Motile and non-motile *Listeria* species adopt distinct genomic and ecological strategies to achieve broad geographic ranges in soil

Running head: Genomic bases of broad geographic ranges

Ying-Xian Goh[1,2,3], Shannon Hepp[1], Kevin J. Cummings[4], Martin Wiedmann[5], Jingqiu Liao[1,2,3]*

[1] Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, United States
[2] Center for Emerging, Zoonotic, and Arthropod-Borne Pathogens, Virginia Tech, Blacksburg, VA 24061, United States
[3] Global Change Center, Virginia Tech, Blacksburg, VA 24061, United States
[4] Department of Public and Ecosystem Health, Cornell University, Ithaca, NY 14853, United States
[5] Department of Food Science, Cornell University, Ithaca, NY 14853, United States

* Corresponding author: Jingqiu Liao
Mailing address: 403 Durham Hall, 1145 Perry Street, Blacksburg, VA 24061.
Email: liaoj@vt.edu

Abstract

Broad geographic ranges indicate high ecological versatility and are generally associated with low extinction risk. Motility is not only a physiological feature, but also a core ecological trait for bacteria. However, the genomic foundations underlying the broad geographic ranges of motile and non-motile bacteria remain poorly understood. To address this, we analyzed 141 and 90 genomes of *Listeria welshimeri* and *L. booriae*, systematically screened from 1,004 soil samples across the United States, representing widespread motile and non-motile species, respectively. We show that *L. welshimeri* exhibits restricted phylogeographic structure and a weak distance-decay relationship, suggesting minimal geographic barriers to dispersal. Its high dispersal capacity likely stems from enhanced motility and effective host colonization, evidenced by strong positive selection on flagellar genes and the close relatedness to isolates from wild birds. In contrast, *L. booriae* displays a regionally endemic distribution and limited dispersal. With a large, open pangenome, *L. booriae* appears strongly adapted to local environments. This is evidenced by pronounced positive selection on genes involved in inorganic ion, amino acid, and coenzyme transport and metabolism, as well as strong associations with abiotic factors, particularly climate, and accompanying bacterial consortia. In summary, to establish widespread distributions, *L. welshimeri* tends to rely on movement and colonization of wildlife hosts to facilitate long-distance dispersal, while *L. booriae* leverages genomic plasticity and metabolic versatility that enable adaptation to diverse environmental conditions. These findings highlight the distinct genomic foundations and ecological strategies that motile and non-motile bacteria use to achieve broad geographic ranges in the environment.

Introduction

Geographic range, the spatial extent over which a species occurs, is a key determinant of microbial ecology, associating with population connectivity, resilience, and ecosystem functioning [1, 2]. Species with broader ranges typically exploit a wider variety of habitats or resources, reflecting greater niche breadth, and are often considered generalists [3–5]. Although geographic range expansion can entail fitness costs, such as reduced performance in the original habitat due to niche-expanding mutations [6] or slower adaptation to novel conditions [7], generalists generally gain a competitive advantage under environmental fluctuations, making them less prone to extinction when habitats are disturbed [8, 9]. Yet microbial communities are often dominated by specialists [10–12], raising the question of why some taxa achieve broad geographic distribution while others remain narrowly restricted. While evolutionary processes such as pangenome expansion [13] and recombination [14] have been proposed to facilitate bacterial survival across diverse environments, large-scale field data that mechanistically link genomic features to their widespread geographic distributions remain limited.

Free-living bacteria can be broadly categorized into motile and non-motile forms. Motility confers multiple ecological advantages, including microhabitat exploration, efficient nutrient acquisition, toxin avoidance, habitat and host colonization, and dispersal across locations [15–17]. Experimental studies show that motile bacteria expand at rates determined by growth and habitat size [18], with chemotaxis enabling directed range expansion [19]. However, these benefits are counterbalanced by substantial energetic costs, including the biosynthesis and maintenance of flagellar structures and ATP expenditure for locomotion [15]. Trade-offs between chemotactic benefits and energetic burden may constrain the net advantage of motility, particularly in complex or turbulent environments [20], leaving open the question of whether mechanistic advantages observed under laboratory conditions translate into broad-scale geographic distribution in natural habitats. In contrast, non-motile bacteria lack active dispersal and may rely on passive mechanisms, such as rain [21], wind [22, 23], or intermicrobial hitchhiking [24, 25], which are likely less effective in penetrating structured environments such as soil due to physical fragmentation of pathways. This suggests that non-motile taxa may evolve alternative strategies to achieve broad geographic ranges, which remain poorly understood.

Listeria is a bacterial genus widespread across diverse environments [26, 27]. Despite sharing a common evolutionary background, we previously found that *Listeria* species have undergone adaptive evolution, showing pronounced differences in geographic distribution in soil [28]. Notably, we observed that *L. welshimeri*, a species that is motile via peritrichous flagella [29], and *L. booriae*, a non-motile species [30], are both widely

distributed in soil across the contiguous United States (US) [28]. The contrasting motility phenotypes but comparable geographic range sizes make these two *Listeria* species an ideal model for investigating the mechanisms underlying ability of motile and non-motile bacteria to disperse to and occupy the broad geographic ranges.

Here, we performed an in-depth comparative genomic and ecological analysis of 141 *L. welshimeri* and 90 *L. booriae* isolates obtained from our systematic screening of 1,004 soil samples across the US [28]. Using phylogeographic, bioinformatic, and statistical approaches, we found that flagellar-associated genes in *L. welshimeri* show evidence of positive selection, and that this species is capable of long-distance dispersal, as indicated by a weak distance-decay relationship. To test whether wildlife may facilitate this dispersal, we performed whole-genome sequencing (WGS) of nine *L. welshimeri* isolates recovered from wild birds and compared them with soil isolates. Their high genomic relatedness supports a potential role for wildlife as a transport vector that enable long-distance dispersal. In contrast, *L. booriae*, which exhibited strong dispersal constraints, appears to offset the absence of motility by maintaining an extremely open pangenome enriched in adaptive signatures related to metabolism, allowing them to exploit a wide range of environmental niches. Indeed, using a suite of statistical analyses, we identified strong correlations between the gene content of *L. booriae* and abiotic factors, particularly climate, as well as accompanying bacterial consortia, particularly Proteobacteria, Chloroflexi, and Actinobacteria, suggesting genetic adaption to local environments. Together, these findings demonstrate that motile and non-motile bacteria achieve widespread geographic distribution in soil through distinct genomic and ecological strategies, offering broader insights into microbial adaptation and persistence in heterogeneous environments.

Materials and Methods

Genomic data of *L. welshimeri* and *L. booriae* isolates, environmental data, and bacterial community data

Existing genome assemblies of 141 *L. welshimeri* and 90 *L. booriae* isolates that we previously obtained from minimally disturbed soil in natural environments across the U.S. [28] were subjected to secondary analysis in this study. The genome assemblies of these two species were previously analyzed along with other species in the context of *Listeria* pangenome evolution but have not been used to directly compare the genomic features between each other. Soil sampling followed a systematic approach that achieved a relatively even spatial coverage across the U.S., and the methods for DNA extraction, WGS and data preprocessing, genome assembly, gene prediction, ortholog gene identification, and functional annotation (i.e., assignment of functional categories for

ortholog genes based on Clusters of Orthologous Groups [COGs] and Kyoto Encyclopedia of Genes and Genomes [KEGG]) for these soil isolates were detailed in Liao et al. (2021) [28].

Additionally, WGS was performed for nine *L. welshimeri* isolates recovered from wild birds, including Rose-breasted Grosbeak and Rock Pigeon, admitted to the Janet L. Swanson Wildlife Hospital at Cornell University, New York [31]. Following our data deposition, these isolates remained the only *L. welshimeri* wildlife isolates with WGS data available in the NCBI database. Genomic DNA was extracted using the NEBExpress T4 Lysozyme and Monarch Spin gDNA Extraction Kit (New England Biolabs) following the manufacturer protocol. DNA quality and concentration were assessed using a NanoDrop spectrophotometer and Qubit fluorometer, respectively. All samples met standard quality thresholds (A260/280 ≈ 1.80; A260/230 ≈ 2.0) and were sequenced on the Illumina MiSeq platform (2 × 250 bp paired-end reads) at the Duke University Center for Genomic and Computational Biology. WGS data preprocessing, genome assembly, gene prediction and gene family identification followed the workflow described in Liao et al. (2021) [28] with minor modification to ensure comparability. In brief, sequencing adapters and low-quality reads were trimmed using fastp v0.23.4 [32], and cleaned reads were assembled de novo with SPAdes v4.0.0 [33]. Contigs <500 bp were removed. Assembly quality was assessed with QUAST v5.2.0 [34] and CheckM2 v1.0.2 [35], and average sequencing depth was estimated using BBmap v39.01 and Samtools v1.17. All assemblies met the following criteria: <300 contigs, N50 >50,000 bp, average coverage >30×, completeness >90%, contamination <5%, and correct taxonomic assignment by Kraken2 v2.1.3 [36]. Quality statistics for wild bird isolates are summarized in **Supplementary Table 1**. Genes of wild bird isolates were predicted using Prodigal v2.6.3 [37], and ortholog genes present in both soil and wild bird isolates were identified using MMseqs2 [38].

The environmental dataset (also referred to as “abiotic factors”) for soil isolates comprised 34 variables: three geolocation variables (latitude, longitude, elevation), 17 soil physicochemical variables (moisture, total nitrogen [TN], total carbon [TC], pH, organic matter [OM], and concentrations of aluminum [Al], calcium [Ca], copper [Cu], iron [Fe], potassium [K], magnesium [Mg], manganese [Mn], molybdenum [Mo], sodium [Na], phosphorous [P], sulfur [S], and zinc [Zn]), four climate (precipitation, wind speed, maximum [Max.] and minimum [Max.] temperatures [temp]), and 10 land use variables (proportional coverage of open water, barren land, forest, shrubland, grassland, cropland, pasture, wetland, and developed [Dev] open spaces with either >20% or <20% impervious [IMP] surfaces in the surrounding landscape). Descriptions of environmental data acquisition and processing were detailed in Liao et al. (2021) [28].

Bacterial community data (also referred to as "biotic factors") for soil isolates, including methods for DNA extraction, 16S rRNA gene amplicon sequencing, sequencing data preprocessing, operational taxonomic unit (OTU) identification, taxonomic classification, and diversity calculation (e.g., weighted UniFrac distances) were reported in our previous study comparing *Listeria* populations from natural and food-associated environments [39]. Both abiotic and biotic factors were subjected to secondary analysis in this present study to assess their differential associations with *L. welshimeri* and *L. booriae* as well as their associations with gene content.

Phylogenetic trees, genetic similarities, and geographic distribution

To investigate phylogeographic structure, phylogenetic trees based on core SNPs were constructed for three groups: (i) *L. welshimeri* soil isolates, (ii) *L. booriae* soil isolates, and (iii) *L. welshimeri* soil and wild bird isolates. Core SNPs were identified using kSNP4 [40], and maximum likelihood trees were generated using RAxML v8.2.12 [41] under the GTR + G + I substitution model with ascertainment bias correction and 1,000 bootstrap replicates. Trees were visualized using iTOL v7 [42]. To assess genomic similarity, pairwise average nucleotide identity (ANI) was computed using pyani v0.3.0-alpha in Python v3.6.8 for these three groups. In addition, Jaccard similarity indices were calculated using the gene presence/absence matrices for groups (i) and (ii).

Geographic distributions of *L. welshimeri* and *L. booriae* soil isolates grouped by major clades on the phylogenetic tree were visualized using a Mercator projection with the Basemap Matplotlib Toolkit v1.2.1 in Python v3.6.8. Pairwise geographic distances between isolates were calculated from GPS coordinates using the geopy module in Python v3.6.8. Mean geographic distances within each major clade were then computed and compared between the two species using a two-sided Mann-Whitney (MW) *U* test.

Pangenome characterization and statistical comparison of genomic features

In this study, core genes are defined as those found in all genomes of a given species, whereas accessory genes are defined as those present in only a subset of genomes [13, 43, 44]. The pangenome refers to the complete repertoire of genes present across all genomes within each species, encompassing both core and accessory genes [13, 45]. To estimate the pangenome size for soil-derived *L. welshimeri* and *L. booriae* based on an assumed 100 genomes each, we applied the power law function cN^{γ} calculated in Liao et al. (2021) [28], which models the relationship between the number of genomes (N) and total ortholog genes (i.e., pangenome size). The specific functions used were $n_{pan} = 2572N^{0.133}$ for *L. welshimeri* and $n_{pan} = 2937N^{0.208}$ for *L. booriae* [28].

To characterize genomic differences between soil-derived *L. welshimeri* and *L. booriae*, we analyzed genome size, gene richness (i.e., total number of unique ortholog genes per genome), nucleotide diversity (π), relative abundance of COGs, relative abundance of putatively functional motility and chemotaxis genes, and KEGG pathway diversity. Species-level π was calculated using popGenStat [46] on gene alignments generated with MUSCLE v3.8.31 [47]. Reference sequences for 31 motility and five chemotaxis genes (**Supplementary Table 2**) were retrieved from the BIGSdb-*Lm* database [48] and used as BLASTN queries against assembled genomes to identify homologs [49]. For each gene, the highest bit-score hit was screened for premature stop codons (excluding terminal codon: TGA, TAG, TAA) and alignment coverage in Python 3.6.8. Motility and chemotaxis genes with >80% coverage and no premature stops were considered putatively functional [39]. Genes were further classified into KEGG pathways and analyzed for classification into four major pathways (i.e., metabolism, genetic information processing, environmental information processing, cellular processes; **Supplementary Table 3**). Shannon-Wiener diversity index was calculated for all pathways combined (referred to as "overall biological pathway") and within each major pathway based on their relative abundance. All genomic features were compared between *L. welshimeri* and *L. booriae* soil isolates using two-sided MW *U* tests. For the comparison of COG relative abundance, Benjamini-Hochberg false discovery rate (BH-FDR) correction was additionally applied to account for multiple testing.

To assess the genomic similarities between monophyletic *L. welshimeri* soil and wild bird isolates, Kruskal-Wallis (KW) tests followed by pairwise two-sided MW *U* tests were used to compare ANI for (i) within-soil, (ii) soil-wild bird, and (iii) within-wild bird isolates. Genome size, gene richness, and GC content were also compared using two-sided MW *U* tests.

Positive selection detection and functional enrichment analysis

Soil isolates of *L. welshimeri* and *L. booriae* were subjected to positive selection detection using the methods detailed in Liao et al. (2021) [28]. Briefly, core and accessory genes of each species were screened for positive selection using the branch-site unrestricted statistical test for episodic diversification (BUSTED) model implemented in Hypothesis Testing Using Phylogenies (HyPhy) [50]. Genes were included if they contained at least three non-identical sequences and exhibited one or more non-synonymous substitutions. Evidence of positive selection was evaluated using a likelihood ratio test (LRT) comparing unconstrained and null models, with significance assessed via a χ^2 test followed by BH-FDR correction. Homologous recombination was also assessed for both core and accessory genes using the Recombination Detection Program v4 (RDP4) [51]. Only genes with an adjusted $P < 0.05$ in the positive selection analysis and no detectable

recombination were considered undergoing positive selection. Functional enrichment analysis was performed using a binomial distribution model, as described in Liao et al. 2023 [39], to identify COGs significantly enriched among core and accessory genes under positive selection for each species.

Environmental characteristics and distance-decay relationships

Multidimensional scaling (MDS) analyses was performed using Euclidean distances for abiotic factors and weighted UniFrac distances for OTUs to compare environmental abiotic conditions and bacterial community compositions between soil samples positive for *L. welshimeri* and *L. booriae*. Clustering significance by species was assessed with PERMANOVA. Differences in individual abiotic and biotic variables between species were evaluated using two-sided MW *U* tests with BH-FDR correction.

To identify the influence of abiotic and biotic factors on the gene richness of soil-derived *L. welshimeri* and *L. booriae*, variation partitioning analysis (VPA) and Spearman's rank correlation analyses were conducted. Two VPA models were constructed: (i) one that included three groups of abiotic variables - soil properties, climate, and surrounding land use - and (ii) another that included two groups of variables, abiotic factors and biotic factors represented by relative abundances of bacterial phyla. To enhance model performance, we evaluated multicollinearity among predictor variables using variance inflation factors (VIF). Variables with the highest VIF were iteratively removed until all remaining predictors had VIF < 10 [52]. This procedure retained 22 and 42 variables for VPA models (i) and (ii), respectively (**Supplementary Table 4**), to quantify the proportion of variance in gene richness explained by each group of selected variables within each species. Adjusted R^2 values were computed using the vegan package 2.6-4 in R and visualized in a Venn diagram. Spearman's rank correlations followed by BH-FDR correction were applied to assess associations between gene richness, both overall and for each COG, and each of the 34 abiotic factors or 25 bacterial phyla.

To infer dispersal limitation, distance-decay relationships were examined by regressing geographic distance and two measures of genomic similarity: ANI and Jaccard similarity. A linear regression was interpreted as evidence of a distance-decay relationship when it met all three criteria: an $R^2 > 0.2$, a negative slope, and $P < 0.05$.

Results

Soil-derived *L. welshimeri* and *L. booriae* exhibits contrasting phylogeography and genomic features.

Our previous work showed that *L. welshimeri* and *L. booriae* were the most common motile and non-motile *Listeria* species detected in systematically collected, minimally disturbed natural soils across the contiguous U.S. [28]. Among 1,004 soil samples, 89 (8.86%) and 67 (6.67%) were positive for *L. welshimeri* and *L. booriae*, respectively [28]. While *L. monocytogenes* is also motile and exhibited a slightly higher overall prevalence (11.75%) than *L. welshimeri*, our previous systematic phylogenomic analysis revealed that *L. monocytogenes* comprises three species-level phylogroups (lineages I, II, and III) which substantially differ in genomic and ecological traits, whereas *L. welshimeri* consists of a single phylogroup [28, 53]. At the same phylogenetic clustering level, *L. welshimeri* was more frequent than the most common *L. monocytogenes* lineage (III; 7.67%) and was therefore selected to represent the most widespread motile *Listeria* species in this study. Both *L. welshimeri* and *L. booriae* exhibited comparable broad geographic distributions (853 km and 850 km, respectively; **Fig. 1A-B**). Core-SNP phylogenies revealed that *L. welshimeri* can be classified into nine major clades (**Fig. 1C**), whose members were geographically intermixed (**Fig. 1A**). *L. booriae* also comprised nine major clades (**Fig. 1D**) but exhibited clear geographic structuring, with major clades endemic to certain regions (**Fig. 1B**). To support this observation, pairwise geographic distances between isolates within each major clade were calculated for each species. The mean geographic distances within major clades ranged from 234-963 km and 33-730 km for *L. welshimeri* and *L. booriae*, respectively (**Supplementary Fig. 1A**) and were significantly larger in *L. welshimeri* (MW U P = 0.022; **Supplementary Fig. 1B**). These findings indicate that although both *L. welshimeri* and *L. booriae* are widespread, they show distinct geographic and population structures.

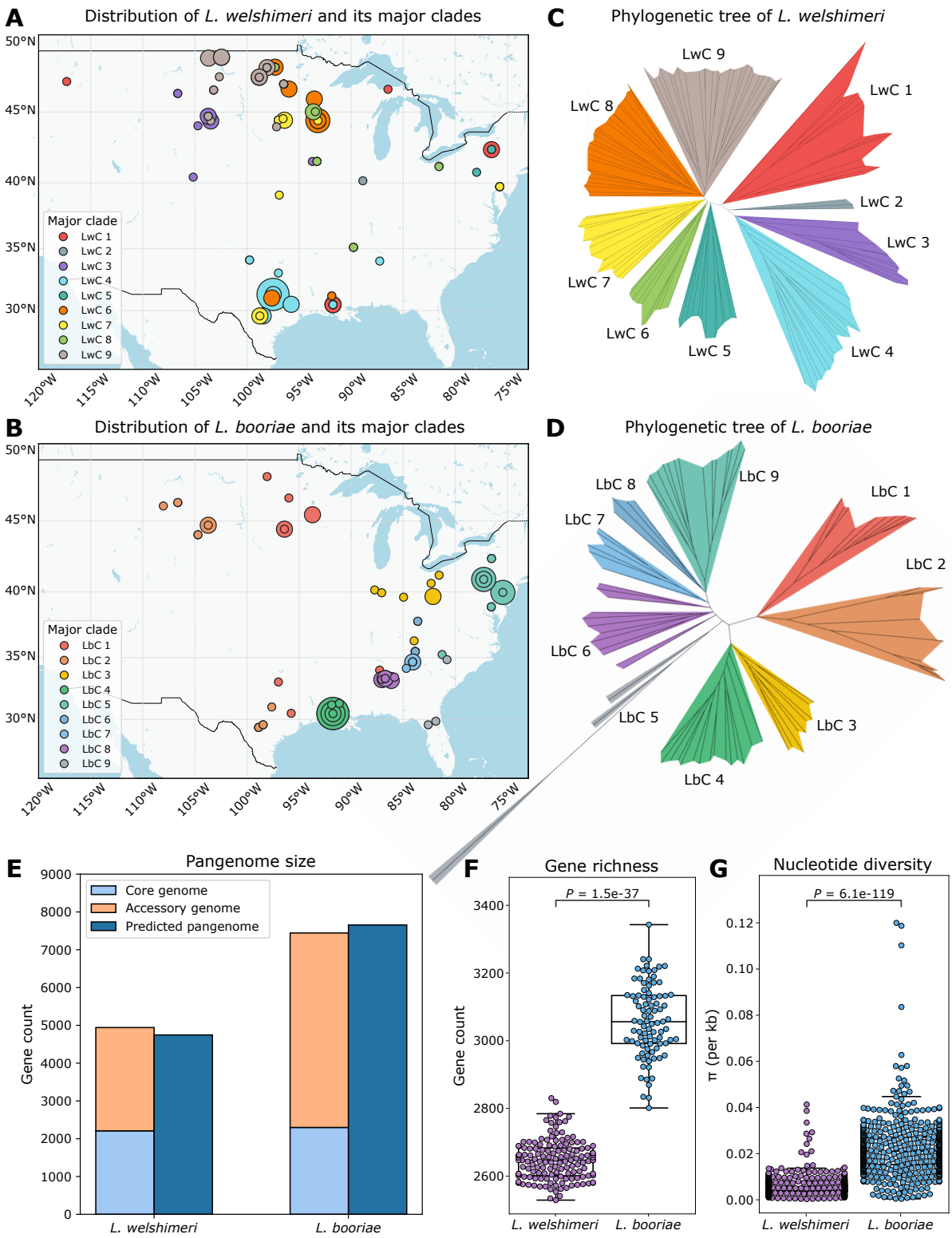


Figure 1. Phylogeographic structure and genomic features of soil-dwelling *L. welshimeri* and *L. booriae*. **(A-B)** Geographic distribution of **(A)** *L. welshimeri* and **(B)** *L. booriae* isolates in soil across the US, color-coded by major clades identified in **C-D**. Circle sizes indicate the number of isolates per clade at each site. **(C-D)** Unrooted maximum likelihood phylogenetic trees of **(C)** 141 *L. welshimeri* soil isolates and **(D)** 90 *L.*

booriae soil isolates based on core SNPs. Branches are color-coded by major clades, with "Lw" and "Lb" indicating *L. welshimeri* and *L. booriae*, respectively. **(E)** Pangenome sizes of *L. welshimeri* and *L. booriae*, divided into core and accessory genomes, with predicted pangenome sizes estimated from 100 genomes per species using the power law function cN^{γ} from Liao et al. (2021) [28]. **(F-G)** **(F)** Gene richness and **(G)** nucleotide diversity (π) compared between *L. welshimeri* and *L. booriae*. Box plots display the interquartile range (IQR) with the median indicated as a line and whiskers extending to 1.5 times the IQR. Two-sided MW *U P* values are annotated.

To assess whether the observed population structures corresponded to differences in genomic content between the two species, we performed pangenome analysis, which identified 4,941 ortholog genes in *L. welshimeri* (2,209 core and 2,732 accessory genes) and 7,443 in *L. booriae* (2,298 core and 5,145 accessory genes) (**Fig. 1E**). To avoid the potential bias caused by unequal sample size, we predicted the pangenome size given 100 genomes for each species based on the pangenome curves, where a more open pangenome was observed in *L. booriae* (**Supplementary Fig. 2A**). The predicted pangenome size was 4,745 for *L. welshimeri* and 7,654 for *L. booriae.* Consistent with this result, *L. booriae* had significantly higher gene richness (MW *U P* = 1.5e-37; **Fig. 1F**) and larger genome size (*P* = 1.5e-37; **Supplementary Fig. 2B**). Also, it had significantly greater nucleotide diversity (*P* = 6.1e-199; **Fig. 1G**). Together, these results show that *L. booriae* has greater genomic diversity and plasticity than *L. welshimeri*.

Soil-derived *L. welshimeri* and *L. booriae* undergo strong positive selection targeting flagellar genes and metabolic genes, respectively.

To assess the genomic differences between *L. welshimeri* and *L. booriae* reflected at the functional level, we analyzed their COG functional profiles. Eleven COGs, including cell motility (N) and replication, recombination, and repair (L), were significantly more abundant in *L. welshimeri*, whereas nine, including cell membrane biogenesis (M), were enriched in *L. booriae* (adjusted MW *U P* < 0.05 for all; **Fig. 2A**). Among these, cell motility (N) genes showed the largest fold change in *L. welshimeri* (adjusted *P* = 5.4e-37; **Fig 2A**), and targeted BLAST searches confirmed significantly more putatively functional motility and chemotaxis genes in this species (*P* = 9.8e-40 and 1.5e-37, respectively; **Supplementary Fig. 3, Fig. 2B**). In contrast, cell membrane biogenesis (M) genes exhibited the largest fold change in *L. booriae*, being significantly more abundant in this species (adjusted *P* = 7.4e-37; **Fig 2A**). These results reveal distinct functional repertoires between *L. welshimeri* and *L. booriae*.

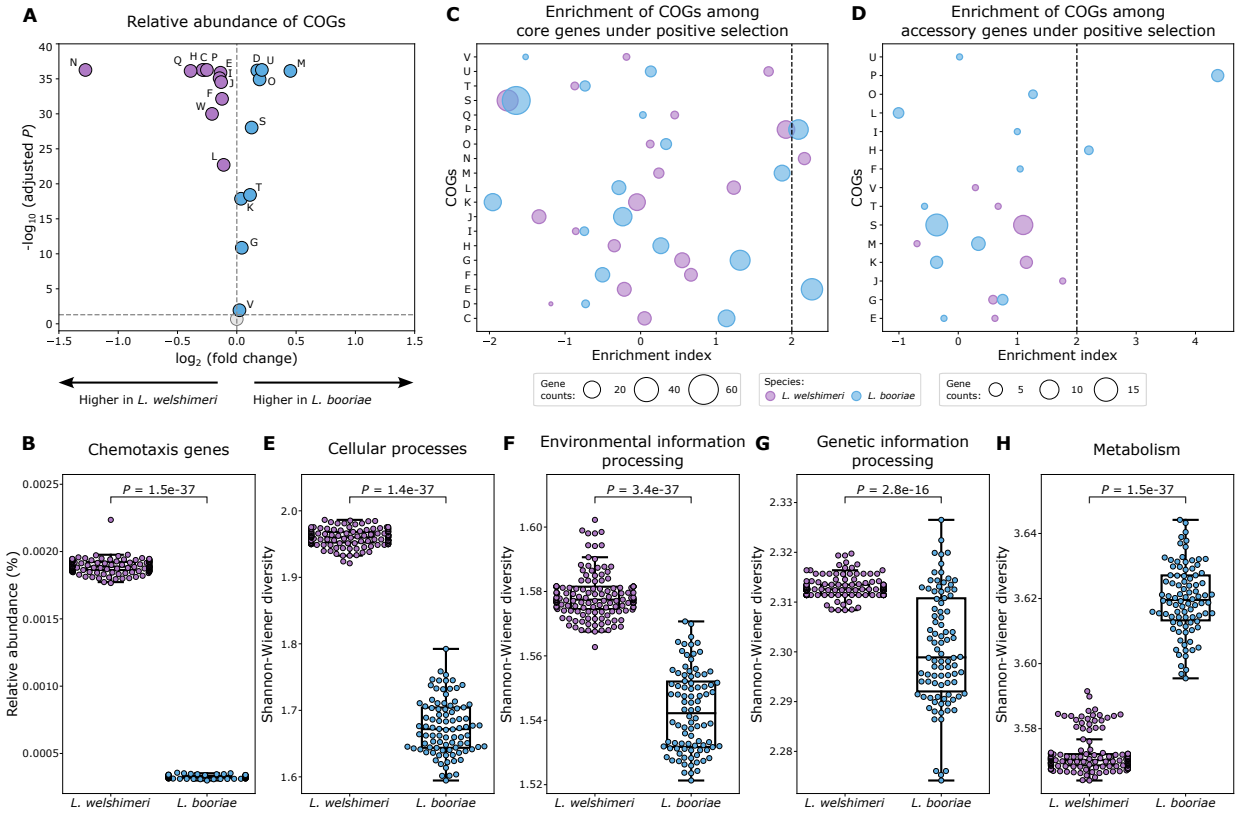


Figure 2. Functional divergence and positive selection in *L. welshimeri* and *L. booriae*. (A) Volcano plot of COG abundance differences between *L. welshimeri* and *L. booriae*, with fold change on the x-axis and significance (two-sided MW *U P*) on the y-axis. Points above the horizontal grey dashed line have adjusted $P < 0.05$. **(B)** Relative abundance of putatively functional chemotaxis genes compared between *L. welshimeri* and *L. booriae.* **(C-D)** Enrichment of COGs among **(C)** core and **(D)** accessory genes with evidence of positive selection. An enrichment index > 2 (black dashed line) indicates significant enrichment (binomial $P < 0.05$). Circle size is proportional to the number of genes annotated to each COG. For **(A), (C),** and **(D),** COGs abbreviations are as follows: C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure, and biogenesis; K: Transcription; L: Replication, recombination, and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport, and catabolism; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms.**(E-H)** Shannon-Wiener diversity of KEGG pathways compared between *L. welshimeri* and *L. booriae* for **(E)** cellular processes, **(F)** environmental information processing, **(G)** genetic information processing, and **(H)** metabolism pathways. For **(B), (E-H),** box plots display the IQR with the median indicated as a line and whiskers extending to 1.5 times the IQR. Two-sided MW *U P* values are annotated.

Because niche breadth in *Listeria* has been linked to positive selection [28], we next identified genes under positive selection in each species using the BUSTED model for gene-wide episodic selection. In *L. welshimeri*, 193 core and 31 accessory genes

374 exhibited evidence of positive selection (adjusted LRT $P < 0.05$ for all; **Supplementary**
375 **Table 5**). Based on the functional enrichment analysis, core genes of *L. welshimeri* under
376 positive selection were significantly enriched for cell motility (N) functions ($P < 0.05$; **Fig.**
377 **2C**), including a chemotaxis protein and several flagellar synthesis genes (e.g., *fliH, fliK,*
378 *flhB, flgC, fliM*; **Supplementary Table 5**). No COGs were significantly enriched among
379 accessory genes under positive selection in this species (**Fig. 2D**). By contrast, *L. booriae*
380 displayed a broader set of positively selected genes, which included 290 core and 64
381 accessory genes (adjusted LRT $P < 0.05$ for all; **Supplementary Table 5**). Core genes
382 under positive selection in this species were enriched for inorganic ion transport and
383 metabolism (P) and amino acid transport and metabolism (E) (**Fig. 2C**), while accessory
384 genes under positive selection were enriched for inorganic ion transport and metabolism
385 (P) and coenzyme transport and metabolism (H) (**Figs. 2D**). Approximately half of the
386 positively selected ion transport genes encoded ATP-binding cassette (ABC)
387 transporters, localized to the cell membrane (**Supplementary Table 5**). These patterns
388 indicate that positive selection has primarily shaped motility capacity in *L. welshimeri* and
389 metabolic versatility in *L. booriae*.
390
391 To determine whether the functional differences extended to higher-level gene networks,
392 we analyzed KEGG pathway diversity. *L. welshimeri* exhibited significantly higher
393 pathway diversity in cellular processes (MW *U P* = 1.4e-37; **Fig. 2E**) and environmental
394 information processing (P = 3.4e-37; **Fig. 2F**). This increased diversity may reflect the
395 presence of flagellar assembly pathways and an expanded repertoire of sensing and
396 regulatory systems (e.g., phosphotransferase system and two-component systems). *L.*
397 *welshimeri* also showed significantly higher pathway diversity in genetic information
398 processing (P = 2.8e-16; **Fig. 2G**), which includes essential modules involved in DNA
399 replication, recombination, and repair (**Supplementary Table 3**). In contrast, *L. booriae*
400 displayed significantly greater diversity in both metabolism and overall biological
401 pathways (P = 1.5e-37 and 4.5e-14, respectively; **Fig. 2H, Supplementary Fig. 4**),
402 indicating broader metabolic potential to thrive in diverse local environments. These
403 findings reveal distinct genomic functional specialization between *L. welshimeri* and *L.*
404 *booriae*.
405
406 ***L. booriae* shows strong genomic associations with soil environments.**
407
408 Because major clades of *L. booriae* exhibited endemism (**Fig. 1B**) and this species
409 possessed a flexible, adaptive pangenome (**Fig. 1E, Supplementary Fig. 2A**), we
410 hypothesized that its genome content is strongly shaped by local environmental
411 conditions, facilitating adaptation to diverse soil habitats and enabling its broad
412 geographic distribution. To test this hypothesis, we compared the abiotic conditions and
413 bacterial communities of *L. welshimeri* and *L. booriae* habitats and assessed their

414 correlations with gene richness of each species. A total of 21 abiotic factors and 15
415 bacterial phyla were found to be significantly different between their habitats (adjusted
416 MW *U P* < 0.05 for all). *L. welshimeri* was primarily isolated from sites with significantly
417 higher minimum temperature, elevation, pH, and concentration of certain soil
418 macronutrients (K, Mg, Ca, P), greater coverage of pasture, wetland, and cropland in
419 surrounding areas (**Supplementary Fig. 5A**), and higher relative abundances of
420 Proteobacteria, Planctomycetes, Elusimicrobia, and Chlamydiae (**Supplementary Fig.**
421 **5B)**. In contrast, *L. booriae* occurred in sites with significantly higher precipitation, greater
422 forest and shrubland coverage in surrounding areas, elevated soil concentrations of Fe,
423 Al, Zn, Mn, and Cu (**Supplementary Fig. 5A**), and enriched with Actinobacteria,
424 Chloroflexi, Bacteroidetes, Firmicutes, and Fibrobacteres (**Supplementary Fig. 5B**)
425 Consistent with these results, both MDS plots based on Euclidian distance of 34 abiotic
426 variables and bacterial community beta diversity indicated by weighted UniFrac distances
427 revealed significant clustering of samples by species (PERMANOVA *P* = 0.001 for both,
428 pseudo-*F* = 18.84 and 32.60, respectively; **Supplementary Fig. 6A and B**), indicating
429 that *L. welshimeri* and *L. booriae* occupied distinct environmental abiotic and biotic
430 spaces.
431

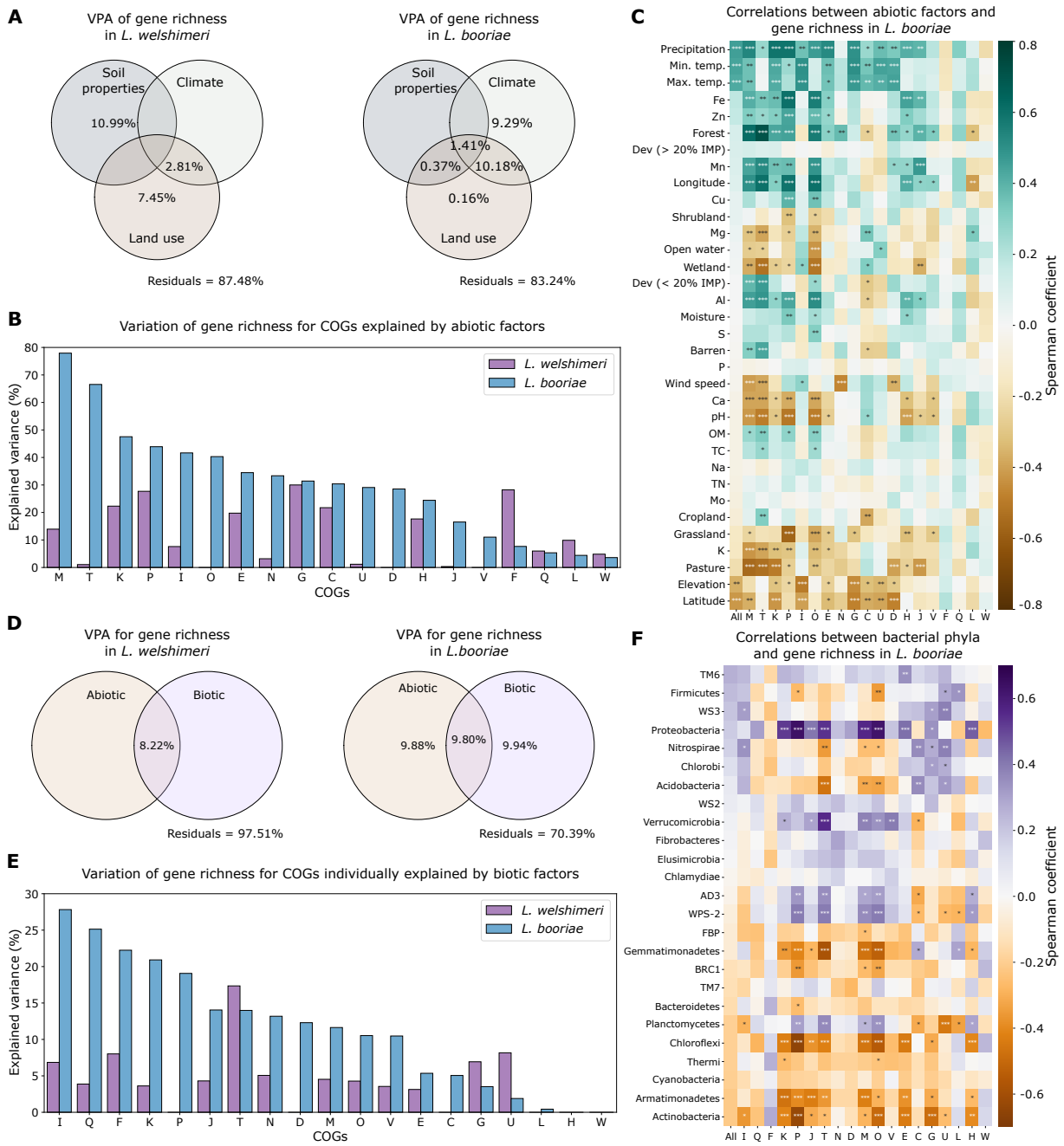


Figure 3. Associations between gene richness in *L. welshimeri* and *L. booriae* and abiotic environmental factors and bacterial community composition. (A) Venn diagrams from variation partitioning analysis (VPA) showing the proportion of overall gene richness variation explained by soil properties, climate, and surrounding land use for *L. welshimeri* and *L. booriae*. **(B)** Variation in gene richness of COG categories explained by abiotic factors based on VPA. **(C)** Spearman's correlations between abiotic factors and gene richness, both overall and for COGs, in *L. booriae*. Abiotic factors are ordered by the descending correlation coefficients of overall ("All") gene richness. Positive and negative correlations are shown in green and brown, respectively. **(D)** Venn diagrams from VPA showing the proportion of gene richness variation explained by abiotic and biotic factors (i.e., surrounding bacterial phyla) for *L. welshimeri* and *L. booriae.* For **(A)** and **(D)**, only positive fractions of explained variance are shown; negative fractions were

omitted, and therefore the total explained variance and the residual does not sum to 1. **(E)** Variation in gene richness of COG categories individually explained by biotic factors based on VPA. **(F)** Spearman's correlations between the relative abundance of bacterial phyla and gene richness, both overall and for COGs, in *L. booriae*. Bacterial phyla are ordered by the descending correlation coefficients of overall ("All") gene richness. TM6, WS3, WS2, AD3, WPS-2, FBP, BRC1, and TM7 represent candidate bacterial phyla that remain uncultured under laboratory conditions. Positive and negative correlations are shown in purple and orange, respectively. For **(C)** and **(F)**, significance levels are indicated as "*", "**", and "***" for adjusted $P < 0.05$, < 0.01, and < 0.001, respectively.

For the genomic associations with abiotic factors, VPA using VIF-selected variables (see Methods) showed that overall abiotic factors explained more gene richness variation in *L. booriae* (16.76%) than *L. welshimeri* (12.52%; **Fig. 3A**). Climate and soil properties contributed the largest unique fraction of explained variation in gene richness in *L. booriae* (9.29%) and *L. welshimeri* (10.99%), respectively (**Fig. 3A**). When stratified by COGs, abiotic factors explained more variation in *L. booriae* than in *L. welshimeri* for 15 of 19 COGs, with the strongest effects observed for cell membrane biogenesis (M; 77.93%), followed by signal transduction mechanism (T; 66.54%) and transcription (K; 47.54%; **Fig 3B**). Consistent with the VPA results, Spearman correlation analyses showed that gene richnesses of *L. booriae* exhibited stronger associations with abiotic factors. Overall gene richness in *L. booriae* was positively associated with precipitation and temperature, and negatively with latitude and elevation (adjusted $P < 0.05$ for all; **Fig. 3C**), whereas no significant association was observed for *L. welshimeri* (**Supplementary Fig. 7A**). Across COGs, an average of ten and five abiotic factors were significantly associated with gene richness in *L. booriae* and *L. welshimeri*, respectively (**Fig. 3C; Supplementary Fig. 7A**). Notably, precipitation showed strong correlations (|Spearman's ρ| > 0.5) with four COGs, including inorganic ion transport and metabolism (P) and amino acid transport and metabolism (E; **Fig. 3C**), both of which also exhibited evidence of positive selection in *L. booriae* (**Fig. 2C and D**). Cell membrane biogenesis (M) and posttranslational modification, protein turnover, chaperones (O) displayed the highest number of significant associations with abiotic factors (22 significant variables; **Fig. 3C**), indicating that genes involved in these functions may be particularly responsive to abiotic variation in *L. booriae*. Collectively, these findings indicate that abiotic factors, particularly climate, exert a stronger influence on gene content in *L. booriae* than *L. welshimeri.*

Additionally, VPA using VIF-selected variables (see Methods) showed that abiotic factors and surrounding bacterial phyla together explained substantially more gene richness variation in *L. booriae* (29.61%) than *L. welshimeri* (2.49%) (**Fig. 3D**). Surrounding bacterial phyla uniquely accounted for 9.94% of the variation in *L. booriae*, while they had minimal explanatory power in *L. welshimeri* (**Fig. 3D**). Notably, in *L. welshimeri*, the total explained variance for overall gene richness decreased from 12.52% (abiotic factors only) to 2.49% when bacterial phyla were included, indicating that bacterial phyla explained little additional variance and that their inclusion resulted in a penalty for increased model

complexity [54]. When stratified by COGs, surrounding bacterial phyla individually explained more variation in *L. booriae* than in *L. welshimeri* for 14 of 19 COGs, with the strongest effects observed for lipid transport and metabolism (I; 27.82%), followed by secondary metabolites biosynthesis, transport and catabolism (Q; 25.14%) and nucleotide transport and metabolism (F; 22.25%; **Fig 3E**). Consistent with the VPA results, Spearman correlation analyses indicated that gene richnesses of *L. booriae* exhibited stronger associations with the relative abundance of bacterial phyla. Although overall gene richness in *L. booriae* showed no significant correlations with any individual bacterial phylum (**Fig. 3F**), stratification by COG revealed substantially more significant associations in this species, with an average of six correlations per phylum, compared with an average of one in *L. welshimeri* (**Fig. 3F; Supplementary Fig. 7B**). In *L. booriae*, Actinobacteria, Proteobacteria, and Chloroflexi showed the highest number of strong correlations with COG-specific gene richness (|Spearman's ρ| > 0.5; **Fig. 3F**). Notably, genes involved in inorganic ion transport and metabolism (P), which exhibited signatures of positive selection (**Fig. 2C and D**), were consistently associated with all three phyla in *L. booriae* (**Fig. 3F**). Consistent with patterns observed for abiotic environmental factors (**Fig. 3C**), genes involved in cell membrane biogenesis (M) and posttranslational modification, protein turnover, chaperones (O) exhibited the highest number of significant associations with phyla (13 and 14, respectively; **Fig. 3F**), indicating their key roles in mediating species-community interactions in *L. booriae*. Collectively, these findings indicate that like abiotic factors, bacterial community composition, particularly Proteobacteria, Chloroflexi, and Actinobacteria, is also more important to the gene content of *L. booriae* than *L. welshimeri*.

Wildlife appears to facilitate long-distance dispersal of *L. welshimeri* in soil.

As *L. welshimeri* had minimal geographically confined clades and was not strongly restricted by environmental conditions, we hypothesized that it frequently disperses across geographic locations, facilitating its broad geographic range. To test this hypothesis, we analyzed the distance-decay relationships for both species. The slope of the linear relationship between ANI and geographic distance were 5.5-fold shallower in *L. welshimeri* (R^2 = 0.15, slope = -8.79e-07) compared with *L. booriae* (R^2 = 0.30, slope = -4.86e-06; **Fig. 4A**), and gene content similarity showed a similar pattern (3.4-fold shallower slopes; *L. welshimeri*: R^2 = 0.05, slope = -7.98e-06; *L. booriae*: R^2 = 0.22, slope = -2.73e-05; **Supplementary Fig. 8**). These results showed that *L. welshimeri* exhibited a much weaker distance-decay relationship than *L. booriae* and support our hypothesis of frequent dispersal across locations in *L. welshimeri*.

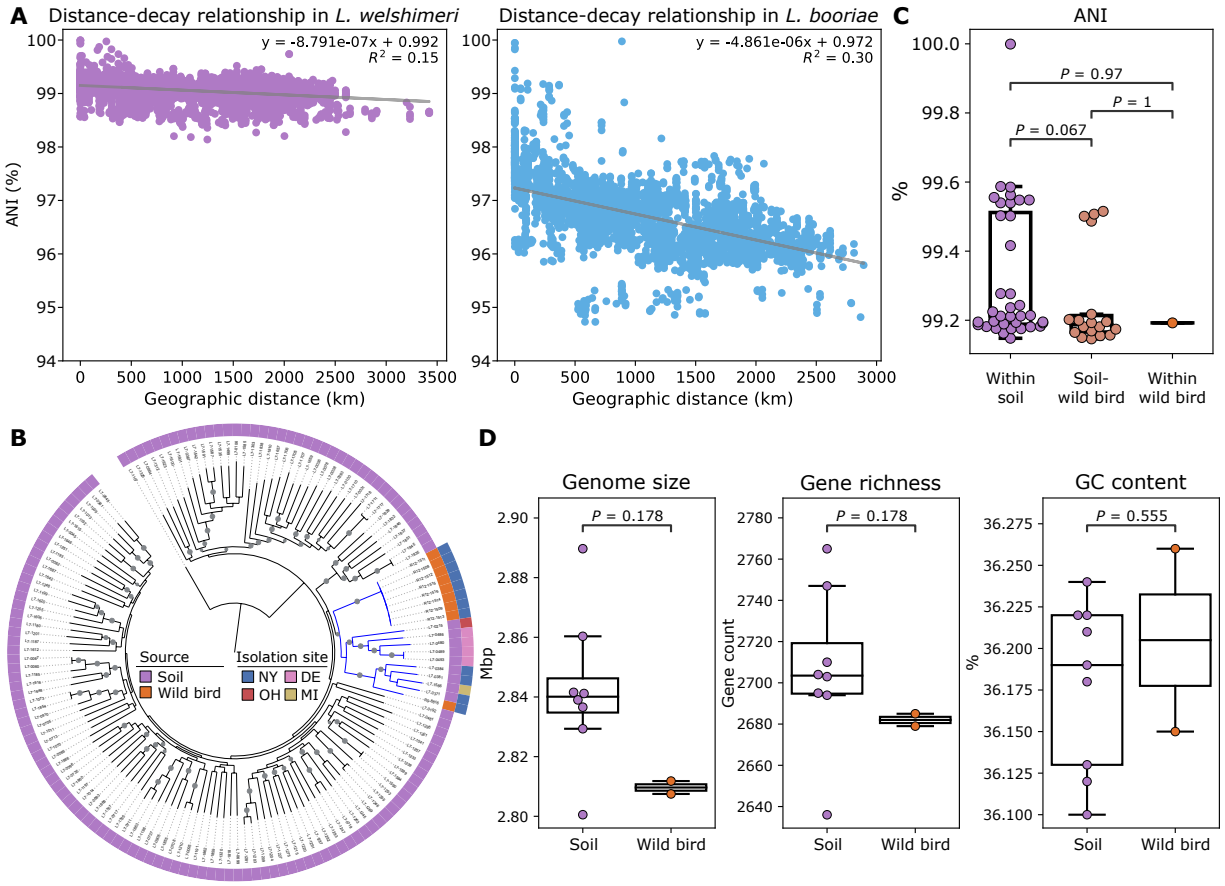
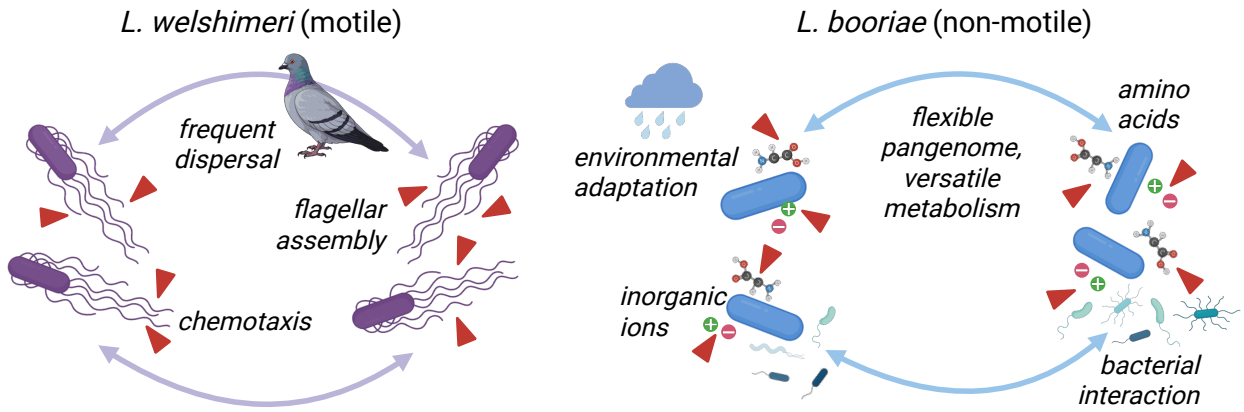


Figure 4. Dispersal patterns of *L. welshimeri* and *L. booriae* and evidence of wild birds as a dispersal vehicle for *L. welshimeri* in soil. (A) Distance-decay relationship in *L. welshimeri* and *L. booriae* inferred by the linear regression for genetic similarities measured by average nucleotide identity (ANI) and geographic distances. A steeper negative slope with a higher R^2 indicates a stronger distance-decay relationship. **(B)** Maximum likelihood phylogenetic tree based on core SNPs of 141 soil *L. welshimeri* isolates (purple) and nine wild bird isolates (orange). The tree was constructed using 1,000 bootstrap replicates, and bootstrap values > 80% are indicated by gray circles on the corresponding branches. The tree was rooted by midpoint. Branches belonging to the monophyletic clade that contains both soil and wild bird isolates are highlighted in blue, and the collection sites for isolates within this clade are annotated. **(C)** Pairwise ANI comparisons between soil and wild bird isolates of *L. welshimeri*. **(D)** Genome size, gene richness, and GC content compared between monophyletic soil and wild bird *L. welshimeri* isolates. For **(C)** and **(D)**, box plots display the interquartile range (IQR) with the median indicated as a line and whiskers extending to 1.5 times the IQR. Adjusted two-sided MW *U P* values are annotated.

As the reported motility speeds for *Listeria* (~0.09 μm/s in cytoplasm [55, 56] and ~10 μm/s on non-charged microscope slide [57]) combined with the physical constraints of soil environments [58, 59] make long-distance dispersal by motility alone improbable, we hypothesized that *L. welshimeri* achieves this via mobile hosts, such as wildlife. To test this hypothesis, we performed WGS of nine *L. welshimeri* isolates recovered from wild birds in New York and compared them with soil isolates. Core-SNP phylogenetic analysis revealed that the wild bird isolates clustered within a monophyletic clade alongside nine

545 soil isolates from nearby regions, including Ohio, Delaware, New York, and Michigan (**Fig.**
546 **4B**). Eight of the nine wild bird isolates formed a tight clade with minimal SNPs,
547 suggesting they represent a single strain. R12-1511 was selected as the representative
548 isolate for this strain alongside the other strain R9-8816 and were included in downstream
549 analyses. The representative wild bird and soil isolates forming a monophyletic clade
550 shared high ANI (mean: 99.25%, range: 99.15%-99.52%), with no significant difference
551 observed among within-soil, soil-wild bird, and within-wild bird groups (KW $P = 0.065$; **Fig.**
552 **4C**). Comparisons of genome size, gene richness, and GC content also revealed no
553 significant differences between soil and wild bird isolates within the monophyletic clade
554 (MW U $P = 0.178$, 0.178, and 0.555, respectively; **Fig. 4D**). These results indicate that *L.*
555 *welshimeri* isolates from soil and wild birds are closely related, supporting the hypothesis
556 that wildlife acts as vehicles for long-distance dispersal of *L. welshimeri*.
557
558 **Discussion**
559
560 Here, we show that motile and non-motile *Listeria* species adopt distinct genomic and
561 ecological strategies to achieve broad geographic ranges in soil (**Fig. 5**). *L. welshimeri,*
562 which is a motile species, appears to achieve a widespread distribution through effective
563 long-distance dispersal mediated by wildlife vectors, enabled by enhanced motility and
564 host colonization. In contrast, *L. booriae*, a non-motile species, appears to have limited
565 dispersal and likely attains a broad geographic range through effective environmental
566 adaptation by maintaining an extremely flexible pangenome responsive to abiotic (e.g.,
567 climate) and surrounding bacterial phyla (e.g., Proteobacteria, Chloroflexi, and
568 Actinobacteria). This genomic plasticity support fitness primarily through membrane
569 modifications and metabolic versatility, including inorganic ion and amino acid transport
570 and metabolism.
571



572
573 **Figure 5. Genomic strategies underlying broad geographic ranges in motile and non-motile *Listeria***
574 **species.** Triangles mark COG functions targeted by positive selection. Created with BioRender.com.
575

576 Flagellar motility and chemotaxis are central to bacterial nutrient sensing and the
577 exploitation of heterogeneous environments [19, 60–63]. In soil, where nutrients are
578 highly patchy and often transient [59], the ability to detect and move toward favorable
579 chemical gradients confers a strong fitness advantage. Flagella-driven motility allows
580 bacteria to navigate toward nutrient hotspots and away from unfavorable conditions,
581 facilitating rapid colonization of resource-rich microenvironments [64]. In *L. welshimeri*,
582 we observed that flagellar genes were targeted by positive selection, which likely
583 contributed to its niche expansion through enhanced motility and chemotaxis. Consistent
584 with this finding, laboratory and evolutionary experiments have shown that, under
585 selective pressure, *Escherichia coli* can evolve enhanced swimming velocity and altered
586 tumbling rates, improving habitat colonization and nutrient acquisition [65]. In addition,
587 the increased diversity of pathways in cellular processes and environmental information
588 processing observed in *L. welshimeri* may reflect an expanded repertoire of sensing and
589 regulatory systems (e.g., phosphotransferase system and two-component systems), and
590 together with the presence of flagellar assembly pathways, enhance its ability to detect
591 and respond to diverse external stimuli, benefiting niche expansion.
592
593 *L. welshimeri* also appears to leverage flagella to expand its ecological niche and broaden
594 its geographic range by enhancing host attachment and colonization, thereby enabling
595 frequent dispersal across landscapes. Flagellar genes can facilitate adhesion, host
596 attachment, and colonization [62, 66–68]. For example, in *Clostridioides difficile*, flagellin
597 is essential for initial attachment to mucosal surfaces [67], while in *Helicobacter pylori*,
598 deletion of *fliK* reduces adhesion to gastric mucosal cells [68]. Both *H. pylori* and
599 *Campylobacter jejuni* regulate flagellar biosynthesis and chemotaxis to penetrate viscous
600 milieus, such as gastrointestinal mucus, facilitating host colonization [62]. In *Listeria*,
601 flagella have also been shown to enhance host cell adhesion and invasion by increasing
602 the bacterium-host contact frequency and facilitating early attachment [69–71]. Such host
603 interactions can help bacteria overcome geographic barriers and promote dispersal over
604 large spatial scales [72]. Indeed, in this study, we observed minimal dispersal limitation
605 in *L. welshimeri*, as indicated by a homogenized population structure across geographic
606 locations. Similar patterns have been reported in *L. monocytogenes* lineage I, *Vibrio*
607 *parahaemolyticus,* and *V. fluvialis,* where frequent movement facilitates gene flow and
608 limits population divergence [53, 73, 74]. Further, using spatial phylogenetics, we
609 identified wild birds as a plausible dispersal vehicle for *L. welshimeri*. Wild birds, such as
610 starlings, pigeons, and migratory gulls, have been reported as carriers and dispersal
611 vectors of *Listeria* [31, 75, 76]. Avian-mediated dispersal has also been shown to shape
612 the biogeographic patterns of *L. monocytogenes* [75] and *E. coli* [77].
613
614 By nature, non-motile bacteria face intrinsic constraints on niche expansion because they
615 lack motility. However, based on our systematic, large-scale survey, the non-motile

species *L. booriae* was found to be highly ecologically successful in soil, occupying a broad range of environmental conditions [28]. In this study, we show that the broad geographic range of *L. booriae* is underpinned by extensive metabolic versatility enabled by a highly flexible pangenome for environmental adaption. Specifically, its pangenome is characterized by a diverse repertoire of genes involved in metabolic pathways. The metabolic network of *L. booriae* may have evolved greater complexity and connectivity as its pangenome expanded. This expansion likely facilitates the exploitation of diverse ecological resources during niche colonization [9]. Also, genes involved in metabolic functions, spanning both the core and accessory genome, show strong signatures of positive selection in *L. booriae*. Major adaptive traits included inorganic ion transport, amino acid and coenzyme metabolism, and protein biosynthesis. These processes are essential for nutrient acquisition, enzymatic activity, and cellular maintenance [78]. Enhanced metabolic efficiency, flexibility, and diversification likely enable *L. booriae* to thrive across nutrient-variable environments, compensating for its lack of motility and limited dispersal capacity.

Our data suggest that key environmental pressures that may shape the pangenome variation and metabolic versatility of *L. booriae* include climatic factors and microbial interactions. Consistent with this result, genes involved in metabolism in soil microbiomes showed strong associations with climatic variation [79–81]. Precipitation, in particular, functions as a dominant selective filter by alternately promoting the solubilization of essential mineral ions and nutrients, such as Fe and amino acids, and driving their depletion [82, 83]. Our results suggest that *L. booriae* responds to these moisture-driven fluctuations through the retention or positive selection of genes involved in inorganic ion, coenzyme, and amino acid transport and metabolism, including high-affinity scavenging and utilization systems for Fe, such as enterochelin esterase, Fe ABC transporters, and siroheme synthase, as well as amino acid transport and metabolism genes, including members of the major facilitator superfamily and pyridoxal phosphate dependent aminotransferases. Similar adaptive strategies have been reported in rhizosphere communities, where Fe metabolism genes are enriched in Actinobacteria under drought conditions that reduce Fe bioavailability [81]. Furthermore, because cofactor scarcity can drive compensatory amino acid uptake to maintain cellular homeostasis [84], the concurrent selection across these functional categories in *L. booriae* suggests metabolic flexibility that likely enables efficient resource scavenging under nutrient limitation while minimizing enzymatic constraints when resources are more abundant, as observed in *E. coli* [85].

In addition to abiotic influences, we identified Proteobacteria, Chloroflexi, and Actinobacteria as bacterial taxa most strongly associated with pangenome flexibility in *L. booriae*, particularly for genes involved in inorganic ion transport and metabolism.

656 Members of these phyla are known to possess diverse systems for inorganic ion
657 acquisition and nutrient transformation [86, 87], suggesting that community composition
658 may exert selective pressures on genes related to these functions in *L. booriae*. For
659 example, *Listeria* does not synthesize siderophores for Fe acquisition but can utilize a
660 range of exogenous organic compounds, including siderophores and catechol ligands, to
661 support growth under iron-limited conditions [88, 89], raising the possibility that *L. booriae*
662 relies, at least in part, on co-occurring bacteria to provide extracellular compounds that
663 facilitate the acquisition of essential inorganic ions.
664
665 Together, our findings delineate distinct genomic and ecological strategies underlying the
666 broad geographic ranges of motile and non-motile *Listeria* species, refining our
667 understanding of how bacteria with contrasting lifestyles achieve broad ecological
668 distributions and resilience to environmental pressures. Notably, our study highlights
669 often-overlooked mechanisms that enable non-motile bacteria to persist and diversify
670 across ecosystems. These insights have important implications for controlling motile and
671 non-motile pathogens and for minimizing environmental transmission. Given the
672 importance of *Listeria* to food safety, and the close monitoring of multiple species beyond
673 the pathogen *L. monocytogenes* in the food industry, our findings also inform risk
674 assessment and targeted interventions throughout food production, processing, and
675 storage.
676
677 **Data availability**
678
679 Whole-genome sequencing data for 141 *L. welshimeri* and 90 *L. booriae* soil isolates used
680 in this study have been submitted to the National Center for Biotechnology Information's
681 (NCBI) BioProject database under accession number PRJNA561882, along with
682 associated environmental metadata, previously published in Liao et al. (2021) [28]. The
683 16S rRNA gene amplicon sequencing data analyzed in this study are available under
684 BioProject accession number PRJNA749132, as published in Liao et al. (2023) [39].
685 Trimmed paired-end reads for the nine *L. welshimeri* wild bird isolates used in this study
686 have been deposited in the NCBI Sequence Read Archive (SRA), and their assembled
687 genomes are available in NCBI GenBank under the accession numbers listed in
688 **Supplementary Table 1**. All data needed to evaluate the conclusions in the paper are
689 present in the paper and/or the Supplementary Information.
690
691 **Code availability**
692 Code to replicate all analyses is publicly available at https://github.com/leaph-
693 lab/Listeria_motility_eco_MS.
694
695 **Acknowledgements**

We are grateful to Dr. Renato Orsi for assistance in identifying *L. welshimeri* wild bird isolates, and to the members of LEAPH (the Liao Laboratory) for their enriching discussions. Part of this work was funded by the Virginia Tech Graduate and Professional Student Senate Graduate Research Development Program (Y.-X.G.) and 4-VA Collaborative Research Grant (J.L.).

Author contributions

Conceptualization, J.L.; Data Curation, Y.-X.G., K.J.C., M.W., J.L.; Investigation, Y.-X.G., S.H., J.L.; Formal Analysis, Y.-X.G., J.L.; Visualization, Y.-X.G.; Writing – original draft preparation, Y.-X.G.; Writing – review & editing, K.J.C., M.W., J.L.; Supervision, J.L.

Competing interest

All authors declare no financial or non-financial competing interests.

References

1. Martiny JBH et al. Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 2006;4:102–112. https://doi.org/10.1038/nrmicro1341
2. Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* 2014;12:263–273. https://doi.org/10.1038/nrmicro3218
3. Gaston KJ et al. Abundance–occupancy relationships. *J Appl Ecol* 2000;37:39–59. https://doi.org/10.1046/j.1365-2664.2000.00485.x
4. Slatyer RA, Hirst M, Sexton JP. Niche breadth predicts geographical range size: a general ecological pattern. *Ecol Lett* 2013;16:1104–1114. https://doi.org/10.1111/ele.12140
5. Brown JH. On the relationship between abundance and distribution of species. *Am Nat* 1984;124:255–279. https://doi.org/10.1086/284267
6. MacArthur RH. Geographical ecology: patterns in the distribution of species. New York: Harper and Row, 1972.
7. Bono LM, Draghi JA, Turner PE. Evolvability costs of niche expansion. *Trends Genet* 2020;36:14–23. https://doi.org/10.1016/j.tig.2019.10.003
8. Pandit SN, Kolasa J, Cottenie K. Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology* 2009;90:2253–2262. https://doi.org/10.1890/08-0851.1
9. Chen Y-J et al. Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. *ISME J* 2021;15:2986–3004. https://doi.org/10.1038/s41396-021-00988-w
10. Logares R et al. Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J* 2013;7:937–948. https://doi.org/10.1038/ismej.2012.168

11. Liao J et al. The importance of neutral and niche processes for bacterial community assembly differs between habitat generalists and specialists. *FEMS Microbiol Ecol* 2016;92:fiw174. https://doi.org/10.1093/femsec/fiw174
12. Guo Q et al. Constraints of seed size on plant distribution and abundance. *Ecology* 2000;81. https://doi.org/10.1890/0012-9658(2000)081[2149:COSSOP]2.0.CO;2
13. McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. *Nat Microbiol* 2017;2:17040. https://doi.org/10.1038/nmicrobiol.2017.40
14. Rosen MJ et al. Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science* 2015;348:1019–1023. https://doi.org/10.1126/science.aaa4456
15. Ottemann KM, Miller JF. Roles for motility in bacterial–host interactions. *Mol Microbiol* 1997;24:1109–1117. https://doi.org/10.1046/j.1365-2958.1997.4281787.x
16. Pernthaler J. Competition and niche separation of pelagic bacteria in freshwater habitats. *Environ Microbiol* 2017;19:2133–2150. https://doi.org/10.1111/1462-2920.13742
17. Ramoneda J et al. Ecological relevance of flagellar motility in soil bacterial communities. *ISME J* 2024;18. https://doi.org/10.1093/ismejo/wrae067
18. Liu W et al. An evolutionarily stable strategy to colonize spatially extended habitats. *Nature* 2019;575:664–668. https://doi.org/10.1038/s41586-019-1734-x
19. Cremer J et al. Chemotaxis as a navigation strategy to boost range expansion. *Nature* 2019;575:658–663. https://doi.org/10.1038/s41586-019-1733-y
20. Taylor JR, Stocker R. Trade-offs of chemotactic foraging in turbulent water. *Science* 2012;338:675–679. https://doi.org/10.1126/science.1219417
21. Joung YS, Ge Z, Buie CR. Bioaerosol generation by raindrops on soil. *Nat Commun* 2017;8:14668. https://doi.org/10.1038/ncomms14668
22. Mayol E et al. Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nat Commun* 2017;8:201. https://doi.org/10.1038/s41467-017-00110-9
23. Smith DJ et al. Intercontinental dispersal of bacteria and archaea by transpacific winds. *Appl Environ Microbiol* 2013;79:1134–1139. https://doi.org/10.1128/AEM.03029-12
24. Muok AR, Briegel A. Intermicrobial hitchhiking: how nonmotile microbes leverage communal motility. *Trends Microbiol* 2021;29:542–550. https://doi.org/10.1016/j.tim.2020.10.005
25. Muok AR, Claessen D, Briegel A. Microbial hitchhiking: how *Streptomyces* spores are transported by motile soil bacteria. *ISME J* 2021;15:2591–2600. https://doi.org/10.1038/s41396-021-00952-8

775 26. Orsi RH et al. Taxonomy, ecology, and relevance to food safety of the genus
776 *Listeria* with a particular consideration of new *Listeria* species described between
777 2010 and 2022. *mBio* 2023. https://doi.org/10.1128/mbio.00938-23
778 27. Orsi RH, Wiedmann M. Characteristics and distribution of *Listeria* spp., including
779 *Listeria* species newly described since 2009. *Appl Microbiol Biotechnol*
780 2016;100:5273–5287. https://doi.org/10.1007/s00253-016-7552-2
781 28. Liao J et al. Nationwide genomic atlas of soil-dwelling *Listeria* reveals effects of
782 selection and population ecology on pangenome evolution. *Nat Microbiol*
783 2021;6:1021–1030. https://doi.org/10.1038/s41564-021-00935-7
784 29. Rocourt J, Grimont PAD. Notes: *Listeria welshimeri* sp. nov. and *Listeria seeligeri*
785 sp. nov. *Int J Syst Bacteriol* 1983;33:866–869. https://doi.org/10.1099/00207713-
786 33-4-866
787 30. Weller D et al. *Listeria booriae* sp. nov. and *Listeria newyorkensis* sp. nov., from
788 food processing environments in the USA. *Int J Syst Evol Microbiol* 2015;65:286–
789 292. https://doi.org/10.1099/ijs.0.070839-0
790 31. Chen T et al. Characterization of *Listeria monocytogenes* isolated from wildlife in
791 central New York. *Vet Med Sci* 2022;8:1319–1329.
792 https://doi.org/10.1002/vms3.758
793 32. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and
794 deduplication using fastp. *iMeta* 2023;2. https://doi.org/10.1002/imt2.107
795 33. Bankevich A et al. SPAdes: A New Genome Assembly Algorithm and Its
796 Applications to Single-Cell Sequencing. *J Comput Biol* 2012;19:455–477.
797 https://doi.org/10.1089/cmb.2012.0021
798 34. Gurevich A et al. QUAST: quality assessment tool for genome assemblies.
799 *Bioinformatics* 2013;29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086
800 35. Chklovski A et al. CheckM2: a rapid, scalable and accurate tool for assessing
801 microbial genome quality using machine learning. *Nat Methods* 2023;20:1203–
802 1212. https://doi.org/10.1038/s41592-023-01940-w
803 36. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2.
804 *Genome Biol* 2019;20:257. https://doi.org/10.1186/s13059-019-1891-0
805 37. Hyatt D et al. Prodigal: prokaryotic gene recognition and translation initiation site
806 identification. *BMC Bioinformatics* 2010;11:119. https://doi.org/10.1186/1471-
807 2105-11-119
808 38. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching
809 for the analysis of massive data sets. *Nat Biotechnol* 2017;35:1026–1028.
810 https://doi.org/10.1038/nbt.3988
811 39. Liao J et al. Comparative genomics unveils extensive genomic variation between
812 populations of *Listeria* species in natural and food-associated environments.
813 *ISME Commun* 2023;3:85. https://doi.org/10.1038/s43705-023-00293-x

814 40. Hall BG, Nisbet J. Building phylogenetic trees from genome sequences with
815 kSNP4. *Mol Biol Evol* 2023;40:msad235. https://doi.org/10.1093/molbev/msad235
816 41. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-
817 analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
818 https://doi.org/10.1093/bioinformatics/btu033
819 42. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the
820 phylogenetic tree display and annotation tool. *Nucleic Acids Res* 2024;52:W78–
821 W82. https://doi.org/10.1093/nar/gkae268
822 43. Rouli L et al. The bacterial pangenome as a new tool for analysing pathogenic
823 bacteria. *New Microbes New Infect* 2015;7:72–85.
824 https://doi.org/10.1016/j.nmni.2015.06.005
825 44. Brockhurst MA et al. The ecology and evolution of pangenomes. *Cur Biol*
826 2019;29:R1094–R1103. https://doi.org/10.1016/j.cub.2019.08.012
827 45. Ku C et al. Endosymbiotic gene transfer from prokaryotic pangenomes: inherited
828 chimerism in eukaryotes. *Procd Natl Acad Sci U S A* 2015;112:10139–10146.
829 https://doi.org/10.1073/pnas.1421385112
830 46. Nabholz B et al. Transcriptome population genomics reveals severe bottleneck
831 and domestication cost in the African rice (*Oryza glaberrima*). *Mol Ecol*
832 2014;23:2210–2227. https://doi.org/10.1111/mec.12738
833 47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
834 throughput. *Nucleic Acids Res* 2004;32:1792–1797.
835 https://doi.org/10.1093/nar/gkh340
836 48. Moura A et al. Whole genome-based population biology and epidemiological
837 surveillance of *Listeria monocytogenes*. *Nat Microbiol* 2016;2:16185.
838 https://doi.org/10.1038/nmicrobiol.2016.185
839 49. Camacho C et al. BLAST+: architecture and applications. *BMC Bioinformatics*
840 2009;10:421. https://doi.org/10.1186/1471-2105-10-421
841 50. Murrell B et al. Gene-wide identification of episodic selection. *Mol Biol Evol*
842 2015;32:1365–1371. https://doi.org/10.1093/molbev/msv035
843 51. Martin DP et al. RDP4: detection and analysis of recombination patterns in virus
844 genomes. *Virus Evol* 2015;1. https://doi.org/10.1093/ve/vev003
845 52. James G et al. Linear Regression. An introduction to statistical learning. New
846 York, NY: Springer, 2021, 99–103.
847 53. Goh Y-X et al. Disentangling the impacts of abiotic and biotic environmental
848 factors and dispersal dynamics on the pangenome fluidity of bacterial pathogens.
849 *bioRxiv* 2025. https://doi.org/10.1101/2025.03.27.645760
850 54. Peres-Neto PR et al. Variation partitioning of species data matrices: estimation
851 and comparison of fractions. *Ecology* 2006;87. https://doi.org/10.1890/0012-
852 9658(2006)87[2614:VPOSDM]2.0.CO;2

853 55. Rafelski SM, Alberts JB, Odell GM. An experimental and computational study of
854 the effect of ActA polarity on the speed of *Listeria monocytogenes* actin-based
855 motility. *PLoS Comput Biol* 2009;5:e1000434.
856 https://doi.org/10.1371/journal.pcbi.1000434
857 56. Robbins JR, Theriot JA. *Listeria monocytogenes* rotates around its long axis
858 during actin-based motility. *Curr Biol* 2003;13:R754–R756.
859 https://doi.org/10.1016/j.cub.2003.09.013
860 57. Wortel IMN et al. *Listeria* motility increases the efficiency of epithelial invasion
861 during intestinal infection. *PLoS Pathog* 2022;18:e1011028.
862 https://doi.org/10.1371/journal.ppat.1011028
863 58. Walters KE et al. Microbial dispersal into surface soil is limited on a meter scale.
864 *ISME J* 2025;19. https://doi.org/10.1093/ismejo/wraf169
865 59. Yang P, van Elsas JD. Mechanisms and ecological implications of the movement
866 of bacteria in soil. *Appl Soil Ecol* 2018;129:112–120.
867 https://doi.org/10.1016/j.apsoil.2018.04.014
868 60. Ottemann KM, Miller JF. Roles for motility in bacterial–host interactions. *Mol*
869 *Microbiol* 1997;24:1109–1117. https://doi.org/10.1046/j.1365-
870 2958.1997.4281787.x
871 61. Staropoli JF, Alon U. Computerized analysis of chemotaxis at different stages of
872 bacterial growth. *Biophys J* 2000;78:513–519. https://doi.org/10.1016/S0006-
873 3495(00)76613-6
874 62. Lertsethtakarn P, Ottemann KM, Hendrixson DR. Motility and chemotaxis in
875 *Campylobacter* and *Helicobacter*. *Annu Rev Microbiol* 2011;65:389–410.
876 https://doi.org/10.1146/annurev-micro-090110-102908
877 63. Zhao K, Liu M, Burgess RR. Adaptation in bacterial flagellar and motility systems:
878 from regulon members to 'foraging'-like behavior in *E. coli*. *Nucleic Acids Res*
879 2007;35:4441–4452. https://doi.org/10.1093/nar/gkm456
880 64. Colin R et al. Multiple functions of flagellar motility and chemotaxis in bacterial
881 physiology. *FEMS Microbiol Rev* 2021;45. https://doi.org/10.1093/femsre/fuab038
882 65. Ni B et al. Evolutionary remodeling of bacterial motility checkpoint control. *Cell*
883 *Rep* 2017;18:866–877. https://doi.org/10.1016/j.celrep.2016.12.088
884 66. Morgan SJ, Chaston JM. Flagellar genes are associated with the colonization
885 persistence phenotype of the *Drosophila melanogaster* microbiota. *Microbiol*
886 *Spectr* 2023;11. https://doi.org/10.1128/spectrum.04585-22
887 67. Sidner B et al. Flagellin is essential for initial attachment to mucosal surfaces by
888 *Clostridioides difficile*. *Microbiol Spectr* 2023;11.
889 https://doi.org/10.1128/spectrum.02120-23
890 68. Tao H et al. The impact of the flagellar protein gene *fliK* on *Helicobacter pylori*
891 biofilm formation. *mSphere* 2025;10. https://doi.org/10.1128/msphere.00018-25

892 69. O'Neil HS, Marquis H. *Listeria monocytogenes* flagella are used for motility, not
893 as adhesins, to Increase host cell invasion. *Infect Immun* 2006;74:6675–6681.
894 https://doi.org/10.1128/IAI.00886-06
895 70. Bigot A et al. Role of FliF and FliI of *Listeria monocytogenes* in flagellar assembly
896 and pathogenicity. *Infect Immun* 2005;73:5530–5539.
897 https://doi.org/10.1128/IAI.73.9.5530-5539.2005
898 71. Dons L et al. Role of flagellin and the two-component CheA/CheY system of
899 *Listeria monocytogenes* in host cell invasion and virulence. *Infect Immun*
900 2004;72:3237–3244. https://doi.org/10.1128/IAI.72.6.3237-3244.2004
901 72. Choudoir MJ, DeAngelis KM. A framework for integrating microbial dispersal
902 modes into soil ecosystem ecology. *iScience* 2022;25:103887.
903 https://doi.org/10.1016/j.isci.2022.103887
904 73. Yang C et al. Recent mixing of *Vibrio parahaemolyticus* populations. *ISME J*
905 2019;13:2578–2588. https://doi.org/10.1038/s41396-019-0461-5
906 74. Zheng H et al. Population genomics of the food-borne pathogen *Vibrio fluvialis*
907 reveals lineage associated pathogenicity-related genetic elements. *Microb Genom*
908 2022;8. https://doi.org/10.1099/mgen.0.000769
909 75. Gan L et al. Carriage and potential long distance transmission of *Listeria*
910 *monocytogenes* by migratory black-headed gulls in Dianchi Lake, Kunming.
911 *Emerg Microbes Infect* 2019;8:1195–1204.
912 https://doi.org/10.1080/22221751.2019.1647764
913 76. Fenlon DR. Wild birds and silage as reservoirs of *Listeria* in the agricultural
914 environment. *J Appl Bacteriol* 1985;59:537–543. https://doi.org/10.1111/j.1365-
915 2672.1985.tb03357.x
916 77. Liao J, Bergholz P, Wiedmann M. Adjacent terrestrial landscapes impact the
917 biogeographical pattern of soil *Escherichia coli* strains in produce fields by
918 modifying the importance of environmental selection and dispersal. *Appl Environ*
919 *Microbiol* 2021;87. https://doi.org/10.1128/AEM.02516-20
920 78. Drapeau GR, Macleod RA. A role for inorganic ions in the maintenance of
921 intracellular solute concentrations in a marine pseudomonad. *Nature*
922 1965;206:531–531. https://doi.org/10.1038/206531a0
923 79. Choudoir MJ et al. Pangenomes suggest ecological-evolutionary responses to
924 experimental soil warming. *mSphere* 2025;10.
925 https://doi.org/10.1128/msphere.00059-25
926 80. Liu J et al. Impacts of rainfall characteristics and slope on splash detachment and
927 transport of loess soil. *Land* 2024;13:189. https://doi.org/10.3390/land13020189
928 81. Xu L et al. Genome-resolved metagenomics reveals role of iron metabolism in
929 drought-induced rhizosphere microbiome dynamics. *Nat Commun* 2021;12:3209.
930 https://doi.org/10.1038/s41467-021-23553-7

931 82. Colombo C et al. Review on iron availability in soil: interaction of Fe minerals,
932 plants, and microbes. *J Soils Sediments* 2014;14:538–548.
933 https://doi.org/10.1007/s11368-013-0814-z
934 83. Lipson DA et al. An empirical model of amino acid transformations in an alpine
935 soil. *Soil Biol Biochem* 2001;33:189–198. https://doi.org/10.1016/S0038-
936 0717(00)00128-0
937 84. Gianoulis TA et al. Quantifying environmental adaptation of metabolic pathways in
938 metagenomics. *Prod Natl Acad Sci U S A* 2009;106:1374–1379.
939 https://doi.org/10.1073/pnas.0808022106
940 85. Basan M et al. Overflow metabolism in *Escherichia coli* results from efficient
941 proteome allocation. *Nature* 2015;528:99–104.
942 https://doi.org/10.1038/nature15765
943 86. Vincze É-B et al. Beneficial soil microbiomes and their potential role in plant
944 growth and soil fertility. *Agriculture* 2024;14:152.
945 https://doi.org/10.3390/agriculture14010152
946 87. Hao X et al. Recent advances in exploring the heavy metal(loid) resistant
947 microbiome. *Comput Struct Biotechnol J* 2021;19:94–109.
948 https://doi.org/10.1016/j.csbj.2020.12.006
949 88. Simon N et al. Utilization of exogenous siderophores and natural catechols by
950 *Listeria monocytogenes*. *Appl Environ Microbiol* 1995;61:1643–1645.
951 https://doi.org/10.1128/aem.61.4.1643-1645.1995
952 89. Coulanges V, Andre P, Vidon DJ-M. Effect of siderophores, catecholamines, and
953 catechol compounds on *Listeria* spp. growth in iron-complexed medium. *Biochem*
954 *Biophys Res Commun* 1998;249:526–530. https://doi.org/10.1006/bbrc.1998.9184
955