genome Sequence Analysis. *Microorganisms* 2020; **8**: 422.
748 100. Sundaramoorthy NS, Suresh P, Selva Ganesan S *et al*. Restoring colistin sensitivity in
749 colistin-resistant E. coli: Combinatorial use of MarR inhibitor with efflux pump inhibitor. *Sci Rep*
750 2019; **9**: 19845.
751 101. Hussain A, Shaik S, Ranjan A *et al.* Risk of transmission of antimicrobial resistant
752 Escherichia coli from commercial broiler and free-range retail chicken in India. *Front Microbiol*
753 2017; **8**: 2120.
754 102. Karade S, Sen S, Shergill SPS *et al*. Whole genome sequence of colistin-resistant Escherichia
755 coli from western India. *Med J Armed Forces India* 2021; **77**: 297–301.
756 103. Beg AZ, Khan AU. Genome analyses of blaNDM-4 carrying ST 315 Escherichia coli isolate
757 from sewage water of one of the Indian hospitals. *Gut Pathog* 2018; **10**: 1–6.
758 104. Hussain A, Shaik S, Ranjan A *et al.* Genomic and Functional Characterization of Poultry
759 Escherichia coli From India Revealed Diverse Extended-Spectrum β-Lactamase-Producing
760 Lineages With Shared Virulence Profiles. *Front Microbiol* 2019; **10**: 2766.

761 105. Ramakrishnan V, Marialouis XA, Sankarasubramanian J *et al*. Whole Genomic analysis of a
762 clinical isolate of Uropathogenic Escherichia coli strain of Sequence Type - 101 carrying the drug
763 resistance NDM-7 in IncX3 plasmid. *Bioinformation* 2021; **17**: 126–31.
764 106. Shaik S, Ranjan A, Tiwari SK *et al.* Comparative genomic analysis of globally dominant
765 ST131 clone with other epidemiologically successful extraintestinal pathogenic Escherichia coli
766 (ExPEC) lineages. *mBio* 2017; **8**: e01596-17.
767 107. Devanga Ragupathi NK, Vasudevan K, Venkatesan M *et al*. First Indian report on B4/H24RxC
768 ST410 multidrug-resistant Escherichia coli from bloodstream infection harbouring blaOXA-181
769 and blaCTX-M-15. *J Glob Antimicrob Resist* 2020; **22**: 568–70.
770 108. Yadav M, Pundir S, Kumari R *et al.* Virulence gene mutations as a differentiator of clinical
771 phenotypes: insights from community-acquired uropathogenic Escherichia coli. *Microbiology*
772 *(United Kingdom)* 2022; **168**: 10.1099/mic.0.001161.
773 109. Patel MA, Pandey A, Patel AC *et al.* Whole genome sequencing and characteristics of
774 extended-spectrum beta-lactamase producing Escherichia coli isolated from poultry farms in
775 Banaskantha, India. *Front Microbiol* 2022; **13**: 996214.
776 110. Bhagwat SS, Hariharan P, Joshi PR *et al.* Activity of cefepime/zidebactam against MDR
777 Escherichia coli isolates harbouring a novel mechanism of resistance based on four-amino-acid
778 inserts in PBP3. *J Antimicrob Chemother* 2020; **75**: 3563–7.
779 111. Chopra M, Bandyopadhyay S, Bhattacharyya D *et al*. Complete genome sequence and
780 comparative analysis of antibiotic resistance plasmids in carbapenem-resistant Escherichia coli
781 from bovine mastitis. *Indian J Biotechnol* 2021; **20**: 13–23.

782 112. Suresh A, Shaik S, Baddam R *et al.* Evolutionary Dynamics Based on Comparative Genomics
783 of Pathogenic Escherichia coli Lineages Harboring Polyketide Synthase (pks) Island. *mBio* 2021;
784 **12**: e03634-20.
785 113. Das Mitra S, Bandopadhyay S, Jadhao S *et al*. Genetic characterization and comparative
786 genomics of a multi drug resistant (MDR) Escherichia coli SCM-21 isolated from a subclinical
787 case of bovine mastitis. *Comp Immunol Microbiol Infect Dis* 2022; **85**: 101799.
788 114. Chopra M, Bandyopadhyay S, Bhattacharya D *et al.* Genome based phylogeny and virulence
789 factor analysis of mastitis causing Escherichia coli isolated from Indian cattle. *Indian J Anim Sci*
790 2021; **90**: 1577–83.
791 115. Moser AI, Campos-Madueno EI, Sendi P *et al.* Repatriation of a patient with COVID-19
792 contributed to the importation of an emerging carbapenemase producer. *J Glob Antimicrob*
793 *Resist* 2021; **27**: 267–72.
794 116. Ranjan A, Shaik S, Mondal A *et al.* Molecular Epidemiology and Genome Dynamics of New
795 Delhi Metallo-β-Lactamase-Producing Extraintestinal Pathogenic Escherichia coli Strains from
796 India. *Antimicrob Agents Chemother* 2016; **60**: 6795–805.
797 117. Kumkar SN, Kamble EE, Chavan NS *et al*. Diversity of resistant determinants, virulence
798 factors, and mobile genetic elements in Acinetobacter baumannii from India: A comprehensive
799 in silico genome analysis. *Frontiers in Cellular and Infection Microbiology* 2022; **12**: 997897.
800 118. Vijayakumar S, Jacob JJ, Vasudevan K *et al.* Genomic Characterization of Mobile Genetic
801 Elements Associated With Carbapenem Resistance of Acinetobacter baumannii From India.
802 *Front Microbiol* 2022; **13**: 869653.

803 119. Naha A, Vijayakumar S, Lal B *et al.* Genome sequencing and molecular characterisation of
804 XDR Acinetobacter baumannii reveal complexities in resistance: Novel combination of
805 sulbactam–durlobactam holds promise for therapeutic intervention. *J Cell Biochem* 2021; **122**:
806 1946–57.
807 120. Vijayakumar S, Anandan S, MS DP *et al.* Insertion sequences and sequence types profile of
808 clinical isolates of carbapenem-resistant A. baumannii collected across India over four year
809 period. *J Infect Public Health* 2020; **13**: 1022–8.
810 121. Vijayakumar S, S BA, Kanthan K *et al*. Whole-genome shotgun sequences of seven colistin-
811 resistant Acinetobacter baumannii isolates from bacteraemia. *J Glob Antimicrob Resist* 2018; **12**:
812 155–6.
813 122. Yadav A, Jain K, Wang Y *et al.* Candida auris on Apples: Diversity and Clinical Significance.
814 *mBio* 2022; **13**: e0051822.
815 123. Desnos-Ollivier M, Fekkar A, Bretagne S. Earliest case of Candida auris infection imported
816 in 2007 in Europe from India prior to the 2009 description in Japan. *Journal of Medical*
817 *Mycology* 2021; **31**: 101139.
818 124. Arora P, Singh P, Wang Y *et al.* Environmental Isolation of Candida auris from the Coastal
819 Wetlands of Andaman Islands, India. *mBio* 2021; **12**: e03181-20.
820 125. Yadav A, Singh A, Wang Y *et al.* Colonisation and Transmission Dynamics of Candida auris
821 among Chronic Respiratory Diseases Patients Hospitalised in a Chest Hospital, Delhi, India: A
822 Comparative Analysis of Whole Genome Sequencing and Microsatellite Typing. *J Fungi (Basel)*
823 2021; **7**: 1–16.

824 126. Lockhart SR, Etienne KA, Vallabhaneni S *et al.* Simultaneous emergence of multidrug-
825 resistant candida auris on 3 continents confirmed by whole-genome sequencing and
826 epidemiological analyses. *Clinical Infectious Diseases* 2017; **64**: 134–40.
827 127. Sharma C, Kumar N, Pandey R *et al*. Whole genome sequencing of emerging multidrug
828 resistant Candida auris isolates in India demonstrates low genetic variation. *New Microbes New*
829 *Infect* 2016; **13**: 77–82.
830 128. Chatterjee S, Alampalli SV, Nageshan RK *et al*. Draft genome of a commonly misdiagnosed
831 multidrug resistant pathogen Candida auris. *BMC Genomics* 2015; **16**: 686.
832 129. Li X, Muñoz JF, Gade L *et al.* Comparing genomic variant identification protocols for
833 Candida auris. *Microb Genom* 2023; **9**: mgen000979.
834 130. Rohit A, Suresh Kumar D, Dhinakaran I *et al.* Whole-genome-based analysis reveals
835 multiclone Serratia marcescens outbreaks in a non-Neonatal Intensive Care Unit setting in a
836 tertiary care hospital in India. *J Med Microbiol* 2019; **68**: 616–21.
837 131. Shankar C, Karunasree S, Manesh A *et al*. First Report of Whole-Genome Sequence of
838 Colistin-Resistant Klebsiella quasipneumoniae subsp. similipneumoniae Producing KPC-9 in
839 India. *Microbial Drug Resistance* 2019; **25**: 489–93.
840 132. Mann A, Malik S, Rana JS *et al*. Whole genome sequencing data of Klebsiella aerogenes
841 isolated from agricultural soil of Haryana, India. *Data Brief* 2021; **38**: 107311.
842 133. Mishra M, Panda S, Barik S *et al*. Antibiotic Resistance Profile, Outer Membrane Proteins,
843 Virulence Factors and Genome Sequence Analysis Reveal Clinical Isolates of Enterobacter Are
844 Potential Pathogens Compared to Environmental Isolates. *Front Cell Infect Microbiol* 2020; **10**:
845 54.

134. Mishra M, Patole S, Mohapatra H. Draft Genome Sequences of Nonclinical and Clinical Enterobacter cloacae Isolates Exhibiting Multiple Antibiotic Resistance and Virulence Factors. *Genome Announc* 2017; **5**: e01218-17.

135. Chanakya PP, Khamari B, Lama M *et al.* Complete genome sequence of an extensively drug resistant (XDR) M. morganii SMM01 isolated from a patient with urinary and fecal incontinence. *BMC Genom Data* 2021; **22**: 27.

136. Mehrotra T, Konar D, Pragasam AK *et al.* Antimicrobial resistance heterogeneity among multidrug-resistant Gram-negative pathogens: Phenotypic, genotypic, and proteomic analysis. *Proceedings of the National Academy of Sciences* 2023; **120**: e2305465120.

137. Bokhary H, Pangesti KNA, Rashid H *et al*. Travel-Related Antimicrobial Resistance: A Systematic Review. *Trop Med Infect Dis* 2021; **6**: 11.

138. Frost I, Van Boeckel TP, Pires J *et al*. Global geographic trends in antimicrobial resistance: the role of international travel. *J Travel Med* 2019; **26**: taz036.

139. Sridhar S, Turbett SE, Harris JB *et al*. Antimicrobial-resistant bacteria in international travelers. *Curr Opin Infect Dis* 2021; **34**: 423–31.

140. Farkas K, Williams R, Alex-Sanders N *et al.* Wastewater-based monitoring of SARS-CoV-2 at UK airports and its potential role in international public health surveillance. *PLOS Global Public Health* 2023; **3**: e0001346.

141. Li J, Hosegood I, Powell D *et al.* A global aircraft-based wastewater genomic surveillance network for early warning of future pandemics. *Lancet Glob Health* 2023; **11**: e791–5.

142. Agrawal S, Orschler L, Tavazzi S *et al*. Genome Sequencing of Wastewater Confirms the Arrival of the SARS-CoV-2 Omicron Variant at Frankfurt Airport but Limited Spread in the City of Frankfurt, Germany, in November 2021. *Microbiol Resour Announc* 2022; **11**: e0122921.

143. Stanley D, Batacan RJ, Bajagai YS. Rapid growth of antimicrobial resistance: the role of agriculture in the problem and the solutions. *Appl Microbiol Biotechnol* 2022; **106**: 6953–62.

144. Noman SM, Shafiq M, Bibi S *et al.* Exploring antibiotic resistance genes, mobile gene elements, and virulence gene factors in an urban freshwater samples using metagenomic analysis. *Environ Sci Pollut Res Int* 2023; **30**: 2977–90.

145. Adelowo OO, Ikhimiukor OO, Knecht C *et al.* A survey of extended-spectrum beta-lactamase-producing Enterobacteriaceae in urban wetlands in southwestern Nigeria as a step towards generating prevalence maps of antimicrobial resistance. *PLoS One* 2020; **15**: e0229451.

146. Datta P, Gupta V, Arora S *et al*. Epidemiology of extended-spectrum β-lactamase, AmpC, and carbapenemase production in Proteus mirabilis. *Jpn J Infect Dis* 2014; **67**: 44–6.

147. Patel NB, Jain G, Chandrakar S *et al*. Ventilator-associated pneumonia due to carbapenem-resistant Providencia rettgeri. *BMJ Case Rep* 2021; **14**: e243908.

148. Karad DD, Somani Y, Khande H *et al*. Molecular characterization of a multidrug-resistant/pandrug-resistant nosocomial polymicrobial infection with Klebsiella pneumoniae, Providencia rettgeri, and Acinetobacter baumannii from Rural Maharashtra, India. *Acta Biochim Pol* 2020; **67**: 387-392.

149. Praharaj AK, Khajuria A, Kumar M *et al*. Phenotypic detection and molecular characterization of beta-lactamase genes among Citrobacter species in a tertiary care hospital. *Avicenna J Med* 2016; **6**: 17–27.

888 150. Jain S, Gaind R, Kothari C *et al.* VEB-1 extended-spectrum β-lactamase-producing
889 multidrug-resistant Proteus mirabilis sepsis outbreak in a neonatal intensive care unit in India:
890 clinical and diagnostic implications. *JMM Case Rep* 2016; **3**: e005056.
891 151. ICMR. *Antimicrobial Resistance Research and Surveillance Network Annual Report*. 2021.
892 152. Mohan S, Agarwal J, Srivastava R *et al*. Observations on Citrobacter species from a tertiary
893 care health center with special reference to multi-drug resistance and presence of CTX-M gene.
894 *Indian J Pathol Microbiol* 2014; **57**: 439–41.
895 153. Chinnam BK, Nelapati S, Tumati SR *et al*. Detection of β-Lactamase–Producing Proteus
896 mirabilis Strains of Animal Origin in Andhra Pradesh, India and Their Genetic Diversity. *J Food*
897 *Prot* 2021; **84**: 1374–9.
898 154. Brockhurst MA, Harrison E, Hall JPJ *et al*. The Ecology and Evolution of Pangenomes.
899 *Current Biology* 2019; **29**: R1094–103.
900 155. Medini D, Donati C, Tettelin H *et al*. The microbial pan-genome. *Curr Opin Genet Dev* 2005;
901 **15**: 589–94.
902 156. Dijkshoorn L, Nemec A, Seifert H. An increasing threat in hospitals: multidrug-resistant
903 Acinetobacter baumannii. *Nat Rev Microbiol* 2007; **5**: 939–51.
904 157. Moubareck CA, Halat DH. Insights into Acinetobacter baumannii: A Review of
905 Microbiological, Virulence, and Resistance Traits in a Threatening Nosocomial Pathogen.
906 *Antibiotics (Basel)* 2020; **9**: 119.
907 158. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical
908 environments. *Can J Microbiol* 2019; **65**: 34–44.