


Evidence of horizontal gene transfer and environmental selection impacting antibiotic resistance evolution in soil-dwelling *Listeria*

Received: 15 May 2024

Accepted: 11 November 2024

Published online: 19 November 2024

 Check for updates

Ying-Xian Goh^{1,2}, Sai Manohar Balu Anupaju³, Anthony Nguyen⁴,
Hailong Zhang⁵, Monica Ponder^{2,6}, Leigh-Anne Krometis^{2,7}, Amy Pruden^{1,2} &
Jingqiu Liao^{1,2} 

Soil is an important reservoir of antibiotic resistance genes (ARGs) and understanding how corresponding environmental changes influence their emergence, evolution, and spread is crucial. The soil-dwelling bacterial genus *Listeria*, including *L. monocytogenes*, the causative agent of listeriosis, serves as a key model for establishing this understanding. Here, we characterize ARGs in 594 genomes representing 19 *Listeria* species that we previously isolated from soils in natural environments across the United States. Among the five putatively functional ARGs identified, *lin*, which confers resistance to lincomycin, is the most prevalent, followed by *mprF*, *sul*, *fosX*, and *norB*. ARGs are predominantly found in *Listeria sensu stricto* species, with those more closely related to *L. monocytogenes* tending to harbor more ARGs. Notably, phylogenetic and recombination analyses provide evidence of recent horizontal gene transfer (HGT) in all five ARGs within and/or across species, likely mediated by transformation rather than conjugation and transduction. In addition, the richness and genetic divergence of ARGs are associated with environmental conditions, particularly soil properties (e.g., aluminum and magnesium) and surrounding land use patterns (e.g., forest coverage). Collectively, our data suggest that recent HGT and environmental selection play a vital role in the acquisition and diversification of bacterial ARGs in natural environments.

As a natural reservoir of antibiotic-resistance genes (ARGs), including those encountered in human pathogens^{1,2}, soil plays a pivotal role in the emergence, evolution, and dissemination of antibiotic resistance across diverse ecosystems^{3,4}. *Listeria*, a bacterial genus that includes two pathogenic members,

L. monocytogenes and *L. ivanovii*, is commonly found in soils⁵. *L. monocytogenes* causes listeriosis in vulnerable human populations with a notable fatality rate of 20–30%^{6,7}, while *L. ivanovii* rarely causes listeriosis in humans and is primarily a pathogen of ruminant animals⁸. *Listeria* can be broadly divided into two groups: *sensu*

¹Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA. ²Center for Emerging, Zoonotic, and Arthropod-Borne Pathogens, Virginia Tech, Blacksburg, VA 24061, USA. ³Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA. ⁴Computational Modeling & Data Analytics Program, Virginia Tech, Blacksburg, VA 24061, USA. ⁵Department of Business Information Technology, Virginia Tech, Blacksburg, VA 24061, USA. ⁶Department of Food Science and Technology, Virginia Tech, Blacksburg, VA 24061, USA. ⁷Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA. ✉e-mail: liao@vt.edu

stricto and *sensu lato*, based on the relatedness of species to *L. monocytogenes*, with *sensu stricto* species being more closely related⁹. The standard treatment for listeriosis is a combination of penicillin and aminopenicillins (ampicillin or amoxicillin)⁹ or ampicillin and gentamicin¹⁰. While the incidence of resistance among clinical *L. monocytogenes* to these antibiotics remains low at present, intrinsic resistance to cephalosporins exists, and increased resistance to penicillin, trimethoprim, and rifampicin has been observed^{9,11–14}. Furthermore, accelerated rates of antibiotic resistance in *L. monocytogenes* have been observed in food-associated environments¹⁵, possibly due to prolonged exposure to sublethal concentrations of antimicrobial agents in food processing and agriculture settings^{16,17}, such as poultry^{15,18} and fresh produce factories^{19,20}. Since *Listeria* can be transmitted from soils directly to humans²¹, or indirectly, via the food production chain²², they could be a key model for understanding how ARGs carried by soil microbes can be potentiated into human pathogens. Also, establishing a fundamental understanding of the ecological and evolutionary drivers of antibiotic resistance among soil-dwelling *Listeria* could help to better interpret current and future trends in antibiotic resistance patterns observed in food and clinical isolates. However, most studies of ARGs in *Listeria* have primarily focused on food-related and clinical isolates of *L. monocytogenes*^{18,23–25}, resulting in an incomplete understanding of the dynamics of ARGs in *Listeria* in the environment.

Previous studies indicate that environmental factors, such as nutrient availability, temperature, pH, and exposure to natural or anthropogenic chemicals, can exert selective pressure favoring antibiotic resistance^{26,27}. Genes essential for metabolism and behavior, including ARGs, have been observed to undergo positive selection (PS) to adapt to varying environments^{28,29}. Apart from PS, environmental pressures can facilitate horizontal gene transfer (HGT), a pivotal pathway for the evolution of new resistant strains³⁰. HGT typically occurs through three mechanisms, transformation (i.e., the uptake of free DNA from the environment), conjugation (i.e., the direct transfer of genetic material from one bacterium to another through physical contact typically encoded by plasmids or transposons), and transduction (i.e., transfer of genetic material via viruses)³¹. Existing evidence suggests that the acquisition of ARGs among *L. monocytogenes* is mediated by HGT. For example, the acquisition of tetracycline and trimethoprim resistances in *L. monocytogenes* has been experimentally linked to transposons like Tn916-Tn1545 and Tn6198^{32,33}. Despite the important role of environmental factors and HGT on the evolution of antibiotic resistance, the prevalence of ARGs, the extent of HGT and PS acting on them, and the influence of environmental factors on the distribution and evolution of ARGs in *Listeria* in the soil environment remains largely unknown.

To bridge these knowledge gaps, we comprehensively examined the distribution of ARGs and associated HGT, PS, and environmental factors using a unique set of whole-genome sequencing data and paired environmental variables for 594 soil-dwelling *Listeria* isolates representing 19 species, including *L. monocytogenes*, that we collected across the United States (US). We identified five putatively functional ARGs, *lin*, *mprF*, *sul*, *fosX*, and *norB*. These ARGs were predominantly found among *Listeria sensu stricto* species and showed evidence of HGT within and/or across *Listeria* species, likely mediated by transformation, rather than conjugation and transduction. With ecological analysis and machine learning models, we also revealed evidence of environmental selection acting on the richness and genetic divergence of ARGs, likely mainly triggered by soil properties and surrounding land use, respectively. This study yields new insights into the dynamics of antibiotic resistance in soils and suggests that environmental disturbance may facilitate the emergence and spread of ARGs among bacterial species.

Results

Prevalence and spatial distribution of ARGs in soil-dwelling *Listeria*

In this study, a gene was defined as a “functional gene” if the sequence coverage exceeded 80% and no premature stop codon was detected, and as a “truncated gene” if the sequence coverage ranged between 30% and 80% or premature stop codon was detected³⁴. A gene being either functional or truncated was referred to as a “present gene” (see Methods). We identified seven distinct ARGs in soil-dwelling *Listeria*: *lin*, *mprF*, *sul*, *fosX*, *norB*, *dfrD*, and *mphB*. Specifically, *lin* confers resistance to lincomycin, *mprF* to defensin, daptomycin, and gallidermin, *sul* to sulfamethoxazole, *fosX* to fosfomycin, *norB* to fluoroquinolones and nalidixic acid, *dfrD* to trimethoprim, and *mphB* to erythromycin, telithromycin, quinupristin, pristinamycin IA, and virginiamycin S^{35–37}. Among these ARGs, the majority of *lin*, *mprF*, *sul*, *fosX*, and *norB* were functional, while *dfrD* and *mphB* were truncated and were each present in only 0.17% of *Listeria* genomes (Fig. 1a). Among the functional ARGs, *lin* was most prevalent among *Listeria* genomes (82.66%) followed by *mprF* (82.32%), *sul* (81.14%), *fosX* (60.77%), and *norB* (58.42%) (Fig. 1a).

Overall, the high richness of functional ARGs was consistently observed in all *sensu stricto* species, especially *L. monocytogenes*, *L. innocua*, *L. marthii*, *L. farberii*, *L. cossartiae*, and *L. swaminathanii*, but not in *sensu lato* species (Fig. 1b). While ARGs were present in *sensu lato* species, nearly all of them were truncated and the overall prevalence was lower than *sensu stricto* species (Fig. 1c). Specifically, *lin*, *mprF*, and *sul* were consistently present in all genomes of *sensu stricto* species ($n = 491$, Supplementary Fig. 1a), with each being functional in 100.00%, 99.59%, and 98.17% of these genomes, respectively (Fig. 1b). Functional *fosX* was found in all *sensu stricto* species, except for *L. seeligeri*, *L. immobilis*, and *L. ivanovii*. Functional *norB* was found in all *sensu stricto* species, except for *L. welshimeri* (Fig. 1b). Among *sensu lato* species, functional ARGs were only detected in *fosX* in one *L. booriae* genome and one *L. rocourtiae* genome (Fig. 1b). Notably, ARG richness in *Listeria* species was highly correlated with their genetic similarity to *L. monocytogenes* for both present and functional ARGs (Spearman's $\rho = 0.88$ and 0.88 , $P = 1.2e-06$ and $1.3e-06$, respectively; Supplementary Fig. 1b, Fig. 1d). This correlation indicates that species more closely related to *L. monocytogenes* tend to manifest a higher richness of ARGs.

The richness of both present and functional ARGs displayed spatial heterogeneity across the US (Supplementary Fig. 1c and Fig. 1e, respectively). Of note, eastern regions exhibited significantly higher present and functional ARG richness compared to western regions (Mann-Whitney $U P = 1e-18$ and $1.6e-12$, respectively; Fig. 1f). The geographic signal of ARG richness appeared to be driven by the distribution of species, as *Listeria sensu stricto* species were more prevalent in the eastern regions, especially *L. monocytogenes*, which harbored high ARG richness (Fig. 1e, Supplementary Fig. 1c).

Evidence of HGT and PS acting on ARGs in soil-dwelling *Listeria*

To gain insights into the phylogenetic origins and HGT of the five functional ARGs (i.e., *lin*, *mprF*, *sul*, *fosX*, and *norB*), a gene tree was constructed for each ARG, depicted in Fig. 2a–e. If a gene is inherited vertically without any HGT, we anticipate the topology of a gene tree aligning with that of the phylogenetic tree, constructed based on more informative genetic variants, such as core single nucleotide polymorphisms (SNPs). Based on this notion, the comparison of gene trees and corresponding core SNP-based trees (Supplementary Figs. 2a–e) suggests that *lin* (Fig. 2a), *fosX* (Fig. 2b), and *norB* (Fig. 2c) likely undergo HGT between *sensu stricto* species. For example, for *lin* (Fig. 2a), we observed clades that included a mix of *L. marthii* and *L. cossartiae* isolates, as well as *L. innocua* and *L. farberii* isolates. Similarly, *fosX* (Fig. 2b) displayed a comparable pattern among *sensu stricto* species, with clades containing a mix of *L. marthii* and *L. cossartiae*

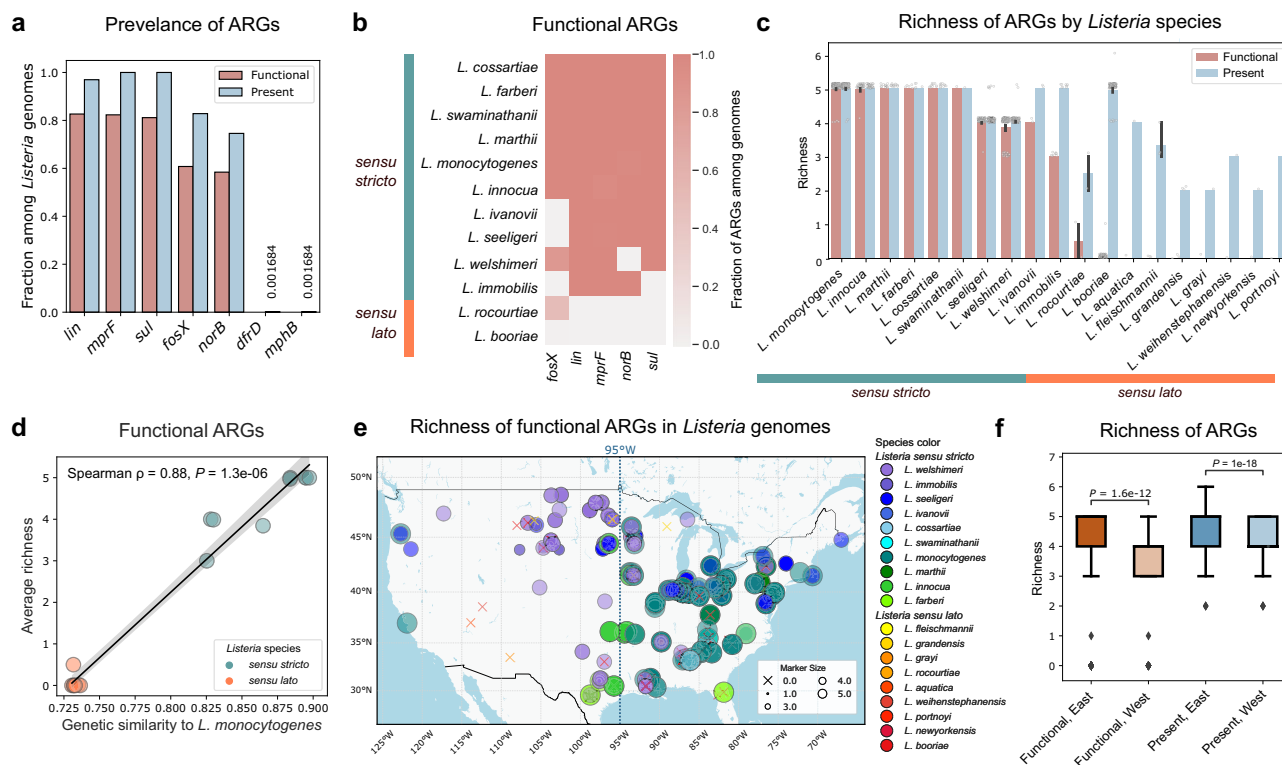


Fig. 1 | National ARG profiles of soil-dwelling *Listeria*. **a** Prevalence of both present (blue) and functional (red) ARGs across *Listeria* genomes. **b** Proportion of functional ARGs among different *Listeria* species, both *sensu stricto* (green) and *sensu lato* (orange). **c** Richness of both present (blue) and functional (red) ARGs in different *Listeria* species. Error bars, mean \pm standard deviation. For *Listeria sensu stricto* (green) species, sample size N is as follows: *L. monocytogenes* (177), *L. innocua* (33), *L. marthii* (14), *L. farberii* (5), *L. cossartiae* (11), *L. swaminathanii* (1), *L. seeligeri* (98), *L. welshimeri* (141), *L. ivanovii* (2), and *L. immobilis* (9). For *Listeria sensu lato* (orange) species, N is as follows: *L. rocourtiae* (2), *L. booriae* (90), *L. aquatica* (1), *L. fleischmannii* (3), *L. grandensis* (3), *L. grayi* (1), *L. weihenstephanensis* (1), *L. portnoyi* (1), and *L. newyorkensis* (1). **d** Spearman's correlation between the genetic similarity to *L. monocytogenes* and average richness of functional ARGs in *Listeria* species. Genetic similarity was calculated based on pairwise average nucleotide identity (ANI) between genomes for a given species and

L. monocytogenes. ρ represents Spearman's rank correlation coefficient. The line and shaded area depict the best-fit trendline and the 95% confidence interval (mean \pm 1.96 standard error of the mean, SEM) for the linear regression. **e** Richness of functional ARGs among *Listeria* genomes across the US. The map was divided into eastern and western regions based on the longitudinal coordinate of the center of the US (-95°). Circles and crosses indicate genomes with and without functional ARGs, respectively, and are color-coded by species. Circle size is proportional to the richness of functional and present ARGs in each genome. **f** Richness of functional and present ARGs among *Listeria* genomes compared between the eastern and western US. A two-sided Mann-Whitney U test $P < 0.05$ was considered statistically significant. $N = 417$ and 177 for the eastern and western regions, respectively, for both present and functional ARG richness. Box plots show the interquartile range (IQR), with the line representing the median and whiskers extending to 1.5 times the IQR.

isolates, as well as *L. welshimeri* and *L. monocytogenes* isolates. Notably, the apparent HGT of *fosX* between *L. welshimeri* and *L. monocytogenes* isolates serves as an example of gene transfer between pathogens (*L. monocytogenes*) and non-pathogens (*L. welshimeri*). Regarding *norB* (Fig. 2c), evidence of HGT was observed between *L. innocua* and *L. farberii* isolates, as well as among *L. monocytogenes* (pathogen), *L. innocua* (non-pathogen), and *L. seeligeri* (non-pathogen) isolates. While gene trees of *mprF* (Fig. 2d) and *sul* (Fig. 2e) did not have major clades of isolates from different species, they may have undergone HGT within a species, which is not easily observable in the tree comparison. Indeed, multiple statistical tests assessing the congruence between trees, including bootstrap proportion using resampling of estimated log-likelihoods (RELL)³⁸, Kishino-Hasegawa (KH)³⁹, Shimodaira-Hasegawa (SH)⁴⁰, expected likelihood weight (ELW)⁴¹, and approximately unbiased (AU)⁴² tests, indicated a significant difference between the gene tree and the corresponding core SNP-based tree for all ARGs ($P < 0.001$ for all; Supplementary Table 1). These results suggest that ARGs in soil-dwelling *Listeria* have evolved differently from the overall species history due to HGT.

To provide more evidence of HGT of ARGs among *Listeria* species, we further detected homologous recombination, a biological process used in HGT by bacteria to exchange genetic material, using nine

methods implemented in Recombination Detection Program v4 (RDP4)⁴³, including RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, PHYLPRO, LARD, and 3SEQ (see Methods). Recombination events were observed for all ARGs (Fig. 2f). A total of 20 recombination events were observed in *lin*, with 14 occurring within species and six across species (Fig. 2f). These included events among *L. marthii*, *L. cossartiae*, and *L. swaminathanii* isolates as well as between *L. farberii* and *L. innocua* isolates (Supplementary Data 1), which were also reflected in the tree comparison (Fig. 2a). Recombination was also detected within *L. monocytogenes* in *lin* (14 events; Supplementary Data 1). For *fosX*, three recombination events were observed, and they all occurred across species (Fig. 2f, Supplementary Data 2), including those among *L. monocytogenes*, *L. welshimeri*, and *L. cossartiae* isolates that were also detected in the tree comparison (Fig. 2b), and between *L. innocua* and *L. farberii* isolates. For *norB*, 18 recombination events were detected, with four occurring within species and 14 across species (Fig. 2f). For instance, recombination detected among *L. monocytogenes*, *L. innocua*, and *L. seeligeri* isolates supports the observation in the tree comparison that HGT may occur between pathogens and non-pathogens (Supplementary Data 3, Fig. 2c). For *mprF*, 14 recombination events (12 within species and two across species) were detected (Fig. 2f), with some involved between *L.*

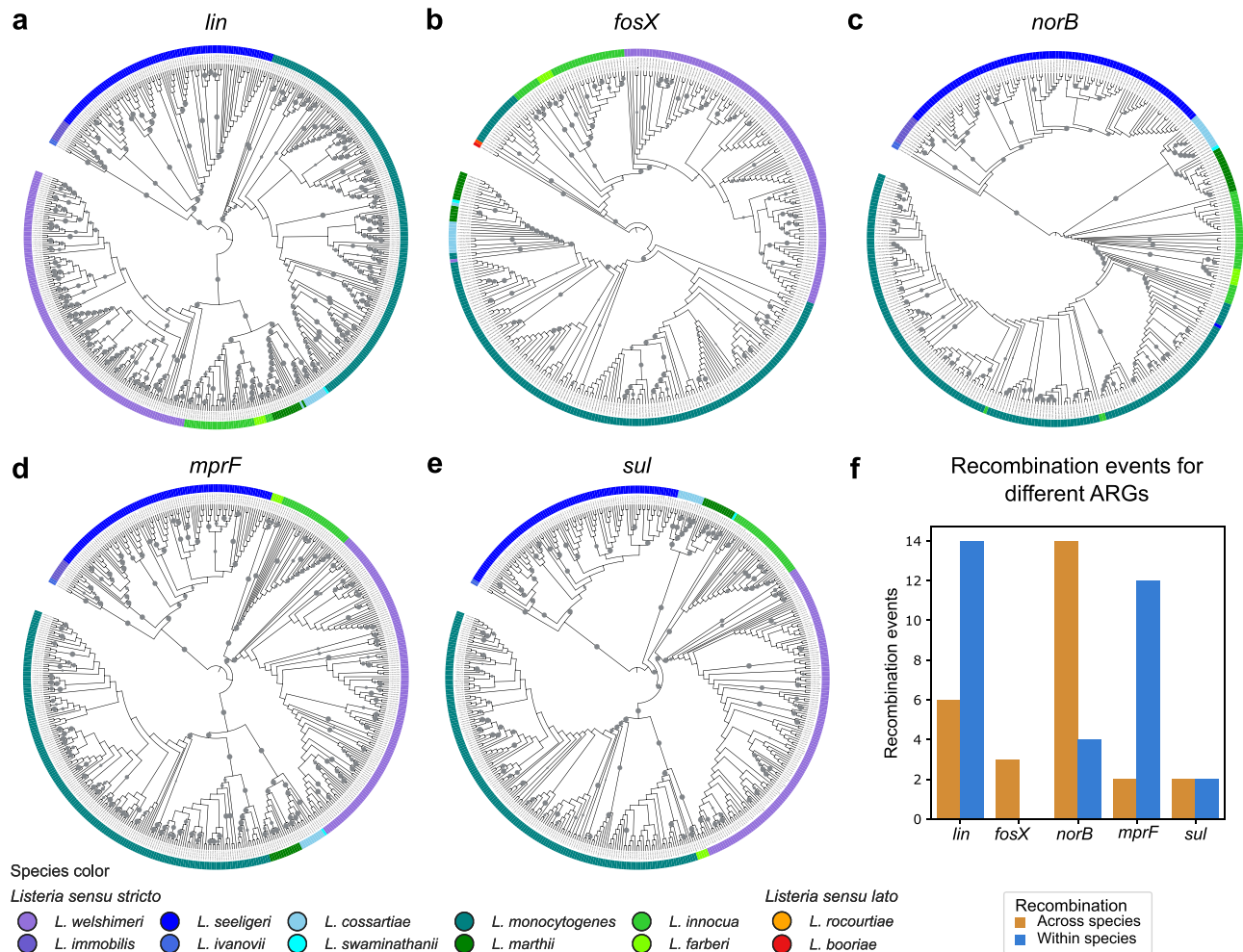


Fig. 2 | Evidence of HGT of functional ARGs among *Listeria* isolates.

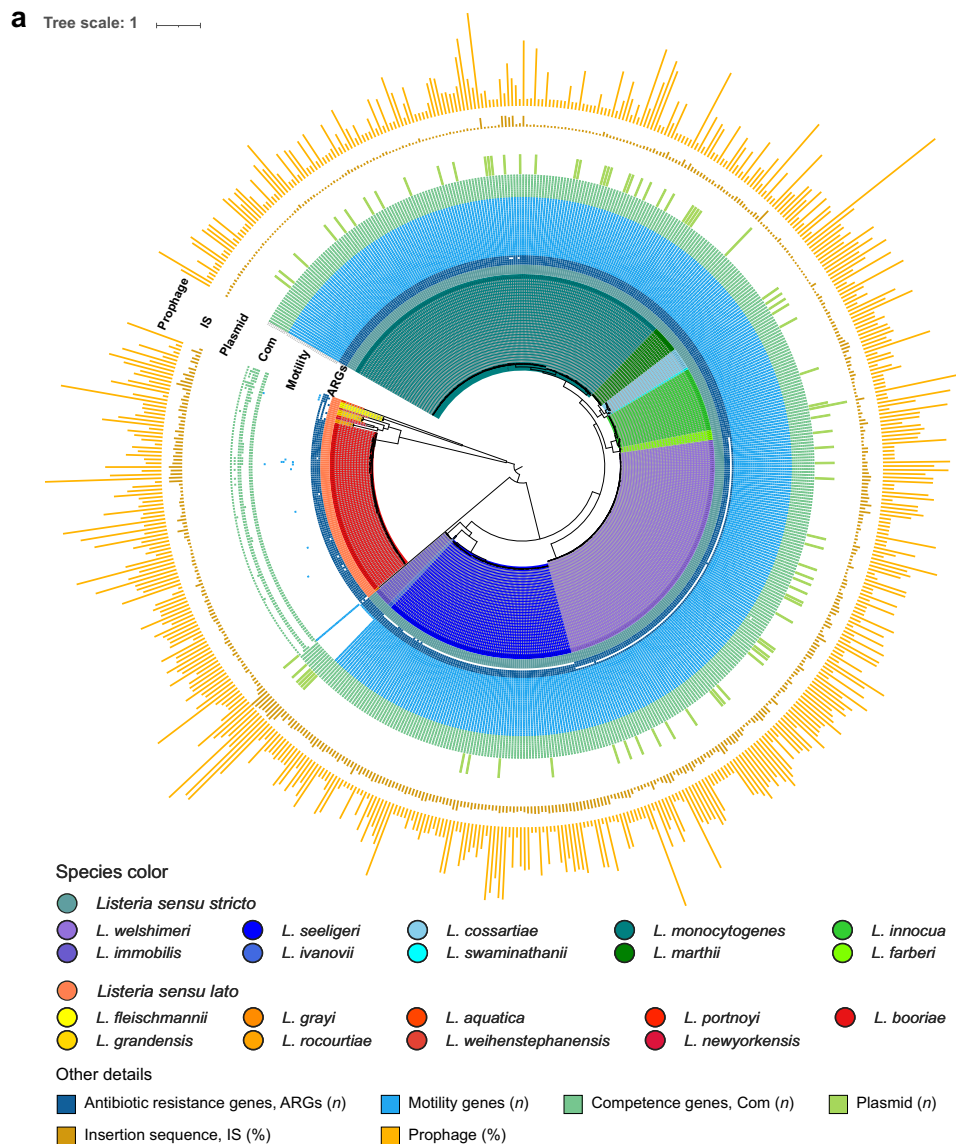
a–e Maximum likelihood gene tree for **a** *lin*, **b** *fosX*, **c** *norB*, **d** *mprF*, and **e** *sul*. Trees were constructed using sequences for *lin*, *fosX*, *norB*, *mprF*, and *sul* detected in 491, 361, 347, 489, and 482 genomes, respectively, with 1,000 bootstraps. The evolutionary models used for constructing the trees were TN + F + I + R4, HKY + F + I + R3,

TVM + F + I + R4, GTR + F + I + R5, and K3Pu + F + I + R3 for *lin*, *fosX*, *norB*, *mprF*, and *sul*, respectively. Trees are rooted by midpoint. Bootstrap values > 80% are indicated by grey circles. **f** Total number of homologous recombination events detected across *Listeria* species (orange) and within species (blue) for each ARG.

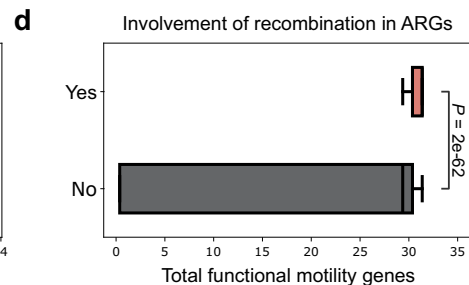
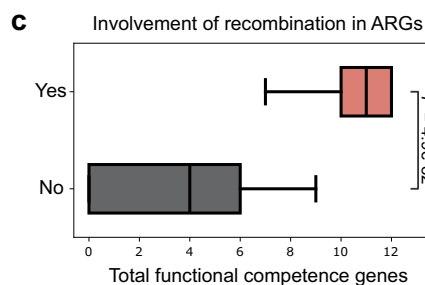
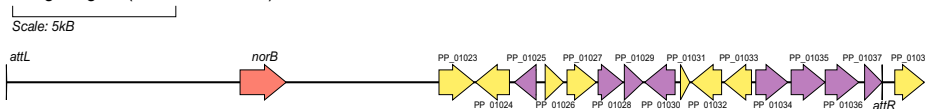
ivanovii (pathogen) and *L. immobilis* (non-pathogen) isolates as well as between *L. marthii* and *L. cossartiae* isolates (Supplementary Data 4). For *mprF*, two recombination events were detected across species (e.g., between *L. ivanovii* and *L. immobilis* isolates and between *L. cossartiae* and *L. marthii* isolates) and 12 events were detected within a species (e.g., *L. innocua*, *L. cossartiae*, *L. welshimeri*, *L. monocytogenes*, and *L. seeligeri*; Supplementary Data 4). For *sul*, four recombination events were detected, with two occurring between two closely related species (i.e., between *L. marthii* and *L. swaminathanii* isolates and between *L. innocua* and *L. welshimeri* isolates; Supplementary Data 5). For the other two recombination events, one was observed within *L. innocua*, and the other was within *L. seeligeri* (Supplementary Data 5). Taken together, these results compellingly suggest that HGT of ARGs occurs both within and across *Listeria* species.

To understand if maintaining ARGs may offer advantages in fitness to adapt to certain environmental stressors, we performed a genome-wide test for PS for each functional ARG. Results showed that *mprF* and *sul* exhibited a better fit to the unconstrained model that allows PS compared to the constrained model (LRT = 13.932 and 17.539, $P = 4.72 \times 10^{-4}$ and 7.77×10^{-5} , respectively; Fig. 2d, e), indicating the presence of PS acting on these two genes. For *lin*, *fosX*, and *norB*, there was no evidence supporting the presence of PS acting on them ($P > 0.05$ for all).

The probable mechanism of HGT of ARGs among *Listeria* species
To investigate the potential mechanisms underlying the HGT of ARGs among *Listeria* species, we employed a predictive approach focusing on mobile genetic elements (MGEs), including prophages, insertion sequences (IS), composite transposons, and plasmids. The presence of ARGs within these MGEs could indicate their role as carriers, informing transduction (via phages) and/or conjugation (via plasmids/transposons) as potential mechanisms of HGT of ARGs in *Listeria*. Among the 594 genomes examined, 1,338 prophages were identified (Fig. 3a). Out of these prophages, 14.7% ($n = 197$) were classified as 'intact'. Of note, we found 12 'incomplete' prophages in *L. booriae*, 11 of which presented *lin* and one of which presented *norB* (Supplementary Data 6). However, all *lin* presented on prophages were found to be truncated (Supplementary Data 7). The presence of remnants of *lin* in the prophages suggests a historical role of bacteriophages in transferring ARGs from other species to *L. booriae*. Apart from these observations, the only instance of recent HGT of a functional ARG potentially mediated by transduction was found in a *norB* gene located within a prophage, PHAGE_Bacill_Spbeta_NC_001884, of *L. ivanovii* isolate L7-1049 (Fig. 3b, Supplementary Data 6). The prophage region, delimited by the left (*attL*) and right (*attR*) attachment sites, comprised the *norB* gene and genes encoding other phage-related proteins and hypothetical proteins (Fig. 3b). The *norB* gene



b Phage region (145981..173528) in *L. ivanovii* FSL L7-1049 that contains an ARG.



identified in this prophage was positioned in the first clade of *L. ivanovii*, *L. seeligeri*, and *L. immobilis* isolates in the gene tree (Fig. 2c). Homologous recombination was also detected among these species in this gene (Supplementary Data 3). These results suggest that overall HGT of functional ARGs mediated by transduction is rare in soil-dwelling *Listeria*, but may have occurred among *L. ivanovii*, *L. seeligeri*, and *L. immobilis* isolates in *norB*.

A total of 4,023 ISs were identified among the 594 genomes (Fig. 3a), with 18.3% ($n = 735$) being classified as 'complete'. IS3 and IS1182 families constituted 66.4% ($n = 488$) and 15.5% ($n = 114$) of the complete ISs, respectively. Of note, we identified an IS-associated ARG that matched the functional *fosX* located on the negative strand in *L. welshimeri* isolate L7-1846, showing 100% identity and coverage. IS3 transposition involves a copy-out-paste-in process that requires at

Fig. 3 | Overview of the phylogeny and prevalence of ARGs, motility genes, competence genes, and MGEs in *Listeria*. **a** The maximum likelihood phylogenetic tree of the 594 *Listeria* genomes using core SNPs with 200 bootstraps. This tree was adapted from a previously published phylogenetic tree constructed using these genomes²⁹. The tree was rooted by midpoint and features branches color-coded by *Listeria* species. The outer ring annotations include the presence/absence of ARGs, motility genes, and competence (Com) genes. A filled box indicates the presence of a functional gene; an empty box indicates the presence of a truncated gene; and white space indicates the absence of a given gene. MGEs are presented either as counts (for plasmids) or as proportions (for ISs and prophages) in the outer ring annotations. **b** Visualization of a prophage carrying a functional *norB*

gene in *L. ivanovii* isolate L7-1049. *norB*, genes encoding phage-related proteins, and genes encoding hypothetical proteins are shown in red, yellow, and purple, respectively. The left and right attachment sites for the phage are referred to as *attL* and *attR*, respectively. **c, d** The number of **c** functional competence genes and **d** functional motility genes compared between genomes with ARGs involved and not involved in homologous recombination events. A two-sided Mann-Whitney U $P < 0.05$ was considered statistically significant. $N = 385$ and 209 for genomes with ARGs involved and not involved in homologous recombination events, respectively, for both functional competence and motility genes. Box plots show the IQR, with the line representing the median and whiskers extending to 1.5 times the IQR.

least two copies of IS3⁴⁴. With the copy of IS3 transposase being located downstream of *fosX* on the positive strand, we consider that the IS3 transposase under this configuration would not be sufficient for gene transfer. Also, no composite transposons carrying ARGs were detected. In addition, a total of 81 plasmids were identified (Supplementary Data 7). The most dominant plasmid incompatibility (Inc) family and group were Inc18 (98.8%, $n = 80$) and repUS25 (96.3%, $n = 78$), respectively. None of these plasmids, however, were found to carry any ARGs. These results suggest that HGT of functional ARGs mediated by conjugation is rare in soil-dwelling *Listeria*.

Given that both conjugation and transduction did not appear to be substantial contributors to the HGT of functional ARGs in soil-dwelling *Listeria*, we further investigated whether natural transformation might play a role. We predicted the presence of 12 competence genes, which signify a bacterium's capacity to uptake foreign DNA, or extracellular DNA (eDNA), from its surroundings, and integrate it into its genome⁴⁵. Results showed that *Listeria sensu stricto* species possessed more competence genes than *sensu lato* species (Fig. 3a). Among *Listeria sensu stricto* species, competence genes were uniformly present, with over 90% functionality observed for several key competence genes, including *comK*, *coiA*, *cinA*, *comEC*, and *comEB* (Fig. 3a). On the contrary, among *Listeria sensu lato* species, specific competence genes such as *comGD*, *comG*, and *comGF* were completely absent (Fig. 3a). The sole functional competence gene identified among *Listeria sensu lato* species was *comEC*, identified in three *L. fleischmannii* isolates (L7-1629, L7-1641, and L7-1645). A further comparison of the count of functional competence genes revealed that genomes with ARGs involved in recombination events had 11 functional competence genes on average, significantly higher than genomes with no recombination events detected (Mann-Whitney U $P = 4.9 \times 10^{-82}$; Fig. 3c). This result suggests the importance of competence in homologous recombination in ARGs of *Listeria*.

As motility plays a crucial role in enabling bacteria to move, providing adaptive advantages in new environments⁴⁶ and facilitating DNA uptake⁴⁷, we further predicted motility genes among *Listeria* genomes. Consistent with competence genes, *Listeria sensu stricto* species possessed more motility genes compared to *sensu lato* species (Fig. 3a). Specifically, all 31 motility genes identified were present in every *Listeria sensu stricto* species, with 90% of these genes being identified as functional in more than 99% of isolates, except for *L. immobilis*, the sole *sensu stricto* species identified thus far that lacks motility⁴⁸. In contrast, almost none isolates of *sensu lato* species harbored any functional motility genes. The high motility in *Listeria sensu stricto* species may increase their chance of being exposed to diverse DNA pools in the environment and facilitate HGT. Indeed, we found that genomes with ARGs involved in homologous recombination have 30 functional motility genes on average, significantly higher than genomes without ARGs involved in recombination (Mann-Whitney U $P = 2 \times 10^{-62}$; Fig. 3d). Collectively, the high prevalence of functional competence and motility genes in *Listeria sensu stricto* species, where functional ARGs predominate, and their associations with homologous recombination suggest an essential role of transformation in HGT of ARGs in *Listeria*.

Environmental factors associated with the richness and genetic divergence of ARGs in soil-dwelling *Listeria*

To assess the potential influence of the environment on ARG acquisition and evolution in *Listeria*, we performed a series of analyses, including Spearman's partial correlation analysis, variation partitioning analysis (VPA), multidimensional scaling (MDS) analysis, Mann-Whitney U tests, and machine learning-based analysis. Considering the high correlation between ARG richness and the genetic similarity of genomes to *L. monocytogenes* (Spearman's $\rho = 0.88$, $P < 0.001$; Fig. 1d, Supplementary Fig. 1b) and a geographic signal likely driven by species (Fig. 1e, Supplementary Fig. 1c), genetic similarity could potentially confound the correlations between environmental variables and ARG richness. Thus, Spearman's partial correlation analysis, controlling for the genetic similarity of *Listeria* species to *L. monocytogenes*, was performed. Thirteen out of 34 environmental variables were significantly correlated with ARG richness (adjusted Spearman's $P < 0.05$ for all). Among these, seven variables, including aluminum, forest, zinc, manganese, iron, longitude, and developed areas with $< 20\%$ impervious surface, exhibited a positive correlation, while the remaining six: copper, wetland, molybdenum, magnesium, pH, and potassium, showed a negative correlation (Fig. 4a). To further quantify the contributions of different environmental variable categories to the observed variation in ARG richness, VPA was conducted. To exclude potential confounding effects, genetic similarity to *L. monocytogenes* was included in this analysis. Of note, the climate category was excluded because no climatic variables were found to be significant (Fig. 4a). As expected, genetic similarity alone accounted for the majority of the variation (adjusted $R^2 = 80.53\%$; Fig. 4b). Soil properties cumulatively explained 7.37% of the variation, similar to surrounding land use, which explained 6.74% (Fig. 4b). Geolocation accounted for less than 1% of variation (Fig. 4b). MDS analysis further showed that isolates with and without functional ARGs formed significantly different clusters based on environmental conditions, which explained 26% of the variation of ARG presence (PERMANOVA $P < 0.001$, $R^2 = 0.260$, pseudo- $F = 6.768$; Fig. 4c). As functional ARGs were predominately detected in *Listeria sensu stricto* species (Fig. 1c), we hypothesized that environmental variables significantly differ between the habitats of *Listeria sensu stricto* and *sensu lato* species. Indeed, a total of 13 environmental variables (five soil property variables, including aluminum, one geolocation variable, three climatic variables, and four surrounding land use variables) were significantly different, and more than half of them (minimum and maximum temperatures, coverage of pasture, cropland, barren, and shrubland, and precipitation) had a fold change of greater than 1 or less than -1 (adjusted Mann-Whitney U $P < 0.05$ for all; Fig. 4d). These results collectively suggest that environmental conditions, particularly soil properties and surrounding land use, are important to the prevalence of ARGs in *Listeria*.

Given the observed correlations between environmental variables and ARG richness, we hypothesize that the presence of ARGs in *Listeria* is predictable from environmental variables using machine learning models. To test this hypothesis, we compared different machine learning algorithms (see Methods). After hyperparameter tuning (Supplementary Data 8), a random forest model,

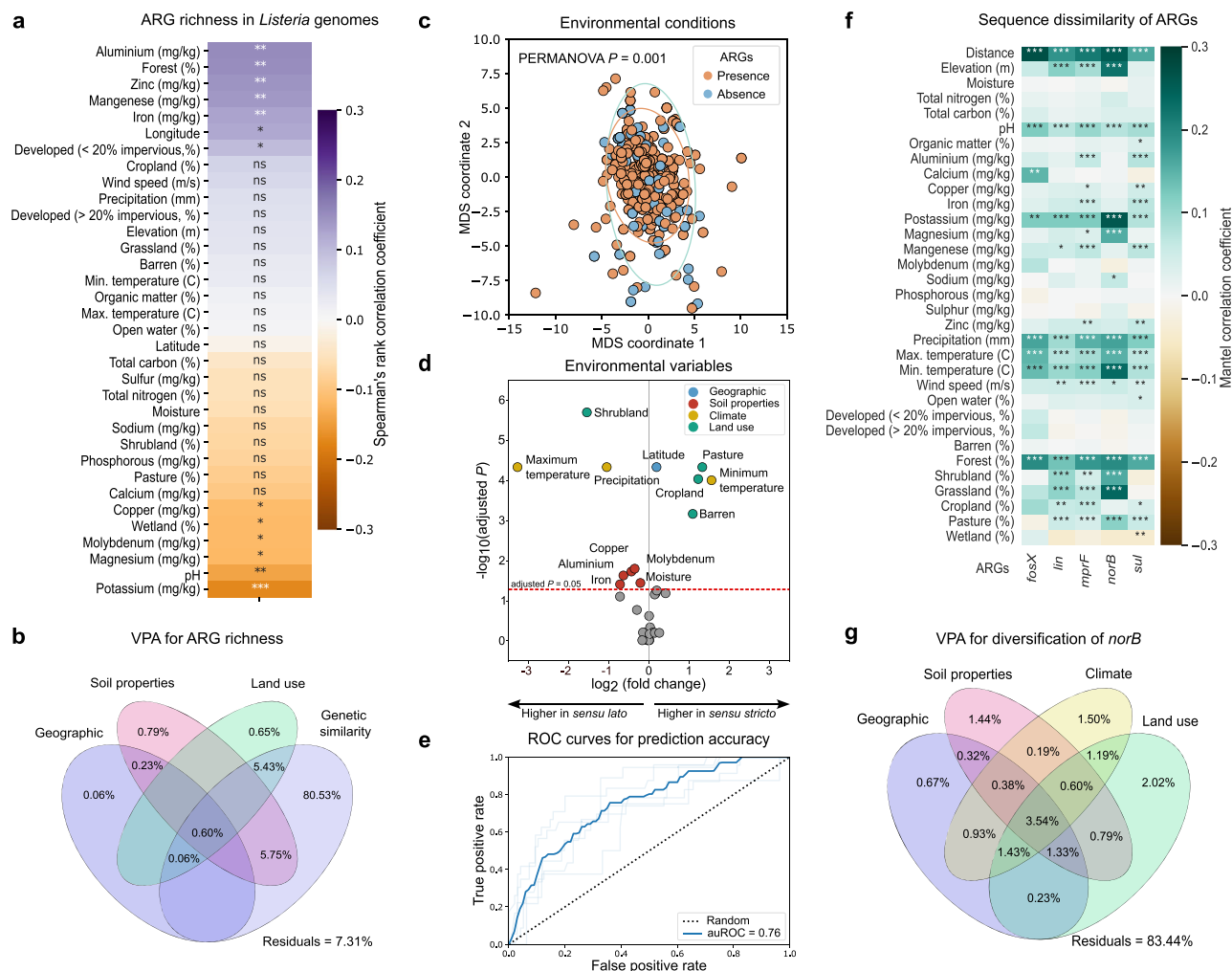


Fig. 4 | Environmental conditions associated with the richness and genetic divergence of ARGs in *Listeria*. **a** Spearman's partial correlation between the richness of ARGs in *Listeria* and environmental variables, controlled for genetic similarity to *L. monocytogenes*. Positive and negative correlation is indicated by purple and orange, respectively. **b** Venn diagram of VPA showing the variation of the functional ARG richness explained by environmental variable categories with significant correlations detected (i.e., geographic locations, soil properties, and surrounding land use) and genetic similarity to *L. monocytogenes*. Residuals indicate unexplained variation. **c** MDS analysis for genomes with (orange) and without (blue) ARGs based on environmental conditions. Each dot represents a genome. PERMANOVA $P = 0.001$ ($R^2 = 0.260$ and pseudo- $F = 6.768$). Ellipse size is determined by two times the standard deviation from the mean. **d** Volcano plot illustrating the significance (two-sided Mann-Whitney; y-axis) of the difference in environmental variables (fold change; x-axis) compared between *Listeria sensu*

stricto and *sensu lato* species. Variables above the red dashed line have an adjusted $P < 0.05$. Dots are color-coded by environmental variable categories. **e** Receiver operating characteristic (ROC) curve for the prediction of ARG presence from environmental variables using a random forest model (auROC = 0.76). Each curve reflects one evaluation using holdout data, repeated 10 times, denoted by light blue lines. The dark blue line represents the mean performance across these repetitions. **f** Mantel tests between the sequence dissimilarity and environmental variables for a given ARG. Positive and negative correlation is indicated by blue and yellow, respectively. **g** Venn diagram of VPA showing the variation of the genetic divergence of *norB* explained by environmental variable categories. For **a** and **f** P -values were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate (BH-FDR) method. Significance levels are denoted by “*”, “**”, “***”, and “****” for $P < 0.05$, $P < 0.01$, $P < 0.001$, and $P < 0.0001$, respectively, and “ns” indicates not significant.

which utilized logistic loss to assess the quality of splits, had a maximum depth of three, considered the logarithm base two of the total number of features to determine the best split, and comprised 100 trees in the forest, was identified as the best model for predicting the presence of ARGs with environmental variables. This model achieved a mean area under the receiver operating characteristic curve (auROC) of 0.76 (Fig. 4e), indicating acceptable discrimination ability⁴⁹, and a mean area under the precision-recall curve (auPR) of 0.95 (Supplementary Fig. 3a). To interpret outputs from the best-performing model, we employed Shapley Additive exPlanations (SHAP)⁵⁰ to evaluate the importance of each feature in the prediction. Among the top ten most predictive features, aluminium emerged as the most influential feature (Supplementary Fig. 3b), aligning with the results from the Spearman's partial

correlation analysis (Fig. 4a). Except for magnesium and developed areas with < 20% impervious surface, the remaining most predictive features (i.e., latitude, organic matter, maximum and minimum temperature, wind speed, calcium, and surrounding shrubland coverage) were not significant in the correlation analysis (Supplementary Fig. 3b, Fig. 4a). This result suggests that machine learning models can capture complex relationships that are often undetectable through conventional statistical methods⁵¹. Despite that the SHAP values of features were overall not high (less than 0.030; Supplementary Fig. 3a), the model effectively used a combination of these features to make relatively accurate predictions, as shown in Fig. 4e. This is common in complex models where the predictive power comes from the collective contribution of many features rather than a few dominant ones^{50,52,53}.

Lastly, to investigate the interplay between the genetic divergence of ARGs and environmental factors, Mantel tests were conducted to assess the correlations between the sequence dissimilarity of each ARG and the distance of each environmental variable. Seven variables—geographic distance, pH, potassium, precipitation, maximum and minimum temperatures, and surrounding forest coverage—exhibited a significant positive correlation with sequence dissimilarity universally in all five functional ARGs (Mantel $P < 0.05$ for all; Fig. 4f). Among these ARGs, *mprF* exhibited significant correlations with the greatest number of environmental variables ($n = 19$). To delineate the contribution of environmental variable categories to the variation of ARG sequence dissimilarity, we further conducted VPA. Among all five ARGs, *norB* sequence divergence was the most affected by environmental variables, collectively accounting for 16.56% of the explained variation (Fig. 4g). For *fosX*, *mprF*, *lin*, and *sul*, environmental factors collectively contributed to 12.99%, 7.93%, 6.79%, and 6.26% of the explained variation, respectively (Supplementary Figs. 4a–d). Despite the varying contributions of environmental variables to the sequence divergence in different ARGs, a consistent pattern emerged that surrounding land use was the most influential factor across all ARGs, independently (and collectively) explaining 2.02% (11.13%), 3.04% (6.48%), 1.35% (4.77%), 1.57% (4.42%), and 1.48% (3.45%) of the variation for *norB*, *fosX*, *mprF*, *lin*, and *sul*, respectively, followed by soil properties. These results suggest that environmental conditions, particularly surrounding land use and soil properties, are important to the genetic diversification of ARGs in *Listeria*.

Discussion

This study investigated the dynamics of ARGs in soil-dwelling *Listeria*, including *sensu stricto* species, which are more closely related to *L. monocytogenes*, and *sensu lato* species. We identified five functional ARGs, *lin*, *mprF*, *sul*, *fosX*, and *norB*, predominantly in *Listeria sensu stricto* species. Most of these five ARGs are still present in *Listeria sensu lato* species but were found to be truncated, suggesting that carrying ARGs may cause metabolic costs in these species⁵⁴. In contrast, maintaining at least some of these ARGs in the genomes might increase fitness in *Listeria sensu stricto* species evidenced by the PS detected in two ARGs, *mprF* and *sul*. The large discrepancy in the prevalence of ARGs between *Listeria sensu stricto* and *sensu lato* species may be partly attributed to the different conditions in soil properties, climate, and surrounding land use these two bacterial groups encounter.

The five functional ARGs identified in this study were previously reported in *L. monocytogenes* and some other *Listeria sensu stricto* species^{23,35,55,56}. The current treatment protocol for listeriosis involves a combination of penicillin and aminopenicillins (ampicillin or amoxicillin)⁹ or ampicillin and gentamicin¹⁰. ARGs conferring resistance to these antibiotics used in clinical treatment, including *ampR*, *aacA4*, and *aadC*, were not detected in this study, which suggests that antibiotics used in clinical settings had limited impact on soil-dwelling *Listeria* in nature. Consistent with our findings, ARG surveillance of *L. monocytogenes* in food-related and clinical settings in France²³, Denmark⁵⁷, and Spain⁵⁸ report that acquired resistance is limited in *L. monocytogenes* and this pathogen remains susceptible to antibiotics over time. Therefore, while emerging resistance in *L. monocytogenes* is observed for certain clinical-use antibiotics, like penicillin¹¹ and rifampicin¹⁴, resistance to the antibiotics used in patients with listeriosis (aminopenicillins and gentamicin) remains rare^{9,23}.

Evidence of HGT was observed for all functional ARGs within and/or between *Listeria sensu stricto* species, but not between *Listeria sensu stricto* and *sensu lato* species. This observation supports the notion that HGT tends to display a bias toward individuals and species that are more closely related⁵⁹. HGT of ARGs has been observed among both clinical and food isolates from various *L. monocytogenes* clonal complexes, with tetracycline resistance identified as the most prevalent acquired resistance phenotype⁶⁰. This has been primarily attributed to

the presence of composite transposons like Tn916-Tn1545 carrying tetracycline resistance (Tn916-carrying *tetM* genes) in *L. monocytogenes*^{23,32,33}. However, we did not identify any tetracycline resistance genes (*tetM* and *tetS*) in soil-dwelling *Listeria* in this study. This is likely attributed to the widespread use of tetracycline in clinical and food-related environments^{61–63}, whereas the baseline tetracycline concentrations in less disturbed environments like soils might be low⁶⁴.

Given the limited instances of acquired resistance observed from transposons, ISs, plasmids, or prophages in this study, we propose that conjugation and transduction may not be the primary mechanisms for the HGT of ARGs in soil-dwelling *Listeria*. Instead, we found that natural transformation is the most likely mechanism. Natural transformation relies on the recipient bacterium that expresses the competence machinery⁶⁵ and largely depends on the uptake and incorporation of exogenous naked DNA from the environment into the genomes of competent recipient organisms⁶⁶. Despite the presence of genes associated with competence machinery, *L. monocytogenes* has not been recognized as naturally transformable in lab settings⁶⁷. The absence of competence in *L. monocytogenes* has been attributed to the truncation of the *comK* gene, which cleaved into two parts by a 42-kb region containing several open reading frames (ORFs) encoding phage-related products⁶⁷. The regulation of the competence system relies on the formation of a functional *comK* gene via prophage excision⁶⁸. In this study, we detected a complete set of competence genes, including *comK* not truncated by prophages, in most genomes of *Listeria sensu stricto* species, where HGT events of ARGs were observed and associated with. Despite unsuccessful attempts to transform *L. monocytogenes* with an intact *comK* gene under lab conditions in a previous study⁶⁷, our results suggest that *Listeria* may require unusual conditions (beyond competence minimal medium, at 37 °C, and selection on Brain Heart Infusion agar supplemented with chloramphenicol) for competence⁶⁷, which complex soil environments may uniquely be able to provide, facilitating HGT of ARGs in this bacterium via natural transformation.

Understanding the associations between environmental factors and ARGs is crucial for unraveling the dynamics and evolution of antibiotic resistance under environmental disturbance. In this study, multiple layers of analysis from different perspectives all suggest that environmental selection, likely triggered by soil properties and surrounding land use, plays a role in the evolution of ARGs in *Listeria*. Soil properties (e.g., pH and nutrients) have been widely reported to influence ARGs in soils^{69–71}. In this study, aluminum and magnesium were found to be the most influential soil properties for ARG richness in *Listeria*, supported by both correlation and machine learning-based analyses. A previous study suggests that nanoalumina can enhance the uptake of ARGs by facilitating the transfer of plasmid-mediated ARGs⁷². Additionally, the persistence of aluminum in soils may exert prolonged selective pressure on bacteria to maintain ARGs through co-selection mechanisms⁷³. These mechanisms can all lead to a positive correlation between ARG richness and aluminum observed in *Listeria* in this study. In contrast, the correlation between magnesium and ARG richness in *Listeria* was negative. This may be due to the competitive dynamics in magnesium-limited environments, where fungal-bacterial competition for magnesium can increase the fitness and survival of bacteria possessing ARGs, as these genes provide a competitive advantage under stress conditions⁷⁴. Furthermore, magnesium-modified biochar has been found to reduce the spread and abundance of ARGs by influencing the bacterial community structure and inhibiting HGT⁷⁵. Given the complexity of natural environments and the observed associations are correlational not causal, further experimental investigations are needed to better understand the influence of soil properties on ARG richness.

Besides soil properties, surrounding land use, particularly forest coverage, was found to be important for the richness of ARGs in soil-

dwelling *Listeria* as well. This is consistent with the finding that forests, irrespective of their location and type (boreal, cold, temperate, or tropical), exhibited the highest richness of ARGs in their soils in a previous global study⁴. Wildlife in forests (e.g., deer⁷⁶, bird⁷⁷, reptiles⁷⁸, and rodents⁷⁹) can serve as carriers of ARGs. These genes can be subsequently excreted into soils through fecal matter as wildlife moves around, contributing to ARG richness in bacteria in adjacent environments. In addition, surrounding land use was identified as the most important factor for the genetic diversification of ARGs in *Listeria*. Prior research has highlighted the contribution of land use to the evolution of antibiotic resistance⁸⁰. For example, in cropland areas, the extensive use of antibiotics to enhance crop productivity selects specific ARGs⁸¹. A plausible explanation for the potential influence of surrounding land use on the genetic diversification of ARGs is that land use could influence soil properties in adjacent natural environments, which indirectly imposes selective pressure on ARGs, leading to genetic diversification. For instance, surrounding cropland and pasture coverage, in which we previously detected a positive correlation with soil magnesium²⁹, were found to be associated with the sequence dissimilarity of *mprF* and *norB*, with the former showing evidence of PS in this study. Overall, the importance of surrounding land use reflects potential anthropogenic effects on the dynamics of antibiotic resistance in the natural environment.

In summary, by leveraging a national reconnaissance of *Listeria* genomes, we showed the genetic and ecological context of their ARGs in the soil environment. ARGs are predominately found in *Listeria sensu stricto* species. Considering the limited occurrence of prophages and plasmids carrying ARGs and the presence of a full set of putatively functional competence genes in *Listeria sensu stricto* species that were correlated with homologous recombination involvement in ARGs, we propose that natural transformation may be the more plausible route for the HGT of ARGs in soil-dwelling *Listeria*. In contrast, HGT of ARGs appears to be often achieved via conjugation in food and clinical isolates, suggesting that *Listeria* isolates from different environments may employ distinct HGT mechanisms for ARG acquisition. We also identified evidence of environmental selection most likely triggered by soil properties and surrounding land use in the acquisition and diversification of ARGs in *Listeria*, highlighting the importance of monitoring the impact of environmental disturbance on the dynamics of antibiotic resistance. Overall, this study provides a baseline knowledge of evolutionary and ecological processes governing the distribution and diversity of ARGs in the soil environment and demonstrates the usage of *Listeria* as a model organism for understanding the impact of environmental changes on ARG mobilization and evolution.

Methods

Listeria genomes and environmental data

The genomic dataset of 594 *Listeria* isolates collected by us from minimally disturbed natural environments across the contiguous US (described in a previous study examining the mechanism underlying bacterial pangenome evolution²⁹) was further analyzed to consider their carriage of ARGs in this study. The genome assemblies met the following quality control criteria: fewer than 300 contigs, N50 greater than 50,000, average coverage exceeding 30X, consistent presence of *sigB* allelic types based on both whole-genome sequencing extraction and PCR-based assays, and no detected contamination using kraken2–2.0.8²⁹. Also, all of these genome assemblies had a high completeness with $99.99\% \pm 0.03\%$ (mean \pm SD) (Supplementary Data 9) assessed by CheckM2⁸². The genomes represent 19 *Listeria* species, predominantly *L. monocytogenes* ($n=177$), followed by *L. welshimeri* ($n=141$), *L. seeligeri* ($n=98$), and *L. booriae* ($n=90$). Other *Listeria* genomes in our dataset included *L. innocua* ($n=33$), *L. marthii* ($n=14$), *L. cossartiae* ($n=11$), *L. immobilis* ($n=9$), *L. farberii* ($n=5$), *L. fleischmannii* ($n=3$), *L. grandensis* ($n=3$), *L. ivanovii* ($n=2$), and *L.*

rocourtiae ($n=2$). Single genomes were available for *L. swaminathanii*, *L. grayi*, *L. aquatica*, *L. weihenstephanensis*, *L. portnoyi*, and *L. new-yorkensis*. Among the species included in this study, *L. monocytogenes*, *L. seeligeri*, *L. marthii*, *L. ivanovii*, *L. welshimeri*, *L. innocua*, *L. cossartiae*, *L. farberii*, *L. immobilis*, and *L. swaminathanii* are classified as *Listeria sensu stricto* species (491 genomes total), while others are *sensu lato* species (103 genomes total)⁸³.

Previously reported environmental data encompassing 34 variables²⁹ paired with this genomic dataset were also examined further in this study. The environmental data include three geolocation (latitude, longitude, and elevation), 17 soil properties (moisture, total nitrogen, total carbon, pH, organic matter, aluminum, calcium, copper, iron, potassium, magnesium, manganese, molybdenum, sodium, phosphorus, sulfur, and zinc), four climatic (precipitation, wind speed, maximum and minimum temperatures), and 10 surrounding land use (open water, barren, forest, shrubland, grassland, cropland, pasture, wetland, and developed open space categorized as >20% and <20% impervious cover) variables²⁹.

Detection of ARGs, competence genes, motility genes, and MGEs

ARGs, competence genes, and motility genes were identified through BLASTN searches⁸⁴, using an E-value of 0.01 and without restrictions on percent identity, against a reference database sourced from the BIGSdb-*Lm* platform³⁵. BIGSdb-*Lm* is a curated bacterial genome sequence database specializing in *L. monocytogenes*³⁵. A total of 25 ARGs, 12 competence genes, and 31 motility genes were extracted from the platform (Supplementary Data 10)³⁵. Gene with the highest bit-score was chosen for further analysis. Subsequently, the presence of premature stop codons (TGA, TAG, TAA), up to but not including the last stop codon⁸⁵, and sequence coverage (%) were assessed for each detected gene using Python 3.6.8. Genes were categorized as putatively functional if their sequence coverage exceeded 80% and no premature stop codon was detected; truncated if its sequence coverage ranged between 30%–80% or premature stop codon was detected; and absent if sequence coverage was less than 30% or no hits were observed in the BLASTN searches³⁴. Based on this categorization, we further simplified the classification as present (including either truncated or functional) and functional³⁴.

To assess possible bias in ARG detection using BIGSdb-*Lm*, we first applied the Comprehensive Antibiotic Resistance Database (CARD)³⁷, which includes a more diverse set of reference ARGs than BIGSdb-*Lm*, to our dataset. We found no significant differences in the overall prevalences for present and functional ARGs between the predictions from BIGSdb-*Lm* and CARD (paired two-sided *t*-test $P=0.81$ and 0.11 , respectively; Supplementary Fig. 5a). Also, a consistent trend was observed where *Listeria sensu stricto* species harbor more functional ARGs than *sensu lato* species (Supplementary Fig. 5b) and the functional ARG richness was strongly positively correlated with the genetic similarity to *L. monocytogenes* (Spearman's $\rho=0.88$ and 0.91 , $P=1.7\text{e-}07$ and $1.3\text{e-}06$; Fig. 1d and Supplementary Fig. 5c for BIGSdb-*Lm* and CARD, respectively). Of note, using CARD, we did not identify any *sul*, an ARG commonly found in *L. monocytogenes*^{23,35}, but it predicted *fosXCC*, an ARG with a similar function to *fosX*, but is commonly identified in *Campylobacter coli* rather than *Listeria* species⁸⁶ (Supplementary Figs. 5a–b). Next, we obtained the reference genome of *L. rocourtiae* FSL F6-920 (accession number: AODK01) from the NCBI database and predicted its ARGs using both BIGSdb-*Lm* and CARD. This isolate was selected because it belongs to *Listeria sensu lato* species, and genotypic and phenotypic antibiotic resistance data for this isolate were published in literature⁸⁷. Using BIGSdb-*Lm*, we identified three ARGs, *sul*, *mprF*, and *fosX*. *fosX* was classified as functional according to our criteria, aligning with its known phenotypic fosfomycin resistance in this isolate⁸⁷, while *sul* and *mprF* were truncated due to the presence of a premature stop codon (Supplementary Data 11). Since this isolate was found to be phenotypically resistant to

sulfamethoxazole (conferred by *sul*)⁸⁷, the functionality of *sul* with a truncated feature might be attributed to its high coverage (94.6%) and the position of stop codons in a later part of the gene (Supplementary Fig. 6). Using CARD, however, none ARGs detected in this genome were functional (Supplementary Data 12). Although it predicted a *fosX* gene, this gene was truncated, and it did not identify any *sul* genes. Taken together, we conclude that our prediction of ARGs in *Listeria* genomes using BIGSdb-*Lm* is not biased, and it is robust and sensitive even for *Listeria sensu lato* species.

To predict MGEs, including plasmids, ISs, transposons, and prophages, we employed specialized programs, including PlasmidFinder⁸⁸ (0.6 cutoff) to identify plasmids; ISEScan⁸⁹ and ISAbR_finder⁹⁰ to detect ISs and ISs associated with ARGs; TnFinder⁹⁰ to identify composite transposons; and PHASTER⁹¹ for prophage prediction. All programs were employed with default settings if not specified. Subsequently, a comparison was made to determine if the functional ARGs were present in (for plasmids and prophages) or near (for IS, within 2000 bp⁹⁰) the predicted MGEs, based on their genomic coordinates. Positive findings were visualized using Gene Graphics⁹².

Richness, diversity, and spatial distribution of ARGs

The richness and Shannon-Wiener diversity index of present and functional ARGs were computed for each genome using the skbio library in Python 3.6.8. Since we observed a strong positive correlation between ARG richness and diversity (Spearman's $\rho = 1$, $P < 10e-30$; Supplementary Fig. 7), and ARG richness is easier to interpret and calculate and has been widely used in other studies^{23,93,94}, subsequent analyses were focused only on ARG richness. To evaluate the association between the ARG richness of *Listeria* species and their genetic similarity to *L. monocytogenes*, we averaged the previously reported pairwise average nucleotide identity (ANI)²⁹ for each genome based on their respective species compared with one *L. monocytogenes* genome and correlated it with the average richness of ARGs (both present and functional) using Spearman's rank correlation analysis. The spatial distribution of ARGs was visualized using Mercator Projection and the Basemap Matplotlib Toolkit v.1.2.1 in Python v.3.6.8. We then compared the ARG richness between the eastern and western regions determined by the longitude of the center of the US (-95°) using two-sided Mann-Whitney *U* test, with $P < 0.05$ indicating a significant difference.

Gene and phylogenetic tree construction, tree topology congruence tests, and detection of homologous recombination and PS

We annotated a previously published core-SNP-based phylogenetic tree of 594 *Listeria* genomes²⁹ with details about species, ARGs, plasmids, competence genes, motility genes, ISs, transposons, and prophages using the Interactive Tree of Life (iTOL) webserver⁹⁵.

Gene tree of each ARG was constructed using a maximum likelihood method with 1,000 bootstraps in IQ-TREE⁹⁶ based on nucleotide sequences aligned using MUSCLE v.3.8.31⁹⁷. IQ-TREE implements ModelFinder to select the best evolutionary model for the phylogenetic estimates⁹⁸. The best-fit models, determined by the Bayesian information criterion (BIC), were TN + F + I + R4, GTR + F + I + R5, K3Pu + F + I + R3, HKY + F + I + R3, and TVM + F + I + R4 for *lin*, *mprF*, *sul*, *fosX*, and *norB*, respectively. A phylogenetic tree for genomes harboring each ARG was constructed based on core SNPs using RAxML v8⁹⁹ with 200 bootstraps and the GTR + G + I substitution model with ascertainment bias correction applied. Core SNPs of genomes included in the construction of each phylogenetic tree were determined using a custom Python script published previously²⁹. All trees were visualized using the iTOL webserver⁹⁵. To test the congruence in the topologies between the core SNP-based tree and gene tree for each ARG, five statistical methods, RELL³⁸, KH³⁹, SH⁴⁰, ELW⁴¹, and AU⁴², were employed. A *P* value < 0.05 indicates significantly different tree topologies.

Nine detection methods (RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, PHYLPRO, LARD, and 3SEQ) implemented in RDP4⁴³ were used to detect homologous recombination events in ARGs. Events detected by at least two methods were considered true positive. To assess whether PS occurs across the entire gene in each functional ARG, we utilized the BUSTED (branch-site unrestricted statistical test for episodic diversification) model in HyPhy (Hypothesis Testing Using Phylogenies)¹⁰⁰. The aligned nucleotide sequences and the ARG trees constructed in the previous steps were used as input files. A likelihood ratio test (LRT) was performed to compare the unconstrained model, which allows PS, and the constrained model, which disallows PS. Statistical significance was determined by approximating the test statistic to a χ^2 distribution. ARGs with a $P < 0.05$ were considered evidence of PS, at least at one specific site in a given ARG.

Assessment of the relationships between environmental variables and the richness and diversification of ARGs

A Spearman's partial correlation analysis, controlling for the genetic similarity to *L. monocytogenes*, was performed to evaluate the associations between ARG richness and each environmental variable followed by a BH-FDR adjustment to account for multiple testing. This approach addresses potential confounding effects caused by genetic similarity to *L. monocytogenes*. Environmental variables with an FDR-adjusted $P < 0.05$ were considered significant. Following this, a VPA was conducted, controlling for genetic similarity to *L. monocytogenes*, to evaluate the relative contribution of each environmental category with significant variables identified (i.e., geolocation, soil property, and land use) to ARG richness. Both genetic similarity and environmental data were structured in matrix format, enabling the use of adjusted R^2 in redundancy analysis (RDA) ordination to partition the variation. The respective adjusted R^2 value in VPA was visualized as a Venn diagram using the 'varpart' function in the vegan package v.2.6-4 in R. Furthermore, MDS based on the Euclidean distance of environmental variables along with a permutational multivariate analysis of variance (PERMANOVA) test was used to compare overall environmental conditions for genomes with and without functional ARGs. PERMANOVA $P < 0.05$ indicates that there are significant differences in the environmental conditions between genomes with and without functional ARGs. Two-sided Mann-Whitney *U* tests were employed to identify significant difference for each of the environmental variables between samples positive for *Listeria sensu stricto* and *sensu lato* species followed by FDR correction.

Mantel tests were conducted to assess the relationships between the distance matrices of environmental variables and ARG sequences followed by FDR correction. Briefly, genetic dissimilarity of a given ARG between genomes was quantified using the Levenshtein distance¹⁰¹, while dissimilarity of a given environmental variable between genomes (excluding longitude and latitude)²⁹ was calculated using Euclidean distance. Geographic distance was computed based on longitude and latitude using the haversine formula. VPA was then performed using the calculated distance matrices to assess the relative contribution of each environmental category with significant variables identified to the sequence dissimilarity of a given ARG.

Machine learning models to predict the presence of ARGs from environmental variables

To predict the presence of ARGs from environmental variables, we developed an end-to-end machine learning-based framework that embodies a series of individual software programs (e.g., scikit-learn and SHAP) written in Python 3.6.8 for data preparation, hyperparameter tuning, model training, and testing, model evaluation through cross-validation, and visualization. Samples were first cleaned and split into the training set (80%) and the testing set (i.e., the holdout set; 20%) in a stratified fashion. The training set was further split into five

stratified folds for cross-validation, in which a collection of predefined models (i.e., classifiers based on decision trees, random forest, multi-layer perceptron, support vector machines, and gradient boosting) were trained and tested with a random set of hyperparameters (Supplementary Data 8). The average auROC score was used to evaluate the performance of the models across the five rounds of cross-validation. To account for stochasticity introduced by the random splitting of samples and division of training data into five folds, we repeated these steps 10 times. We selected the best model and its hyperparameter set with the highest interquartile mean of the auROC scores out of the 10 repetitions among the predefined models. The interquartile means of the auROC and auPR scores of the best model that was exclusively trained on the training set were reported based on a single evaluation of the holdout data from each of the 10 repetitions. The importance of the features was quantified using SHAP⁵⁰.

Reporting summary

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

Data availability

The data used in this study, including *Listeria* genomic data and environmental data, were previously published in Liao et al., 2021²⁹. The GenBank accession numbers for the genomic data are listed in Supplementary Data 9. All processed/source data generated in this study are available on Zenodo at <https://doi.org/10.5281/zenodo.14027083>¹⁰².

Code availability

Code to replicate all analyses is available at https://github.com/leaph-lab/Soil_Listeria_ARG_manuscript¹⁰².

References

- Nesme, J. & Simonet, P. The soil resistome: a critical review on antibiotic resistance origins, ecology and dissemination potential in telluric bacteria. *Environ. Microbiol.* **17**, 913–930 (2015).
- Forsberg, K. J. et al. The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
- Larsson, D. G. J. & Flach, C.-F. Antibiotic resistance in the environment. *Nat. Rev. Microbiol.* **20**, 257–269 (2022).
- Delgado-Baquerizo, M. et al. The global distribution and environmental drivers of the soil antibiotic resistome. *Microbiome* **10**, 219 (2022).
- Orsi, R. H. & Wiedmann, M. Characteristics and distribution of *Listeria* spp., including *Listeria* species newly described since 2009. *Appl. Microbiol. Biotechnol.* **100**, 5273–5287 (2016).
- U. S. Food and Drug Administration. Get the facts about *Listeria*. <https://www.fda.gov/animal-veterinary/animal-health-literacy/get-facts-about-listeria> (2020).
- Freitag, N. E., Port, G. C. & Miner, M. D. *Listeria monocytogenes* — from saprophyte to intracellular pathogen. *Nat. Rev. Microbiol.* **7**, 623–628 (2009).
- Guillet, C. et al. Human listeriosis caused by *Listeria ivanovii*. *Emerg. Infect. Dis.* **16**, 136–138 (2010).
- Koopmans, M. M., Brouwer, M. C., Vázquez-Boland, J. A. & van de Beek, D. Human listeriosis. *Clin. Microbiol. Rev.* **36**, e00060–19 (2023).
- Schlech, W. F. Epidemiology and clinical manifestations of *Listeria monocytogenes* infection. *Microbiol. Spectr.* **7**, 10–112 (2019).
- Morvan, A. et al. Antimicrobial resistance of *Listeria monocytogenes* strains isolated from humans in France. *Antimicrob. Agents Chemother.* **54**, 2728–2731 (2010).
- Bertsch, D. et al. Antimicrobial susceptibility and antibiotic resistance gene transfer analysis of foodborne, clinical, and environmental *Listeria* spp. isolates including *Listeria monocytogenes*. *Microbiologyopen* **3**, 118–127 (2014).
- Chenal-Francisque, V. et al. Highly rifampin-resistant *Listeria monocytogenes* isolated from a patient with prosthetic bone infection. *Antimicrob. Agents Chemother.* **58**, 1829–1830 (2014).
- Isnard, C. et al. In vivo emergence of rifampicin resistance by *rpoB* mutation in *Listeria monocytogenes* during therapy of prosthetic joint infection. *Int. J. Antimicrob. Agents* **48**, 572–574 (2016).
- Alonso-Hernando, A., Prieto, M., García-Fernández, C., Alonso-Calleja, C. & Capita, R. Increase over time in the prevalence of multiple antibiotic resistance among isolates of *Listeria monocytogenes* from poultry in Spain. *Food Control* **23**, 37–41 (2012).
- Rodríguez-López, P., Rodríguez-Herrera, J., Vázquez-Sánchez, D. & López Cabo, M. Current knowledge on *Listeria monocytogenes* biofilms in food-related environments: incidence, resistance to biocides, ecology and biocontrol. *Foods* **7**, 85 (2018).
- Wales, A. & Davies, R. Co-selection of resistance to antibiotics, biocides and heavy metals, and its relevance to foodborne pathogens. *Antibiotics* **4**, 567–604 (2015).
- Panera-Martínez, S., Rodríguez-Melcón, C., Serrano-Galán, V., Alonso-Calleja, C. & Capita, R. Prevalence, quantification and antibiotic resistance of *Listeria monocytogenes* in poultry preparations. *Food Control* **135**, 108608 (2022).
- Smith, A., Moorhouse, E., Monaghan, J., Taylor, C. & Singleton, I. Sources and survival of *Listeria monocytogenes* on fresh, leafy produce. *J. Appl. Microbiol.* **125**, 930–942 (2018).
- Kayode, A. J. & Okoh, A. I. Antibiotic resistance profile of *Listeria monocytogenes* recovered from ready-to-eat foods surveyed in South Africa. *J. Food Prot.* **85**, 1807–1814 (2022).
- Godshall, C. E., Suh, G. & Lorber, B. Cutaneous listeriosis. *J. Clin. Microbiol.* **51**, 3591–3596 (2013).
- Bengtsson-Palme, J. Antibiotic resistance in the food supply chain: where can sequencing and metagenomics aid risk assessment? *Curr. Opin. Food Sci.* **14**, 66–71 (2017).
- Moura, A. et al. Phenotypic and genotypic antimicrobial resistance of *Listeria monocytogenes*: an observational study in France. *Lancet Reg. Health Eur.* **37**, 100800 (2024).
- Olaimat, A. N. et al. Emergence of antibiotic resistance in *Listeria monocytogenes* isolated from food products: a comprehensive review. *Compr. Rev. Food Sci. Food Saf.* **17**, 1277–1292 (2018).
- Rostamian, M., Kooti, S., Mohammadi, B., Salimi, Y. & Akya, A. A systematic review and meta-analysis of *Listeria monocytogenes* isolated from human and non-human sources: the antibiotic susceptibility aspect. *Diagn. Microbiol. Infect. Dis.* **102**, 115634 (2022).
- Fletcher, S. Understanding the contribution of environmental factors in the spread of antimicrobial resistance. *Environ. Health Prev. Med.* **20**, 243–252 (2015).
- Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiol. Rev.* **42**, fux053 (2018).
- Martínez, J. L. Bottlenecks in the transferability of antibiotic resistance from natural ecosystems to human bacterial pathogens. *Front. Microbiol.* **2**, 265 (2012).
- Liao, J. et al. Nationwide genomic atlas of soil-dwelling *Listeria* reveals effects of selection and population ecology on pangenome evolution. *Nat. Microbiol.* **6**, 1021–1030 (2021).
- Liu, G., Thomsen, L. E. & Olsen, J. E. Antimicrobial-induced horizontal transfer of antimicrobial resistance genes in bacteria: a mini-review. *J. Antimicrob. Chemother.* **77**, 556–567 (2022).
- Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
- Bertrand, S. et al. Detection and characterization of *tet(M)* in tetracycline-resistant *Listeria* strains from human and food-

- processing origins in Belgium and France. *J. Med. Microbiol.* **54**, 1151–1156 (2005).
33. Bertsch, D. et al. Tn6198, a novel transposon containing the trimethoprim resistance gene *dhfrG* embedded into a Tn916 element in *Listeria monocytogenes*. *J. Antimicrob. Chemother.* **68**, 986–991 (2013).
 34. Liao, J. et al. Comparative genomics unveils extensive genomic variation between populations of *Listeria* species in natural and food-associated environments. *ISME Commun.* **3**, 85 (2023).
 35. Moura, A. et al. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nat. Microbiol.* **2**, 16185 (2016).
 36. Wang, M. et al. VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens. *Nucleic Acids Res.* **50**, W768–W773 (2022).
 37. Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–D699 (2023).
 38. Kishino, H., Miyata, T. & Hasegawa, M. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **31**, 151–160 (1990).
 39. Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* **29**, 170–179 (1989).
 40. Shimodaira, H. & Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1116 (1999).
 41. Strimmer, K. & Rambaut, A. Inferring confidence sets of possibly misspecified gene trees. *Proc. R. Soc. Lond. B. Biol. Sci.* **269**, 137–142 (2002).
 42. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).
 43. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).
 44. Chandler, M., Fayet, O., Rousseau, P., Ton Hoang, B. & Duval-Valentin, G. Copy-out-paste-in transposition of IS911: a major transposition pathway. *Microbiol. Spectr.* **3**, <https://doi.org/10.1128/microbiolspec.mdna3-0031-2014> (2015).
 45. Blokesch, M. Natural competence for transformation. *Curr. Biol.* **26**, R1126–R1130 (2016).
 46. Piskovsky, V. & Oliveira, N. M. Bacterial motility can govern the dynamics of antibiotic resistance evolution. *Nat. Commun.* **14**, 5584 (2023).
 47. Dubnau, D. & Blokesch, M. Mechanisms of DNA uptake by naturally competent bacteria. *Annu. Rev. Genet.* **53**, 217–237 (2019).
 48. Carlin, C. R. et al. *Listeria cossartiae* sp. nov., *Listeria immobilis* sp. nov., *Listeria portnoyi* sp. nov. and *Listeria rustica* sp. nov., isolated from agricultural water and natural environments. *Int. J. Syst. Evol. Microbiol.* **71**, (2021).
 49. Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. Assessing the Fit of the Model. (Wiley, 2013). <https://doi.org/10.1002/9781118548387>.
 50. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. in *Advances in Neural Information Processing Systems* vols 2017–December (2017).
 51. Yu, Q. et al. Study becomes insight: ecological learning from machine learning. *Methods Ecol. Evol.* **12**, 2117–2128 (2021).
 52. Wang, H., Liang, Q., Hancock, J. T. & Khoshgoftaar, T. M. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *J. Big Data* **11**, 44 (2024).
 53. Aboze, B. J. A comprehensive guide into SHAP values. Deepchecks <https://deepchecks.com/a-comprehensive-guide-into-shap-shapley-additive-explanations-values/> (2023).
 54. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
 55. Luque-Sastre, L. et al. Antimicrobial resistance in *Listeria* species. *Microbiol. Spectr.* **6**, 10–1128 (2018).
 56. Scotti, M. et al. Epistatic control of intrinsic resistance by virulence genes in *Listeria*. *PLoS Genet.* **14**, e1007525 (2018).
 57. Hansen, J. M., Gerner-Smidt, P. & Bruun, B. Antibiotic susceptibility of *Listeria monocytogenes* in Denmark 1958–2001. *APMIS* **113**, 31–36 (2005).
 58. Marco, F. et al. In vitro activities of 22 antimicrobial agents against *Listeria monocytogenes* strains isolated in Barcelona, Spain. *Diagn. Microbiol. Infect. Dis.* **38**, 259–261 (2000).
 59. Andam, C. P. & Gogarten, J. P. Biased gene transfer in microbial evolution. *Nat. Rev. Microbiol.* **9**, 543–555 (2011).
 60. Baquero, F., Lanza V, F., Duval, M. & Coque, T. M. Ecogenetics of antibiotic resistance in *Listeria monocytogenes*. *Mol. Microbiol.* **113**, 570–579 (2020).
 61. Chopra, I. & Roberts, M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232–260 (2001).
 62. Thaker, M., Spanogiannopoulos, P. & Wright, G. D. The tetracycline resistome. *Cell. Mol. Life Sci.* **67**, 419–431 (2010).
 63. Conde-Cid, M. et al. Tetracycline and sulfonamide antibiotics in soils: presence, fate and environmental risks. *Processes* **8**, 1479 (2020).
 64. Bengtsson-Palme, J. et al. Towards monitoring of antimicrobial resistance in the environment: for what reasons, how to implement it, and what are the data needs? *Environ. Int.* **178**, 108089 (2023).
 65. Michaelis, C. & Grohmann, E. Horizontal gene transfer of antibiotic resistance genes in biofilms. *Antibiotics* **12**, 328 (2023).
 66. Chen, I., Christie, P. J. & Dubnau, D. The ins and outs of DNA transfer in bacteria. *Science* **310**, 1456–1460 (2005).
 67. Borezee, E., Msadek, T., Durant, L. & Berche, P. Identification in *Listeria monocytogenes* of MecA, a homologue of the *Bacillus subtilis* competence regulatory protein. *J. Bacteriol.* **182**, 5931–5934 (2000).
 68. Rabinovich, L., Sigal, N., Borovok, I., Nir-Paz, R. & Herskovits, A. A. Prophage excision activates *Listeria* competence genes that promote phagosomal escape and virulence. *Cell* **150**, 792–802 (2012).
 69. Shi, H. et al. Soil component: a potential factor affecting the occurrence and spread of antibiotic resistance genes. *Antibiotics* **12**, 333 (2023).
 70. Wang, Z., Zhang, N., Li, C. & Shao, L. Diversity of antibiotic resistance genes in soils with four different fertilization treatments. *Front. Microbiol.* **14**, 1291599 (2023).
 71. Han, B. et al. The source, fate and prospect of antibiotic resistance genes in soil: A review. *Front. Microbiol.* **13**, 976657 (2022).
 72. Ding, C. et al. Enhanced uptake of antibiotic resistance genes in the presence of nanoalumina. *Nanotoxicology* **10**, 1051–1060 (2016).
 73. Murray, L. M. et al. Co-selection for antibiotic resistance by environmental contaminants. *npj Antimicrob. Resist.* **2**, 9 (2024).
 74. Hsieh, Y.-Y. P. et al. Widespread fungal–bacterial competition for magnesium lowers bacterial susceptibility to polymyxin antibiotics. *PLoS Biol.* **22**, e3002694 (2024).
 75. Liu, H. et al. Effects of magnesium-modified biochar on antibiotic resistance genes and microbial communities in chicken manure composting. *Environ. Sci. Pollut. Res.* **30**, 108553–108564 (2023).
 76. Ballash, G. A. et al. Temporal trends in antimicrobial resistance of fecal *Escherichia coli* from deer. *Ecohealth* **18**, 288–296 (2021).
 77. Bonnedahl, J. & Järhult, J. D. Antibiotic resistance in wild birds. *Ups. J. Med. Sci.* **119**, 113–116 (2014).
 78. Wheeler, E., Hong, P.-Y., Bedon, L. C. & Mackie, R. I. Carriage of antibiotic-resistant enteric bacteria varies among sites in Galápagos reptiles. *J. Wildl. Dis.* **48**, 56–67 (2012).

79. Gilliver, M. A., Bennett, M., Begon, M., Hazel, S. M. & Hart, C. A. Antibiotic resistance found in wild rodents. *Nature* **401**, 233–234 (1999).
80. Nolan, T. M. et al. Land use as a critical determinant of faecal and antimicrobial resistance gene pollution in riverine systems. *Sci. Total Environ.* **871**, 162052 (2023).
81. Mann, A., Nehra, K., Rana, J. S. & Dahiya, T. Antibiotic resistance in agriculture: perspectives on upcoming strategies to overcome upsurge in resistance. *Curr. Res. Micro. Sci.* **2**, 100030 (2021).
82. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023).
83. Orsi, R. H., Liao, J., Carlin, C. R. & Wiedmann, M. Taxonomy, ecology, and relevance to food safety of the genus *Listeria* with a particular consideration of new *Listeria* species described between 2010 and 2022. *mBio* **15**, e00938–23 (2023).
84. Ye, J., McGinnis, S. & Madden, T. L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res.* **34**, W6–W9 (2006).
85. Wong, T.-Y. et al. Role of premature stop codons in bacterial evolution. *J. Bacteriol.* **190**, 6718–6725 (2008).
86. Wang, Y. et al. Identification of a novel *fosXCC* gene conferring fosfomycin resistance in *Campylobacter*. *J. Antimicrob. Chemother.* **70**, 1261–1263 (2015).
87. Leclercq, A. et al. *Listeria rocourtiae* sp. nov. *Int. J. Syst. Evol. Microbiol.* **60**, 2210–2214 (2010).
88. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
89. Xie, Z. & Tang, H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* **33**, 3340–3347 (2017).
90. Ross, K. et al. TnCentral: a prokaryotic transposable element database and web portal for transposon analysis. *mBio* **12**, <https://doi.org/10.1128/mbio.02060-21> (2021).
91. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
92. Harrison, K. J., Crécy-Lagard, V. & Zallot, R. Gene Graphics: a genomic neighborhood data visualization web application. *Bioinformatics* **34**, 1406–1408 (2018).
93. Feodorova, V. A. et al. Complete genome of the *Listeria monocytogenes* strain AUF, used as a live listeriosis veterinary vaccine. *Sci. Data* **11**, 643 (2024).
94. Alvarez-Molina, A. et al. Unraveling the emergence and population diversity of *Listeria monocytogenes* in a newly built meat facility through whole genome sequencing. *Int. J. Food Microbiol.* **340**, 109043 (2021).
95. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
96. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
97. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
98. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
99. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
100. Murrell, B. et al. Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **32**, 1365–1371 (2015).
101. Berger, B., Waterman, M. S. & Yu, Y. W. Levenshtein distance, sequence comparison and biological database search. *IEEE Trans. Inf. Theory* **67**, 3287–3294 (2021).
102. Goh, Y.-X. & Liao, J. Data and code for ‘Evidence of horizontal gene transfer and environmental selection impacting antibiotic resistance evolution in soil-dwelling *Listeria*’. <https://doi.org/10.5281/zenodo.14027083> (2024).

Acknowledgements

We thank all members of LEAPH (the Liao laboratory) for their enriching discussions. This work was funded by the Virginia Tech Center for Emerging, Zoonotic, and Arthropod-borne Pathogens (CeZAP) Interdisciplinary Team-building Pilot Grant (J.L.) and the Interdisciplinary Graduate Education Program in Infectious Disease (ID IGEP) fellowship (Y.-X.G.).

Author contributions

J.L. designed the study. Y.-X.G., S.M.B.A., A.N., H.Z., and J.L. analyzed the data. Y.-X.G. wrote the paper with input from J.L., M.P., L.-A.K., and A.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54459-9>.

Correspondence and requests for materials should be addressed to Jingqiu Liao.

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

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

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

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

© The Author(s) 2024