

Bloominess defines an ecological axis of sporadic dominance that shapes antimicrobial resistance dynamics in the microbiome

Melanie Hay

mchay@rvc.ac.uk

Royal Veterinary College <https://orcid.org/0000-0003-3544-3584>

Bhoomika Joshi

Khushbu Rana

Ankit Hinsu

<https://orcid.org/0000-0002-8985-4541>

Rameshchandra Pandit

Kelyn Seow

Androniki Psifidi

The Royal Veterinary College <https://orcid.org/0000-0003-4095-1452>

Richard Stabler

<https://orcid.org/0000-0002-2402-6630>

Madhvi Joshi

Haidaruliman Paleja

Subhash Jakhesara

Chaitanya Joshi

Gujarat Biotechnology Research Centre

Pham Hoa

Pham Ngoc

Paritosh Biswas

Rukan Hoque

Mohammed Samad

Ludovic Pelligand

Guillaume Fournie

Prakash Koringa

Patricia Conway

Fiona Tomley

Royal Veterinary College <https://orcid.org/0000-0003-2188-8013>

Dong Xia

The Royal Veterinary College <https://orcid.org/0000-0003-4571-2776>

Damer Blake

Royal Veterinary College

Article

Keywords: Antimicrobial resistance, AMR, poultry, One Health, plasmids, metagenome assembled genomes, MAGs, caecal microbiome

Posted Date: July 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10195821/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Bloominess defines an ecological axis of sporadic dominance that shapes antimicrobial resistance dynamics in the microbiome

Authors:

Melanie C. Hay¹, Bhoomika Joshi², Khushbu Rana², Ankit Hinsu¹, Ramesh Pandit³, Kelyn Seow⁵, Androniki Psifidi¹, Richard Stabler⁴, Madhvi Joshi³, Haidaruliman Paleja², Subhash Jakhesara², Chaitanya Joshi³, Hoa Pham⁶, Pham Thi Ngoc⁷, Paritosh Kumar Biswas⁸, Rukan Hoque⁸, Mohammed Abdus Samad⁹, Guillaume Fournie¹⁰, Ludovic Pelligand¹, Prakash Koringa², Patricia Conway⁵, Fiona Tomley¹, Dong Xia¹, **Damer P. Blake**¹

First author: Melanie C. Hay

Corresponding author: Melanie C. Hay and Damer Blake

Institutions:

¹ Royal Veterinary College, Hatfield, United Kingdom,

² Kamdhenu University, Gujarat, India,

³ Gujarat Biotechnology Research Centre (GBRC), Gujarat, India,

⁴ London School of Hygiene & Tropical Medicine, London, United Kingdom,

⁵ Nanyang Technological University, Singapore, Singapore,

⁶ CIRAD, the French Agricultural Research Centre for International Development, Hanoi, Vietnam,

⁷ National Institute of Veterinary Research, Ha Noi, Vietnam,

⁸ Chittagong Veterinary and Animal Sciences University, Chittagong, Bangladesh,

⁹ Bangladesh Livestock Research Institute, Dhaka, Bangladesh

¹⁰ INRAE, the National Research Institute for Agriculture, Food and the Environment, Lyon, France

Keywords: Antimicrobial resistance, AMR, poultry, One Health, plasmids, metagenome assembled genomes, MAGs, caecal microbiome

Data availability Raw sequencing data are deposited in the European Nucleotide Archive (ENA) under Study accession PRJEB101827.

Code availability Custom R code used for all analyses described in this study is available at <https://github.com/MelanieCHay/Bloominess>. The repository is currently private and will be made publicly available upon acceptance of this manuscript.

Ethics Ethical approval for sample collection was obtained from the National Institute of Veterinary Research, Vietnam (reference 020-433/DD-YTCC), the Chattogram Veterinary and Animal Sciences University, Bangladesh (reference EC/2020/165/2/1), and the Royal Veterinary College Ethics and Welfare Committee, United Kingdom (reference URN:20201983-3, covering sample collection in Gujarat, India).

Funding This work was supported by the UKRI GCRF One Health Poultry Hub, funded through the Global Challenges Research Fund (GCRF) from UK Research and Innovation (UKRI) [BB/S011269/1].

Author contributions (CRediT)

Melanie Hay: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualisation, Writing – original draft, Writing – review & editing. Bhoomika Joshi: Data curation, Investigation, Project administration, Resources, Writing – review & editing. Khushbu Rana: Data curation, Investigation, Project administration, Resources, Writing – review & editing. Ankit Hinsu: Data curation, Investigation, Writing – review & editing. Ramesh Pandit: Investigation, Writing – review & editing. Kelyn Seow: Data curation, Investigation, Software, Writing – review & editing. Androniki Psifidi: Investigation, Project administration, Resources, Writing – review & editing. Richard Stabler: Funding acquisition, Investigation, Project administration, Resources, Writing – review & editing. Madhvi Joshi: Investigation, Project administration, Resources, Writing – review & editing. Haidaruliman Paleja: Data curation, Investigation, Project administration, Resources, Writing – review & editing. Subhash Jakhesara: Data curation, Investigation, Project administration, Resources, Writing – review & editing. Chaitanya Joshi: Funding acquisition, Investigation, Project administration, Writing – review & editing. Hoa Pham: Investigation, Project administration, Resources, Writing – review & editing. Pham Thi Ngoc: Investigation, Project administration, Resources, Writing – review & editing. Paritosh Kumar Biswas: Investigation, Project administration, Resources, Writing – review & editing. Rukan Hoque: Investigation, Project administration, Resources, Writing – review & editing. Mohammed Abdus Samad: Investigation, Project administration, Resources, Writing – review & editing. Guillaume Fournie: Investigation, Supervision, Writing – review & editing. Ludovic Pelligand: Investigation, Supervision, Writing – review & editing. Prakash Koringa: Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. Patricia Conway: Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. Fiona Tomley: Conceptualisation, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. Dong Xia: Conceptualisation, Formal analysis, Investigation, Software, Supervision, Writing – review & editing. Damer P. Blake: Conceptualisation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing.

73 ABSTRACT

74 Microbiome studies commonly characterise taxa by their prevalence across samples and their
75 average abundance within samples. However, these descriptors do not capture how
76 abundance is distributed among hosts, and therefore cannot distinguish taxa that maintain
77 stable abundance from those that are typically rare but occasionally undergo extreme
78 expansion. Such sporadic dominance events may have disproportionate functional
79 consequences, including transient elevation of ARG abundance.

80 Here, we introduce *bloominess*, a metric that quantifies sporadic dominance by measuring the
81 divergence between mean and median abundance across samples in which a taxon is present.
82 Using a genome-resolved atlas of chicken caecal microbiomes reconstructed from shotgun
83 metagenomic sequencing, we applied this framework to thousands of metagenome-assembled
84 genomes (MAGs).

85 Bloominess revealed a structured ecological landscape in which sporadic dominance was
86 constrained at both low and high prevalence, but varied widely among taxa of intermediate
87 prevalence. Taxonomic groups differed markedly in bloominess distributions, with several
88 clinically relevant taxa exhibiting strong sporadic dominance behaviour. When integrated
89 with antimicrobial resistance annotations, bloominess distinguished stable background
90 reservoirs of resistance from taxa capable of transiently elevating resistance signals.

91 Together, these results establish bloominess as a complementary ecological axis for
92 interpreting microbiome structure, and show that antimicrobial resistance risk reflects both
93 gene content and the potential for sporadic ARG elevation.

1. INTRODUCTION

1.1 Sporadic dominance as an ecological phenomenon

Microbial communities are characterised by long-tailed abundance distributions in which most taxa are rare and a small number dominate at any given time ¹. This macroecological pattern is consistent across environments from soils to animal-associated microbiomes and reflects a fundamental asymmetry in ecological impact. Taxa are not equivalent in their abundance or their contribution to community function. Importantly, this asymmetry is not static. Ecological theory predicts, and empirical observations confirm, that taxa can shift dramatically in relative abundance across hosts or timepoints. A taxon that is typically rare may occasionally undergo extreme expansion and become a transient dominant ².

We use the term *sporadic dominance* to describe this population-level pattern: the observation that certain taxa, while typically at low or background abundance across most hosts, reach substantially elevated abundance in a subset of individuals within a cohort. Critically, this is a cross-sectional definition, derived from a single metagenomic snapshot per host. It captures variability in abundance across hosts rather than temporal dynamics within individual birds. Cross-sectional data cannot distinguish between a taxon that undergoes genuine temporal bloom events within certain hosts and one whose abundance differs stably between individuals for host-specific reasons; both scenarios produce equivalent right-skewed distributional signatures. Sporadic dominance is thus a property of a taxon's across-host abundance distribution, quantified here without any assumption about within-host temporal trajectories. It is also worth distinguishing sporadic dominance from the related concept of extreme dominance, which describes the magnitude of peak abundances relative to a taxon's typical level. A taxon may be sporadically dominant without reaching extreme absolute abundances, while a consistently high-abundance core taxon may be considered extremely dominant without being sporadic. Bloominess captures both dimensions simultaneously: taxa that are both sporadically present at high abundance and reach extreme levels when present exhibit the largest divergence between mean and median.

Such abundance fluctuations are not random noise. Microbial communities follow predictable macroecological relationships in which variance scales with mean abundance ³, suggesting that the propensity for sporadic dominance is a biologically meaningful and potentially characteristic property of individual taxa.

When a taxon undergoes sporadic dominance, its functional contribution to the community is elevated far beyond what its average abundance would predict. This matters wherever microbial function is dose-dependent, including nutrient cycling, pathogen dynamics, and the local concentration of antimicrobial resistance genes (ARGs). A taxon that rarely blooms but carries resistance genes is ecologically distinct from one that carries the same genes as a stable, low-abundance resident. Despite this, no quantitative framework exists for formalising sporadic dominance as a taxon-level trait.

1.2 The limits of current descriptors

The most widely used dimensions for describing microbial community members are prevalence and mean abundance. Prevalence, defined as the proportion of samples in which a taxon is detected, captures ecological persistence. High-prevalence taxa represent stable community members repeatedly supported across hosts, whereas low-prevalence taxa tend to occupy more transient or conditional niches^{4,5}. Both metrics are ecologically informative, but neither resolves how abundance is distributed among the samples in which a taxon is present.

To illustrate, consider two taxa each detected in 60% of samples. Taxon A is consistently detected at low abundance, with a narrow distribution centred on a stable median. Taxon B is also usually detected at low abundance, but in a subset of samples reaches orders of magnitude higher abundance. Despite identical prevalence, and potentially similar cohort-level mean abundance, their ecological roles are distinct. Taxon A represents stable background presence, whereas Taxon B exhibits context-dependent elevation in abundance. Standard abundance summaries collapse this distinction, treating stable background presence and rare dominance events as equivalent⁶.

This analytical limitation parallels early RNA-seq analysis, where explicit modelling of biological dispersion transformed interpretation of gene-expression data⁷. Likewise, microbiome studies may benefit from complementary descriptors that quantify distributional shape, rather than relying solely on prevalence and average abundance.

Large-scale, genome-resolved MAG atlases are powerful tools for identifying what taxa are present and how prevalent they are^{8,9}. However, they do not resolve distributional shape. What is missing is a dimension that distinguishes taxa with stable abundance profiles from those characterised by right-skewed distributions driven by rare but extreme peak-abundance events.

1.3 The chicken gut as a model system

Chicken caecal microbiomes represent a tractable and consequential system for investigating these questions. They are genome-resolvable at scale, shaped by the same ecological forces that structure gut communities more broadly, and directly relevant to global One Health concerns. Poultry production is a major reservoir and transmission interface for antimicrobial resistance, particularly in regions where dense human-animal-environmental interactions facilitate resistance emergence and dissemination^{10,11}. Resistome composition in these communities is shaped by both microbial ecology and antimicrobial use practices¹².

Genome-resolved studies have established a rich catalogue of the chicken gut microbiome and its resistance gene content^{13–17}. Notably, Feng et al. (2021) assembled 12,339 metagenome-assembled genomes from 799 samples across ten countries, constructing a gene catalogue that revealed geographic structuring of the chicken gut resistome and highlighted the role of mobile genetic elements in ARG dissemination¹⁵. This work established which ARG-carrying taxa are present, but did not address how those taxa behave ecologically.

Whether a taxon carrying resistance genes is a stable community resident or a sporadic bloomer has direct implications for how resistance risk should be interpreted. A transiently dominant taxon may generate brief but intense spikes in local ARG concentration that are not captured by mean abundance-based surveillance.

1.4 Introducing bloominess and study aims

Here, we introduce bloominess, a metric that formalises sporadic dominance as a quantitative taxon-level trait. Although the mean-to-median ratio is a well-established measure of distributional asymmetry, its application as a taxon-level ecological trait, combined with prevalence to define a structured two-dimensional community space, has not, to our knowledge, been formalised in this context; the novelty of bloominess lies in its ecological framing and application rather than in the formula itself. Bloominess quantifies the divergence between mean and median abundance across samples in which a taxon is present, capturing the skew of the within-cohort abundance distribution. By combining bloominess with prevalence, we define a two-dimensional ecological space in which taxa can be positioned according to both their persistence and their abundance behaviour when present.

Using a genome-resolved atlas of chicken caecal microbiomes, we (i) characterise the prevalence landscape of the MAG catalogue, (ii) quantify bloominess across taxa, and (iii) examine how antimicrobial resistance gene burden is distributed across prevalence–bloominess space. Together, these analyses establish bloominess as a complementary ecological axis for interpreting microbiome structure and resistance dynamics, and demonstrate its value for distinguishing stable background reservoirs of resistance from taxa capable of transiently elevating resistance signals.

2. METHODS

2.1 Study design and sample collection

This study analysed genome-resolved metagenomic data generated as part of the One Health Poultry Hub (<https://www.onehealthpoultry.org/>), a large collaborative project investigating the structure and function of poultry gut microbiomes across diverse production systems in India, Bangladesh and Vietnam. The present analysis focuses on the construction and ecological interpretation of a metagenome-assembled genome (MAG) catalogue derived from chicken caecal microbiomes, rather than on sample-level comparisons.

Caecal content samples were collected from poultry across multiple production contexts and processed using standardised protocols within the wider project framework. Detailed descriptions of cohort design, sampling strategies, animal metadata, and ethical approvals have been reported previously^{18–20}. In this manuscript, samples are treated as independent observations contributing to a shared genomic catalogue, enabling the characterisation of microbial taxa and their ecological properties across hosts. Four samples were excluded from ecological analyses prior to prevalence and abundance estimation, due to low read-mapping recovery (**Supplementary Figure S1**). A high number of *Eimeria* spp. in several excluded

samples suggests that a high proportion of apicomplexan reads in these samples confounded microbial abundance estimates and rendered them unsuitable for inclusion in downstream analyses. All prevalence, bloominess, and ARG analyses were therefore conducted across 116 samples.

2.2 DNA extraction and sequencing

Metagenomic DNA was extracted from the caecal content using the Qiagen QIAamp Fast DNA Stool Mini Kit following manufacturer's protocol^{21,22}. Extracted DNA was quantified using Qubit and checked on Agarose gel for quality.

2.2.1 Illumina shotgun metagenomic sequencing

Illumina short-read shotgun metagenomic sequencing was performed on 120 chicken caecal DNA samples at Nanyang Technological University (Singapore). Libraries were sequenced using paired-end 2 × 150 bp chemistry across eight sequencing runs. Each biological sample was distributed across multiple flow cells, and raw sequencing reads were merged across runs to generate a single Illumina dataset per sample. Sequencing accession numbers and run metadata are provided in **Supplementary Table S1**.

2.2.2 Oxford Nanopore long-read sequencing

Long-read sequencing was performed on a subset of 56 samples at the Royal Veterinary College (UK) using Oxford Nanopore Technologies platforms. Two sequencing batches were generated using different chemistries. A pilot batch of 15 samples was prepared using Kit 12 (SQK-NBD112.24) chemistry and sequenced on FLO-MIN112 flowcells using an Oxford Nanopore Mk1c platform, while a second batch of 41 samples was prepared using Kit 14 chemistry (FLO-PRO114M) and sequenced on FLO-PRO114M flowcells on a PromethION P2 Solo. Individual samples were sequenced across multiple flow cells; reads were merged by barcode within and across runs to obtain a single long-read dataset per sample. Details of sequencing batches, chemistries, and flow cells are provided in **Supplementary Table S2**.

2.3 Read processing and quality control

2.3.1 Illumina read processing

Illumina sequencing reads were received as quality-filtered FASTQ files from the sequencing provider. Adapter trimming and initial quality filtering were performed prior to downstream analysis. Summary quality metrics were assessed using standard tools (e.g. FastQC), and reads were retained for assembly without further aggressive filtering to preserve low-abundance signal.

2.3.2 Nanopore read processing

Nanopore reads were filtered based on read quality using Chopper (v0.9.1)²³. Given differences in sequencing chemistry and raw read quality between batches, quality filtering was performed separately for each dataset. For Kit 12 pilot samples, a lower mean Phred

score threshold (≥ 12) was applied to retain sufficient read depth, while a stricter threshold (≥ 15) was used for Kit 14 samples sequenced on the PromethION platform. The impact of chemistry-specific filtering thresholds on read retention is summarised in **Supplementary Figure S2**.

To address duplicated read identifiers arising from multi-run sequencing, read names were checked and standardised using a custom deduplication script (**described in Supplementary Methods SM1**). Filtered and standardised Nanopore reads were used for downstream hybrid assembly, scaffolding, and polishing steps.

2.4. Metagenomic assembly strategies

To capture complementary layers of genomic diversity within the chicken caecal microbiome, we employed multiple assembly strategies designed to balance completeness, strain resolution, and recovery of low-abundance taxa. These strategies were applied in parallel and subsequently integrated to construct a comprehensive MAG catalogue. Detailed assembly methods and parameters are provided in **Supplementary Methods SM2**.

2.4.1 Illumina short-read co-assembly

Illumina short-read data from 120 samples were combined and co-assembled using MEGAHIT²⁴. A multi-k-mer strategy was employed to optimise assembly across a range of coverage depths, prioritising recovery of shared genomic content while maintaining sensitivity to moderately abundant taxa. The resulting co-assembly comprised tens of millions of contigs and provided a common reference for downstream binning and comparative analyses.

To improve recovery of low-prevalence and low-coverage genomic fragments, a secondary assembly step was performed. Illumina reads that did not map to the initial co-assembly MAG collection were extracted and re-assembled using more sensitive MEGAHIT presets and smaller k-mer values. This step aimed to rescue genomic content that may be systematically under-represented in large co-assemblies, particularly from rare or conditionally present taxa.

2.4.2 Single-sample hybrid assemblies

In parallel, hybrid assemblies were generated for individual samples for which both Illumina and Nanopore sequencing data were available ($n=56$). Hybrid assembly was performed using hybridSPAdes²⁵, integrating short-read accuracy with long-read continuity. These assemblies prioritised the recovery of strain-specific variation and accessory genome content that may be diluted or collapsed in co-assembly approaches.

Hybrid assemblies were particularly informative for recombination-prone and panmictic species of public health relevance, such as *Escherichia coli* and *Campylobacter* spp., where strain-level variation and chromosomally integrated antimicrobial resistance genes are of

interest. Hybrid assemblies were treated as complementary to co-assembled MAGs rather than as replacements, contributing additional genomic diversity to the overall catalogue.

2.4.3 Rationale for complementary strategies

Co-assemblies emphasise shared genomic content and facilitate cross-sample comparisons, while single-sample hybrid assemblies preserve fine-scale genomic variation and accessory elements. By combining these strategies, we aimed to maximise recovery of both stable, high-prevalence taxa and variable or strain-specific genomic features, providing a robust foundation for downstream ecological and functional analyses.

2.5. Binning, dereplication, and MAG curation

2.5.1 Genome binning and ensemble integration

Assembled contigs were grouped into putative genomes using an ensemble binning strategy that combined multiple binning algorithms with complementary strengths. Contigs from the Illumina co-assemblies were binned independently using MetaBAT2²⁶, VAMB²⁷ and SemiBin2²⁸. These tools span graph-based, deep learning, and semi-supervised binning paradigms, respectively, enabling robust recovery of genomes across a range of coverage profiles and community complexities.

Outputs from all binning approaches were integrated using DAS Tool²⁹, which scores and selects a non-redundant set of high-confidence bins based on marker gene completeness and contamination. This ensemble approach mitigates method-specific biases and improves overall genome recovery relative to any single binning strategy (**Supplementary Methods S3**).

2.5.2 Dereplication and quality filtering

To reduce redundancy and select representative genomes, the 5,393 DAS Tool bins were dereplicated using dRep v3.6.2³⁰ with a primary clustering ANI threshold = 0.90, and a secondary clustering ANI threshold = 0.995. This collapsed near-identical genomes to remove duplicate strains, leaving 4,858 strain-level MAGs. A second, dereplication at a lower ANI threshold (0.95) was applied to define species-level MAGs (n=1374), facilitating analyses at both strain and species resolution.

Genome quality was assessed using CheckM2³¹, and bins were filtered based on completeness and contamination thresholds consistent with community standards³². Only genomes meeting quality criteria (completion > 80%, contamination < 10 %) were retained for downstream analyses. Quality statistics for the final MAG catalogue are provided in **Supplementary Figure S3 and Supplementary Table S3**. Because bins from both the co-assembly and hybrid assembly strategies were pooled and dereplicated jointly, each species-level cluster is represented by a single MAG regardless of which assembly strategy produced it. Abundance estimation was therefore performed against this non-redundant catalogue, precluding double-counting of organisms recovered by multiple assembly approaches. The

resulting species-level MAG set forms the primary basis for prevalence, bloominess, and AMR analyses presented in this study.

2.6. Taxonomic annotation and MAG abundance estimation

2.6.1 Taxonomic classification of MAGs

All curated MAGs were taxonomically classified using GTDB-Tk (v2.4.1; release 226), which assigns genomes to a standardised, genome-based taxonomy using phylogenomic placement against the Genome Taxonomy Database^{33,34}. This approach provides consistent classification across bacterial and archaeal lineages and avoids biases associated with marker-gene or reference-limited taxonomies.

Taxonomic assignments were retained at all ranks from phylum to species. Analyses in this study primarily use class- and order-level taxonomy, which provide sufficient resolution to detect ecological patterns while avoiding sparsity at finer taxonomic levels. Taxonomic summaries of the MAG catalogue are provided in **Supplementary Table 4** and **Supplementary Figure S4**.

2.6.2 MAG abundance estimation across samples

MAG abundance was quantified by mapping Illumina short reads from each sample back to the curated MAG catalogue. Reads were aligned using minimap2³⁵, and coverage statistics were computed using anvi'o³⁶.

To obtain a robust estimate of MAG abundance that is less sensitive to uneven coverage, repetitive regions, or local amplification artifacts, we used the mean per-base coverage calculated over the interquartile range (Q2-Q3) of covered positions. This metric captures the central tendency of coverage while excluding extreme low- and high-coverage regions that may distort abundance estimates.

This abundance measure was calculated independently for each MAG in each sample and used consistently across all downstream analyses, including prevalence estimation, abundance distribution visualisation, and bloominess calculations. Read mapping was performed competitively against the full concatenated MAG catalogue, such that each read was assigned to a single best-matching reference; this prevents the same read from contributing to the coverage estimates of multiple MAGs simultaneously.

2.7. Prevalence and sporadic dominance metrics

2.7.1 Definition of MAG presence and prevalence

For each MAG, presence within a sample was defined using a joint criterion combining coverage and detection thresholds. A MAG was considered present in a given sample if:

1. Mean Q2-Q3 coverage exceeded a specified coverage threshold, and

2. The proportion of the MAG detected within the sample exceeded a minimum detection threshold.

Unless otherwise stated, analyses presented in the main text use a coverage threshold of **0.5**× and a detection threshold of **0.2**, which were selected to balance sensitivity to low-abundance taxa with robustness to spurious mapping. Sensitivity analyses exploring alternative thresholds are presented in **Supplementary Figure S5**.

MAG **prevalence** was then defined as the fraction of samples in which that MAG met the presence criterion:

$$\text{Prevalence}_i = \frac{\text{number of samples where MAG}_i \text{ is present}}{\text{total number of samples}}$$

Prevalence reflects consistent presence across the sampled cohort rather than dominance within any individual sample.

2.7.2 Prevalence classes

For interpretability, the continuous prevalence axis was discretised into three prevalence classes:

- **Core:** MAGs present in $\geq 85\%$ of samples
- **Variable:** MAGs present in $>10\%$ and $<85\%$ of samples
- **Rare:** MAGs present in $\leq 10\%$ of samples

These thresholds were chosen to delineate broad ecological regimes rather than to imply strict biological boundaries. These boundaries were selected a priori to reflect broadly recognised ecological regimes analogous to core and accessory partitions used in pangenome analyses, and were not optimised to the specific distributional properties of this dataset. Prevalence classes are used throughout the manuscript as a descriptive framework for comparing taxonomic composition, abundance distributions, and antimicrobial resistance characteristics.

2.7.3 Definition of bloominess (sporadic dominance)

To quantify asymmetry in abundance distributions across samples, we introduce **bloominess**, a MAG-level metric capturing sporadic dominance.

Bloominess was calculated independently for each MAG, using only the subset of samples in which the MAG was classified as present (**Supplementary Methods S5**). Conditioning the calculation on presence ensures that bloominess reflects variation in abundance when the organism is biologically supported, rather than conflating true absence with non-detection or sequencing depth effects.

Within this subset of present samples, bloominess was defined by comparing the mean and median abundance of the MAG:

$$\text{Bloominess}_i = \log_{10} \left(\frac{\text{mean}_i + \varepsilon}{\text{median}_i + \varepsilon} \right)$$

where ε is a small constant added to stabilise values near zero.

The median abundance represents the typical background contribution of a taxon, while the mean is sensitive to rare, high-abundance events. When abundance is stably distributed, mean and median values are similar and bloominess approaches zero. In contrast, taxa exhibiting rare but extreme dominance events show elevated bloominess values as the mean is disproportionately inflated relative to the median.

2.7.4 Interpretation and properties of bloominess

Bloominess is a shape-based metric that captures the asymmetry of abundance distributions across samples. It is scale-independent and does not depend on absolute abundance or prevalence (see comparison of coverage vs abundance-based calculations in **Supplementary Figure S6**). Because bloominess is computed within each MAG across its present-sample distribution, rather than relative to other MAGs within a sample, it does not require that total microbial biomass or sequencing depth be consistent across samples. Consequently, taxa with similar prevalence can exhibit markedly different bloominess values.

Importantly, bloominess is distinct from commonly used diversity or dispersion metrics. Variance quantifies overall spread, and evenness describes within-sample community structure. Bloominess instead captures across-sample asymmetry driven specifically by rare, extreme high-abundance events. It therefore represents an ecological axis orthogonal to prevalence.

Together, prevalence and bloominess define a two-dimensional ecological space in which taxa can be positioned according to persistence across hosts and propensity for sporadic dominance. Bloominess is a continuous trait; the term “bloomer” is used colloquially throughout this paper to refer to taxa at the high end of the bloominess distribution, without implying a discrete threshold. To aid interpretation, bloominess values translate directly to fold differences between mean and median abundance: a score of 0.3 corresponds to an approximately two-fold difference ($10^{0.3} \approx 2\times$), while a score of 1.0 indicates a ten-fold difference ($10^{1.0} = 10\times$) between mean and median. These thresholds are used as visual reference lines in Figure 4 to anchor interpretation of the magnitude of distributional asymmetry across the MAG atlas.

2.8 Detection and characterisation of antimicrobial resistance genes

2.8.1 Prediction of protein-coding genes

Protein-coding genes were predicted from all curated MAGs using Prodigal (v2.6.3)³⁷ in metagenomic mode. Gene prediction was performed independently for each MAG to ensure consistent annotation across assemblies derived from different strategies. The resulting protein sequences formed the input for antimicrobial resistance gene (ARG) detection.

2.8.2 Identification of antimicrobial resistance genes

ARGs were identified using AMRFinderPlus (v4.0.23)³⁸ with database version 2025-07-16.1, which detects both acquired resistance genes and resistance-associated point mutations through curated reference sequences and hidden Markov models.

AMRFinderPlus was run on the predicted protein sequences from each MAG. Hits were retained using default confidence thresholds, and each detected ARG was annotated with its resistance class, subclass, mechanism, and target drug category.

Only chromosomally encoded ARGs detected within MAGs were considered in this analysis; plasmid-associated resistance elements were analysed separately and are reported elsewhere.

2.8.3 Aggregation of ARGs at the MAG level

For each MAG, we summarised ARG content using multiple complementary metrics:

1. **Raw ARG count:** the total number of ARG annotations detected within the MAG
2. **ARG presence:** a binary indicator denoting whether at least one ARG was detected
3. **ARG class composition:** the distribution of resistance classes encoded by each MAG

MAGs with no detected ARGs were explicitly retained and assigned an ARG count of zero.

Because AMRFinderPlus reports individual genes rather than functional operons, some resistance systems (e.g. *van* operons) may contribute multiple ARG annotations within a single genome. Raw ARG counts therefore reflect the genetic burden of resistance-associated loci, rather than the number of independent resistance phenotypes. This distinction is considered explicitly in downstream analyses.

2.8.4 ARG burden across prevalence and bloominess gradients

ARG summaries were integrated with MAG-level prevalence and bloominess metrics by merging on standardised MAG identifiers. MAGs were grouped according to prevalence class (core, variable, rare), and ARG distributions were examined across these ecological categories.

To explore whether sporadic dominance is associated with resistance potential, bloominess distributions were compared across MAGs stratified by ARG burden. MAGs were binned

according to the number of detected ARGs (e.g. 0, 1, 2, 3, ≥ 4), allowing assessment of whether taxa encoding multiple resistance determinants tend to exhibit different abundance dynamics when present.

Analyses focusing on ARG-bloominess relationships were restricted to variable MAGs, where sufficient prevalence ensures robust estimation of abundance distributions. Detailed descriptions of all steps are in **Supplementary Methods SM6**.

2.8.5 Interpretation of ARG counts

ARG counts were interpreted as a genome-level resistance load, not as a direct proxy for phenotypic resistance. Multiple ARGs within a MAG may represent:

- multi-gene resistance operons
- resistance to multiple drug classes
- accumulation of resistance determinants within opportunistic or mobile taxa

Accordingly, results are presented primarily in terms of relative patterns across ecological classes, rather than absolute thresholds of risk.

2.9. Statistical analysis and visualisation

All analyses were conducted in R (v4.4.1). Data wrangling and statistical summaries were performed using the tidyverse suite of packages. Visualisations were generated using ggplot2, with additional extensions for colour palettes and layout where required.

2.9.1 Prevalence, abundance, and bloominess analyses

Prevalence, abundance, and bloominess were treated as MAG-level properties. Unless otherwise stated, analyses were conducted on the species-level dereplicated MAG set.

Comparisons of bloominess across prevalence classes, taxonomic groups, and ARG burden categories were performed descriptively using distributional summaries (violin plots, boxplots, density plots). Given the large number of MAGs and the non-normal, heavy-tailed nature of several distributions, emphasis was placed on effect size and distributional shape, rather than formal hypothesis testing.

Where thresholds were applied (e.g. prevalence class boundaries, presence criteria for bloominess calculations), these were defined a priori and held constant across analyses.

2.9.2 Handling of zero-inflated and skewed distributions

Several variables in this study, including MAG abundance, bloominess, and ARG counts, exhibited zero inflation and right-skewed distributions. These properties were handled through:

- log-transformation of abundance metrics where appropriate
- conditioning bloominess calculations on presence

- explicit retention of zero values for ARG counts

No parametric assumptions of normality were imposed.

2.9.3 ARG-associated analyses

ARG counts per MAG were treated as discrete variables. For exploratory analyses relating ARG burden to bloominess, MAGs were grouped into bins based on the number of detected ARGs (e.g. 0, 1, 2, 3, ≥ 4). These bins were chosen to balance interpretability and sample size, and were not intended to imply discrete biological categories.

2.9.4 Taxonomic stratification

Taxonomic comparisons were performed using GTDB-Tk classifications. Analyses were conducted at multiple taxonomic ranks (phylum, class, order) depending on the question of interest. For visual clarity, taxonomic groups represented by few MAGs were aggregated into an “Other” category where appropriate.

Taxonomic analyses were descriptive and exploratory in nature, aimed at identifying broad ecological patterns rather than testing lineage-specific hypotheses.

2.9.5 Visualisation principles

Figures were designed to emphasise distributional structure and comparative trends rather than point estimates. Mean and median values were explicitly distinguished in abundance distribution plots using consistent colour and line-type conventions throughout the manuscript.

Smoothing lines (e.g. LOESS) were used only for visual guidance and were not interpreted as predictive models. All plots were generated directly from processed analysis tables to ensure traceability.

Portions of this manuscript were drafted and refined with the assistance of large language model tools, including Claude (Anthropic) and ChatGPT (OpenAI). These tools were used to support text drafting, editing, and literature organisation. All content was critically reviewed, verified, and approved by the authors. No AI tool is listed as an author or holds responsibility for the work presented.

3. RESULTS

3.1 Prevalence landscape and taxonomic composition of the MAG Atlas

Across the full MAG atlas ($n=1374$), prevalence exhibited a strongly right-skewed distribution, with a small number of taxa detected in nearly all samples and a long tail of taxa occurring sporadically across the cohort (**Fig. 1A**). Based on this structure, MAGs were partitioned into three prevalence classes: core (≥ 0.85 prevalence), variable ($0.1-0.85$), and rare (≤ 0.1), reflecting highly conserved community members, variably present taxa, and infrequently detected taxa, respectively (**Fig. 1A, B**). These classes were defined to capture ecological roles rather than to balance class sizes, resulting in a compact core and a substantially larger variable and rare pool.

Taxonomic composition differed markedly across prevalence classes (**Fig. 1C**). The core was taxonomically constrained, comprising exclusively members of the phylum Bacillota and a small number of genera, consistent with a limited set of broadly adapted gut generalists. In contrast, the variable and rare classes encompassed a wide range of phyla and genera, with extensive taxonomic diversity distributed across many low-prevalence taxa. Together, these patterns establish a structured ecological space in which prevalence captures broad taxonomic and abundance trends, while leaving substantial heterogeneity unresolved within the large variable fraction.

Variable-class MAGs accounted for the majority of mapped reads in all samples (median 85.1%, IQR 78.9–88.1%), with core MAGs contributing a smaller but consistent fraction (median 8.5%, IQR 5.9–10.7%) and rare-class MAGs making up the remainder (median 5.2%, IQR 2.8–9.9%; **Supplementary Figure S7; Supplementary Method S7**). Despite their near-universal prevalence, the 18 core MAGs were not uniformly present: only 66 of 116 samples (57%) harboured all 18 simultaneously, and Ward hierarchical clustering of co-absence patterns identified three clusters of MAGs with coordinated absence behaviour, suggesting ecological co-dependencies among subsets of the core community (**Supplementary Figure S8; Supplementary Method S7**).

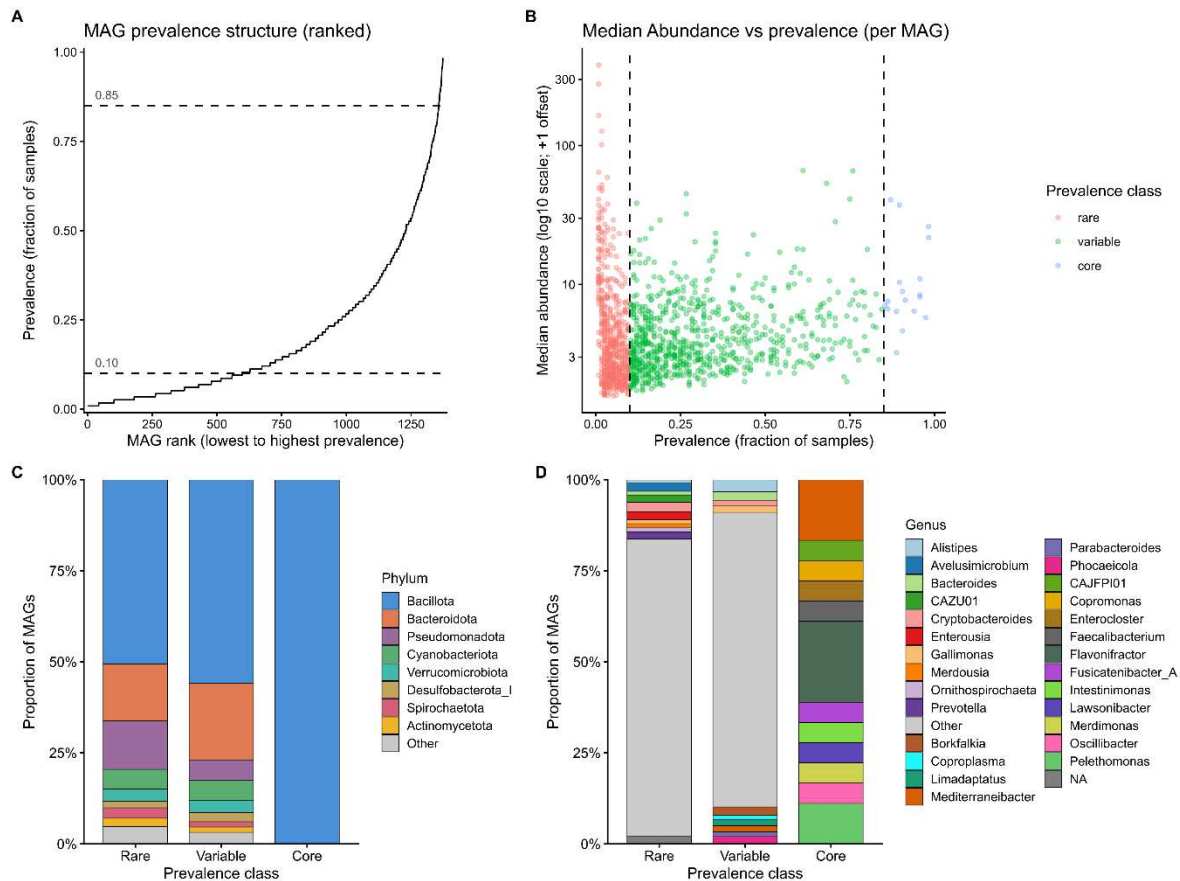


Figure 1 Prevalence structure and taxonomic composition of the 1,374 MAG poultry gut microbiome catalogue. (A) MAGs ranked from lowest to highest prevalence (fraction of 116 samples in which each MAG was detected). Dashed lines indicate the thresholds defining prevalence classes: core (≥ 0.85 ; present in $\geq 85\%$ of samples) and rare (< 0.10 ; present in $< 10\%$ of samples). MAGs between the two thresholds are classified as variable. (B) Median relative abundance versus prevalence for each of the 1,374 MAGs, coloured by prevalence class. Median abundance was calculated across samples in which each MAG was detected (Q2Q3 coverage $\geq 0.5\times$ and detection ≥ 0.2). Abundance values are shown on a log₁₀ scale with a pseudocount of +1. (C) Phylum-level taxonomic composition of MAGs within each prevalence class, expressed as the proportion of MAGs in that class. Taxonomic assignments from GTDB-Tk (release 226). The eight most represented phyla are shown; the remaining 10 phyla (3.8% of the catalogue) are grouped as 'Other'. (D) Genus-level composition of MAGs within each prevalence class, showing the ten most represented genera per class. Remaining genera are grouped as 'Other'.

3. 2 Archetypal abundance distributions

Across samples in which a taxon was detected, abundance distributions exhibited distinct and recurrent shapes that were not captured by prevalence alone (Fig 2). While prevalence describes how frequently a taxon occurs across a cohort, it does not specify how that taxon distributes its abundance across the samples in which it is present. As a result, taxa with similar prevalence can differ substantially in their abundance behaviour.

Conceptually, abundance distributions can be broadly distinguished by their shape (Fig. 2A). Some taxa display relatively narrow distributions, characterised by modest variation around a low or intermediate abundance across most samples. In these cases, abundance is relatively

even when the taxon is present, and summary statistics such as the mean and median abundance closely coincide. Other taxa exhibit heavy-tailed distributions, in which abundance remains low in most samples but reaches extremely high values in a small subset. Although both distribution types may correspond to the same prevalence, their ecological implications differ markedly.

These archetypal patterns were evident in the empirical data (**Fig. 2B, D**). Among taxa with comparable prevalence, some showed consistently constrained abundance across samples, whereas others were dominated by rare but pronounced abundance peaks. These differences were not attributable to detection alone, as all distributions were computed using only samples in which the taxon met fixed coverage and detection thresholds. Instead, they reflected intrinsic differences in how taxa behave across the cohort when present.

The divergence between narrow and heavy-tailed abundance distributions has direct consequences for commonly used summary statistics (**Fig. 2C, E**). In narrow distributions, the mean and median abundance provide similar estimates of a taxon's typical abundance. In contrast, heavy-tailed distributions generate a pronounced separation between these metrics: a small number of high-abundance samples disproportionately elevate the mean, while the median remains anchored to the low-abundance majority. This divergence highlights a fundamental distinction between typical abundance and sporadic dominance that is obscured when abundance is summarised using a single statistic.

Together, these observations demonstrate that prevalence alone does not capture the full structure of taxon abundance across a cohort. Even within a single prevalence class, taxa may differ in whether they contribute consistently at background levels or participate in rare but extreme dominance events. This motivates the need for a complementary measure that explicitly quantifies variability in abundance distributions across samples, independent of prevalence.

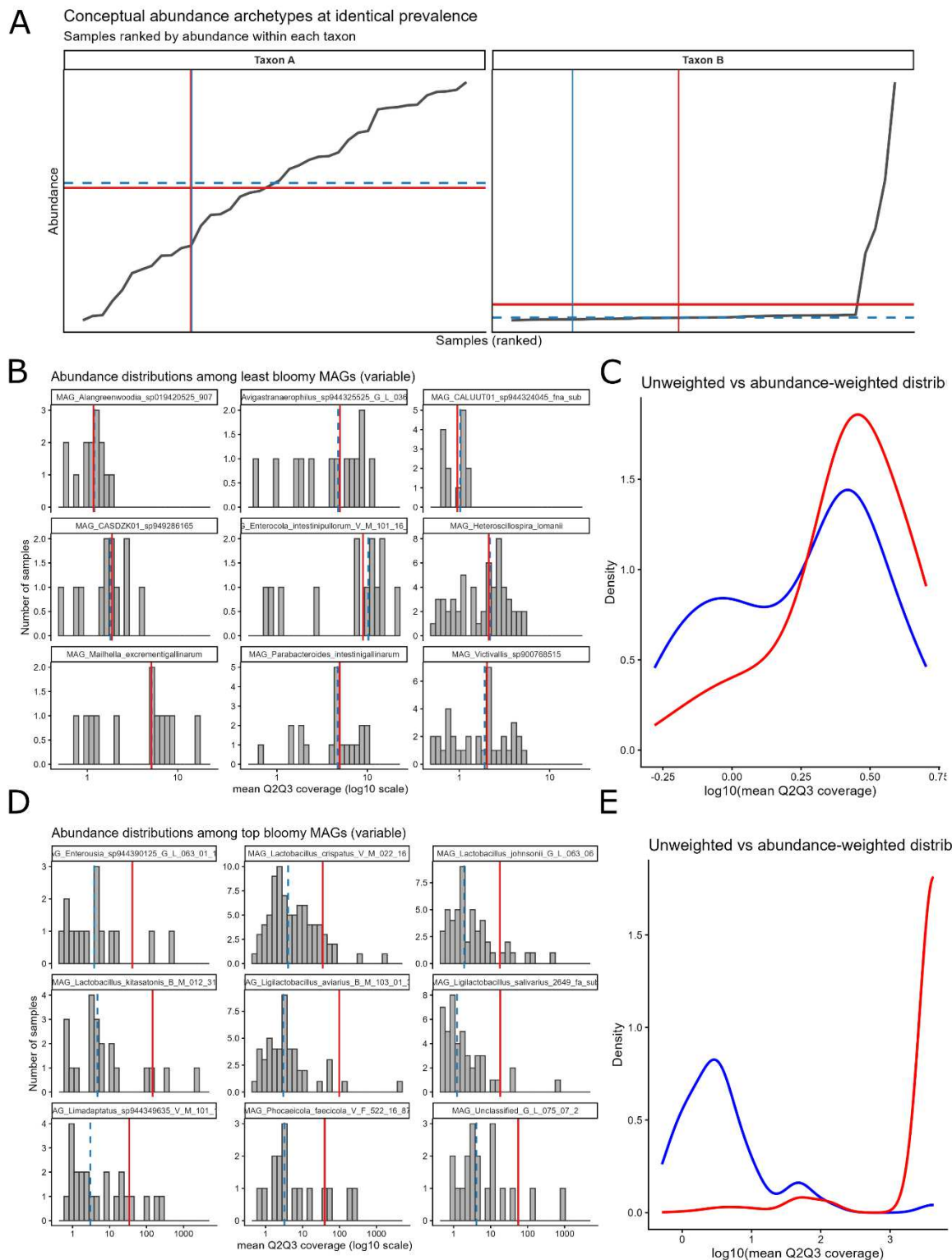


Figure 2. (A) Conceptual abundance archetypes. Two taxa detected in the same number of samples (identical prevalence) can exhibit markedly different abundance distributions across those samples. Taxon A shows relatively even abundance, with mean and median closely aligned, whereas Taxon B displays rare but extreme dominance events that disproportionately elevate the mean relative to the median. (B) Abundance distributions for selected variable MAGs showing relatively even abundance. (C) Mean and median abundance closely coincide, indicating stable abundance distributions with little influence from extreme dominance events. (D) Abundance distributions for selected variable MAGs showing episodic dominance. (E) High bloominess arises from a small number of high-abundance samples that disproportionately shift the mean relative to the median. Histograms show the distribution of mean Q2Q3 coverage across samples in which each MAG is detected. Dashed blue lines indicate the median abundance, and solid red lines indicate the mean abundance.

3.3 Quantifying sporadic dominance: bloominess

Bloominess formalises this distributional divergence as a single, reproducible metric (Methods section 2.7.3). Together, prevalence and bloominess define a two-dimensional ecological space in which taxa are positioned according to persistence across samples and propensity for sporadic dominance; this contrast is illustrated for two archetypal taxa in **Figure 3**.

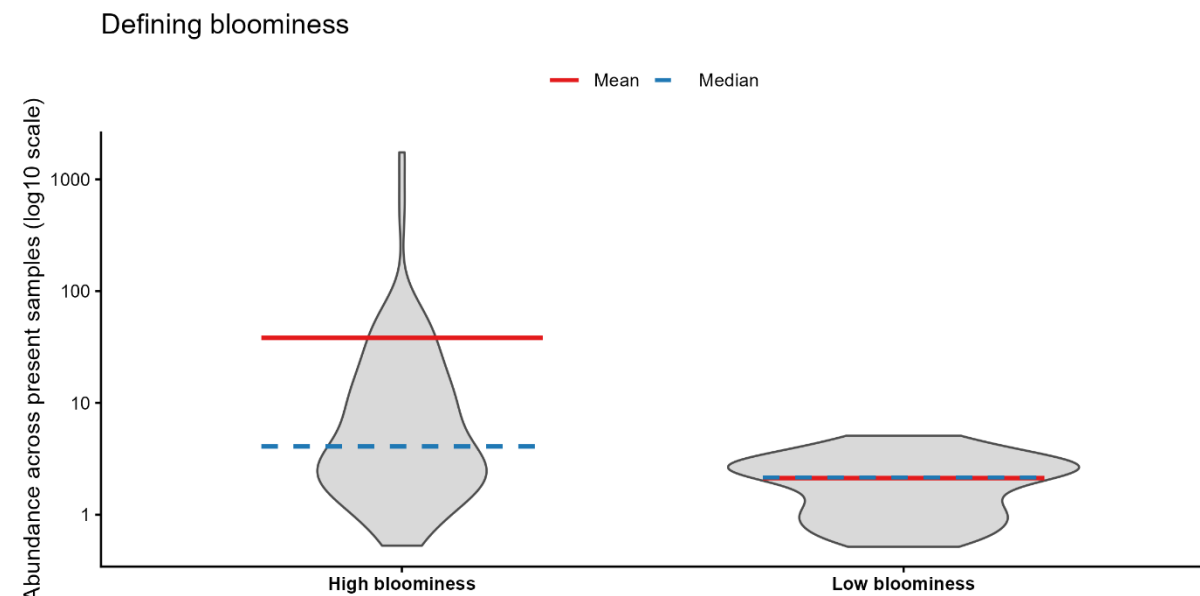


Figure 3. Conceptual illustration of bloominess as the divergence between mean and median abundance across samples. Distributions of abundance across samples are shown for two archetypal taxa on a $\log_{10}$ scale. In a high-bloominess taxon (left, *MAG_Lactobacillus_crispatus_V_M_022_16*), rare but extreme high-abundance events inflate the mean relative to the median, producing a large divergence between these statistics. In contrast, a low-bloominess taxon (right, *MAG_Heteroscollospira_lomanii*) exhibits more consistent abundance across samples, with mean and median closely aligned. Bloominess is defined as the log-transformed ratio of mean to median abundance across samples in which the taxon is present, capturing the degree of sporadic dominance independent of prevalence.

3.4 Bloominess in the context of prevalence

When bloominess is considered alongside prevalence, taxa occupy a structured two-dimensional ecological space defined by persistence across samples and sporadic dominance when present **Fig 4**.

Prevalence imposes clear constraints on bloominess (**Fig 4A**). At very high prevalence (core taxa), bloominess values are uniformly low. Taxa detected in nearly all samples tend to exhibit stable abundance distributions, with limited divergence between mean and median abundance. Conversely, at very low prevalence, bloominess is also constrained: rare taxa are supported by few samples, limiting the opportunity for extreme dominance events to disproportionately influence their cohort-level distribution.

Between these extremes lies the variable region, where prevalence alone does not determine behaviour. Taxa within this intermediate prevalence range span the full spectrum of bloominess values, from highly stable to strongly sporadic. Importantly, variable taxa are not uniformly bloomy. Rather, this region represents an ecological space in which sporadic dominance is possible, but not guaranteed.

This structure reflects two distinct layers of ecological conditionality (**Fig 4B**). First, factors governing presence determine whether a taxon is detected in a given host or environment at all. Second, conditional on presence, additional factors may influence whether that taxon remains at background abundance or shows elevated abundance. Bloominess captures variability at this second layer, independent of the first.

As a result, the probability of observing high bloominess is elevated within the variable region, not because intermediate prevalence causes blooms, but because it defines a space in which dominance dynamics are not tightly constrained. This region therefore represents a flexible ecological zone in which context-dependent elevation in abundance may occur.

Together, prevalence and bloominess reveal a structured landscape in which stable core taxa, constrained rare taxa, and a permissive variable edge can be distinguished without conflating persistence with dominance.

Within this two-dimensional ecological space, a distinct region emerges at intermediate prevalence and moderate-to-high bloominess. We refer to this region as the variable edge. Taxa occupying the variable edge are neither ubiquitously stable members of the core microbiome nor sporadic detections constrained by limited support. Instead, they represent a flexible ecological component: taxa whose presence varies across hosts and whose abundance, when present, may remain stable or occasionally show elevated abundance.

The variable edge does not imply inherent instability or risk. Rather, it denotes a zone of ecological responsiveness, where both persistence and dominance are context-dependent. In this region, shifts in host environment, management practices, microbial interactions, or selective pressures may influence not only whether a taxon is detected, but also whether it expands disproportionately within individual microbiomes. By distinguishing the variable edge from the stable core and the constrained rare tail, the joint interpretation of prevalence and bloominess enables a structured understanding of where context-dependent dynamics are most likely to occur.

Core taxa are constrained in dominance. Rare taxa are constrained in support. Variable taxa are constrained by neither.

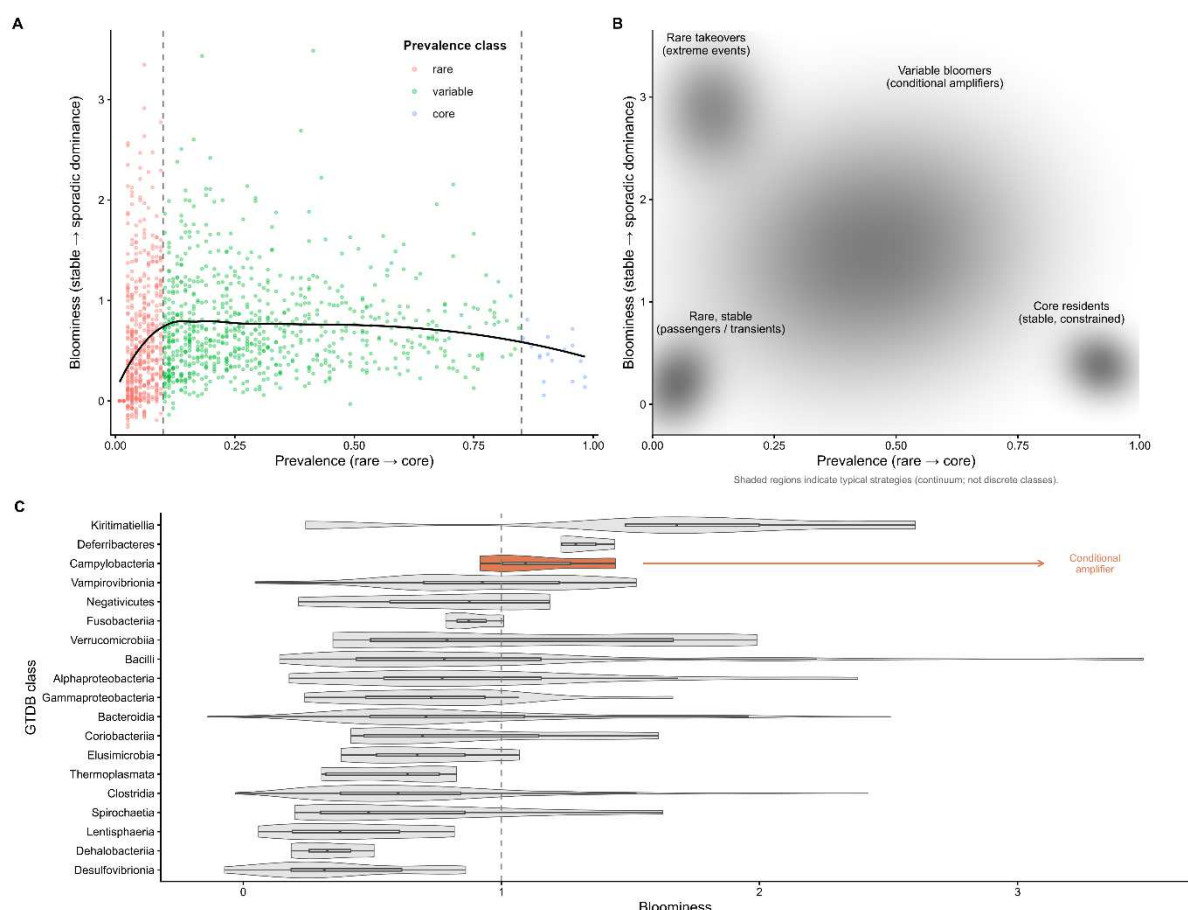


Figure 4 Prevalence-bloominess structure and abundance distributions underlying sporadic dominance. (A)

Relationship between taxon prevalence across samples and bloominess. Points represent MAGs coloured by prevalence class (rare, variable, core). Dashed vertical lines indicate prevalence thresholds defining classes. The smoothed trend highlights a wedge-shaped structure in which bloominess is constrained at both low and high prevalence, with greatest variability at intermediate prevalence. (B). Conceptual map of the prevalence × bloominess space illustrating four biological strategies. Shaded regions represent typical occupancy zones for each strategy (continuum; boundaries are not discrete classes). Rare, stable taxa (low prevalence, low bloominess) behave as passengers or transients; rare takeovers (low prevalence, high bloominess) represent extreme stochastic events; variable bloomers (intermediate prevalence, high bloominess) are conditional amplifiers that transiently dominate; and core residents (high prevalence, low bloominess) are stable community members with constrained abundance dynamics. (C) Bloominess distributions across GTDB classes for variable MAGs (classes with ≥ 3 MAGs shown; ordered by median bloominess). *Campylobacter* (orange) ranks among the most bloomy classes, consistent with their role as conditional amplifiers capable of episodic dominance within the chicken gut microbiome. Dashed line indicates bloominess = 1.0.

Ecological anchor taxa sanity check. Bloominess was computed across all MAGs without reference to taxonomy. As a qualitative ecological sanity check, we inspected the placement of taxa with well-described sporadic dominance behaviour in poultry microbiomes. Within the variable prevalence class, bloominess varied systematically across taxonomic groups (Fig. 4C). Violin plots of bloominess values stratified by GTDB class revealed clear class-level differences in both central tendency and distributional shape. Several classes exhibited broad, right-skewed bloominess distributions with pronounced upper tails, indicating frequent sporadic dominance events among a subset of taxa. In contrast, other classes showed narrower distributions centred at lower bloominess values, consistent with more stable abundance profiles across samples. These patterns were not driven by prevalence, which was

constrained by definition within the variable, but instead reflected taxon-specific differences in the degree to which rare high-abundance events inflated cohort-level means relative to medians.

Notably, **Campylobacteria** displayed consistently elevated bloominess values with limited overlap with low-bloominess regimes, indicating strong sporadic dominance behaviour across samples. In contrast, **Gammaproteobacteria**, which include *Escherichia coli*, showed broader and more heterogeneous bloominess distributions, spanning both stable and highly bloomy regimes. **Bacilli**, encompassing taxa such as *Enterococcus cecorum*, exhibited extended right tails despite moderate median bloominess, suggesting occasional but substantial dominance events within otherwise stable populations. Together, these patterns indicate that bloominess captures ecologically and clinically relevant differences in abundance dynamics among taxa occupying similar prevalence niches.

3.5. Bloominess associates with AMR

Having established bloominess as an axis of ecological variation orthogonal to prevalence and structured by taxonomy, we next considered its potential functional implications. Sporadic dominance events concentrate biomass into a small number of samples and may therefore elevate the contribution of associated genes to the resistome. This motivated us to examine how antimicrobial resistance genes are distributed across taxa with differing bloominess, and whether sporadically dominant taxa contribute disproportionately to ARG signal relative to their prevalence.

ARG burden varied strongly across prevalence classes (**Figure 5A**). Interestingly, core taxa carried several ARGs, whereas rare taxa exhibited a broader and higher ARG count distribution, despite their low prevalence. Variable taxa showed intermediate but heterogeneous ARG burdens, consistent with their position at the variable ecological edge.

Within variable taxa, bloominess was not monotonically associated with ARG count (**Figure 5B**). MAGs carrying an intermediate number of ARGs (approximately three) exhibited elevated bloominess, whereas the highest ARG burden group showed lower bloominess. This suggests that sporadic dominance is most pronounced among taxa with moderate resistance potential, while highly resistant taxa may persist more stably once established.

To explore how resistance burden interacts with ecological dynamics, ARGs were collapsed into 10 kb loci and analysed across bloominess tertiles within variable taxa. Overall ARG locus burden was evenly distributed across bloominess classes, indicating that sporadic dominance is not intrinsically associated with increased resistance acquisition. However, drug class composition differed markedly across ARG burden and bloominess strata (**Figure 5D**). β -lactam resistance loci were most frequently observed as solitary resistance elements, consistent with widespread low-cost resistance. In contrast, glycopeptide and lincosamide resistance loci preferentially occurred in multi-ARG MAGs and were enriched in low-bloominess taxa, suggesting stable lineage-associated resistance architectures, potentially reflecting intrinsic resistance or higher fitness costs. These patterns highlight that resistance

risk emerges from the interaction between ecological behaviour and resistance composition,
rather than from bloominess or ARG burden alone.

The distribution of ARG loci across the bloominess $\times$ total ARG burden grid revealed contrasting patterns by drug class: beta-lactam resistance loci were concentrated in high-bloominess MAGs carrying a single total ARG locus, while glycopeptide resistance loci were disproportionately found in low-bloominess MAGs with higher overall ARG burden (**Fig. 5D**). Chi-squared goodness-of-fit tests against a null of proportional MAG distribution confirmed that glycopeptide ($p_{\text{adj}} = 0.001^{**}$) and tetracycline ($p_{\text{adj}} = 0.001^{**}$) carrier MAGs were distributed non-randomly across the bloominess $\times$ ARG burden grid (**Supplementary Method S6**). Wilcoxon rank-sum tests comparing bloominess between drug-class carriers and non-carriers among variable MAGs confirmed that tetracycline-resistance carriers were significantly bloomier than non-carriers ($\Delta_{\text{median}} = +0.314$, $p_{\text{adj}} = 0.0017^{**}$), while glycopeptide-resistance carriers were significantly less bloomy ($\Delta_{\text{median}} = -0.117$, $p_{\text{adj}} = 0.028^{*}$; **Supplementary Figure S9**).

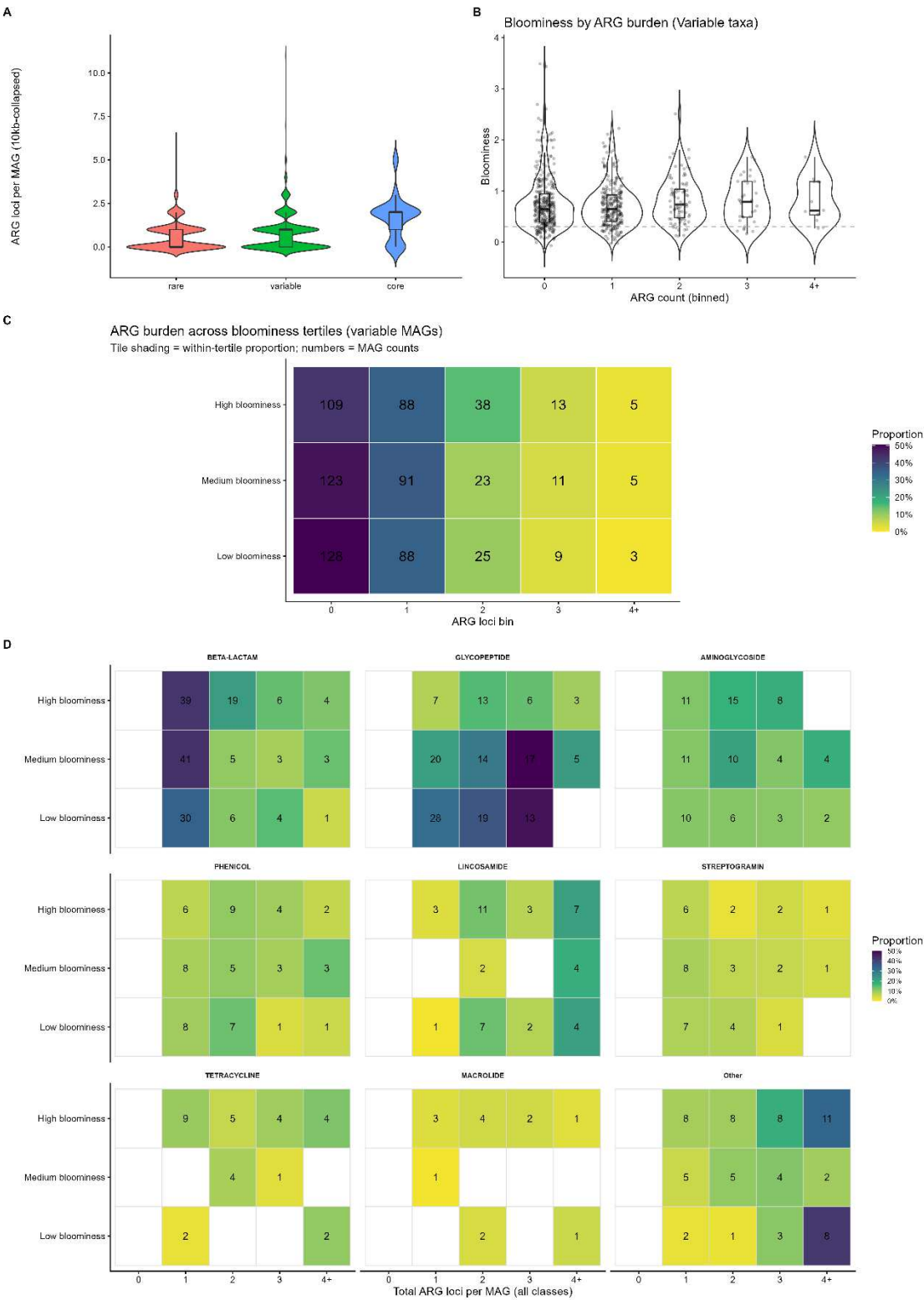


Figure 5 (A) Distribution of ARG loci per MAG (ARGs collapsed into 10 kb loci) across prevalence classes (rare, variable, core). Core taxa harbour a higher median ARG locus burden than rare or variable taxa. (B) Bloominess distributions across binned ARG locus burden for variable taxa. Bloominess is broadly similar across ARG burden bins, indicating that sporadic dominance is not intrinsically associated with increased resistance burden. (C) Heatmap showing ARG locus burden across bloominess tertiles for variable taxa. Tile shading represents within-row proportions, and numbers indicate MAG counts.

ARG burden is evenly distributed across bloominess classes. **(D)** Drug class composition of ARG loci across bloominess tertiles and ARG burden bins for variable taxa. Values represent the proportion of ARG loci within each bloominess × burden category. Drug classes differ markedly in their association with ARG burden and bloominess.

Although overall ARG locus burden was evenly distributed across bloominess tertiles, a small subset of variable taxa combined high bloominess with elevated ARG locus counts. These MAGs were numerically rare but represent a distinct ecological risk category, coupling sporadic dominance with multi-locus resistance. In contrast to stable core taxa, which harbour resistance as a persistent background feature, these high-bloominess, high-burden taxa have the potential to transiently elevate resistance gene abundance. We therefore define these taxa as a “watchlist” group: infrequent, but of disproportionate concern when present.

Ecological-resistance mode	Description
Background reservoirs	Core taxa with stable, often intrinsic resistance
Silent carriers	Low-bloominess taxa with high ARG burden
Watchlist taxa	High-bloominess taxa with multi-locus resistance

3.6 Bloominess is a stable ecological axis with structured uncertainty

Across alternative coverage thresholds, bloominess rankings were highly conserved, indicating that relative positioning along the bloominess axis was robust to reasonable changes in presence criteria (**Fig. 6A**). Instability was not uniformly distributed across the MAG atlas. Instead, variation in bloominess estimates was concentrated among low-support taxa detected in relatively few samples, whereas well-supported taxa remained stable across threshold settings (**Fig. 6B**). Importantly, the broader ecological structure of the dataset was preserved: the wedge-shaped prevalence-bloominess distribution remained evident across alternative parameterisations, with bloominess constrained at both low and high prevalence and greatest variability in the variable region (**Fig. 6C**). Together, these results indicate that uncertainty in bloominess reflects structured limitations of detection and support rather than failure of the underlying ecological axis.

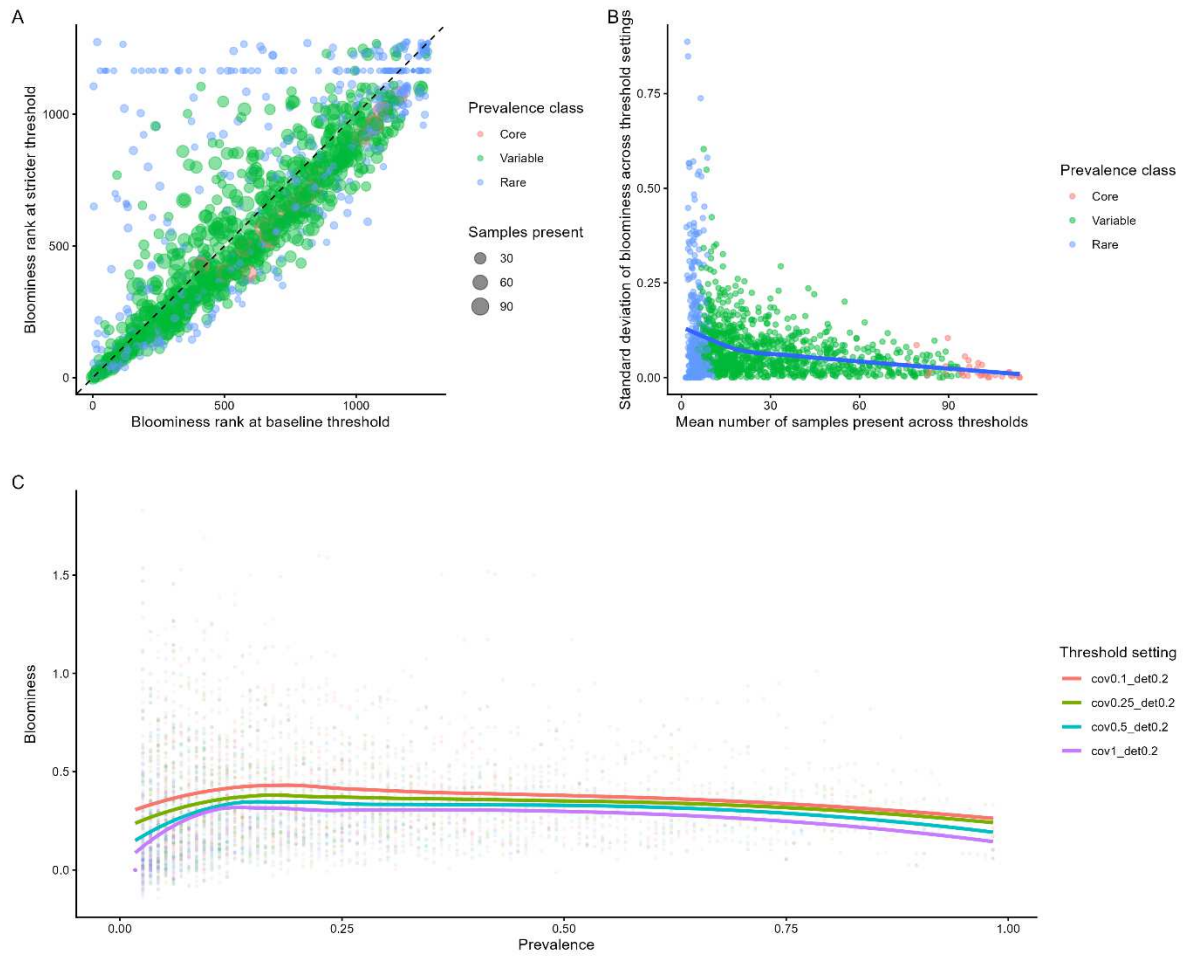


Figure 6: Bloominess is a stable ecological axis with structured uncertainty. (A) Bloominess rankings are highly consistent across alternative presence thresholds, with most MAGs remaining close to the diagonal. (B) Instability in bloominess ranking is concentrated among low-support MAGs detected in relatively few samples, whereas well-supported taxa remain stable across threshold choices. (C) The prevalence–bloominess landscape is preserved across threshold settings, indicating that the wedge-shaped ecological structure is robust even when individual bloominess values shift modestly.

4. DISCUSSION

4.1 Bloominess as an ecological axis

In this study, we introduce bloominess as a taxon-level ecological trait describing the propensity for sporadic dominance across samples, independent of prevalence. By separating how often a taxon is detected from how unevenly it distributes its abundance when present, bloominess defines an additional axis of microbiome structure that is not captured by prevalence or mean abundance alone.

Across the poultry caecal microbiome, taxa exhibited a continuum of abundance distributions ranging from stable, low-variance profiles to highly skewed distributions characterised by rare but extreme dominance events (**Fig 2**). These distributional shapes were not predictable from prevalence: both widely distributed and infrequently detected taxa could exhibit either stable or highly sporadic abundance behaviour (**Fig 4**). This decoupling highlights a key limitation of prevalence-based frameworks and demonstrates the need for descriptors that explicitly capture abundance variability.

Importantly, bloominess is defined independently of absolute abundance and sequencing depth and is calculated only across samples in which a taxon is present. As such, it captures an intrinsic ecological property rather than a technical artefact or sampling noise. We interpret bloominess as a measure of ecological responsiveness, capturing a taxon's capacity to expand under transient conditions such as host perturbation, dietary change, or competitive release^{2,39}.

4.2 Ecological structure of abundance behaviour

When mapped into prevalence-bloominess space, taxa formed a structured distribution in which bloominess was constrained at both extremes of prevalence (**Fig 4**). Rare taxa, observed in only a small number of samples, exhibited limited bloominess due to restricted opportunity for extreme events to influence their abundance distributions. At the opposite end of the spectrum, highly prevalent taxa displayed consistently low bloominess, reflecting stable abundance profiles across hosts.

Between these extremes, taxa occupying intermediate prevalence exhibited the greatest diversity of behaviours. This region, corresponding to the variable or “variable” microbiome, contained taxa with both stable and highly sporadic abundance profiles. As a result, bloominess expanded into a wedge-shaped distribution across prevalence, with increasing variance toward intermediate prevalence and strong constraint at both ends (**Figure 4B**).

This structure demonstrates that abundance behaviour is not determined by prevalence alone. Instead, taxa with similar prevalence can exhibit fundamentally different ecological dynamics, ranging from stable background presence to context-dependent dominance. The

prevalence-bloominess framework therefore resolves a dimension of microbiome organisation that is not captured by prevalence or mean abundance independently.

4.3 ARG burden: load versus ecological visibility

Bloominess provides a framework for distinguishing between the presence of antimicrobial resistance genes (ARGs) and their ecological expression within microbial communities. While ARG burden is typically quantified in terms of abundance or prevalence, these measures do not capture the extent to which resistance is concentrated within dominant taxa at a given time. Importantly, bloominess does not act as a proxy for resistance load; taxa with similar ARG burdens can exhibit markedly different abundance dynamics, reinforcing that bloominess captures ecological visibility rather than underlying resistance capacity.

Across the dataset, ARGs were distributed across taxa with a wide range of bloominess values, indicating that resistance is not confined to either stable or highly sporadic members of the microbiome. However, taxa with elevated bloominess have a disproportionate potential to influence community-level resistance profiles, as rare but extreme increases in their abundance can transiently elevate the representation of associated ARGs.

This distinction highlights a critical difference between ARG load and ecological visibility. Taxa with low bloominess may carry substantial ARG repertoires but contribute consistently and predictably to the resistome. In contrast, highly bloomy taxa may carry fewer ARGs on average but can intermittently dominate the community, elevating their resistance signal when they dominate.

As a result, measures based solely on average abundance or prevalence may underestimate the impact of taxa capable of sporadic dominance. Incorporating bloominess into resistome analyses therefore provides a means of identifying taxa whose contribution to antimicrobial resistance is conditional, context-dependent, and potentially underestimated in cross-sectional studies.

4.4 Resistance composition and ecological context

Beyond differences in overall ARG burden, resistance profiles also varied in composition across taxa occupying different regions of prevalence-bloominess space. This suggests that ecological behaviour is linked not only to the magnitude of resistance but also to the types of resistance present.

Taxa with low bloominess, particularly those in the highly prevalent core microbiome, tended to carry consistent and predictable resistance profiles. These patterns are consistent with intrinsic or chromosomally encoded resistance mechanisms that are stably maintained within well-adapted commensal populations⁴⁰.

In contrast, taxa occupying more variable regions of the prevalence-bloominess spectrum displayed greater heterogeneity in resistance composition. These taxa often exhibited combinations of resistance traits that varied across samples, consistent with more dynamic genomic architectures and potential contributions from mobile genetic elements.

This ecological stratification of resistance composition suggests that different regions of microbiome space may be associated with distinct modes of resistance maintenance and transmission. Stable, low-bloom taxa may contribute to a persistent background resistome, whereas more variable and episodic taxa may reflect contexts in which resistance is more dynamic, transferable, or responsive to environmental conditions.

At a finer scale, resistance composition also varied across ecological space, with certain drug classes showing differential associations with bloominess and ARG burden. For example, some resistance types were broadly distributed across taxa regardless of ecological behaviour, whereas others were more frequently associated with either stable or highly sporadic taxa, suggesting that resistance architecture may be shaped by both genomic context and ecological dynamics.

4.5 Watchlist taxa: conditional risk

The prevalence-bloominess framework enables the identification of taxa that may be disproportionately important for antimicrobial resistance despite limited average abundance or prevalence. In particular, taxa exhibiting both elevated bloominess and substantial ARG burden represent candidates for conditional risk, as their contribution to the resistome may be minimal under baseline conditions but elevated during sporadic dominance.

These taxa often occupy intermediate or low-prevalence regions of microbiome space, where their presence is sporadic but their capacity for expansion is high. As a result, they may be overlooked in analyses that prioritise consistently abundant or prevalent community members, despite having the potential to transiently dominate and reshape resistome composition.

This behaviour is consistent with the concept of conditionally rare taxa, which remain at low abundance under typical conditions but can rapidly increase in response to environmental or host perturbations^{2,41}. Within the context of antimicrobial resistance, such taxa may act as taxa capable of sporadically elevating ARG signals, linking ecological dynamics to transient increases in community-level risk.

Identifying these “watchlist” taxa provides a complementary perspective to traditional surveillance approaches focused on dominant or persistent organisms. By incorporating bloominess, it becomes possible to prioritise taxa whose risk is context-dependent, potentially improving early detection of shifts in resistome structure associated with management practices, antimicrobial exposure, or environmental change.

4.6 Integrating gene-centric and ecology-centric risk

Current frameworks for assessing antimicrobial resistance risk are largely gene-centric, prioritising ARGs based on factors such as mobility, clinical relevance, and host range⁴². These approaches have been instrumental in identifying high-risk resistance determinants but typically consider genes independently of the ecological dynamics of their host organisms.

Our results suggest that integrating ecological context provides an additional and complementary dimension to AMR risk assessment. While gene-centric frameworks identify which resistance determinants have the potential to spread or impact clinical outcomes, bloominess captures the ecological conditions under which those genes may become dominant within a microbial community.

This distinction highlights that the impact of an ARG is not solely determined by its inherent properties but also by the behaviour of the taxa in which it resides. Resistance genes carried by highly stable, low-bloom taxa may contribute to a consistent background resistome, whereas the same or similar genes carried by highly bloomy taxa may show sporadic elevation in abundance, with transient but potentially significant increases in ARG representation.

In this sense, bloominess can be understood as an ecological modulator of gene-level risk. By incorporating both gene-centric and ecology-centric perspectives, it becomes possible to distinguish between resistance that is persistently present and resistance that is conditionally elevated. This integrated view provides a more nuanced framework for identifying when and where antimicrobial resistance is likely to become most impactful. We propose that antimicrobial resistance risk is jointly determined by gene-level properties and potential for ARG elevation.

4.7 Future directions: linking ecological behaviour to resistance dynamics

The framework introduced here opens several avenues for future work. Integrating bloominess with genomic context will allow direct investigation of how resistance genes are embedded within taxa exhibiting distinct ecological behaviours. Pangenomic analyses of variable taxa may reveal whether mobile genetic elements are preferentially associated with taxa capable of sporadic dominance, providing a mechanistic link between ecological responsiveness and resistance dissemination.

More broadly, bloominess enables a more nuanced interpretation of taxon-environment relationships by distinguishing between absence, stable presence, and sporadic dominance. This provides a framework for disentangling the conditions associated with colonisation from those that drive expansion, allowing ecological drivers of presence and bloom behaviour to be investigated separately. The MAGs underlying this analysis are linked to extensive metadata describing host, environmental, and management factors. Leveraging these data will enable future work to identify the conditions associated with transitions between absence,

stable presence, and bloom states, providing a route to resolving the ecological drivers of conditionally emergent taxa and their associated resistance profiles. This provides a route to resolving the ecological drivers of conditionally emergent taxa and their associated resistance profiles, and ultimately informing risk-based surveillance strategies.

Linking bloominess to gene-level variation within taxa also provides an opportunity to investigate the functional consequences of resistance acquisition. By comparing ARG coverage across samples for the same MAG, it may be possible to identify associations between resistance gene presence and bloom behaviour. For example, ARGs that are absent in high-abundance (“bloom”) samples but present in low-abundance states may impose a fitness cost, whereas ARGs enriched specifically during bloom events may confer a selective advantage under certain conditions.

Extending this framework to longitudinal datasets would enable direct observation of bloom events and their temporal dynamics. This would allow bloominess to be validated as a predictive trait, linking cross-sectional abundance distributions to real-time ecological expansion and contraction.

Translating prevalence-bloominess structure into lower-resolution profiling methods, such as 16S rRNA gene sequencing or targeted AMR panels, could enable scalable surveillance strategies. Identifying taxonomic proxies for high-risk ecological configurations may provide a route to monitoring conditionally emergent resistance in settings where genome-resolved metagenomics is not feasible.

Finally, while this study focuses on poultry microbiomes, sporadic dominance is a general ecological phenomenon. Applying this framework across host-associated and environmental microbiomes will help determine the extent to which bloominess represents a conserved ecological trait and whether similar prevalence-bloominess structures underpin functional dynamics in diverse systems.

4.8 Limitations and uncertainty

Several limitations should be considered when interpreting these results. First, bloominess is calculated based on taxa classified as present according to coverage and detection thresholds, and therefore depends on the sensitivity of metagenomic sequencing and the criteria used to define presence. In particular, low-abundance taxa near the detection limit may be inconsistently classified across samples, which could influence estimates of prevalence and bloominess.

Second, bloominess is derived from cross-sectional data and reflects the distribution of abundance across samples rather than directly observed temporal dynamics. While this provides a useful proxy for sporadic dominance, longitudinal data will be required to confirm whether taxa identified as highly bloomy indeed exhibit recurrent bloom events over time³⁹.

Third, the estimation of bloominess is constrained by the number of samples in which a taxon is detected. Rare taxa, present in only a small number of samples, have limited opportunity to exhibit extreme abundance variation, while highly prevalent taxa are similarly constrained by consistently stable abundance profiles. These boundary effects contribute to the observed structure in prevalence-bloominess space and should be considered when interpreting patterns at the extremes.

An additional limitation arises from the use of metagenome-assembled genomes (MAGs) as the primary unit of analysis. While this approach captures chromosomally integrated resistance and enables linkage between ARGs and their host taxa, it may not fully resolve plasmid-borne or other mobile genetic elements that are not confidently assigned to specific genomes. As a result, some taxa classified as lacking ARGs may in fact harbour unobserved mobile resistance determinants, and associations between bloominess and resistance may be partially influenced by genomic elements not captured in MAG-based analyses. Incorporating approaches that explicitly link mobile genetic elements to their host taxa, such as proximity ligation (e.g. Hi-C) or advanced co-association methods, will be important for fully characterising the relationship between ecological behaviour and mobile resistance.

Finally, while associations between bloominess and resistance composition suggest potential links between ecological behaviour and resistance dynamics, this study does not directly resolve the genomic context or functional impact of individual ARGs. In particular, distinguishing between intrinsic and mobile resistance mechanisms, or assessing fitness costs and benefits associated with resistance acquisition, will require targeted genomic and experimental approaches.

4.9 Conclusions

In this study, we introduce bloominess as a taxon-level ecological trait capturing the propensity for sporadic dominance across samples. By separating prevalence from abundance variability, this framework reveals a structured dimension of microbiome organisation in which taxa differ not only in how frequently they are detected but in how they distribute their abundance when present.

Applying this framework to the poultry caecal microbiome demonstrates that ecological behaviour shapes the expression of antimicrobial resistance. While ARGs are broadly distributed across taxa, their impact on the resistome depends on the dynamics of their hosts, with highly bloomy taxa acting as potential sources of elevated ARG signal during sporadic dominance.

Together, these findings support a model in which antimicrobial resistance is jointly determined by gene-level properties and potential for ARG elevation. Incorporating ecological behaviour into resistome analysis provides a complementary perspective to gene-centric frameworks and offers a route to identifying context-dependent risks that may be overlooked by approaches based solely on abundance or prevalence.

References

1. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5970–5975 (2016).
2. Shade, A. *et al.* Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. *mBio* **5**, 10.1128/mbio.01371-14 (2014).
3. Grilli, J. Macroecological laws describe variation and diversity in microbial communities. *Nat. Commun.* **11**, 4743 (2020).
4. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
5. Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12 (2012).
6. Vandeputte, D. *et al.* Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017).
7. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
8. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649–662.e20 (2019).
9. Wang, Y. *et al.* The multi-kingdom microbiome catalog of the chicken gastrointestinal tract. *Biosaf. Health* **6**, 101–115 (2024).
10. Van Boeckel, T. P. *et al.* Global trends in antimicrobial resistance in animals in low- and middle-income countries. *Science* **365**, eaaw1944 (2019).
11. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655 (2022).
12. Forslund, K. *et al.* Country-specific antibiotic use practices impact the human gut resistome. *Genome Res.* **23**, 1163–1169 (2013).

- 1005 13. Glendinning, L., Stewart, R. D., Pallen, M. J., Watson, K. A. & Watson, M. Assembly of
1006 hundreds of novel bacterial genomes from the chicken caecum. *Genome Biol.* **21**, 34
1007 (2020).
- 1008 14. Glendinning, L. *et al.* Altitude-dependent agro-ecologies impact the microbiome diversity of
1009 scavenging indigenous chicken in Ethiopia. *Microbiome* **12**, 138 (2024).
- 1010 15. Feng, Y. *et al.* Metagenome-assembled genomes and gene catalog from the chicken gut
1011 microbiome aid in deciphering antibiotic resistomes. *Commun. Biol.* **4**, 1–9 (2021).
- 1012 16. Gilroy, R. *et al.* Extensive microbial diversity within the chicken gut microbiome revealed by
1013 metagenomics and culture. *PeerJ* **9**, e10941 (2021).
- 1014 17. Segura-Wang, M., Grabner, N., Koestelbauer, A., Klose, V. & Ghanbari, M. Genome-Resolved
1015 Metagenomics of the Chicken Gut Microbiome. *Front. Microbiol.* **12**, 2390 (2021).
- 1016 18. Xin, C. *et al.* Prevalence of *Campylobacter* and non-typhoidal *Salmonella* along broiler
1017 chicken production and distribution networks, Northern Vietnam. *PLoS Negl. Trop. Dis.* **19**,
1018 e0013615 (2025).
- 1019 19. Lehri, B. *et al.* Geographic clustering and population structures of *Campylobacter jejuni* and
1020 *Campylobacter coli* in South and Southeast Asian poultry systems. *Microb. Genomics* **12**,
1021 001706 (2026).
- 1022 20. Pelligand, L. *et al.* High prevalence of antimicrobial residues in broiler meat in Bangladesh,
1023 India and Vietnam. *Npj Sci. Food* <https://doi.org/10.1038/s41538-026-00806-7> (2026)
1024 doi:10.1038/s41538-026-00806-7.
- 1025 21. Pandit, R. J. *et al.* Microbial diversity and community composition of caecal microbiota in
1026 commercial and indigenous Indian chickens determined using 16s rDNA amplicon
1027 sequencing. *Microbiome* **6**, 115 (2018).
- 1028 22. Hay, M. C. *et al.* *Chicken Caecal Enterotypes in Indigenous Kadaknath and Commercial*
1029 *Cobb Chicken Lines Are Associated with Campylobacter Abundance and Influenced by*

1030 *Farming Practices*. <https://www.researchsquare.com/article/rs-2381640/v1> (2023)

1031 doi:10.21203/rs.3.rs-2381640/v1.

1032 23. De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read
 1033 sequencing data. *Bioinformatics* **39**, btad311 (2023).

1034 24. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node
 1035 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
 1036 *Bioinformatics* **31**, 1674–1676 (2015).

1037 25. Antipov, D., Korobeynikov, A., McLean, J. S. & Pevzner, P. A. hybridSPAdes: an algorithm for
 1038 hybrid assembly of short and long reads. *Bioinformatics* **32**, 1009–1015 (2016).

1039 26. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately
 1040 reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165
 1041 (2015).

1042 27. Nissen, J. N. *et al.* Improved metagenome binning and assembly using deep variational
 1043 autoencoders. *Nat. Biotechnol.* **39**, 555–560 (2021).

1044 28. Pan, S., Zhao, X.-M. & Coelho, L. P. SemiBin2: self-supervised contrastive learning leads to
 1045 better MAGs for short- and long-read sequencing. *Bioinformatics* **39**, i21–i29 (2023).

1046 29. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication,
 1047 aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).

1048 30. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate
 1049 genomic comparisons that enables improved genome recovery from metagenomes through
 1050 de-replication. *ISME J.* **11**, 2864–2868 (2017).

1051 31. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and
 1052 accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods*
 1053 **20**, 1203–1212 (2023).

1054 32. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a
1055 metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**,
1056 725–731 (2017).

1057 33. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify
1058 genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).

1059 34. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a
1060 phylogenetically consistent, rank normalized and complete genome-based taxonomy.
1061 *Nucleic Acids Res.* **50**, D785–D794 (2022).

1062 35. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–
1063 4574 (2021).

1064 36. Eren, A. M. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data.
1065 *PeerJ* **3**, e1319 (2015).

1066 37. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
1067 identification. *BMC Bioinformatics* **11**, 119 (2010).

1068 38. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination
1069 of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci.*
1070 *Rep.* **11**, 12728 (2021).

1071 39. Faust, K., Lahti, L., Gonze, D., de Vos, W. M. & Raes, J. Metagenomics meets time series
1072 analysis: unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **25**, 56–66
1073 (2015).

1074 40. Martinez, J. L. Environmental pollution by antibiotics and by antibiotic resistance
1075 determinants. *Environ. Pollut.* **157**, 2893–2902 (2009).

1076 41. Lennon, J. T. & Jones, S. E. Microbial seed banks: the ecological and evolutionary
1077 implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119–130 (2011).

1078 42. Zhang, A.-N. *et al.* An omics-based framework for assessing the health risk of antimicrobial
1079 resistance genes. *Nat. Commun.* **12**, 4765 (2021).

Supplementary Materials

Supplementary Methods

Supplementary Methods SM1. Nanopore read preprocessing and deduplication

Description of Chopper filtering thresholds, chemistry-specific read retention, and custom read-name deduplication procedures.

Supplementary Methods SM2. Metagenomic co-assembly and rare-contig recovery pipeline

Full parameters for MEGAHIT assemblies, including k-mer schedules and secondary assembly of unmapped reads used to recover low-coverage taxa.

Supplementary Methods SM3. Ensemble binning workflow

Implementation details for MetaBAT2, VAMB, and SemiBin2 binning procedures, including coverage matrices and integration using DAS Tool.

Supplementary Methods SM4. Dereplication and genome quality filtering

Detailed description of dRep clustering thresholds, species-level clustering criteria, CheckM2 completeness/contamination filtering.

Supplementary Methods SM5. Bloominess robustness and sensitivity analysis

Evaluation of bloominess stability across alternative detection thresholds, coverage thresholds, and subsampling regimes. Includes rank stability analyses and threshold-sensitivity grids.

Supplementary Method SM6. Drug class bloominess heatmap and statistical analysis

Describes construction of the drug class $\times$ prevalence class bloominess heatmap and the chi-squared framework used to test for non-random associations between ARG drug class and bloominess. Includes the carrier versus non-carrier Wilcoxon rank-sum test and violin plot methodology.

Supplementary Method SM7. Prevalence class relative abundance and core MAG co-absence analysis

Describes calculation of the proportional contribution of core, variable, and rare MAGs to total community coverage per sample, and the co-absence analysis used to identify pairs of core MAGs that are consistently absent together across samples.

1110 **Supplementary Tables**

1111 **Supplementary Table S1. Illumina sequencing runs and accession numbers**

1112 Sequencing metadata, run identifiers, and accession numbers for all Illumina datasets.

1113 **Supplementary Table S2. Nanopore sequencing runs and flow cell metadata**

1114 Sequencing platform, chemistry, flow cell information, and sample identifiers for long-read
1115 datasets.

1116 **Supplementary Table S3. Quality statistics for curated MAG catalogue**

1117 Completeness, contamination, genome size, and N50 values for all retained MAGs.

1118 **Supplementary Table S4. Taxonomic assignments for MAG catalogue**

1119 GTDB-Tk classifications for all MAGs across taxonomic ranks.

1120 **Supplementary Table S5. MAG prevalence and bloominess metrics**

1121 Complete dataset of prevalence values, bloominess scores, abundance metrics, and
1122 prevalence class assignments for each MAG.

1123

1124 **Supplementary Figures**

1125 **Supplementary Figure S1. Illumina read mapping rates across all 120 samples**

1126 Bar chart showing the percentage of Illumina reads mapping to the 1374-MAG catalogue for
1127 each sample, sorted by mapping rate. Four samples with <10% mapping (Eimeria-dominated
1128 libraries) are highlighted in red and were excluded from downstream abundance profiling.

1129 **Supplementary Figure S2. Nanopore read retention after quality filtering**

1130 Distribution of retained reads following chemistry-specific Chopper filtering.

1131 **Supplementary Figure S3. Genome quality distribution of MAG catalogue**

1132 Completeness and contamination distributions following dereplication and quality filtering.

1133 **Supplementary Figure S4. Taxonomic composition of the MAG atlas**

1134 Distribution of MAGs across phyla and classes according to GTDB-Tk taxonomy.

1135 **Supplementary Figure S5. Sensitivity of prevalence estimates to detection thresholds**

1136 Comparison of prevalence estimates across alternative coverage and detection thresholds.

1137 **Supplementary Figure S6. Stability of bloominess rankings across parameter settings**

1138 Rank correlation of bloominess estimates across alternative presence thresholds.

1139 **Supplementary Figure S7. Relative abundance of prevalence classes across samples**

1140 Stacked bar chart showing the proportional contribution of core, variable, and rare MAGs to
1141 total community coverage in each of the 116 retained samples, ordered by region and
1142 sampling source.

1143 **Supplementary Figure S8. Co-absence patterns among core MAGs**
1144 Heatmap of pairwise co-absence scores for the 18 core MAGs (prevalence $\geq 85\%$),
1145 highlighting pairs that are consistently absent together across samples more often than
1146 expected by chance.

1147 **Supplementary Figure S9. Violin plot of bloominess in carriers vs noncarriers by**
1148 **resistance drug class.**
1149 Wilcoxon rank-sum tests comparing bloominess between drug-class carriers and non-carriers
1150 among variable MAGs.

1151 [Supplementary Data](#)

1152 **Supplementary Dataset 1. Species-level MAG catalogue**
1153 FASTA files and metadata for representative species-level MAGs. (To be hosted on Zenodo,
1154 link after acceptance)

1155 **Supplementary Dataset 2. ARG locus catalogue**
1156 Genomic loci of antimicrobial resistance genes identified across MAGs. (To be hosted on
1157 Zenodo, link after acceptance)

1158

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS5coveredetectionqc.png](#)
- [FigS4taxonomiccomposition.png](#)
- [SM1nanoporepreprocessing.pdf](#)
- [SM7prevalenceclassabundanceandcorecoabsence.md](#)
- [SupplementaryTableS4MAGtaxonomy.xlsx](#)
- [FigS7prevalenceclassrelabund.png](#)
- [SupplementaryTableS3MAGquality.xlsx](#)
- [FigS1readmapping.png](#)
- [FigS6bloominessmetriccomparison.png](#)
- [SM2assembly.pdf](#)
- [SM5coverageprofilingandbloominess.pdf](#)
- [SM3readmapping.pdf](#)
- [FigS3MAGquality.png](#)
- [SM6drugclassbloominessheatmapandstatistics.pdf](#)
- [SM4binning.pdf](#)
- [FigSbloominesscarriersvsnoncarriersviolin.png](#)
- [FigS8coreMAGcoabsence.png](#)
- [SupplementaryTableS2Nanoporesequencing.xlsx](#)
- [SupplementaryTableS1Illuminasequencing.xlsx](#)
- [FigS2nanoporereadretention.png](#)
- [RS630.pdf](#)