

Longitudinal dynamics of farmer and livestock nasal and faecal microbiomes and resistomes

Received: 3 July 2023

Accepted: 14 February 2024

Published online: 3 April 2024

 Check for updates

Bejan Mahmud^{1,2}, Rhiannon C. Vargas^{1,2}, Kimberley V. Sukhum^{1,2}, Sanket Patel^{1,2}, James Liao^{1,2}, Lindsey R. Hall¹, Akhil Kesaraju¹, Thao Le³, Terrie Kitchner⁴, Erik Kronholm⁵, Kyle Koschalek⁶, Casper G. Bendixsen^{1,2}, Jeffrey J. VanWormer⁵, Sanjay K. Shukla^{1,2,7,8}✉ & Gautam Dantas^{1,2,9,10,11}✉

Globally, half a billion people are employed in animal agriculture and are directly exposed to the associated microorganisms. However, the extent to which such exposures affect resident human microbiomes is unclear.

Here we conducted a longitudinal profiling of the nasal and faecal microbiomes of 66 dairy farmers and 166 dairy cows over a year-long period. We compare farmer microbiomes to those of 60 age-, sex- and ZIP code-matched people with no occupational exposures to farm animals (non-farmers). We show that farming is associated with microbiomes containing livestock-associated microbes; this is most apparent in the nasal bacterial community, with farmers harbouring a richer and more diverse nasal community than non-farmers. Similarly, in the gut microbial communities, we identify more shared microbial lineages between cows and farmers from the same farms. Additionally, we find that shared microbes are associated with antibiotic resistance genes. Overall, our study demonstrates the interconnectedness of human and animal microbiomes.

A substantial fraction of the population in the United States, and globally, is involved in animal agriculture^{1,2}, routinely coming into contact with livestock. As a consequence, farm workers are frequently exposed to microbes of farm and animal origins^{3,4}, with documented instances of microbial acquisition by farmers in farm settings^{5,6}. This microbial exposure is of importance for the health of farm residents. In some children, exposure to farm animals and associated microbes correlates with more rapid maturation of the gut microbiome⁷ and lower rates

of asthma, atopy and allergies^{7–10}; however, childhood exposure to livestock microbes has also been associated with stunted growth¹¹. In adults, a positive association between asthma and exposure to farm animals is observed^{12,13}, particularly in the absence of early-life farm exposure^{14–16}. Long-term contact throughout a person's life with farm animals is also associated with elevated risks of other respiratory disorders^{17,18} and blood cancers^{19,20}, further highlighting the risks of occupational exposures on farmer health.

¹The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, MO, USA. ²Department of Pathology and Immunology, Division of Laboratory and Genomic Medicine, Washington University School of Medicine, St Louis, MO, USA. ³Integrated Research Development Laboratory, Marshfield Clinic Research Institute, Marshfield, WI, USA. ⁴Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI, USA. ⁵National Farm Medicine Center, Marshfield Clinic Research Institute, Marshfield, WI, USA. ⁶Computational Informatics in Biology and Medicine program, University of Wisconsin-Madison, Madison, WI, USA. ⁷Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI, USA. ⁸Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO, USA. ¹⁰Department of Biomedical Engineering, Washington University in St Louis, St Louis, MO, USA. ¹¹Department of Pediatrics, Washington University School of Medicine, St Louis, MO, USA. ¹²These authors contributed equally: Bejan Mahmud, Rhiannon C. Vargas. ✉e-mail: shukla.sanjay@marshfieldresearch.org; dantas@wustl.edu

The exposure of farmers to livestock microbes is also important in the context of the growing antibiotic resistance crisis. The increasing demand for dietary animal protein sources has been met by an expansive use of antibiotics in agriculture aimed at disease prevention and growth promotion of food animals²¹. Remarkably, 73% of the global antimicrobial use in 2017 was directed towards use in animals²². This abundant use of antibiotics has resulted in the selection for and enrichment of antimicrobial-resistant organisms (AROs) in livestock^{23,24}. The selection for antimicrobial resistance in farm animals is expected to grow even stronger, as the global antimicrobial usage in livestock is estimated to increase by 8% between 2020 and 2030²⁵. Farmers may thus be at an increasing risk of acquiring AROs of zoonotic origins^{26–29} and may serve as subsequent vectors of dissemination of AROs and antibiotic resistance genes (ARGs) to other members of the community^{30,31}, highlighting the intimate connection between antibiotic use in agriculture and public health.

In addition to diet and pharmaceutical exposures (that is, antibiotics), human–animal interfaces are increasingly recognized for their capacity to impact the long-term dynamics of human microbiomes³². Such interfaces are particularly noteworthy within agricultural settings. As detailed above, occupational exposures associated with livestock farming are important for their direct implications on the health of half a billion people and their potential role in the spread of antimicrobial resistance. Despite this global public health importance, the effects of occupational exposures on the taxonomic and functional (that is, ARG) compositions of native microbiomes of farmers are not well understood. Although there is evidence for farming-associated microbiome signatures^{33–35}, it is unclear how biogeographically distinct microbial communities (that is, colonizing different body habitats) are differentially affected by occupational exposures. This knowledge is critical for elucidating the mechanistic basis of the occupational health risks faced by farmers and assessing the contributions of commonplace farming practices towards global ARO dissemination.

Results

To address these gaps in knowledge, we conducted a large, longitudinal investigation of the resident microbial communities of dairy farmers and cows from 37 dairy farms located in central Wisconsin in the United States. Our investigation focused on nasal and gut microbiomes, as both microbial communities have important health implications but would be differentially exposed to the farm environment. To determine the occupational signatures of the farmer microbiomes, we also profiled the nasal and gut communities of an age-, sex- and ZIP code-matched human cohort with no occupational exposures to the farm environment (that is, non-farmers). In all, our year-long (13 months) sampling yielded 712 faecal ($n_{\text{farmer}} = 171$, $n_{\text{non-farmer}} = 114$, $n_{\text{cow}} = 427$) and 726 nasal ($n_{\text{farmer}} = 171$, $n_{\text{non-farmer}} = 137$, $n_{\text{cow}} = 418$) samples from 292 ($n_{\text{farmer}} = 66$, $n_{\text{non-farmer}} = 60$, $n_{\text{cow}} = 166$) longitudinally sampled subjects (Supplementary Table 1), allowing an in-depth and multidimensional characterization of the resident microbial communities.

Characterization of the farmer nasal microbiome

Occupational exposures shift the farmer nasal microbiome to be compositionally more similar to the nasal microbiome of dairy cows. Profiling of nasal microbiome compositions using 16S rRNA gene sequencing showed that dairy farmers persistently harboured a more diverse nasal microbiome relative to non-farmers (Fig. 1a and Extended Data Fig. 1a). Similarly, exposure to pig-farm environments has been associated with diversification of pig-farmer nasal microbiomes³⁴. Increased nasal microbial diversity is associated with reduced mucosal inflammation³⁶ and may thus play a role in the reported protective effects of livestock exposure against atopy and allergy^{9,10}. The cow nasal microbiome was similarly more diverse relative to the non-farmer nasal community. Conversely, the Shannon diversity metrics of the cow and farmer nasal microbiomes did not differ significantly from each other in three of the

four seasons (spring, $P = 0.40$; summer, $P = 0.37$; autumn, $P = 0.081$; Wilcoxon rank-sum test; Fig. 1a). However, the microbial richness was consistently higher in the cow nasal community relative to that of farmers (Extended Data Fig. 1a). As expected, the nasal microbiomes of cows and non-farmers were compositionally distinct ($P < 0.001$, permutational multivariate analysis of variance (PERMANOVA); Fig. 1b,c), with 83 differentially abundant microbial families (75 enriched in cows; Supplementary Table 2) and 154 differentially abundant genera (140 enriched in cows; Supplementary Table 3) between the two hosts. Interestingly, the farmer nasal microbiome is compositionally distinct from the nasal communities of cows ($P < 0.001$, PERMANOVA) and non-farmers ($P < 0.001$, PERMANOVA), and occupies a compositionally intermediate state (Fig. 1b). Relative to the nasal microbiome of non-farmers, the farmer nasal microbiome overlapped with that of cows to a much larger extent (Fig. 1c). Indeed, the farmer communities had lower Bray–Curtis distances (that is, more compositionally similar) to the cow samples than the non-farmer communities did to the cow samples (Fig. 1d). The cow nasal microbiome was also compositionally more similar to the nasal communities of farmers from other farms than to the nasal microbiome of non-farmers residing within the same ZIP code (Fig. 1d), suggesting that there is a general dairy-farm effect on the nasal microbial community that extends beyond individual farms. Furthermore, 58 microbial families (57 enriched in farmers; Extended Data Fig. 2) and 79 genera (71 enriched in farmers; Fig. 1e) were differentially abundant between the farmer and non-farmer nasal microbiomes. Among the bacterial families overrepresented in farmers are Lactobacillaceae, Aerococcaceae and Enterococcaceae, all of which are associated with protection against mucosal inflammation and sinus infection³⁶. Notably, of the taxa enriched in the farmer nasal communities, 56 families (98.2%) and 70 genera (98.6%) were also enriched in the cow nasal communities relative to the non-farmer nasal microbiome (Fig. 1e and Extended Data Fig. 2). Taken together, these data suggest that occupational exposures coincide with the restructuring of the farmer nasal microbiome primarily through the acquisition of microbes enriched in the cow nasal community. It is important to note that although these overlaps in the nasal microbiome compositions of cows and farmers could emerge from microbial transmission between the two mammalian hosts, it is also possible that a common environmental source underlies the microbial spread to both hosts. As such, the origins of the similarities between the nasal communities of cows and farmers remain to be elucidated in future investigations.

Farm-specific gut microbial signatures

We next investigated whether the nasal microbial acquisition in farmers affects the composition of their gut microbiomes. As with nasal microbiome profiling, we first used 16S rRNA gene sequencing to characterize the cow and human faecal community compositions. In contrast to the observed difference in nasal communities, cows persistently harboured a more diverse gut community than both human groups, and the farmer and non-farmer gut microbiomes did not differ from each other significantly in diversity or richness (Fig. 2a and Extended Data Fig. 1b). Compositionally, microbiomes cluster by the mammalian host type (that is, cow versus human): the primary axis of the principal coordinate analysis, which explains over half of the observed compositional variation, separates the cow and human samples (Fig. 2b and Supplementary Tables 4 and 5), suggesting that the mammalian host type is the predominant determinant of gut microbiome composition. Unlike the nasal microbiome, occupational exposure to livestock did not correlate significantly with the farmer gut microbiome taxonomic structure at the genus level (Extended Data Fig. 3a–d), and both farmer and non-farmer gut communities were equally dissimilar in their compositions to the gut community of cows (Extended Data Fig. 4a). These observations may be interpreted to suggest that either occupational exposures have no effects on the farmer gut microbiome or these effects are apparent only at higher taxonomic or functional resolutions.

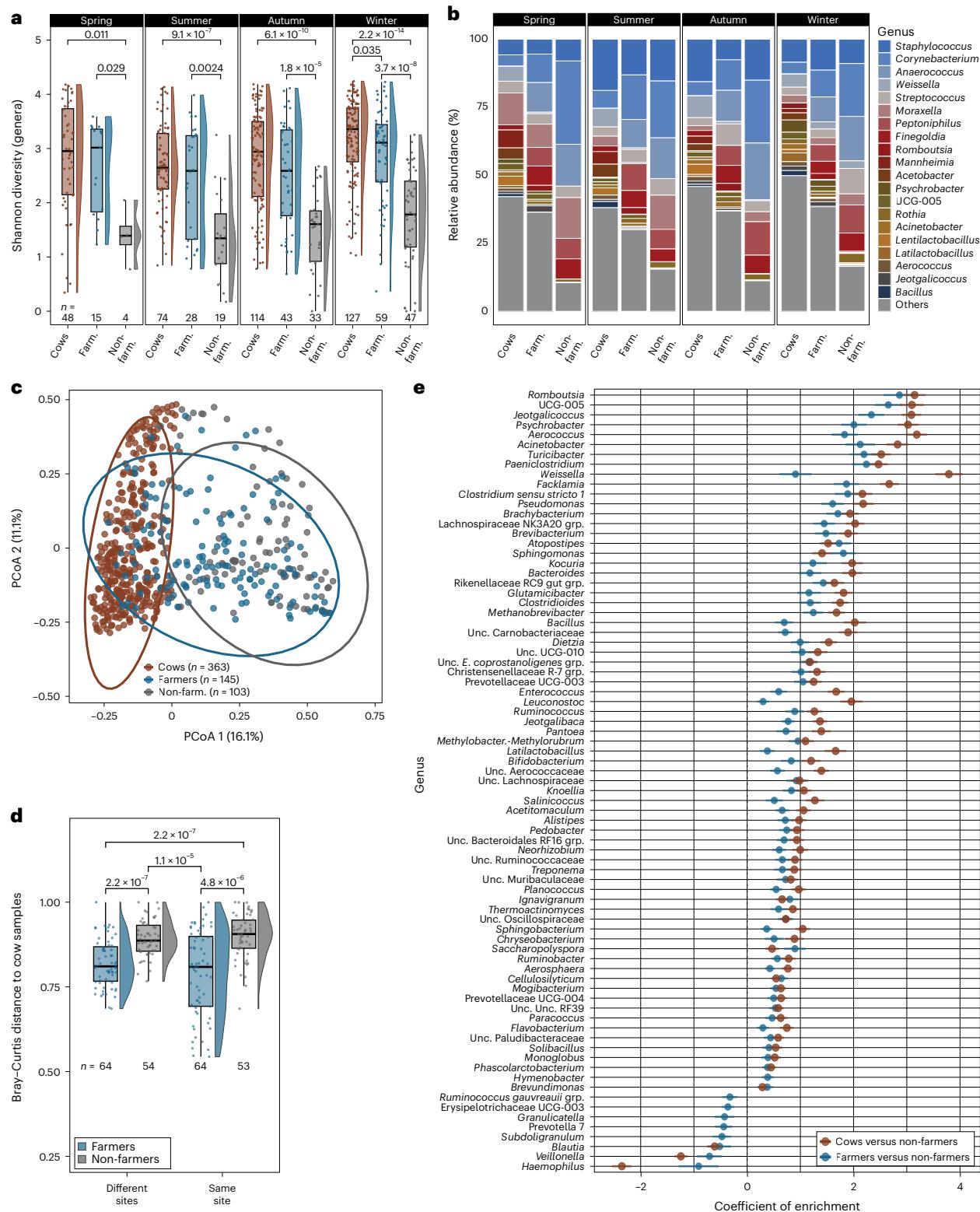


Fig. 1 | Farmer nasal microbiome comprises cow-associated microbes.

a, Genus-level Shannon diversity of the cow, farmer and non-farmer nasal microbiomes across seasons. **b**, The average nasal microbiome composition of cows, farmers (farm.) and non-farmers (non-farm.) across seasons. Top 20 genera by relative abundance are coloured. Sample numbers are as in **a**. **c**, Principal coordinate analysis (PCoA) of Bray–Curtis dissimilarities of the genus compositions of nasal samples. **d**, Average nasal Bray–Curtis distances of farmers and non-farmers residing in the same or different collection sites. Beta diversity is based on genus compositions. **e**, Genera with significant ($P < 0.05$) differential abundances between farmer and non-farmer nasal microbiomes.

For each genus, when significant, the corresponding coefficient in cows relative to non-farmers is also shown. Points denote mean coefficients, and whiskers correspond to standard error. Enrichment was tested using MaAsLin2 (Methods), with Benjamini–Hochberg correction for multiple hypotheses. Sample numbers are as in **a** and **d**, the boxplots show the median (centre line), quartiles (box limits) and $1.5 \times$ interquartile range (whiskers). Dots correspond to individual samples. Half-violins show the data distribution. P values were calculated using the two-tailed Wilcoxon rank-sum test, with subsequent Benjamini–Hochberg correction for multiple hypotheses.

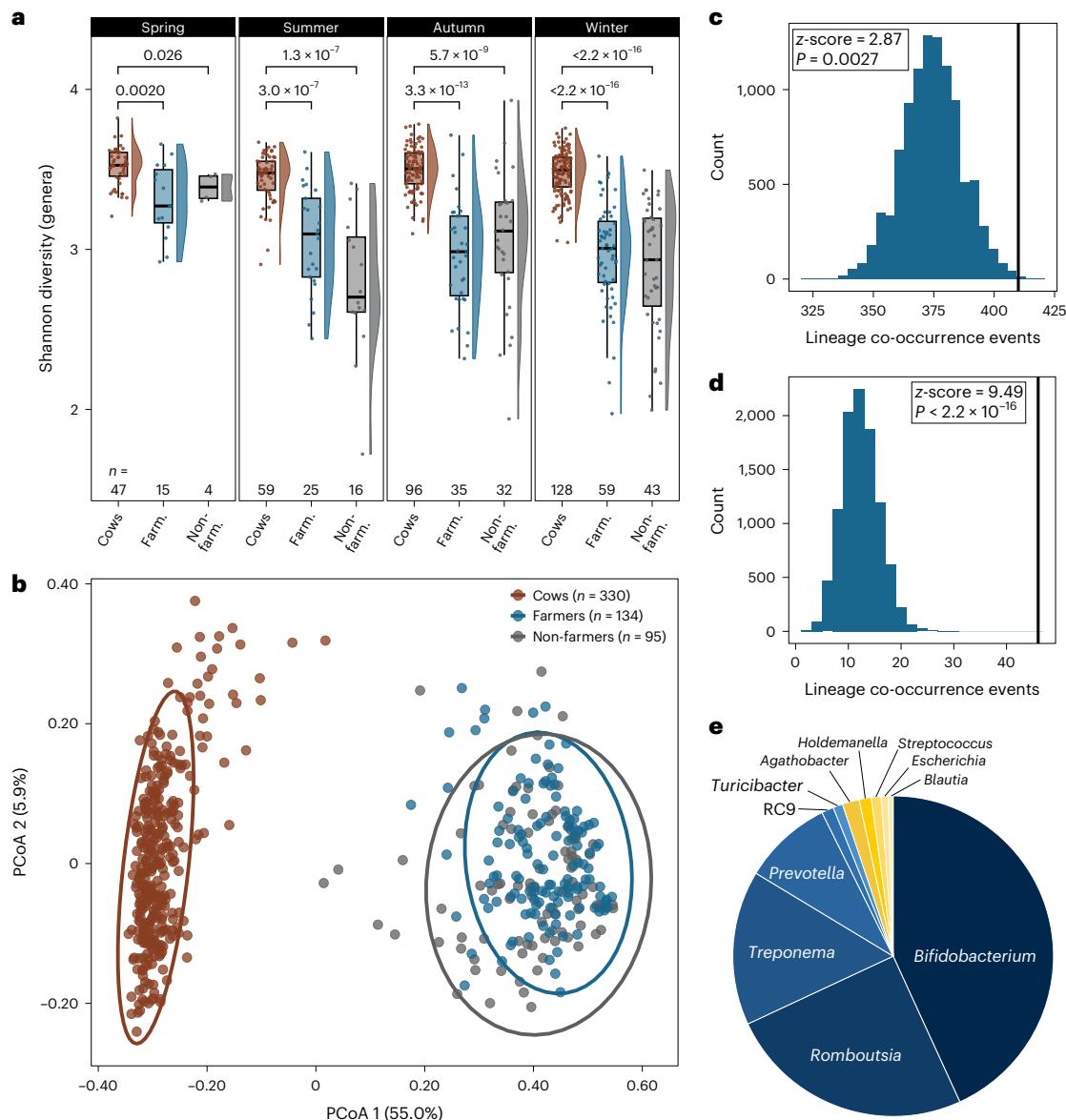


Fig. 2 | Occupational impacts on the gut microbiome are apparent at higher resolutions. a, Genus-level Shannon diversity of cow, farmer and non-farmer faecal microbiomes across seasons. The boxplots show the median (centre line), quartiles (box limits) and $1.5 \times$ interquartile range (whiskers). Dots correspond to individual samples. Half-violins show the data distribution. *P* values were calculated using the two-tailed Wilcoxon rank-sum test, with subsequent Benjamini–Hochberg correction for multiple hypotheses. **b,** Principal coordinate analysis of the Bray–Curtis dissimilarities of the genus compositions of faecal samples. **c,d,** Permutational analysis of the enrichment in microbial lineage

co-occurrences between cows and farmers independent of site (**c**) or from the same farms (**d**). The histograms depict the distribution of the expected outcome. Vertical lines denote the observed counts. One-tailed *P* values were calculated using *z*-scores (Methods). The Benjamini–Hochberg method was subsequently used to correct for multiple hypotheses. **e,** Genera represented among microbial lineages co-occurring in cow and farmer guts. The ‘slice’ sizes are proportional to the number of co-occurrence events involving the corresponding genera. Genera enriched in the cow and farmer nasal microbiomes relative to those of non-farmers are coloured in shades of blue; the rest are in shades of yellow.

We were prompted to test the latter possibility based on the reported lower rates of gastrointestinal distress (that is, diarrhoea, constipation and dyspepsia) among dairy farmers relative to non-farmers³⁷. Accordingly, we employed shotgun metagenomic sequencing of the faecal samples, which enables microbial community profiling at the species, sub-species and functional levels. We shotgun-sequenced all 712 faecal samples, generating 22.7 Gb of paired-end sequencing data, and profiled the species and metabolic pathway compositions using MetaPhlAn 4³⁸ and HUMAN N3³⁹, respectively. We found that the farmer and non-farmer gut microbiomes did indeed differ significantly in their compositions at the species level (Extended Data Fig. 3e–l). The differences between the gut communities of the two human groups

also emerge at the level of individual taxa (Extended Data Fig. 5a,b). Namely, farmer guts were enriched in butyrate-producing Firmicutes (for example, *Coprococcus eutactus* and *Roseburia faecis*), the depletion of which is associated with diarrhoeal diseases⁴⁰. Last, we observed differences in the metabolic capacities of the farmer and non-farmer guts; the farmer microbiomes were enriched in degradation pathways for simple sugars (for example, lactose and galactose), whereas the metabolic pathways for the degradation of complex sugars (for example, starch and glycogen) were overrepresented in the non-farmer guts (Extended Data Fig. 5c). In all, these data suggest that occupational exposures impact the nasal and gut microbiomes of farmers to varying degrees, with the associated effects on the gut community being only

apparent at higher taxonomic and functional resolution. We also note that, although exposure to the farm setting could underlie the reported occupational signatures, other factors (for example, diet) could also contribute towards the observed differences in the gut microbiomes of farmers and non-farmers.

To assess the extent of horizontal dissemination of microbes within dairy-farm settings, we used *inStrain*⁴¹ to investigate the co-occurrences of microbial sub-species (that is, lineages) in human and cow guts. We first assembled 15,005 metagenome-assembled genomes (MAGs), with 993 high- and 3,891 medium-quality assemblies. DerePLICATION then yielded a set of 978 representative MAGs that were used as references for microbial lineage tracking. We found an enrichment in the number of lineages present in farmers and cows relative to the lineages present in cows and non-farmers ($P = 0.0027$, permutational analysis; Fig. 2c). When looking only at farmer–cow pairs, the frequency of lineage sharing was substantially higher in subject pairs residing on the same farms ($P < 2.2 \times 10^{-16}$, permutational analysis; Fig. 2d), suggesting that cohabitation with livestock is the predominant driver of acquisition of livestock-associated microbial lineages. We note, however, that the presence of shared lineages can have aetiologies other than direct transfer (for example, a common environmental source), and future investigations are needed to establish the mechanisms underlying the co-occurrence of microbes in farmers and cows. The lineages co-occurring in both farmer and cow guts represent 11 genera, three of which (*Prevotella*, *Holde manella* and *Blautia*) are overrepresented in the farmer gut relative to that of non-farmers (Extended Data Fig. 5a,b). Moreover, only six genera account for 94.9% of all lineage-sharing events: *Bifidobacterium*, *Romboutsia*, *Treponema*, *Prevotella*, RC9 and *Turicibacter* (Fig. 2e). Notably, all these genera were found to be enriched in both farmer and cow nasal microbiomes, relative to the nasal community of non-farmers (Fig. 1e). Put together, these data suggest that, although farmer nasal microbial acquisition of dairy-cow-associated microbes is not correlated with large-scale restructuring of farmer gut microbiome architecture, it is associated with intestinal harbouring of specific livestock-associated microbial lineages.

Functional screening of novel ARGs

To understand the scope of ARG sharing within a farm setting, we used the faecal shotgun metagenomic sequencing data to profile the gut antibiotic resistome compositions of cows, farmers and non-farmers. High-throughput resistome characterization relies on mapping of sequencing data to a reference ARG set. However, such reference datasets are biased towards human hosts⁴² and would thus underrepresent the resistome composition of non-human subjects (for example, cows). To address these shortcomings, we identified novel gut-resident ARGs (that is, not present in existing databases) using functional metagenomic selections, which enables the identification of functionally validated ARGs without reliance on sequence similarity to known genes^{43–47}. We constructed 13 metagenomic libraries using stool samples representing every study participant (humans and cows) and screened them against a panel of 17 antimicrobials and combinations, including those commonly used in farm animals⁴⁸ (Supplementary Table 6). Our screens yielded 2,049 unique ARGs, with 1,202 originating from libraries constructed using cow samples (Fig. 3a). These ARGs are underrepresented in current ARG databases (median 32.8% amino acid identity), but are mostly similar to proteins within the comprehensive NCBI non-redundant (nr) database (median 92.5% amino acid identity) (Fig. 3a). This suggests that, although the nucleic-acid sequences of the functionally screened ARGs have been described, they have not been previously annotated to confer antimicrobial resistance⁴⁵. Notably, there is a disparity across mammalian hosts in how similar their ARGs are to known proteins, with cow ARGs being significantly underrepresented (median 84.8% amino acid identity) compared to those of

human origin (median 98.8% amino acid identities for both farmer and non-farmers) (Fig. 3a). These observations provide further support to the human bias noted in ARG databases⁴² and demonstrate that this bias also exists in the most comprehensive, non-functional reference catalogues.

Profiling the cow gut resistome

We combined the 2,049 functionally screened ARGs with those present in the Comprehensive Antibiotic Resistance Database (CARD)⁴⁹ and the NCBI antimicrobial resistance gene catalogue⁵⁰. To this combined set, we also added 17,292 ARGs resulting from past functional selections on samples with diverse origins (preterm infants, soil, non-human primates)^{44–47}, yielding, to our knowledge, the most extensive reference ARG set and enabling a more comprehensive resistome profiling. Using this dataset, we find a higher ARG burden (that is, relative abundance) in human guts than in cows (Fig. 3b and Extended Data Fig. 6). However, resistance determinants against lincosamide, phenicol, macrolide and tetracycline classes of antibiotics were overrepresented in the cow gut resistome relative to human resistomes (Fig. 3c). The cow resistome also had lower richness than the gut resistomes of both human groups (Fig. 3d). Mimicking the gut taxonomic profiles, the gut resistome compositions primarily separated by mammalian host type (cow versus human; Fig. 3e). Relative to the human gut resistome, the cow gut is enriched in 413 ARGs, conferring resistance to tetracycline, trimethoprim, glycopeptide and beta-lactam antibiotics, among other drug classes (Fig. 3f and Extended Data Fig. 7). We note that, of the ARGs enriched in the cow gut, 94.9% (392 of 413) were identified through our functional screens (Fig. 3f), emphasizing the value of functional metagenomics in enabling more comprehensive resistome profiling. To enable broader investigations of ARG burdens, particularly in under-studied environments, we provide the full set of functionally selected ARGs (Methods).

Use of antimicrobials leads to the expected expansion in the relative abundance of ARGs conferring resistance to the treatment agent⁵¹. However, post-treatment increases in the levels of ARGs unrelated to the administered drug(s) are also commonly observed^{51,52}, contributing towards the ongoing depletion of drug options available for the treatment of bacterial infections²³. The enrichment of ARGs unrelated to the treatment agent is a consequence of co-selection of diverse ARGs carried by the same bacterial hosts⁴⁵ and/or encoded within multidrug resistance gene clusters⁵³. We investigated the instances of ARG linking within the cow gut, as this knowledge could guide more informed treatment choices. We hypothesized that linked ARGs would be correlated in their relative abundances (RPKMs) across samples. We identified 7,960 unique pairs of highly correlated ($r > 0.7$, $P < 0.05$) ARGs, and we found numerous examples of these linked genes encoded in close proximity (within -1 kbp) to each other (Fig. 3g), providing further support for their capacity to undergo co-selection upon antibiotic challenge. Most commonly, sets of strongly correlated ARGs conferred resistance to combinations of trimethoprim, glycopeptide, tetracycline and beta-lactam classes of antibiotics (Fig. 3h). Consequently, treatments with these drugs are predicted to have high probabilities of enrichment for collateral resistance mechanisms, whereas this predicted probability is relatively low for fosfomycins and quinolones, suggesting these as potentially more desirable treatment options for this cohort of dairy cows. Noteworthy among the set of linked ARGs is the pair of *mef(En2)* and *Inu(AN2)*, conferring resistance to macrolide and lincosamide drug classes, respectively. Although the instances of linked genes conferring resistance to macrolides and lincosamides are relatively rare, with only six other such gene pairs found among the 7,960 total, *mef(En2)* and *Inu(AN2)* were strongly correlated ($r = 0.86$, $P < 1 \times 10^{-15}$) in their relative abundances (Fig. 3g). Furthermore, the two contiguous genes were enriched in the cow gut relative to the human resistomes and were highly prevalent, present in 94.6% and 91.8% of all cow samples, respectively. The *mef(En2)*-*Inu(AN2)* pair has also

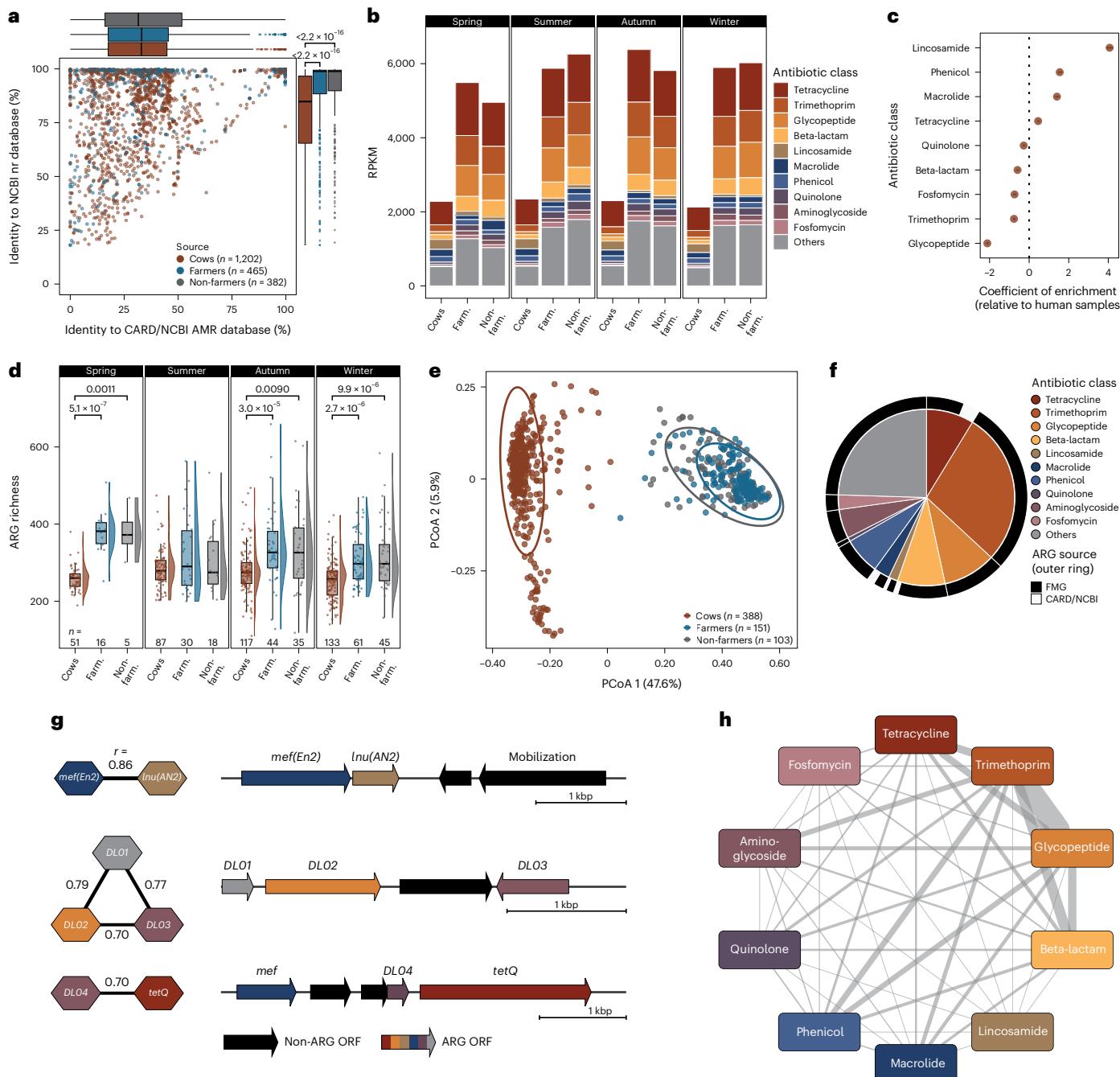


Fig. 3 | Functional metagenomics enables cow gut resistome profiling. **a**, Similarities (percent amino acid identity) of functionally screened ARGs to top hits from ARG and NCBI nr databases. **b**, Average gut resistome composition of cows, farmers and non-farmers across seasons. The top 20 antibiotic classes by combined relative abundances (RPKMs) of corresponding ARGs are coloured. Sample numbers are provided in **d**. **c**, Antibiotic classes differentially represented (combined RPKMs of corresponding ARGs) in the cow gut resistome relative to that of humans. Whiskers correspond to s.e. Enrichment was tested using MaAsLin2 (Methods). **d**, Gut resistome richness in cows, farmers and non-farmers. **e**, Principal coordinate analysis of Bray–Curtis dissimilarities of gut ARG compositions. **f**, Piechart of the ARGs enriched in the cow gut relative to the human gut resistome. The inner circle depicts the distribution of antibiotic classes among the enriched genes. The outer ring indicates the database source

of the ARG. FMG, functional metagenomics. **g**, Diamonds depict individual ARGs, and correlating genes are connected with lines, with corresponding Pearson correlation coefficients indicated. On the right, example contigs encoding the correlating ARGs are shown. ARGs are coloured by antibiotic class, as in **f**. **h**, Network of correlating ARGs. Nodes indicate antibiotic classes corresponding to the ARGs and are coloured as in **f**. Edges are proportional in size to the number of unique, correlating ARG pairs corresponding to the connected antibiotic classes. In **a** and **d**, the boxplots show the median (centre line), quartiles (box limits) and 1.5× interquartile range (whiskers). Dots correspond to individual samples. Half-violins show the data distribution. *P* values were calculated using the two-tailed Wilcoxon rank-sum test, with subsequent Benjamini–Hochberg correction for multiple hypotheses.

been reported to be carried by other domesticated animals (ducks⁵⁴, pigs⁵⁵ and dogs⁵⁶), where it is encoded by distinct microbial taxa. This suggests that this pair might exist within a mobilizable cassette that

enables its dissemination among diverse microbial and mammalian hosts. Indeed, the contig encoding the ARGs also includes a proximal gene with annotated mobilization function (Fig. 3g).

Profiling the farmer gut resistome

As previously stated, the mammalian host (cow versus human) was the predominant driver of gut resistome composition (Fig. 3e), with the farmer and non-farmer resistomes clustering together and being similarly distinct in their compositions from the makeup of the cow resistome (Extended Data Fig. 4b). Furthermore, the farmer and non-farmer resistomes did not differ significantly in the number of unique ARGs encoded (Fig. 3d), suggesting that occupational exposure to farm animals does not result in a large-scale restructuring of the human gut resistome. As this mirrors the trends in gut taxonomic composition (Fig. 2a,b), we wondered whether the occupational signatures of the gut resistome might similarly become apparent only at higher resolutions of analysis. With a view to assessing this potential, we first focused on the six genera that were enriched in the cow and farmer nasal communities and represented the vast majority of the microbial lineages co-occurring in both cow and farmer guts, that is, *Bifidobacterium*, *Romboutsia*, *Treponema*, *Prevotella*, RC9 and *Turicibacter* (Figs. 1e and 2e). We hypothesized that the ARGs carried by these taxa within the cow gut would be overrepresented in farmers relative to those without routine exposure to cows. Our approach to identifying such ARGs was based on the expected correlation between the relative abundances of microbes and their encoded genes⁴⁵. We identified 29 ARGs whose relative abundances in the cow gut correlated significantly ($p > 0.5$, $P < 0.05$) with those of the six genera of interest (Fig. 4a), giving a set of candidate genes encoded by the members of these taxa. More than half of these genes (17 of 29) were found to be encoded within corresponding cow MAGs (Fig. 4a), providing further support to the ARGs being encoded by these genera. Notably, these 29 ARGs were seasonally overrepresented ($P = 0.031$, Wilcoxon rank-sum test) in the farmer guts relative to the non-farmer resistomes (Extended Data Fig. 8). Extending the analysis to all 11 genera constituting all shared microbial lineages (Fig. 2e) yielded a larger set of correlating ARGs (Supplementary Table 7), which were similarly significantly overrepresented ($P = 0.0040$, Wilcoxon rank-sum test) in the farmer resistome relative to that of non-farmers (Fig. 4b). These data indicate that farmers acquire ARGs via occupational exposures and that this acquisition is mediated by microbial lineages circulating within the farm environment. In support of this, many of these ARGs were also found to be encoded by the farmer MAGs corresponding to the genera of interest (Fig. 4a). This is further illustrated by the example of two *Bifidobacterium* MAGs that both encoded six ARGs (all enriched in cows relative to humans) within highly similar genomic contexts (Fig. 4c); the MAGs in question were assembled from the samples of a cow and a farmer residing within the same farm, demonstrating the ARG dissemination through acquisition of microbial lineages.

Horizontally disseminating ARGs

We identified several examples of ARGs within conserved, putatively mobile cassettes present in otherwise distantly related MAGs of cow and farmer origins (Fig. 4d). ARGs can disseminate to new gut-resident bacterial hosts through mobile genetic elements (for example, transposons and plasmids). We aimed to broadly investigate the distribution of mobile elements within the farm setting, with a particular interest in the elements shared between distinct microbes of cow and farmer origins, given their role in the dissemination of ARGs to diverse microbes upon acquisition by the mammalian host. Following an established protocol for identifying horizontally transferred regions in bacterial genomes⁵⁷, we identified all near-identical DNA fragments ($\geq 99\%$ identity by BLAST) present in pairs of otherwise distantly related high- and medium-quality MAGs ($< 95\%$ pairwise average nucleotide identity) from distinct origins (cow versus farmer) (Methods). We identified 18,288 instances of such horizontally disseminating fragments, most commonly shared by members of the Lachnospiraceae, *Bifidobacteriaceae*, *Bacteroidaceae* and *Acutalibacteraceae* families (Fig. 4e). These horizontally disseminating regions were enriched in ARGs

relative to the remainder of the microbial genomes ($P = 7.32 \times 10^{-8}$, χ^2 test), indicating their associated role in ARG spread. Nonetheless, only 4.2% of such fragments encoded ARGs (Fig. 4e), suggesting that the role of these horizontally transferred elements is not limited to resistance dissemination. Considering the ARG-encoding fragments, genes targeting tetracyclines, lincosamides and macrolides were the most common resistance elements (Fig. 4f), with these three drug classes also being significantly overrepresented in the cow gut resistome relative to that of humans (Fig. 3c). Of the instances of horizontal transfer of ARGs, 85.8% involve ARGs enriched in the cow gut, suggesting that, within the farm environment, the resistance spread is predominantly driven by cow-associated ARGs. The aforementioned *mef(En2)* and *Inu(AN2)* were among the most common horizontally disseminating ARGs and were frequently transferred together (Fig. 4e), providing further evidence for their genetic linkage and presence within a single, mobilizable cassette (Fig. 4d). As stated earlier, both *mef(En2)* and *Inu(AN2)* are enriched and highly prevalent in the cow gut (relative to the human gut) and have been found in the resistomes of other animals^{54–56}. In the cow gut, the *mef(En2)*–*Inu(AN2)* cassette was found in microbes belonging to four microbial families, most commonly in Lachnospiraceae (57.1% of instances); the host range for this cassette was narrower within the farmer gut, with the gene pair found in three families and predominantly (83.3% of instances) carried by members of Bacteroidaceae (Fig. 4f). Furthermore, as determined through NCBI BLASTn ($\geq 99\%$ identity, $\geq 90\%$ query coverage), the *mef(En2)*–*Inu(AN2)* cassette has also been identified in 42 isolates reported elsewhere (Supplementary Table 8), most (34 of 42) of human origins. Similar to our findings, the reported human isolates carrying this mobilizable cassette are predominantly within the family Bacteroidaceae (33 of 34) and genus *Bacteroides* (30 of 34), including clinical *Bacteroides fragilis* isolates from patient wound and blood samples⁵⁸. These observations may be interpreted to suggest that *Bacteroides* spp. are the primary carriers of the *mef(En2)*–*Inu(AN2)* cassette in the environment and that, upon introduction into the cow gut, this gene pair is enriched through selection, resulting in expansion of its microbial host range⁵⁹. Conversely, it might be suggested that the enrichment of the mobilizable fragment within the cow gut facilitates the broader dissemination of the genes, including acquisition by humans. The dissemination patterns of the *mef(En2)*–*Inu(AN2)* cassette identified via our metagenomic analyses highlight the interconnected nature of human and livestock resistomes, emphasizing the suitability of the One Health approach for better understanding the transmission dynamics of antibiotic resistance⁶⁰.

Conclusion

In one of the largest studies of its kind, we have investigated the effects of exposure to the dairy-farm environment on the microbial communities of dairy farmers. We find that the occupational effects of farming on farmer microbiomes are based on the acquisition of livestock-associated bacterial lineages. This is most evident at the nasal level, with farmers harbouring a significantly more diverse nasal microbiome than non-farmers. Furthermore, in the farmer nasal microbiome, we find an enrichment of microbial taxa that have putative anti-inflammatory activity³⁶ and may thus be implicated in the reported protective effects of farming against allergy and atopy^{7–10}. Relative to the observed differences at the nasal level, the effects of farming on the gut microbiome only emerge at higher taxonomic and functional resolution. We observe an enrichment of taxa associated with protection from diarrhoeal diseases⁴⁰, which is in line with the reported lower rates of gastrointestinal ailments among dairy farmers³⁷. However, we emphasize that mechanistic elucidation of the impacts of identified microbial signatures on farmer health and disease was beyond the scope of this work.

Towards elucidating the occupational effects on the antimicrobial resistome, we used functional metagenomics to identify 2,049

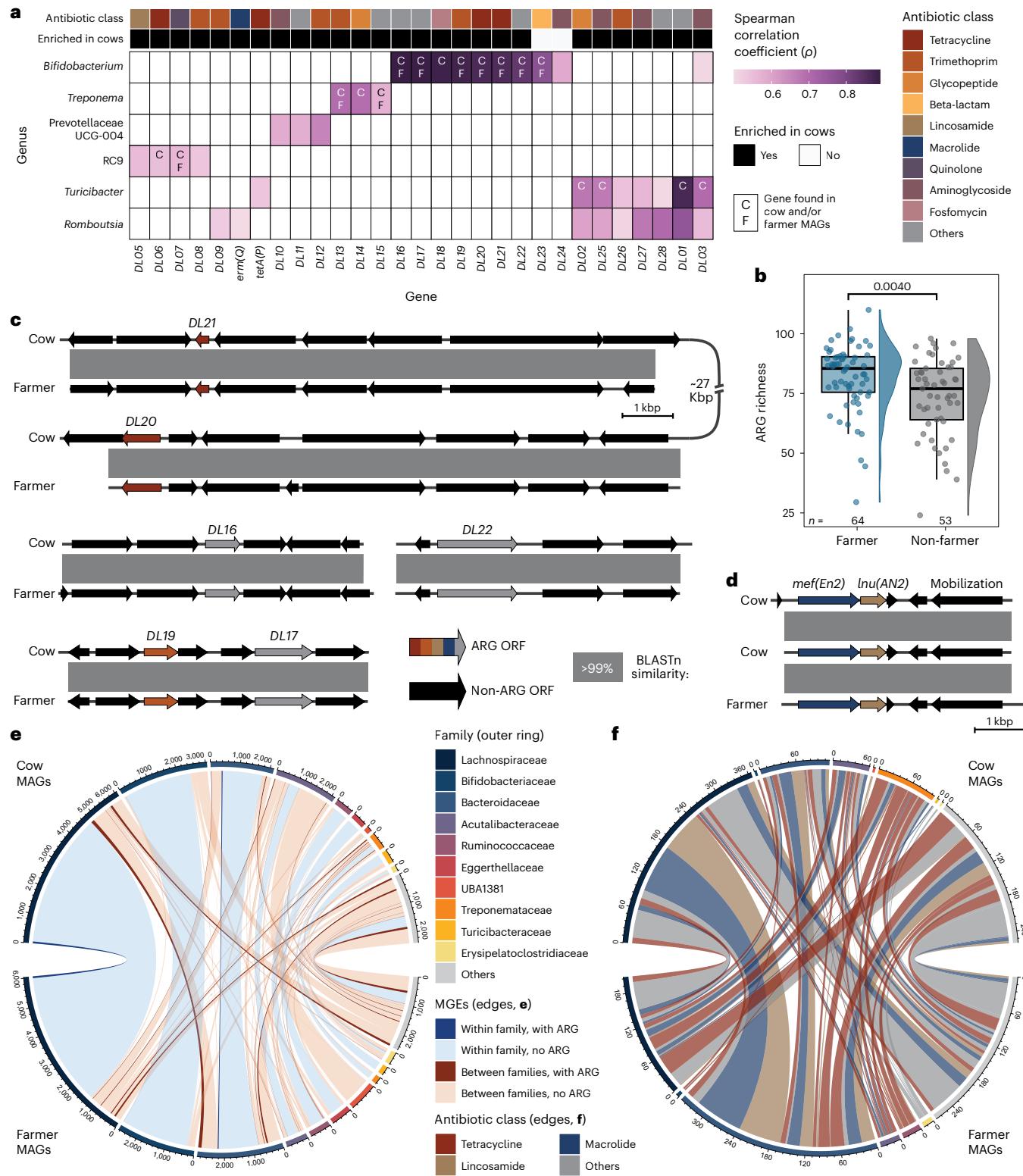


Fig. 4 | Spread of cow-associated ARGs to farmers. **a**, ARGs correlating in relative abundance within the cow gut with the relative abundances of genera of interest (see text), with the corresponding boxes coloured according to the Spearman correlation coefficient. The corresponding antibiotic classes are shown at the top. Enrichment of the ARGs in the cow gut relative to human resistomes is shown as an annotation bar on top. ARGs located in corresponding genera MAGs from cows and farmers are indicated with 'C' and 'F', respectively. **b**, Average richness of ARGs correlated with genera of interest (see text) in farmer and non-farmer guts. The boxplots show the median (centre line), quartiles (box limits) and $1.5 \times$ interquartile range (whiskers). Dots correspond to individual samples. Half-violins show the data distribution. The P -value was calculated

using the two-tailed Wilcoxon rank-sum test. **c,d**, Similarity of contigs from two *Bifidobacterium* (**c**) and three *Prevotella* (**d**) MAGs from cow and farmer faecal samples. ARGs are coloured according to antibiotic class, as in **a,e,f**. Chord diagrams of all (**e**) or ARG-encoding (**f**) horizontally disseminating fragments (Methods). The outer rings correspond to MAG family assignments and are proportional in size to the number of fragment-sharing events involving those families. The edges indicate instances of DNA fragments identified in distant MAGs. In **e**, the edges are coloured by family pairs and ARG-encoding status. In **f**, the edges are coloured by the antibiotic class of the encoded ARG. Note, for fragments encoding ARGs of multiple classes, each unique antibiotic class is represented with a separate edge.

functionally screened ARGs, most of which are novel and underrepresented in the available resistance gene databases. By combining the functionally screened ARGs from this and our previous studies with the existing ARG databases, we generated a large reference set of ARGs, which we provide here for future use by the research community. This combined set allowed us to find evidence for the transfer of cow-enriched ARGs by microbes shared between cows and farmers, and these ARGs were overrepresented in the farmer gut relative to that of non-farmers. Notably, we find examples of these transferred ARGs encoded within mobile genetic elements and identify these cassettes in human clinical isolates reported elsewhere, demonstrating the connection between agriculture and public health. Finally, through our functional screens, we provide definitive evidence of human-origin bias in existing genomic databases, suggesting that the biological relevance of non-human resistomes may have remained obscured by the limitations of available reference sets.

In our profiling of the occupational impacts of livestock farming on human-resident microbiomes, we used age-, sex- and, importantly, ZIP code-matched non-farmers to account for the established role of these factors in shaping human microbial communities. Inclusion of cow samples further allowed us to interrogate which of the observed differences between the two human groups could be attributed to exposures to the farm environment, and which may arise from factors not explicitly measured in our study (for example, diet, medications). As such, how occupational exposures interact with other environmental factors to shape farmer microbiomes remains to be elucidated. Specifically, future studies are warranted to determine whether and how diet and antibiotics facilitate the acquisition, maintenance or amplification of farm- and livestock-associated microbes, as well as their ARGs. Ultimately, such a comprehensive understanding of how human–animal interfaces shape the resident human microbiomes could give rise to novel strategies for mitigating the risks of animal agriculture on farmers and broader public health.

Methods

Design and participants

Details on the overarching design of the base study are described elsewhere³⁷. Briefly, the Dairy Microbiome (DOME) Study used a prospective matched cohort design to examine seasonal microbiome changes in dairy-farm (and non-farm) workers and cows. The source population included residents in and near the Marshfield Epidemiologic Study Area (MESA) in central Wisconsin⁶¹. Participants included dairy-farm workers and a comparison group of non-farm individuals, matched on age, sex and residential ZIP code. All study procedures were approved by the Marshfield Clinic Institutional Review Board (SHU10117) and included written informed consent from all participants. Inclusion criteria for farmers were (1) active/living status in or adjacent to MESA Central, (2) confirmed dairy-farm worker (at least 6 h of daily dairy-farm activities and minimum of 20 h per week), (3) age 18 years and older, (4) assigned a Marshfield Clinic Health System (MCHS) Medical History Number, indicating they have an MCHS account and (5) conversational competence in English or, if Spanish speaking only, comfortable in enrolling with the help of an English- and Spanish-speaking bilingual interpreter. Inclusion criteria for non-farmers were (1) active/living status in or adjacent to MESA Central, (2) age 18 years and older, (3) has not lived or worked in a farm in the past three years, (4) no current occupational exposure to a farm and engagement in any daily farm activities and (5) living at least 0.5 miles away from a dairy farm. For all study participants, the extents of farm exposure were assessed using a questionnaire provided during recruitment and screening.

Recruitment

Potential farm workers were identified after contacting known active farms in and near MESA. All farms were on the register of licensed dairy producers from Wisconsin's Department of Agriculture, Trade

and Consumer Protection. In all, we recruited farmers from 37 dairy farms, ranging in size from 24 to 1,700 milking cows (median = 110, mean = 195.5 ± 270.1). Person-time follow-up, or cohort entry, for individuals began on the day of informed consent. Contact information for farm residences were extracted from MCHS administrative records. Residents of these farm addresses were contacted by letter and telephone to receive the study description/invitation, screen for eligibility criteria, and set up a study enrolment visit with eligible individuals. Study coordinators attempted to reach potential participants over a four-week timeframe, with up to four telephone attempts made after the invitation letter until the individual was reached. Respondents were able to request to opt out of future contacts for this study at any time. Age- and sex-matched non-farm, office-based workers were recruited from the same ZIP code and lived 0.5–1.0 mile away from the matched farmers. The study enrolment visit was scheduled with a trained research coordinator at an agreed-upon time and location. At the enrolment visit, written informed consent was obtained. Participants were incentivized through monetary compensation at the completion of their study enrolment visit. Participants also consented to have their symptoms survey data linked to their stored medical/dental Electronic Health Record data for study analyses.

Sample collection

All human faecal and nasal samples were collected by the study participants. To enable self-collections, all study participants were trained in proper sample-collection techniques by designated research coordinators, emphasizing the importance of aseptic techniques and upholding the integrity of the samples. The study participants received sterile and sealed collection kits, which included collection swabs, collection tubes, gloves and written instructions to ensure technical consistency and sample integrity. Nasal samples from humans were collected using Copan ESwabs (Copan Diagnostics, catalogue no. 480C) by inserting one swab in each anterior nares and making one complete, clockwise turn along the inner nostrils. Nasal secretions and faecal samples from cows were collected either by research coordinators or by farmers, who, as before, were trained in proper sample collection techniques. Specifically, the collection of cow nasal samples involved using an ESwab to do one complete, clockwise turn of the swab two to three inches inside the nares in one of the nostrils. Swabs from the (human and cow) sample collections were inserted back into the Copan ESwab transport tubes. All nasal swabs were collected in duplicate. Faecal samples were collected in collection cups and aliquoted into 5-ml cryovials under aseptic conditions. Immediately upon collection, all samples were kept on ice for up to 12 h and subsequently stored at -80°C until further processing. Our sampling took place between March 2019 and March 2020, spanning five seasons (that is, spring 2019 to spring 2020). However, due to the onset of the COVID-19 pandemic, we halted all sample collection efforts in March 2020, so our analyses of seasonal differences between the subjects excluded the samples collected during spring 2020, given the incomplete sampling within this season. March 2020 samples were included in enrichment analyses, where collection season was controlled for as a random effect, and MAG assembly, which was done at the subject level and was independent of season.

DNA extraction and quantification

Approximately 100 mg of frozen faecal material was collected using sterile spatulas in a biosafety cabinet, while exchanging gloves and cleaning the space with 10% bleach solution between samples. Metagenomic DNA was extracted from the faecal aliquot using the DNeasy Powersoil kit (QIAGEN, catalogue no. 47014). The kit protocol was used with a modification where samples were lysed using a Mini-Beadbeater 24 system (BioSpec Products, catalogue no. 112011) rather than a vortex adapter^{46,62,63}. Metagenomic DNA was extracted from nasal ESwabs (Copan Diagnostics, catalogue no. 480C) by defrosting samples on ice and subsequently processing the 1-ml Liquid Amies

preservation medium with the QiAamp DNA Blood Mini kit (QIAGEN, catalogue no. 51104). We quantified DNA concentration using Qubit fluorometer double-strand DNA (dsDNA) assays (Thermo Fisher Scientific, catalogue no. Q32851).

16S rRNA library preparation sequencing and processing

16S rRNA gene-sequencing libraries were created with 515F-806R barcoded primers that target the V4 region of the 16S rRNA gene⁶⁴. Amplicons were pooled and sequenced on an Illumina MiSeq platform to obtain 2 × 250-bp paired-end reads. Towards determining the minimum sequencing depths, we performed a rarefaction analysis on a representative subset of faecal ($n_{\text{cow}} = 41$, $n_{\text{farmer}} = 42$, $n_{\text{non-farmer}} = 42$) and nasal ($n_{\text{cow}} = 42$, $n_{\text{farmer}} = 42$, $n_{\text{non-farmer}} = 42$) samples, which involved subsampling in steps of 2,000 reads (starting at 2,000 reads) using seqtk (v.0.5.0; <https://github.com/lh3/seqtk>) and calculating genus richness at each subsampling depth. For each sample type, the target sequencing depth was defined as the subsampling step above which no significant increases in richness were identified (Extended Data Fig. 9). Only faecal ($n_{\text{cow}} = 368$, $n_{\text{farmer}} = 156$, $n_{\text{non-farmer}} = 105$) and nasal ($n_{\text{cow}} = 393$, $n_{\text{farmer}} = 163$, $n_{\text{non-farmer}} = 120$) samples sequenced to depths at or above the corresponding thresholds were included in the subsequent analysis. For the included samples, no reads were discarded, even if the corresponding sequencing depths were higher than the calculated rarefaction thresholds. Illumina paired-end reads were demultiplexed by index, and samples were processed using DADA2 (v.1.14.0)⁶⁵, generating a table of amplicon sequence variants (ASVs). Taxonomic classification of ASVs was done using the SILVA database (v.138.1)⁶⁶, and mitochondrial and chloroplast ASVs were manually removed. A set of kit-only, negative control samples ($n = 70$) were similarly sequenced and utilized to remove contaminants using the R package decontam (v.1.6.0, thresh = 0.5)⁶⁷. Finally, we imposed a 0.1% relative abundance threshold on the taxonomic abundance data. Taxonomic richness and Shannon diversity were calculated using the specnumber and diversity (index = 'shannon') functions, respectively, within the R package vegan (v.2.6.4)⁶⁸. Bray–Curtis distances were calculated using the vegdist function (method = 'bray') within vegan as well. The National Parks Palettes R package (v.0.2; <https://github.com/kevinsblake/NatParksPalettes>) was used for data visualization.

Short-read sequencing and processing

Given their relatively low microbial burden, nasal samples are particularly susceptible to contamination with host DNA. Indeed, an average of 95.1% of reads from short-read (that is, shotgun) sequencing of a preliminary subset of nasal samples were of cow or human origins⁶⁹. Given this inefficiency, whole metagenome sequencing was applied only to faecal samples, where the average share of read of host origin was only 0.12%. For each faecal sample, the extracted DNA was diluted to 0.5 ng μl^{-1} and subsequently used as the input for Illumina sequencing library preparation using the Nextera kit (Illumina, catalogue no. 20060060)⁷⁰. The libraries were purified using AMPure XP beads (Beckman Coulter, catalogue no. A63881), pooled, and sequenced on the NovaSeq 6000 platform (Illumina) to obtain 2 × 150-bp paired-end reads. As before the sequencing depths were informed through a rarefaction analysis involving a representative subset of faecal samples ($n_{\text{cow}} = 40$, $n_{\text{farmer}} = 20$, $n_{\text{non-farmer}} = 20$). The samples were subsampled in steps of 500,000 reads (starting at 500,000 reads) using seqtk (v.0.5.0), and ARG richness was calculated for each step. For each subject type, the minimum sequencing depth was defined as the subsampling step above which no significant increases in resistome richness were identified (Extended Data Fig. 10); all 712 faecal samples were sequenced to depths at or above the corresponding thresholds. Raw sequencing files underwent adapter trimming and quality filtering using Trimmomatic (v.0.38; ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:1:true, SLIDINGWINDOW:4:20, LEADING:10, TRAILING:10, MINLEN:60)⁷¹. Reads of human and cow origins were subsequently removed using DeconSeq (v.4.3;

-dbs hsref38,cow)⁷². Finally, unpaired reads were removed using the repair.sh script within BBMap (v.38.82; sourceforge.net/projects/bbmap/), yielding clean metagenomic reads to be used for succeeding analyses. Community taxonomic profiling was accomplished via MetaPhlAn 4³⁸, with 0.1% relative abundance threshold. Taxonomic richness and Shannon diversity were calculated using the specnumber and diversity (index = 'shannon') functions, respectively, within vegan. Bray–Curtis distances were calculated using the vegdist function (method = 'bray') within vegan as well. Assessment of the encoded microbial pathways was done using HUMAN N 3³⁹.

Assembly of MAGs

MAGs were generated at the subject level. To this end, for a given subject (cow or human), sequencing reads from all time points were combined into a single file before assembly. Assembled contigs were generated using MEGAHIT (v.1.1.4; --min-contig-len 1000)⁷³, which were subsequently binned using MaxBin (v.2.2.7; -in_contig_length 1500)⁷⁴, MetaBAT (v.2.11.2; --minContig 1500)⁷⁵ and CONCOCT (v.1.1.0)⁷⁶. We then used DAS Tool (v.1.1.4; --search_engine diamond, --score_threshold 0.1)⁷⁷ to generate an integrated and optimized set of bins for each subject. This resulted in a total of 15,005 MAGs across all subjects, with 993 high-quality (completeness > 90%, contamination < 5%, strain heterogeneity = 0%) and 3,891 medium-quality (completeness > 50%, contamination < 10%,) MAGs, as assessed with CheckM (v.1.1.3)⁷⁸. Taxonomic classification of the MAGs was done using GTDB-Tk (v.1.7.0)⁷⁹. Annotation of open reading frames (ORFs) was done using Bakta (v.1.5.1; --min-contig-length 200)⁸⁰.

Microbial lineage sharing

We used inStrain (v.1.5.7)⁴¹ to determine instances of microbial lineage sharing between cows and humans. To this end, we first dereplicated the assembled MAGs using dRep (v.3.2.2; -sa 0.98, -nc 0.1)⁸¹, resulting in a set of 1,105 MAGs (354 high-quality, 624 medium-quality). These MAGs were then used as representative genomes for aligning sample reads. Due to the large memory requirements of inStrain, lineage tracking was completed in two steps. First, we determined the presence of each MAG in all samples using inStrain Profile (breadth > 50%). Next, for each MAG, we ran inStrain Compare only with samples where the MAG was present. Co-occurrence of microbial lineages in sample pairs was defined as popANI ≥ 99.5% and percent genome compared ≥ 50%.

Functional ARG screening

We generated functional metagenomics libraries using a previously described protocol⁴³, with the following modifications. Extracted metagenomic DNA from the latest stool sample from each unique study subject (cow and human, $n = 283$) was pooled in sets of ~20, yielding 13 pools; importantly, samples were pooled within subject groups (cows, farmers and non-farmers), resulting in eight cow, three farmer and two non-farmer pools. Pooled DNA was diluted to 25–100 ng μl^{-1} in 200 μl Buffer EB and subsequently sheared using the Covaris E220 sonicator set to the following parameters: bath at 19–21 °C, 20% duty cycle, intensity of 0.1, 1,000 cycles per burst, and treatment time of 600 s. Upon shearing, the DNA was purified using the QIAquick PCR purification kit (QIAGEN, catalogue no. 28104) and eluted in 40 μl of Buffer EB. Fragments 2–5 kb in size were then isolated in the BluePippin system (Sage Science) using the 0.75% agarose dye-free gel cassette as well as 0.75% agarose, 1–6 kb, marker S1 (target 1 ph, mA = 0.60). Size-selected fragments were end-repaired using the End-It DNA End-Repair Kit (Biosearch Technologies, catalogue no. ERO720), purified using the QIAquick PCR purification kit (QIAGEN, catalogue no. 28104), and ligated into the pZE21-MCS-1 vector via the Fast-Link DNA Ligation kit (Lucigen, catalogue no. LK0750H) at a 5:1 insert-to-vector ratio. Ahead of ligation, pZE21 was linearized through HindIII digestion, dephosphorylated using rSAP (New England Biolabs, catalogue no. M0371S) and purified through gel extraction. The ligated product was purified by

dialysis with a 0.025- μm cellulose membrane (Millipore Sigma, catalogue no. VSWP02500) and electroporated into the *E. coli* 10-G ELITE cells (Lucigen, catalogue no. 60052-1), in accordance with the manufacturer's protocol. Post-transformation, the cells were recovered in 1 ml recovery medium (Invitrogen, catalogue no. 12648010) with shaking at 37 °C for an hour, after which the transformants were grown in 50 ml of Lysogeny Broth (LB) containing 50 $\mu\text{g ml}^{-1}$ of kanamycin (LB-Kan) until reaching an optical density of 0.6–1.0. The cells were then pelleted and resuspended in LB-Kan containing 15% glycerol for storage at –80 °C. Aliquots of the resulting metagenomic libraries were plated on LB-Kan plates for titration; successful insert ligation was assessed by colony polymerase chain reaction (PCR). The libraries were plated on LB plates with one of the 17 antibiotics listed in Supplementary Table 6 and incubated at 37 °C for 24–48 h. For each antibiotic screen, we also plated *E. coli* 10-G ELITE cells with unmodified pZE21, as the vector-only control, to ensure adequate antibiotic selection pressure. The surviving colonies were grown in LB-Kan, and the resistance-conferring plasmids were extracted using the QIAprep Spin Miniprep kit (QIAGEN, catalogue no. 27104) and sequenced using the NovaSeq 6000 platform, as described above. Reads mapping to *E. coli* or the pZE21 backbone were removed. The remaining reads were assembled into contigs using PAR-FuMS (v.1.1.0)⁴⁴, and the encoded ORFs were determined with Prodigal (v.2.6.3)⁸². The annotation of ORFs took place in a stepwise manner: first, ORFs were matched against BLAST-based ARG databases (CARD⁴⁹, ResFinder⁸³ and AMRFinder-Prot⁸⁴) with high identity (>95%) and coverage (>95%), then the remaining ORFs were annotated via HMM-based ARG databases (Resfams⁸⁵, AMRFinder-fam⁸⁴). This process yielded a total of 2,049 unique ARGs. To determine the similarity of the identified ARGs to those within existing ARG databases, we first used BLASTp within blast-plus (v.2.11.0; -matrix BLOSUM45) to find the best match for each ARG to the joint entries within CARD (v.3.2.2)⁴⁹ and the NCBI antimicrobial resistance (AR) gene catalogue (v.2022-04-04.1)⁵⁰. We then performed a global sequence alignment between each ARG and the corresponding best database hit using the needle algorithm within EMBOSS (v.6.6.0; -gapopen 10, -gapextend 0.5) to obtain the percent sequence identity. The percent identity to the top hits within the NCBI nr database were calculated in an identical manner. Finally, the antibiotic class assignments to the functionally screened ARGs were made based on their top hits to the CARD and NCBI gene catalogue entries.

Resistome profiling

The relative abundances of ARGs in stool metagenomic datasets were calculated using ShortBRED (v.0.9.4)⁸⁶. We first curated a reference ARG set that included the 2,049 functionally screened ARGs from this study, as well as 17,292 ARGs identified through functional metagenomics in our previous studies^{44–46,62,87–92} and the ARGs within CARD (v.3.2.2)⁴⁹ and the NCBI AR gene catalogue (v.2022-04-04.1)⁵⁰, yielding a total of 31,333 reference protein sequences. From this reference set, we generated an ARG marker sequence database using shortbred_identify.py (~clustid 0.95) and the UniRef90 set (downloaded 29 May 2022)⁹³. Through this process, we generated a comprehensive ARG marker database containing 22,637 markers for 9,293 unique ARG families. The functionally screened gene sets and the resulting ARG marker sequence database are available at <https://github.com/dantaslab/DOME/tree/main/ShortBRED>. This marker database was subsequently used to quantify ARG relative abundances using shortbred_quantify.py with default settings. The resistome richness was calculated using the specnumber function within the R package vegan. As before, the Bray–Curtis dissimilarity was calculated using the vegdist function (method = 'bray'). The Pearson correlation coefficients between the relative abundances of ARG pairs in cow samples were calculated using the rcorr command within the R package Hmisc R package (v.4.8.0). The Spearman correlation coefficients between ARGs and genera in the cow stool samples were calculated using the cor function (method = c('spearman')) within the stats package (v.4.1.3).

Assessment of mobile genetic elements

We followed a previously established protocol for the identification of horizontally disseminating genomic regions⁵⁷. We used BLASTn (blast-plus) to align contigs of the cow high- and medium-quality MAGs to those of farmers in an all-versus-all manner. The pairwise average nucleotide identities (ANIs) were calculated for all MAG pairs using FastANI (v.1.33)⁹⁴. Instances of horizontal transfer were defined as ≥500-bp fragments present with >99% identity in MAG pairs with ANI < 95% and/or classified as disparate species (through GTDB-Tk).

Statistical analysis

The subsamples in the rarefaction analyses were compared using Dunn's test through the dunnTest function within the R package FSA (v.0.9.4). Community composition (taxonomic and ARG) differences were determined using PERMANOVA in R with the adonis2 function of vegan. Differences in Shannon diversity, richness (taxonomic and ARG), pairwise Bray–Curtis distances and amino acid identities were calculated using the Wilcoxon rank-sum test with the compare_means function within the R package ggpibr (v.0.6.0). Differentially abundant taxa, microbial pathways, ARGs and antibiotic classes were determined using MaAsLin2⁹⁵, setting groups (that is, farmer versus non-farmer) as the fixed effect, and subject ID, sampling seasons and collection sites as random effects. In all cases, *P* values were corrected for multiple hypotheses using the Benjamini–Hochberg method via the p.adjust command (method = 'BH') in the stats R package (v.4.1.3).

To test the enrichment of lineage co-occurrence events between cow and farmer stool samples, we first calculated the numbers of possible farmer–cow and non-farmer–cow sample pairs. We then distributed the total number of observed lineage-sharing events between farmers and non-farmers using the choices method within the Python random module, setting the calculated total possible sample pairs as weights. We repeated this process 10,000 times, resulting in a theoretical distribution of lineage co-occurrences, allowing us to calculate a z-score for the observed number of lineage co-occurrences using the statistics module in Python. The z-score for lineage-sharing events between farmers and cows from the same farms relative to the pairs from disparate farms was calculated in a similar fashion. The calculated z-scores were used to obtain *P* values using the pnorm function (lower.tail = FALSE) within the stats package. As before, the Benjamini–Hochberg method was used to correct for multiple hypotheses. Finally, to test for the enrichment of ARGs within the horizontally disseminating regions, we ran a χ^2 test using the chi2_contingency command within scipy.stats (v.1.11)⁹⁶.

Reporting summary

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

Data availability

All 16S and shotgun sequencing data pertaining to this study are available from the NCBI SRA under BioProject ID PRJNA964705. The databases used in this study include the SILVA database (v.138.1) (<https://www.arb-silva.de/documentation/release-1381/>), DeconSeq (v.4.3; -dbs hsref38,cow) (<https://deconseq.sourceforge.net/>), shortBRED (v.0.9.4) (<https://github.com/biobakery/biobakery/wiki/shortbred>), CARD (v.3.2.2) (<https://card.mcmaster.ca/download>), NCBI AR gene catalogue (v.2022-04-04.1) (<https://www.ncbi.nlm.nih.gov/pathogens/refgene/#>), UniRef90 (v.2022-05-29) (<https://ftp.uniprot.org/pub/databases/uniprot/uniref/uniref90>) and BLASTn (blast-plus) (<https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html#downloadblastdata>). Source data are provided with this paper.

Code availability

The code for all computational analyses is available at <https://github.com/dantaslab/DOME/tree/main/Scripts>.

References

1. *Moving Towards Sustainability: The Livestock Sector and the World Bank* (The World Bank, 2020); <https://www.worldbank.org/en/topic/agriculture/brief/moving-towards-sustainability-the-livestock-sector-and-the-world-bank>
2. *Farming and Farm Income* (US Department of Agriculture, 2023); <https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/farming-and-farm-income/>
3. Sowiak, M. et al. An assessment of potential exposure to bioaerosols among swine farm workers with particular reference to airborne microorganisms in the respirable fraction under various breeding conditions. *Aerobiologia* **28**, 121–133 (2012).
4. Stein, M. M. et al. Innate immunity and asthma risk in Amish and Hutterite farm children. *N. Engl. J. Med.* **375**, 411–421 (2016).
5. Levy, S. B., FitzGerald, G. B. & Macone, A. B. Spread of antibiotic-resistant plasmids from chicken to chicken and from chicken to man. *Nature* **260**, 40–42 (1976).
6. Angen, O. et al. Transmission of methicillin-resistant *Staphylococcus aureus* to human volunteers visiting a swine farm. *Appl. Environ. Microbiol.* **83**, e01489-17 (2017).
7. Depner, M. et al. Maturation of the gut microbiome during the first year of life contributes to the protective farm effect on childhood asthma. *Nat. Med.* **26**, 1766–1775 (2020).
8. Ege, M. J. et al. Exposure to environmental microorganisms and childhood asthma. *N. Engl. J. Med.* **364**, 701–709 (2011).
9. Illi, S. et al. Protection from childhood asthma and allergy in Alpine farm environments—the GABRIEL advanced studies. *J. Allergy Clin. Immunol.* **129**, 1470–1477 (2012).
10. Steiman, C. A. et al. Patterns of farm exposure are associated with reduced incidence of atopic dermatitis in early life. *J. Allergy Clin. Immunol.* **146**, 1379–1386 (2020).
11. Chen, D. et al. *Campylobacter* colonization, environmental enteric dysfunction, stunting and associated risk factors among young children in rural Ethiopia: a cross-sectional study from the *Campylobacter* Genomics and Environmental Enteric Dysfunction (CAGED) Project. *Front. Public Health* **8**, 615793 (2020).
12. Jenkins, P. L., Earle-Richardson, G., Bell, E. M., May, J. J. & Green, A. Chronic disease risk in central New York dairy farmers: results from a large health survey 1989–1999. *Am. J. Ind. Med.* **47**, 20–26 (2005).
13. Eduard, W., Douwes, J., Omenaas, E. & Heederik, D. Do farming exposures cause or prevent asthma? Results from a study of adult Norwegian farmers. *Thorax* **59**, 381–386 (2004).
14. Carnes, M. U. et al. House dust endotoxin levels are associated with adult asthma in a US farming population. *Ann. Am. Thorac. Soc.* **14**, 324–331 (2017).
15. Omland, O., Hjort, C., Pedersen, O. F., Miller, M. R. & Sigsgaard, T. New-onset asthma and the effect of environment and occupation among farming and nonfarming rural subjects. *J. Allergy Clin. Immunol.* **128**, 761–765 (2011).
16. Radon, K., Schulze, A. & Nowak, D. Inverse association between farm animal contact and respiratory allergies in adulthood: protection, underreporting or selection? *Allergy* **61**, 443–446 (2006).
17. Portengen, L., Preller, L., Tielen, M., Doekes, G. & Heederik, D. Endotoxin exposure and atopic sensitization in adult pig farmers. *J. Allergy Clin. Immunol.* **115**, 797–802 (2005).
18. Smit, L. A. et al. Exposure–response analysis of allergy and respiratory symptoms in endotoxin-exposed adults. *Eur. Respir. J.* **31**, 1241–1248 (2008).
19. Fisher, J. A. et al. Residential proximity to intensive animal agriculture and risk of lymphohematopoietic cancers in the agricultural health study. *Epidemiology* **31**, 478–489 (2020).
20. t'Mannetje, A., Eng, A. & Pearce, N. Farming, growing up on a farm and haematological cancer mortality. *Occup. Environ. Med.* **69**, 126–132 (2012).
21. Van Boeckel, T. P. et al. Global trends in antimicrobial use in food animals. *Proc. Natl. Acad. Sci. USA* **112**, 5649–5654 (2015).
22. Van Boeckel, T. P. et al. Reducing antimicrobial use in food animals. *Science* **357**, 1350–1352 (2017).
23. Van Boeckel, T. P. et al. Global trends in antimicrobial resistance in animals in low- and middle-income countries. *Science* **365**, eaaw1944 (2019).
24. Pitta, D. W. et al. The distribution of microbiomes and resistomes across farm environments in conventional and organic dairy herds in Pennsylvania. *Environ. Microbiome* **15**, 21 (2020).
25. Mulchandani, R., Wang, Y., Gilbert, M. & Van Boeckel, T. P. Global trends in antimicrobial use in food-producing animals: 2020 to 2030. *PLoS Glob. Public Health* **3**, e0001305 (2023).
26. Van Den Broek, I. V. F. et al. Methicillin-resistant *Staphylococcus aureus* in people living and working in pig farms. *Epidemiol. Infect.* **137**, 700–708 (2009).
27. Pirolo, M. et al. Unidirectional animal-to-human transmission of methicillin-resistant *Staphylococcus aureus* ST398 in pig farming; evidence from a surveillance study in southern Italy. *Antimicrob. Resist. Infect. Control* **8**, 187 (2019).
28. Garcia-Graells, C. et al. Dynamic of livestock-associated methicillin-resistant *Staphylococcus aureus* CC398 in pig farm households: a pilot study. *PLoS ONE* **8**, e65512 (2013).
29. Cuny, C. et al. Nasal colonization of humans with methicillin-resistant *Staphylococcus aureus* (MRSA) CC398 with and without exposure to pigs. *PLoS ONE* **4**, e6800 (2009).
30. Larsen, J. et al. Meticillin-resistant *Staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. *Euro Surveill.* <https://doi.org/10.2807/1560-7917.ES.2015.20.37.30021> (2015).
31. Liu, C. M. et al. *Escherichia coli* ST131-H22 as a foodborne uropathogen. *mBio* **9**, e00470-18 (2018).
32. Kuthyar, S. & Reese, A. T. Variation in microbial exposure at the human-animal interface and the implications for microbiome-mediated health outcome. *mSystems* **6**, e0056721 (2021).
33. Sun, J. et al. Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. *Nat. Commun.* **11**, 1427 (2020).
34. Kraemer, J. G., Aebi, S., Hilty, M. & Oppiger, A. Nasal microbiota composition dynamics after occupational change in animal farmers suggest major shifts. *Sci. Total Environ.* **782**, 146842 (2021).
35. Kraemer, J. G., Ramette, A., Aebi, S., Oppiger, A. & Hilty, M. Influence of pig farming on the human nasal microbiota: key role of airborne microbial communities. *Appl. Environ. Microbiol.* **84**, e02470-17 (2018).
36. Abreu, N. A. et al. Sinus microbiome diversity depletion and *Corynebacterium tuberculostearicum* enrichment mediates rhinosinusitis. *Sci. Transl. Med.* **4**, 151ra124 (2012).
37. VanWormer, J. J., Bendixsen, C. G. & Shukla, S. K. Dairy farm work and protection from gastrointestinal illness. *J. Agromed* **28**, 640–646 (2023).
38. Blanco-Miguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
39. Beghini, F. et al. Integrating taxonomic, functional and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088 (2021).
40. Antharam, V. C. et al. Intestinal dysbiosis and depletion of butyrogenic bacteria in *Clostridium difficile* infection and nosocomial diarrhea. *J. Clin. Microbiol.* **51**, 2884–2892 (2013).
41. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736 (2021).

42. Manara, S. et al. Microbial genomes from non-human primate gut metagenomes expand the primate-associated bacterial tree of life with over 1,000 novel species. *Genome Biol.* **20**, 299 (2019).
43. Mahmud, B., Boolchandani, M., Patel, S. & Dantas, G. Functional metagenomics to study antibiotic resistance. *Methods Mol. Biol.* **2601**, 379–401 (2023).
44. Forsberg, K. J. et al. The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
45. Gibson, M. K. et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.* **1**, 16024 (2016).
46. Gasparrini, A. J. et al. Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. *Nat. Microbiol.* **4**, 2285–2297 (2019).
47. Campbell, T. P. et al. The microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. *ISME J.* **14**, 1584–1599 (2020).
48. 2021 Summary Report on Antimicrobials Sold or Distributed for Use in Food-producing Animals (US Food and Drug Administration, 2022); <https://www.fda.gov/media/163739/download>
49. 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).
50. Feldgarden, M. et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response and virulence. *Sci. Rep.* **11**, 12728 (2021).
51. Anthony, W. E. et al. Acute and persistent effects of commonly used antibiotics on the gut microbiome and resistome in healthy adults. *Cell Rep.* **39**, 110649 (2022).
52. de Nies, L. et al. Evolution of the murine gut resistome following broad-spectrum antibiotic treatment. *Nat. Commun.* **13**, 2296 (2022).
53. Johnson, T. A. et al. Clusters of antibiotic resistance genes enriched together stay together in swine agriculture. *mBio* **7**, e02214–e02215 (2016).
54. Xihui, Z. et al. Antibiotic resistance of *Riemerella anatipestifer* and comparative analysis of antibiotic-resistance gene detection methods. *Poult. Sci.* **102**, 102405 (2023).
55. Fenske, G. J., Ghimire, S., Antony, L., Christopher-Hennings, J. & Scaria, J. Integration of culture-dependent and independent methods provides a more coherent picture of the pig gut microbiome. *FEMS Microbiol. Ecol.* **96**, fiaa022 (2020).
56. Cusco, A., Perez, D., Vines, J., Fabregas, N. & Francino, O. Novel canine high-quality metagenome-assembled genomes, prophages and host-associated plasmids provided by long-read metagenomics together with Hi-C proximity ligation. *Micro. Genom.* **8**, 000802 (2022).
57. Brito, I. L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).
58. Wallace, M. J., Jean, S., Wallace, M. A., Burnham, C. D. & Dantas, G. Comparative genomics of *Bacteroides fragilis* group isolates reveals species-dependent resistance mechanisms and validates clinical tools for resistance prediction. *mBio* **13**, e0360321 (2022).
59. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).
60. Hernando-Amado, S., Coque, T. M., Baquero, F. & Martinez, J. L. Defining and combating antibiotic resistance from One Health and Global Health perspectives. *Nat. Microbiol.* **4**, 1432–1442 (2019).
61. Kieke, A. L. et al. Validation of health event capture in the Marshfield Epidemiologic Study Area. *Clin. Med. Res.* **13**, 103–111 (2015).
62. Sukhum, K. V. et al. Manure microbial communities and resistance profiles reconfigure after transition to manure pits and differ from those in fertilized field soil. *mBio* **12**, e00798-21 (2021).
63. D'Souza, A. W. et al. Cotrimoxazole prophylaxis increases resistance gene prevalence and α -diversity but decreases β -diversity in the gut microbiome of Human Immunodeficiency Virus-exposed, uninfected infants. *Clin. Infect. Dis.* **71**, 2858–2868 (2020).
64. Caporaso, J. G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
65. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
66. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
67. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
68. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
69. Lloyd-Price, J. et al. Erratum: Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **551**, 256 (2017).
70. Staff, P. O. Correction: Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE* **10**, e0131262 (2015).
71. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
72. Schmieder, R. & Edwards, R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* **6**, e17288 (2011).
73. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
74. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
75. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
76. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
77. Sieber, C. M. K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).
78. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
79. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
80. Schwengers, O. et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Micro. Genom.* **7**, 000685 (2021).
81. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
82. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
83. Zankari, E. et al. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).

84. Feldgarden, M. et al. Validating the AMRFinder Tool and Resistance Gene Database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* **63**, e00483-19 (2019).
85. Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2015).
86. Kaminski, J. et al. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput. Biol.* **11**, e1004557 (2015).
87. Moore, A. M. et al. Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PLoS ONE* **8**, e78822 (2013).
88. Moore, A. M. et al. Gut resistome development in healthy twin pairs in the first year of life. *Microbiome* **3**, 27 (2015).
89. Pehrsson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).
90. Forsberg, K. J. et al. Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616 (2014).
91. Clemente, J. C. et al. The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183 (2015).
92. Tsukayama, P. et al. Characterization of wild and captive baboon gut microbiota and their antibiotic resistomes. *mSystems* **3**, e00016–e00018 (2018).
93. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
94. Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
95. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* **17**, e1009442 (2021).
96. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

Acknowledgements

We thank all study participants, particularly the dairy farmers and dairy-farm owners, who, in spite of their busy schedules and harvesting season, participated in the study and also allowed their cows to be enrolled in the study. We thank and acknowledge R. Pilsner, M. Presson and N. Esser of Marshfield Agricultural Station for helping with biospecimen collection from dairy farms. We would like to acknowledge the IACUC of the University of Wisconsin-Madison for reviewing our animal protocol and guidance. We also thank the staff at The Edison Family Center for Genome Sciences & Systems Biology at the Washington University School of Medicine in St Louis, including E. Martin and B. Koebbe for computational support, J. Hoisington-López and M. Crosby for managing the high-throughput sequencing core, and B. Dee, K. Matheny, J. Theodore and K. Page for administrative support. Finally, we would like to thank the members of the Dantas laboratory for helpful general discussions and comments on the manuscript. We thank the National Institute for

Occupational Safety and Health of the US Centers for Disease Control and Prevention (grant no. R01OH011578 to G.D. and S.K.S.) for funding support, as well as Marshfield Clinic Research Institute and Weber Endowment Fund (to S.K.S.), the Society for Healthcare Epidemiology of America Research Scholar Award (to K.V.S.), the Initiative for Maximizing Student Development R25 (grant no. GM103757 to R.C.V.) and the Genome Analysis Training Program T32 (grant no. HG000045 to R.C.V.).

Author contributions

S.K.S., G.D., C.G.B. and J.J.V. designed the study and secured the funding. S.K.S., G.D., C.G.B., J.J.V. and K.K. strategized the sample collection method. T.K., T.L. and E.K. performed the logistics of sample collection, processing and management of the sample databases. B.M., R.C.V., K.V.S., S.P. and J.L. processed the samples and prepared the sequencing libraries. B.M., R.C.V., L.R.H. and A.K. conducted the computational analysis. B.M. wrote the initial draft of the manuscript, with subsequent review and editing by R.C.V., L.R.H., C.G.B., J.J.V., S.K.S. and G.D. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41564-024-01639-4>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-024-01639-4>.

Correspondence and requests for materials should be addressed to Sanjay K. Shukla or Gautam Dantas.

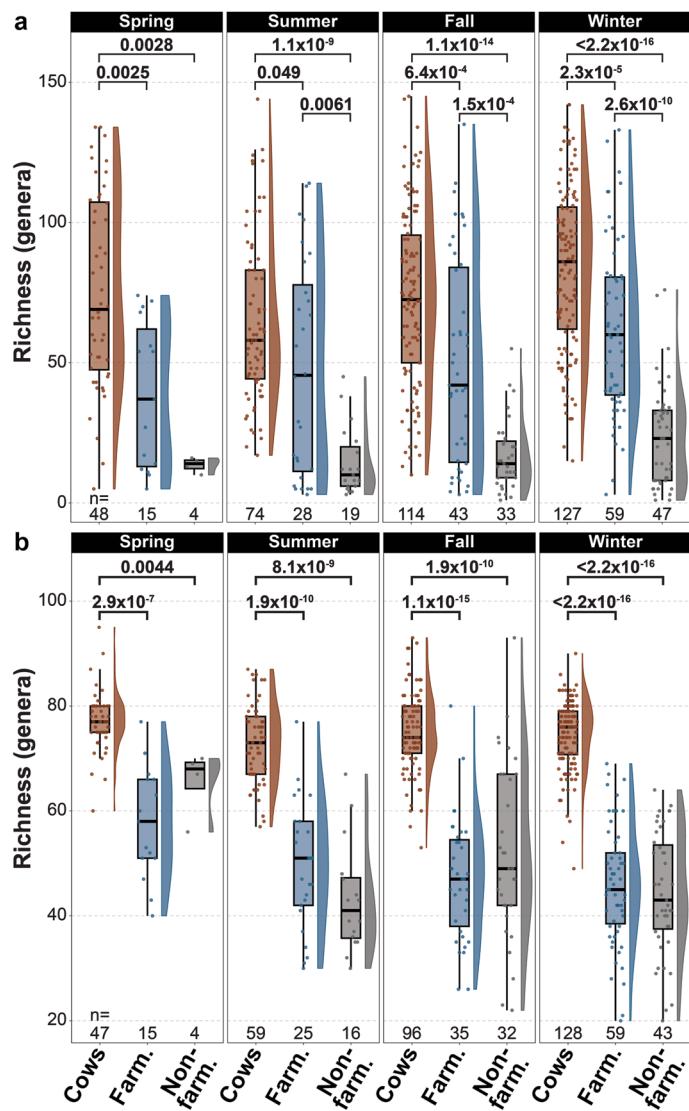
Peer review information *Nature Microbiology* thanks Timothy Walsh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

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

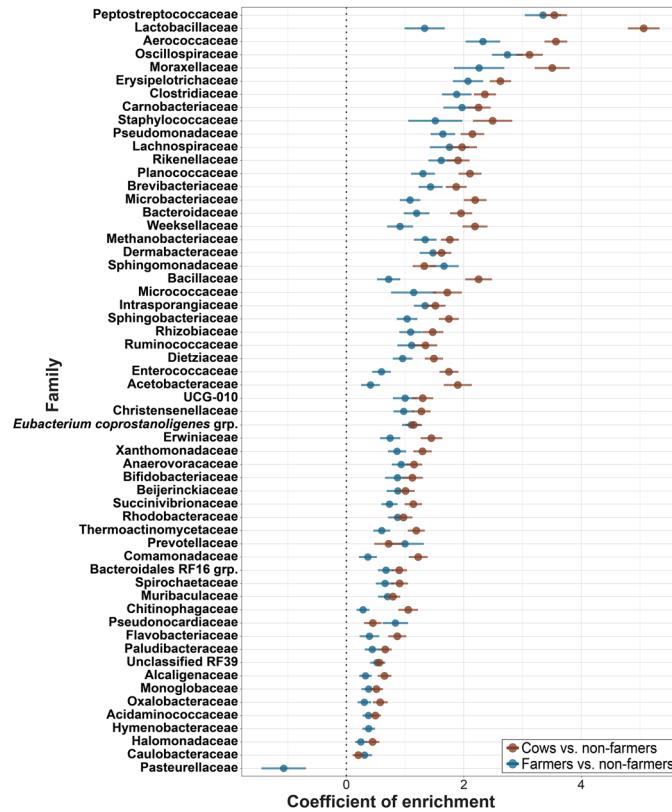
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024


Extended Data Fig. 1 | Genus-level richness of nasal and faecal samples.

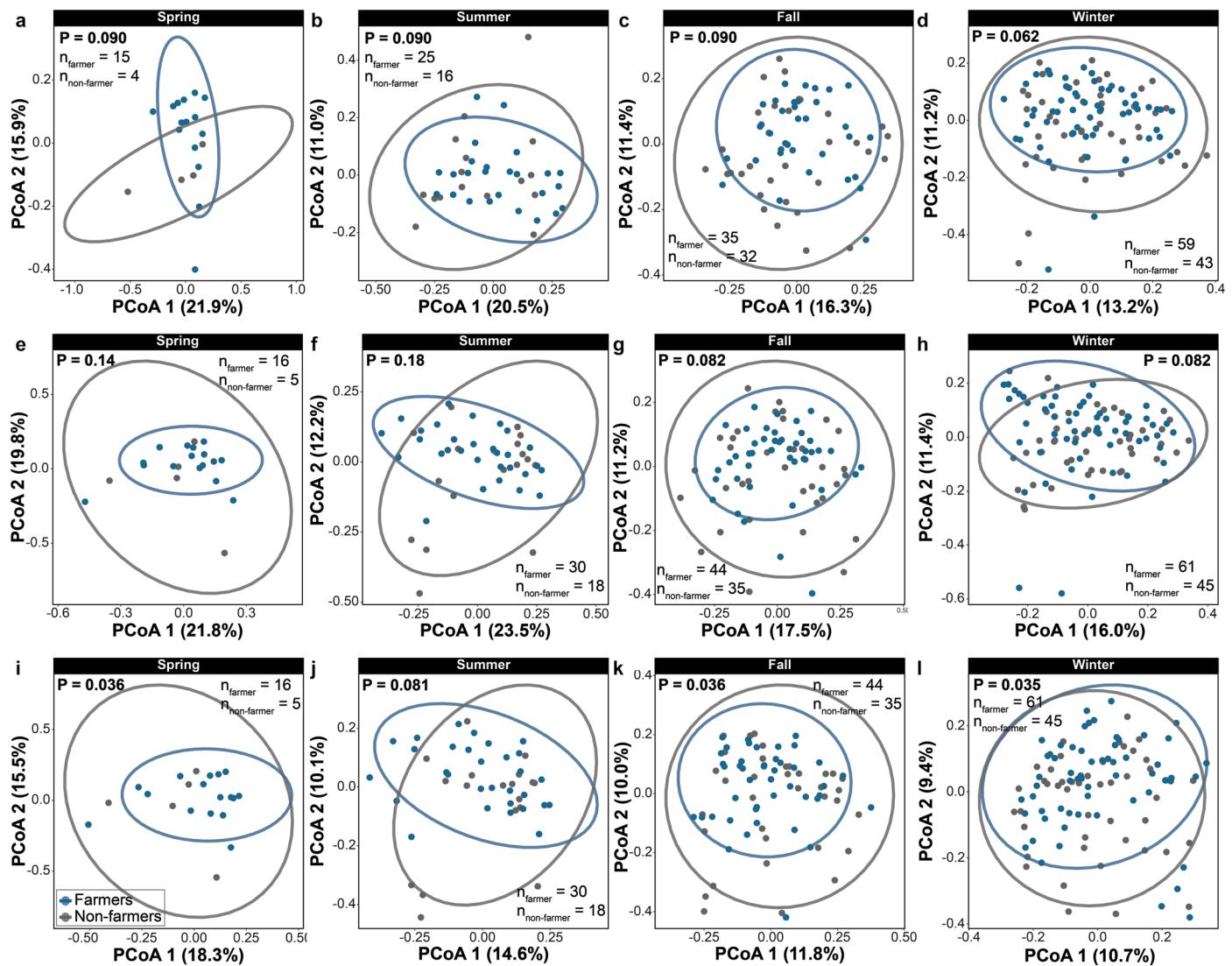
Genus-level richness of the cow, farmer, and non-farmer nasal (**a**) and faecal (**b**) microbiomes across seasons. Boxplots show median (center line), quartiles (box limits), and 1.5x interquartile range (whiskers). Dots correspond to

individual samples. Half-violins show the data distribution. P values were calculated using the two-tailed Wilcoxon rank-sum test, with subsequent Benjamini-Hochberg correction for multiple hypotheses.



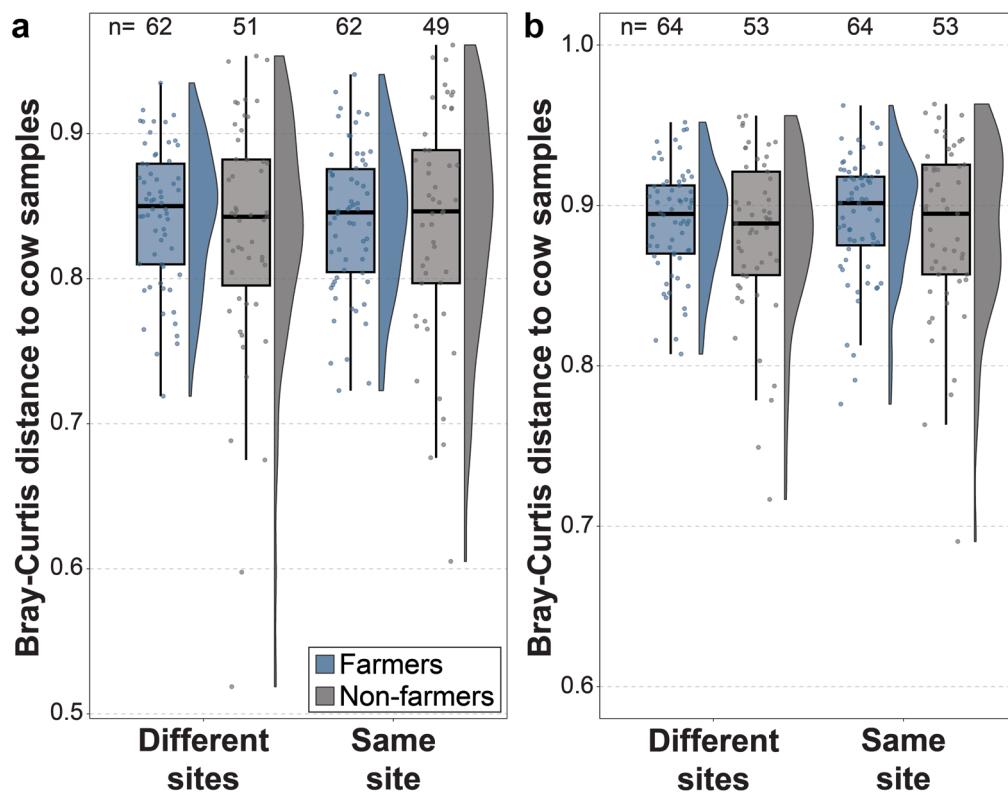
Extended Data Fig. 2 | Microbial families differentially abundant in nasal samples. Microbial families with significant ($P < 0.05$) differential abundances between farmer ($n = 145$) and non-farmer ($n = 103$) nasal microbiomes are indicated. For each family, when significant, the corresponding coefficient in cow

samples ($n = 363$) relative to those of non-farmers is also shown. Points denote mean coefficients; whiskers correspond to standard error. Enrichment tested using MaAsLin2 (see Methods), Benjamini-Hochberg correction for multiple hypotheses.



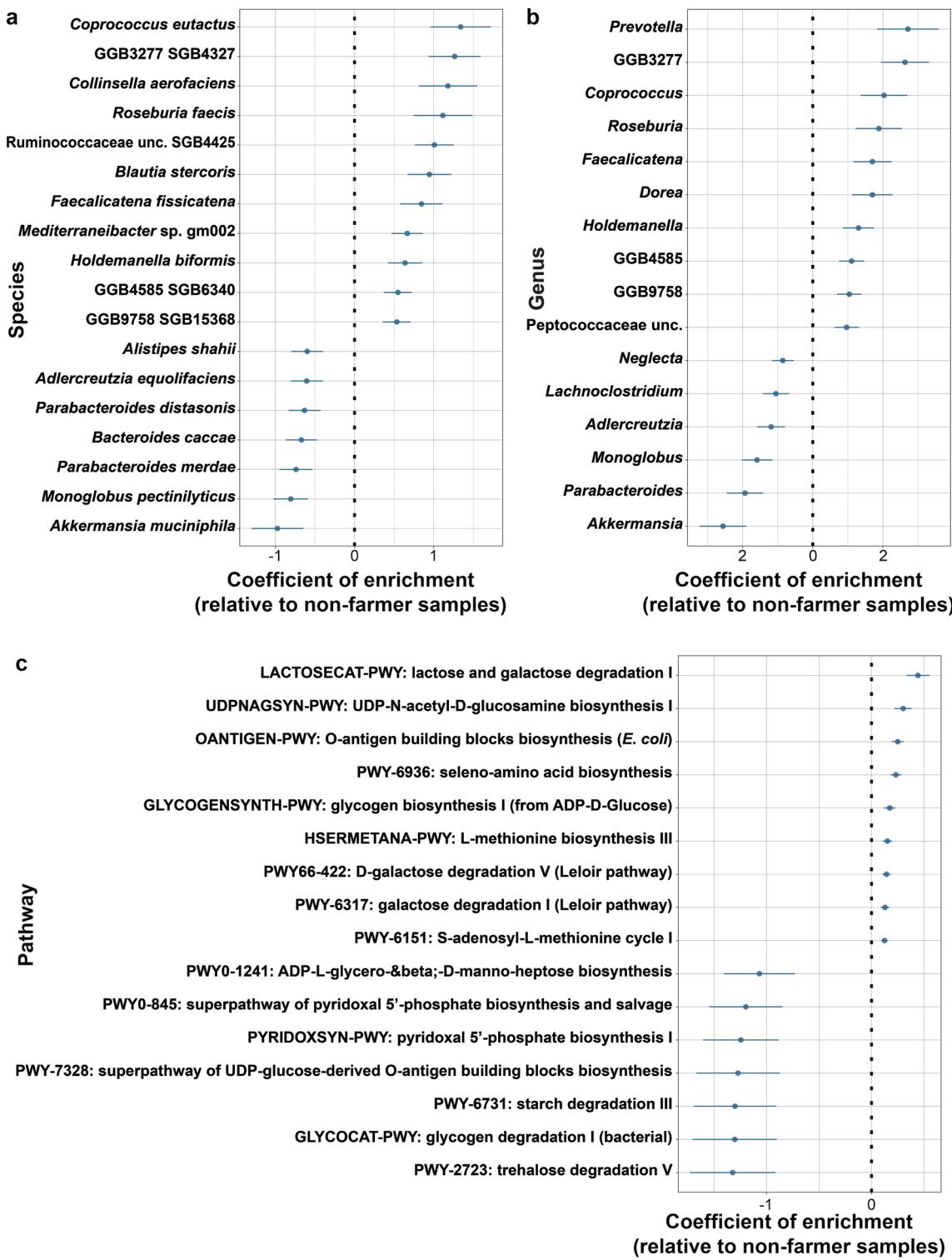
Extended Data Fig. 3 | Principal coordinate analysis of Bray-Curtis dissimilarities. Analysis consists of genus (a–h) and species (i–l) compositions of human fecal samples across seasons. Taxonomic profiling is based on 16S

(a–d) and shotgun metagenomic (e–l) sequencing data. P values were calculated using PERMANOVA and adjusted for multiple hypotheses using the Benjamini-Hochberg method.



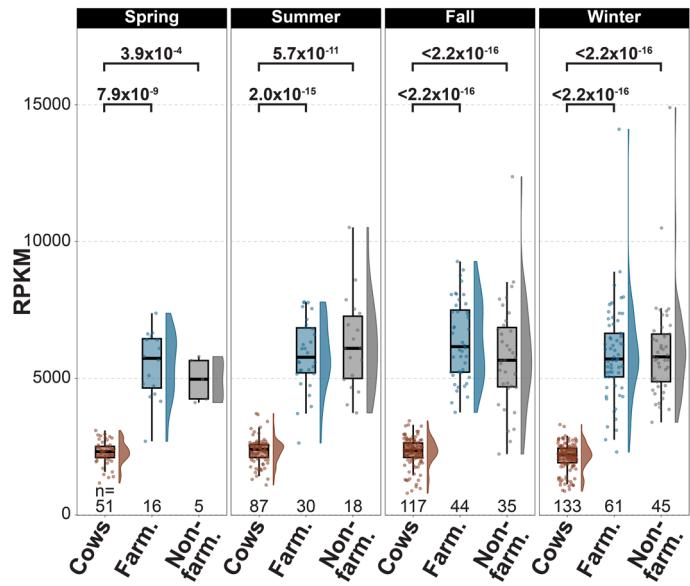
Extended Data Fig. 4 | Average fecal Bray-Curtis distance of farmers and non-farmers to cows residing in the same or different collection site. Beta diversities are based on genus (a) or resistome (b) compositions. Boxplots show median (center line), quartiles (box limits), and 1.5x interquartile range

(whiskers). Dots correspond to individual samples. Half-violins show the data distribution. P values were calculated using the two-tailed Wilcoxon rank-sum test, with subsequent Benjamini-Hochberg correction for multiple hypotheses. No significant differences ($P < 0.05$) were identified.



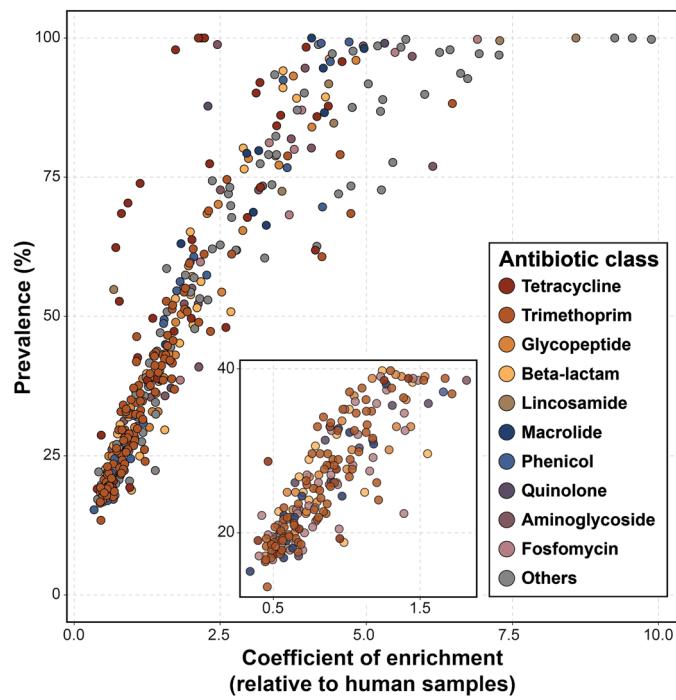
Extended Data Fig. 5 | Differential abundances of microbiome characteristics across farmers and non-farmers. Species (a), genera (b), and microbial pathways (c) with significant ($P < 0.05$) differential abundances between farmer ($n = 134$) and non-farmer ($n = 95$) fecal microbiomes are indicated. Points denote

mean coefficients; whiskers correspond to standard error. Enrichment tested using MaAsLin2 (see Methods), Benjamini-Hochberg correction for multiple hypotheses.

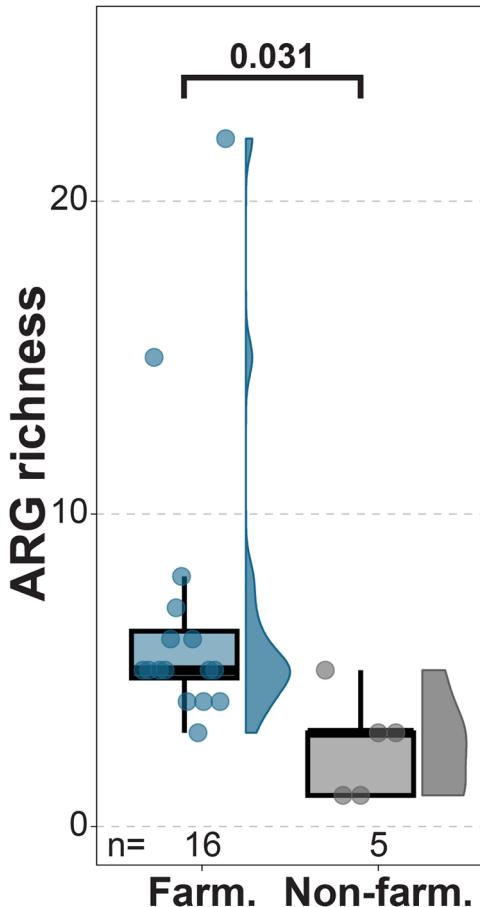


Extended Data Fig. 6 | Total ARG relative abundances of cow, farmer, and non-farmer fecal microbiomes across seasons. Boxplots show median (center line), quartiles (box limits), and 1.5x interquartile range (whiskers).

Dots correspond to individual samples. Half-violins show the data distribution. P values were calculated using the two-tailed Wilcoxon rank-sum test, with subsequent Benjamini-Hochberg correction for multiple hypotheses.

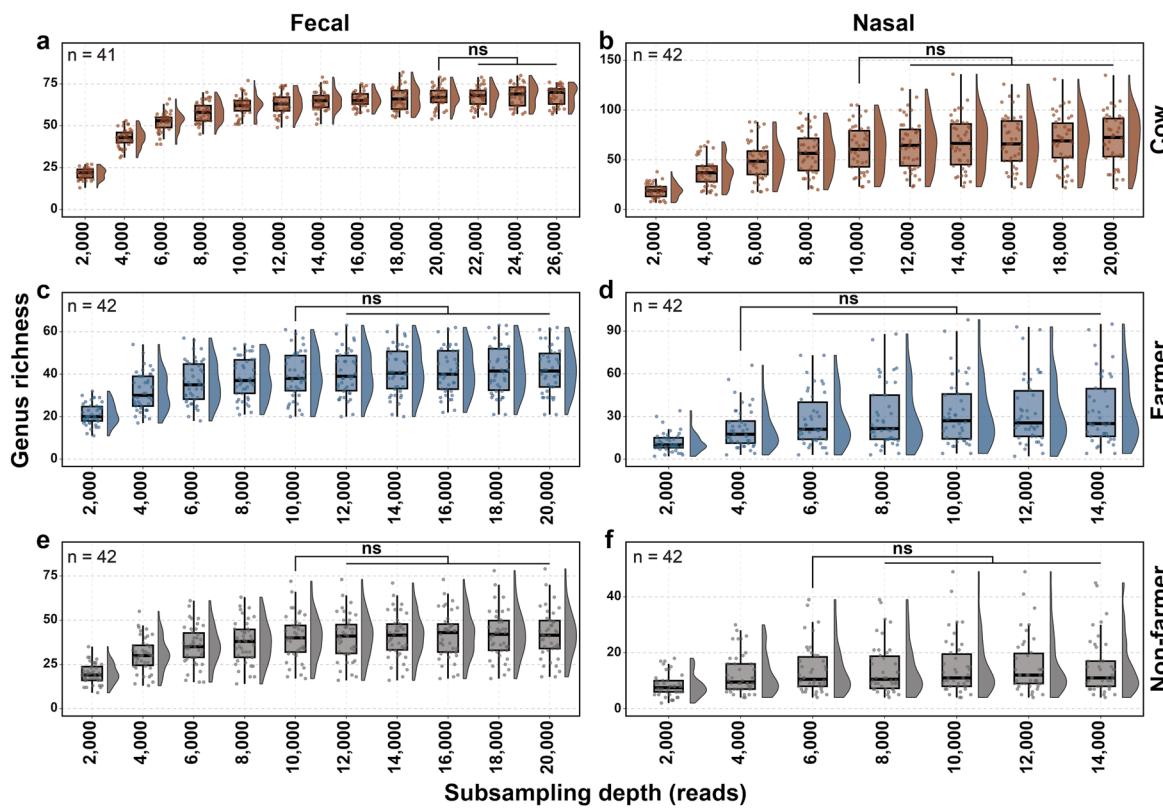


Extended Data Fig. 7 | ARGs overrepresented in the cow gut. ARGs enriched ($P < 0.05$) in the cow gut resistome ($n = 330$) relative to that of humans ($n = 229$). The enrichment coefficients were determined through MaAsLin2 (see Methods), with Benjamini-Hochberg correction for multiple hypotheses. The ARGs are colored according to the corresponding antibiotic classes.



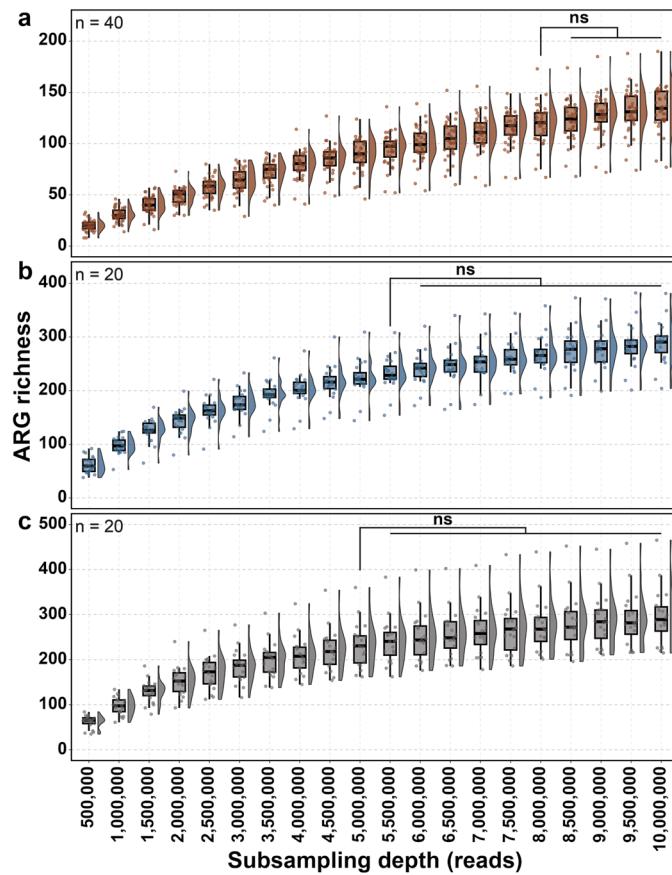
Extended Data Fig. 8 | Farmer and non-farmer gut ARG richness in spring.
The analysis of richness involved only ARGs correlated with genera that i) are enriched in the farmer and cow nasal microbiomes relative to that of non-farmers, and ii) represent the microbial lineages cooccurring the farmer

and cow guts. Boxplots show median (center line), quartiles (box limits), and 1.5x interquartile range (whiskers). Dots correspond to individual samples. Half-violins show the data distribution. The P value was calculated using the two-tailed Wilcoxon rank-sum test.



Extended Data Fig. 9 | Rarefaction analysis for 16S rRNA sequencing. Analysis consisted of fecal (a,c,e) and nasal (b,d,f) samples of cows (a,b), farmers (c,d), and non-farmers (e,f). The analysis was based on genus richness. Boxplots show median (center line), quartiles (box limits), and 1.5x interquartile range

(whiskers). Dots correspond to individual subsamples. Half-violins show the data distribution. The differences among subsamples were tested for significance using Dunn's test, and the P values adjusted for multiple hypotheses using Benjamini-Hochberg. ns, not significant.



Extended Data Fig. 10 | Rarefaction analysis for shotgun sequencing. Analysis consisted of cow (a), farmer (b), and non-farmer (c) fecal samples. The analysis was based on ARG richness. Boxplots show median (center line), quartiles (box limits), and 1.5x interquartile range (whiskers). Dots correspond to

individual subsamples. Half-violins show the data distribution. The differences among subsamples were tested for significance using Dunn's test, and the P values adjusted for multiple hypotheses using Benjamini-Hochberg. ns, not significant.

Rhiannon Vargas, Sanjay Shukla, Gautam Dantas

Corresponding author(s): Last updated by author(s): Feb 5, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

DADA2 (v1.14.0), SILVA database (v138.1), Trimmomatic (v0.38; ILLUMINACLIP:Nexterape-PE.fa:2:30:10:1:true,702 SLIDINGWINDOW:4:20, LEADING:10,.TRAILING:10, MINLEN:60), DeconSeq (v4.3; -dbs hsref38,cow), BBMap (v38.82), MetaPhlAn 4 (v4.0), HUMAnN 3 (v3.0), MEGAHIT (v1.1.4; --min-contig-len 1000), MaxBin (v2.2.7; - min_contig_length 1500), MetaBAT (v2.11.2; --minContig 1500), CONCOCT (v1.1.0) DAS Tool (v1.1.4; --search_engine diamond, --score_threshold 0.1), CheckM (v1.1.3) GTDB-Tk (v1.7.0), Bakta (v1.5.1; --min-contig-length 200), dRep (v3.2.2; -sa 0.98, -nc 0.1), PARFuMS (v1.1.0), Prodigal (v2.6.3), EMBOSS (v6.6.0; -gapopen 10, -gapextend 0.5), ShortBRED (v0.9.4), CARD (v3.2.2), NCBI AR gene catalogue (v2022-04-04.1), UniRef90 (v2022-05-29), FastANI (v1.33), BLASTn (blast-plus)

Data analysis

seqtk (v0.5.0), decontam (v1.6.0, thresh = 0.5), R package vegan (v2.6.4), Hmisc R package (v4.8.0), stats R package (v4.1.3), R package FSA (v0.9.4), R package ggpibr (v0.6.0), R package MaAsLin 2 (v2.0), scipy.stats(v1.11), National Parks Palettes R package (v0.2)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All 16S and shotgun sequencing data pertaining to this study are available from the NCBI SRA under BioProject ID PRJNA964705. The databases used in this study include SILVA database (v138.1) (<https://www.arb-silva.de/documentation/release-1381/>), DeconSeq (v4.3; -dbs hsref38.cow) (<https://deconseq.sourceforge.net/>), shortBRED (v0.9.4) (<https://github.com/biobakery/biobakery/wiki/shortbred>), CARD (v3.2.2) (<https://card.mcmaster.ca/download>), NCBI AR gene catalogue (v2022-04-04.1) (<https://www.ncbi.nlm.nih.gov/pathogens/refgene/#>), UniRef90 (v2022-05-29) (<https://ftp.uniprot.org/pub/databases/uniprot/uniref90>), and BLASTn (blast-plus) (<https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html#downloadblastdata>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

[Information was not used in the analyses.](#)

Reporting on race, ethnicity, or other socially relevant groupings

[Information was not used in the analyses.](#)

Population characteristics

Inclusion criteria for farmers were: 1) active/living status in or adjacent to MESA Central, 2) confirmed dairy farm worker (at least 6 hours of daily dairy farm activities and minimum of 20 hours per week), 3) age 18 years and older, 4) assigned a Marshfield Clinic Health System (MCHS) Medical History Number, indicating they have an MCHS account, and 5) conversational competence in English or, if Spanish speaking only, comfortable in enrolling with the help of an English and Spanish speaking bilingual interpreter. Inclusion criteria for non-farmers were: 1) active/living status in or adjacent to MESA Central, 2) age 18 years and older, 3) has not lived or worked in a farm in the past 3 years, 4) no current occupational exposure to a farm and engagement in any daily farm activities, and 5) living at least 0.5 miles away from a dairy farm. For all study participants, the extents of farm exposure were assessed using a questionnaire provided during recruitment and screening.

Recruitment

Potential farm workers were identified after contacting known active farms in and near Marshfield Epidemiologic Study Area (MESA). All farms were on the register of licensed dairy producers from Wisconsin's Department of Agriculture, Trade and Consumer Protection. In all, we recruited farmers from 37 dairy farms. Person-time follow-up, or cohort entry, for individuals began on the day of informed consent. Contact information for farm residences were extracted from MCHS administrative records. Residents of these farm addresses were contacted by letter and telephone to receive the study description/invitation, screen for eligibility criteria, and set up a study enrollment visit with eligible individuals. Study coordinators attempted to reach potential participants over a 4-week timeframe, with up to four telephone attempts made after the invitation letter until the individual is reached. Respondents were able to request to opt-out of future contacts for this study at any time. Age- and sex-matched non-farm, office-based workers were recruited from the same ZIP code and lived 0.5-1.0 mile away from the matched farmers. The study enrollment visit was scheduled with a trained Research Coordinator at an agreed upon time and location. At the enrollment visit, written informed consent was obtained. Participants were incentivized through monetary compensation at the completion of their study enrollment visit. Participants also consented to have their symptoms survey data linked to their stored medical/dental EHR data for study analyses.

Ethics oversight

All study procedures were approved by the Marshfield Clinic Institutional Review Board (SHU10117) and included written informed consent from all participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

The Dairy Microbiome (DOME) study used a prospective matched cohort design to examine seasonal microbiome changes in dairy farm (and non-farm) workers and cows. The source population included residents in and near the Marshfield Epidemiologic Study

Area in central Wisconsin 58. Participants included dairy farm workers and a comparison group of non-farm individuals, matched on age, sex, and residential ZIP code. All study procedures were approved by the Marshfield Clinic Institutional Review Board and included written informed consent from all participants. Inclusion criteria for participants were: 1) active/living status in or adjacent to MESA Central, 2) confirmed dairy farm worker (at least 6 hours of daily dairy farm activities, 3) age 18 years and older, 4) assigned an Marshfield Clinic Health System (MCHS) Medical History Number, indicating they have an MCHS account, and 5) conversational competence in English or, if Spanish speaking only, comfortable in enrolling with the help of an English and Spanish speaking bilingual interpreter.

Research sample

Nasal samples from humans were collected either by trained and experienced research coordinators or through self-collection using Copan ESwabs (Copan Diagnostics) by inserting one swab in each anterior nares and making one complete, clockwise turn along the inner nostrils. Nasal secretions from cows were collected using an ESwab by doing one complete, clockwise turn of the swab 2-3 inches inside from the nares in one of the nostrils. Swabs from respective (human and cow) sample collection were inserted back into the Copan ESwab transport tube and put into a cooler containing ice. All nasal swabs were collected in duplicate. Fecal samples were collected in collections cups and aliquoted into 5-mL cryovials under sterile conditions. Our sampling took place between 03/2019 and 03/2020, spanning five seasons (i.e., spring 2019 through spring 2020). However, due to the onset of the COVID-19 pandemic, we halted all sample collection efforts in mid-March 2019; as such, our analyses of seasonal differences between the subjects excluded the samples collected during spring 2020, given the incomplete sampling within this season. March 2020 samples were included in enrichment analyses, where collection season was control for as a random effect, and MAG assembly, which was done at the subject level and was independent of season. All human fecal and nasal samples were collected by the study participants. To enable self-collections and reduce self-collection bias, all study participants were trained in proper sample collection techniques by designated research coordinators, emphasizing the importance of aseptic techniques and upholding the integrity of the samples. The study participants received sterile and sealed collection kits, which included collection swabs, collection tubes, gloves, and written instructions to ensure technical consistency and sample integrity.

Sampling strategy

See above.

Data collection

See above.

Timing and spatial scale

See above.

Data exclusions

Due to the onset of the COVID-19 pandemic, we halted all sample collection efforts in mid-March 2019; as such, our analyses of seasonal differences between the subjects excluded the samples collected during spring 2020, given the incomplete sampling within this season. March 2020 samples were included in enrichment analyses, where collection season was control for as a random effect, and MAG assembly, which was done at the subject level and was independent of season.

Reproducibility

Duplicates of each sample collection was saved and stored. Similar study designs in the Marshfield Epidemiologic Study Area has been performed.

Randomization

Allocation was not random however, collection season and collection sites were used as random effects

Blinding

Blinding was not relevant to the study based on the collection methods.

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

Study did not involve laboratory animals.

Wild animals	Study did not involve wild animals.
Reporting on sex	Information was not used in analyses.
Field-collected samples	All farms were on the register of licensed dairy producers from Wisconsin's Department of Agriculture, Trade and Consumer Protection. In all, we recruited farmers from 37 dairy farms, ranging in size from 24 to 1,700 milking cows (median = 110, mean = 195.5 ± 270.1). The study participants received sterile and sealed collection kits, which included collection swabs, collection tubes, gloves, and written instructions to ensure technical consistency and sample integrity. Nasal secretions and fecal samples from cows were collected either by research coordinators or by farmers, who, as before, were trained in proper sample collection techniques. Immediately upon collection, all samples were kept on ice for up to 12 hours and subsequently stored at -80 °C until further processing.
Ethics oversight	All study procedures were approved by the Marshfield Clinic Institutional Review Board (SHU10117) and included written informed consent from all participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>