

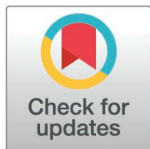
RESEARCH ARTICLE

A metabolic atlas of the *Klebsiella pneumoniae* species complex reveals lineage-specific metabolism and capacity for intra-species co-operation

Ben Vezina^{1,2*}, Helena B. Cooper^{1,2}, Christopher K. Barlow³, Martin Rethoret-Pasty⁴, Sylvain Brisse⁴, Jonathan M. Monk⁵, Kathryn E. Holt^{1,6}, Kelly L. Wyres^{1,2,6*}

1 Department of Infectious Diseases, School of Translational Medicine, Monash University, Melbourne, Victoria, Australia, **2** Centre to Impact AMR, Monash University, Clayton, Victoria, Australia, **3** Monash Proteomics and Metabolomics Platform, Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia, **4** Institut Pasteur, Université Paris Cité, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, France, **5** Department of Medicine, University of California, San Diego School of Medicine, San Diego, California, United States of America, **6** Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom

* benjamin.vezina@monash.edu (BV); Kelly.wyres@monash.edu (KLW)



OPEN ACCESS

Citation: Vezina B, Cooper HB, Barlow CK, Rethoret-Pasty M, Brisse S, Monk JM, et al. (2025) A metabolic atlas of the *Klebsiella pneumoniae* species complex reveals lineage-specific metabolism and capacity for intra-species co-operation. PLoS Biol 23(12): e3003559. <https://doi.org/10.1371/journal.pbio.3003559>

Academic Editor: Sebastian E. Winter, University of California Davis School of Medicine, UNITED STATES OF AMERICA

Received: March 4, 2025

Accepted: November 26, 2025

Published: December 12, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pbio.3003559>

Copyright: © 2025 Vezina et al. This is an open access article distributed under the terms of

Abstract

The *Klebsiella pneumoniae* species complex inhabits a wide variety of hosts and environments, and is a major cause of antimicrobial resistant infections. Genomics has revealed the population comprises multiple species/sub-species and hundreds of distinct co-circulating sub-lineage (SLs) that are associated with distinct gene complements. A substantial fraction of the pan-genome is predicted to be involved in metabolic functions and hence these data are consistent with metabolic differentiation at the SL level. However, this has so far remained unsubstantiated because in the past it was not possible to explore metabolic variation at scale. Here, we used a combination of comparative genomics and high-throughput genome-scale metabolic modeling to systematically explore metabolic diversity across the *K. pneumoniae* species complex ($n=7,835$ genomes). We simulated growth outcomes for each isolate using carbon, nitrogen, phosphorus, and sulfur sources under aerobic and anaerobic conditions ($n=1,278$ conditions per isolate). We showed that the distributions of metabolic genes and growth capabilities are structured in the population, and confirmed that SLs exhibit unique metabolic profiles. In vitro co-culture experiments demonstrated reciprocal commensalistic cross-feeding between SLs, effectively extending the range of conditions supporting individual growth. We propose that these substrate specializations may promote the existence and persistence of co-circulating SLs by reducing nutrient competition and facilitating commensal interactions. Our findings have implications for understanding the eco-evolutionary dynamics of *K. pneumoniae*

the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All genomic data used is publicly available—accessions and genomes used can be found in [S1 Data](#). Raw metabolomics files are found at <https://doi.org/10.5281/zenodo.17705463>. All other raw data can be found in [S2 Data–S6 Data](#). Phylogenetic trees, R, and MICOM code can be found at Figshare (<https://doi.org/10.6084/m9.figshare.24503737>). Interactive phylogeny can be found at <https://microreact.org/project/pcTMwQZAGCsqqhbEZzaj11-a-metabolic-atlas-of-the-klebsiella-pneumoniae-species-complex>.

Funding: This work was funded by the Australian Research Council Discovery Project (DP200103364, awarded to K.L.W., K.E.H., S.B., and J.M.M.). Additionally, K.L.W. was supported by a National Health and Medical Research Council of Australia Investigator Grant (APP1176192). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: K.E.H. is a member of PLOS Biology's Editorial Board. The other authors declare that no competing interests exist.

Abbreviations: ANI, average nucleotide identity; COG, clusters of orthologous genes; GFA, graphical fragment assembly; HILIC, hydrophilic interaction liquid chromatography; HRMS, high-resolution mass spectrometry; KpSC, *Klebsiella pneumoniae* species complex; LC-MS, Liquid Chromatography-Mass Spectrometry; LIN, life identifier number; SL, sub-lineages; STs, sequence types.

and for the design of novel strategies to prevent opportunistic infections caused by this World Health Organization priority antimicrobial resistant pathogen.

Introduction

Klebsiella pneumoniae is a ubiquitous gram-negative bacterium and major cause of opportunistic healthcare-associated infections with significant global burden [1]. Novel prevention and control strategies are urgently required to target this organism, in particular the growing numbers of multi-drug resistant strains that cause infections that are extremely difficult to treat [2]. However, our capacity to design effective control measures is complicated by the extraordinary diversity among *K. pneumoniae* and gaps in our knowledge about how key types of variation, such as metabolic variation, intersect with ecology, and population structure.

Isolates identified as *K. pneumoniae* in clinical laboratories include *K. pneumoniae sensu stricto* in addition to six other taxa (species and sub-species) collectively known as the *K. pneumoniae* species complex (KpSC) [3]. These species are separated by 3%–4% nucleotide divergence in their core genes, and each is further subdivided into hundreds or thousands of phylogenetically distinct sub-lineage (SLs) separated by ~0.5% nucleotide divergence [3,4], but with access to a highly diverse shared gene pool. It is well known that gene content differences play a role in driving differential SL epidemiologies, i.e., where a subset is recognized as globally-distributed multi-drug resistant agents of healthcare-associated infections or as 'hypervirulent' community-acquired pathogens (usually drug-susceptible) [3]. It has also been shown that total gene content is nonrandomly distributed between SLs and that up to 37% of the total KpSC gene pool encodes proteins contributing to cellular metabolism [4].

We therefore hypothesized that SLs differ in terms of metabolic capabilities, which likely contributes to their epidemiological behaviors and may facilitate the maintenance of diverse SLs within the population. A similar hypothesis was proposed for the gram-positive bacterial pathogen, *Streptococcus pneumoniae* [5], and is consistent with genomic studies of other organisms that have implicated a nonrandom distribution of metabolic traits within the population [6–10]. However, the complexities of cellular metabolism make systematic prediction of phenotypes from genome data difficult, limiting capacity for large-scale analyses of population metabolism.

Genome-scale metabolic models attempt to encapsulate the total metabolic potential of an organism and can be used to predict metabolic phenotypes [11]. Comparative metabolic modeling analyses can be utilized to probe diversity within bacterial species populations, and recent larger-scale studies ($n=50$ –3,083 genomes) have highlighted the power of these approaches to identify industrially relevant, ecology-, drug-resistance- and pathogenicity- associated metabolic traits [12–18]. While a recent study of 2,773 metabolic models confirmed that *Escherichia coli* phylogroups (which are approximately as diverse as KpSC species) are associated with unique metabolic reactions [19]. However, none of these works have attempted to investigate the interplay between phenotypic diversity and fine-scale population structure, i.e., to understand how diversity is distributed within and between phylogenetic SLs.

Such information is crucial to understanding the interaction between metabolism, ecology and evolution within species, and in turn for design of broad-acting control strategies that seek to exploit metabolic vulnerabilities.

Here, we describe a population-level analysis of 7,835 *KpSC* genomes to evaluate the distribution and diversity of metabolism in the context of the population structure, first at the level of species, then SLs within species. We performed quantitative analyses of metabolic gene content, and constructed metabolic models from which we predicted 1,278 metabolic phenotypes for each isolate. This approach allowed us to directly infer the phenotypic impact of genetic variation and explore phenotype variation within a single species in the context of the established *KpSC* population genomics framework [3] and at a scale that was not previously possible. We validated a subset of these phenotypes via experimental assays. Our data reveal considerable variation between and within species that is nonrandomly distributed across SLs. Notably, we show for the first time that *KpSC* SLs, including multi-drug resistant and hypervirulent clones, are associated with distinct substrate usage profiles that appear to facilitate commensal interactions and which we hypothesize may promote co-existence.

Materials and methods

Genome acquisition, quality filtering, and genotyping

Genomes were included from 29 medium-large-scale studies [4,20–46] that comprised diverse *KpSC* collections (S1 Data). Outbreak investigations were explicitly excluded to reduce clonal oversampling and genomes were de-replicated to minimize sampling biases for overrepresented SLs (see method: Genome dereplication). Reads were acquired from the European Nucleotide Archive using the `enaDataGet` and `enaGroupGet` scripts from `enaBrowserTools` v1.6 [47], trimmed with `Trim Galore` v0.5.0 using `-q 20` option [48], then assembled with `Unicycler` v0.4.7 with the `--keep 0` option [49]. Assemblies failing the established KlebNET-Genomic Surveillance Platform inclusion criteria were excluded (>500 contigs, total length <4,969,898 or >6,132,846 bp). Graphical Fragment Assembly (GFA) files were analyzed to count GFA dead ends using `getUnicyclerGraphStats.py` from `Unicycler` [49].

Assemblies were annotated with `Bakta` v1.1.1 [50] with the `'--gram --'` option and `Bakta` database (date accessed: 1/09/2021), while `Kleborate` v2.0.4 [51] was used to identify species and Sequence Types (STs). `Pathogenwatch` (<https://pathogen.watch/>) and `BIGSdb` (<https://bigsdb.pasteur.fr/>) [52,53] were used to assign LIN codes [54] and SLs. LIN codes are assigned using core genome MLST profiles that consider variation in 629 core *KpSC* genes. The third position in the LIN code is used to infer SL membership [54] and results in groupings that correspond to well-recognized deep-branching monophyletic clades within species. Notably, these SL assignments do not rely on *de novo* cluster analysis, and are stable with the addition of new genomes [55]. Pairwise ANI was calculated using `FastANI` v1.33 [56] using `--fragLen 3000`.

Genome dereplication

We noticed that several SLs were vastly overrepresented in the dataset (e.g., due to dense sampling in specific geographies or inclusion of transmission clusters within larger collections). In order to minimize the impact of these sampling biases we de-replicated the dataset using a similar approach to that described in [51]. Initially, `popPUNK` v2.4.0 [57] was used to finely cluster isolates. The `'create-db'` function was used with the following options: `'--sketch-size 1000000 --min-k 15 --max-k 29 --qc-filter prune'`. Then the `'fit-model'` function was used with the following options: `'dbscan --ranks 1,2,3,5 --graph-weights'`. Finally, the `'fit-model'` function was used again, with the options: `'refine --graph-weights --unconstrained'`. Additionally, the `'poppunk_visualise'` function was used, with the `'--distances'` and `'--previous-clustering'` utilizing the refined model fit, to output a neighbor-joining core tree.

Isolates which shared the same `popPUNK` cluster, isolation source and country were sorted using `newfangled_untangler.py` v1.0.4 (https://github.com/bananabenana/newfangled_untangler/) then de-replicated via `Assembly-Dereplicator` v0.1.0 [58] with the following options: `'--threshold 0.0003'`. Code also available at Figshare (<https://doi.org/10.6084/m9.figshare.24503737>).

Pan-genome analysis

The 7,835 filtered and dereplicated *KpSC* genomes were split into species, then Panaroo v1.2.8 [59] was run with the following options: ‘--clean-mode sensitive -a core --aligner mafft --no_clean_edges --core_threshold 0.98 --merge_paralogs --remove-invalid-genes’. Due to computational limits, the 6,676 *K. pneumoniae* genomes were split into two groups of 3,338 (based on closest genetic distance) and run separately on Panaroo. The output graphs for all species were then merged using the panaroo-merge command with the following options: ‘--merge_paralogs’ to obtain the final pan-genome.

Metabolic gene identification and analysis

To analyze metabolic pathways, the pan_genome_reference.fa file from Panaroo was translated using the transeq tool from EMBOSS v6.6.0 [60] with the following options: “-table 11 -frame 1.” Translated sequences were analyzed using kofam_scan v1.3.0 (https://github.com/takaram/kofam_scan), the command line version of KEGG’s [61] kofamKOALA [62]. KEGG release 94.0 ver. 2020-04-02.

Genes with an e-value of ≤ 0.001 were retained and analyzed using KEGG Mapper (<https://www.genome.jp/kegg/mapper/reconstruct.html>) to identify pathway information. Using the BRITE tab, genes which fell under the ‘metabolism’ subheading were used to classify genes as metabolic or nonmetabolic. KEGG ortholog assignments were used to convert the Panaroo gene presence/absence matrix into a metabolic ortholog matrix. This allowed grouping of functional isozymes which would otherwise have an identical KEGG function. Pairwise metabolic ortholog Jaccard similarities were calculated using the vegdist function from R package vegan v2.5-7 [63].

Twilight [64] was used to classify distribution of metabolic orthologs among *KpSC* taxa and 48 common *K. pneumoniae* SLs ($n \geq 15$ genomes each, defined by LIN codes) using the -s 5 and -s 20 commands, respectively. ‘Core’: $\geq 95\%$ prevalence. ‘Intermediate’: $<95\%$ and $>15\%$ prevalence. ‘Rare’: $\leq 15\%$ and $>0\%$ prevalence. ‘Absent’: 0% prevalence.

Metabolic modeling

Metabolic models were generated via Bactabolize v1.0.1 [65] using the *draft_model* command with the following options: “--media_type m9 --atmosphere_type aerobic --min_coverage 25 --min_pident 80.” The *KpSC*-pan v2.0 model [66] was used as the input reference. One-hundred and sixty of 7,835 models (2.0%) required gap-filling to enable simulation of growth on M9 minimal media with glucose (median: 1 reaction added, range: 1–23), using flux balance analysis to optimize the most recent version of the biomass objective function “BIOMASS_Core_Feb2022.” We have previously shown that models gap-filled to simulate growth in minimal media plus glucose can accurately predict a range of other substrate growth phenotypes [65].

Flux balance analysis was used to predict growth phenotypes for all possible carbon, nitrogen, phosphorus, and sulfur sources supported by the reference model as sole sources of carbon, nitrogen, phosphorus, or sulfur in M9 minimal media defined in silico, in both aerobic and anaerobic conditions. These analyses were performed using Bactabolize’s *fba* command with the following option: “--fba_spec_name m9.” All growth lower bounds and default substrate sources used in this analysis are available at <https://github.com/kelwyres/Bactabolize>. Models were considered capable of simulating growth when the optimized biomass value was ≥ 0.0001 .

Pairwise Jaccard similarities and population distribution of positive growth phenotypes were calculated as described above for metabolic orthologs.

To evaluate co-occurrence of metabolic traits and control for population structure, genomes were randomly subsampled to a maximum of 10 per SL ($n = 2,427$ genomes), then predicted aerobic growth phenotypes analyzed using Coinfinder v1.2.0 [67] with the --associate option.

In vitro growth experiments

To validate substrate usage predictions from the metabolic models, 13 isolates (S2 Data) were selected from our in-house collections [4,26,68] and tested in triplicate on each of nine substrates in minimal media, aerobic conditions. Two

experiments were performed: i) endpoint OD₆₀₀ at 24- and 48-hours; ii) 3-day substrate coaxing to induce gene expression [69] for predicted false-negative growth results in the endpoint experiment. Isolates were coaxed onto the testing substrate via sub-culturing into progressively lower amounts of D-glycerol, which all isolates can use as a carbon source (10 mM on subculture 1, 5 mM on subculture 2, 0 mM on subculture 3).

Nine growth substrates were tested: seven as sole sources of carbon (methanol, L-hydroxyproline, xylitol, butyrate, acetoacetate, D-glycerol, and D-glucose), one as a sole source of nitrogen (allantoin) and one as a sole source of sulfur (methanesulfonate).

For endpoint experiments, isolates were grown overnight in 2 mL M9 media + 10 mM D-glycerol. M9 media was prepared by using 1X M9 Minimal Salts (Sigma) and adding 2 mM MgSO₄ and 0.1 mM CaCl₂ after autoclaving. 1 mL of culture was pelleted at 8,000 x g for 10 min, supernatant discarded, then centrifugally washed using 1 mL 0.9% NaCl to remove excess carbon/nitrogen/sulfur. Isolates were then diluted 1:100 using 0.9% NaCl and 5 µL sub-cultured into 96-well plates (CLS3603, Corning) containing 200 µL M9 + substrate. Substrates were tested at the following concentrations: 400 mM methanol, 20 mM L-hydroxyproline, 20 mM xylitol, 20 mM sodium butyrate, 20 mM acetoacetate, 20 mM D-glucose, 10 mM D-glycerol 20 mM methanesulfonate, 60 mM allantoin, all at pH 7.0 and 0.2 µm filter sterilized. For allantoin as a nitrogen source, custom nitrogen-negative M9 was prepared (34 mM NaH₂PO₄, 64 mM K₂HPO₄, 1 µM FeSO₄, 2 mM MgSO₄, and 0.1 mM CaCl₂). For methanesulfonate as a sulfur source, sulfur-negative M9 was prepared (2 mM MgCl₂ instead of MgSO₄). Isolate negative and carbon positive controls were performed for each media. Plates were incubated at 37 °C, shaking at 150 RPM and optical density read every 24 hours at OD₆₀₀ absorbance using an infinite 200Pro and i-control v2.0.10.0 (Tecan). Each experiment was performed in biological triplicate. Each well was blanked with a no-isolate control, then the mean used to determine growth, with a cutoff of 0.05 (limit of detection).

Coaxing experiments were performed as above, over 3 days. Isolates were added to media containing the tested substrate plus D-glycerol 10 mM on subculture day 1, 5 mM on subculture day 2, and 0 mM on subculture day 3. Prior to sub-culturing, cells were centrifugally washed as above and 5 µL transferred to 200 µL total media. Only subcultures on day 3 were used for analysis (0 mM D-glycerol).

Co-culture experiments

We selected three pairs of SLs and three corresponding pairs of substrates (four total substrates) for which one was predicted as a SL-specific core trait in SL A and the other predicted as a SL-specific core trait in SL B. For each SL, three isolates were selected from our in-house collections [4,26,70] and treated as biological replicates (S3 Data).

Isolate pairs were grown in single and co-culture using 0.4 µm pore Transwells (Sigma) and 24-well plates (Costar, Corning). Co-culture results were compared to single cultures and log₂-fold-change OD₆₀₀ values calculated.

To rule out strain-strain antagonism prior to co-culturing, a primary inoculum was streaked onto LB agar plates and grown for 8 hours at 37 °C. Then, primary inoculums of additional isolates were streaked perpendicularly. Plates were incubated for 16 hours at 37 °C and no growth inhibition of the perpendicular streaks was observed.

Co-culture results were compared to single cultures using OD₆₀₀ of pairs of isolates (S3 Data) that were physically separated by 0.4 µm pores using 24-well Transwells (Sigma). All cultures were grown at 37 °C, 100 RPM. Substrates were used in 1X M9 media at the following concentrations: 20 mM for galactitol, L-sorbose, L-tartrate, and 5 mM for L-hydroxyproline.

M9 media plus the tested substrate was added to the Transwell (100 µL) and the main well (800 µL). Isolates were grown overnight in 1X M9 containing 10 mM D-glycerol, then centrifugally washed as above. Cultures were diluted 1:100 in 0.9% NaCl, and 5 µL added to wells. The prototroph was added to the Transwell while auxotroph added to the main well. Cultures were grown and 75 µL was taken from wells at time point for OD₆₀₀ measurements. For isolate-substrate combinations where we observed slow growth rates, 24, 48, and 72-hour time points were used, otherwise time points at 0, 24, and 36 hours were used. OD₆₀₀ values were blanked using no-isolate media controls. For log₂ fold-change analysis,

values of 0 were transformed to 1×10^{-5} , but kept raw for other statistical analyses. After co-culture, isolates were then plated onto LB agar, confirmed uniform, distinct colony morphology, then re-grown under the same nutrient conditions in single culture.

Co-culture simulations

In silico co-culture experiments were performed using MICOM v0.32.2 [71] with the relevant *K. pneumoniae* metabolic models. Community models were built using the build function, then communities were grown in growth media matching co-culture experiments (M9 minimal media with various carbon sources) with tradeoff set to 0.5. Metabolite and reaction flux were then analyzed to determine metabolite sharing from prototrophs to auxotrophs.

Preparation of samples for metabolomics analysis

We used an untargeted Liquid Chromatography-Mass Spectrometry (LC-MS) approach to explore the metabolites present in the culture supernatant of putative cross-feeding prototrophs. We selected one representative prototroph for each SL-growth substrate combination: AJ155, INF359, INF120 (galactitol), INF171 (L-hydroxyproline), INF225 (L-sorbose), and INF354 (L-tartrate) and compared to no-isolate controls for each substrate. Briefly, overnight cultures in M9 minimal media + 20 mM glucose were washed twice in 0.9% NaCl, centrifuged at $8,000 \times g$ for 10 min. Cultures were then standardized to McFarland Standard 0.45–0.55 in 0.9% NaCl, then inoculated into 5 mL M9 minimal media + substrates (as previously described in co-culture experiments). Performed with four biological replicates except negative controls which were performed in triplicate. Independent cultures were grown aerobically at 37 °C, shaking at 200 RPM for 18 hours. Cells were pelleted at $8,000 \times g$ for 10 min, then supernatant filtered through a Acrodisc Supor PES Membrane 0.2 μ m syringe filter (Pall). Supernatant was transferred to a new vial and frozen at –80 °C prior to transfer to the Monash Proteomics and Metabolomics Platform for LC-MS analysis, where 20 μ L of cell culture supernatant was treated with 180 μ L of ice-cold methanol containing a mixture standard isotope labeled amino acids acting as internal standards. The sample was then mixed at 4 °C for 10 min on a thermomixer before being subjected to centrifugation at $21,000 \times g$ for 10 min at 4 °C. 900 μ L of the supernatant was then transferred to an LC-MS vial for analysis.

LC-MS analysis of metabolites

We used both hydrophilic interaction liquid chromatography (HILIC) coupled to high-resolution mass spectrometry (HRMS). Samples were analyzed on a Vanquish Horizon coupled to Exploris 480 orbitrap mass spectrometer (Thermo Scientific). Chromatography was performed using an iHILIC-(P) Classic, (PEEK, 150 x 4.6 mm, 5 μ m, 200Å; HILICON, Sweden) with a gradient elution of 20 mM ammonium carbonate (A) and acetonitrile (B) (linear gradient time-%B as follows: 0 min-80%, 15 min-50%, 18 min-5%, 21 min-5%, 24 min-80%, 32 min-80%) at a flow rate of 500 μ L/min. The mass spectrometer was operated in polarity switching modes with a resolution at 120,000. The electrospray conditions were: ionization voltage was 3.5 kV in positive and –2.5 kV in negative mode, ion transfer tube temperature = 325 °C; sheath gas = 60; Aux gas = 15; sweep gas = 2; vaporizer temp = 350 °C.

For accurate metabolite identification, a standard library of ~400 metabolites was analyzed before sample testing and accurate retention time for each standard was recorded. This standard library also forms the basis of a retention time prediction model used to provide putative identification of metabolites not contained within the standard library [72]. Acquired LC-MS/MS data was processed in an untargeted fashion using opensource software IDEOM [73], which initially used msConvert (ProteoWizard) [74] to convert raw LC-MS files to mzXML format and XCMS [75] to pick peaks to convert to peakML files. Mzmatch was subsequently used for sample alignment and filtering [76]. Peak intensities were compared for each prototroph versus the matched no-isolate control using a Wilcox Rank Sum test with Benjamini-Hochberg multiple testing correction. Missing and zero peak intensity data were imputed as 20% of the minimum peak intensity for each prototroph-control pair.

Data visualization and code availability

Statistical analysis and graphical visualization were performed using R v4.0.3 [77], RStudio v1.3.1093 [78], with the following software packages: tidyverse v1.3.1 [79], viridis v0.5.1 [80], RColorBrewer v1.1-2 [81], ggpubr v0.4.0 [82], ggpmisc v0.4.4 [83], aplot v0.1.6 [84], colorspace v2.0-2 [85], ggtree v2.4.1 [86], vegan v2.5-7 [63], rstatix v0.7.0 [87], and ggnewscale v0.4.5 [88].

All code used to generate metabolic models and simulate growth phenotypes, including media definitions, is available at <https://github.com/kelwyres/Bactabolize>. The *KpSC* pan-reference metabolic model is available at <https://github.com/kelwyres/KpSC-pan-metabolic-model>. All code used to generate figures, simulations, and perform statistical analysis can be found at (<https://doi.org/10.6084/m9.figshare.24503737>), alongside all metabolic models used in analyses.

Results

Metabolic traits are bimodally distributed across the population

The study dataset included 7,835 high-quality *KpSC* genomes covering all seven taxa in the complex (Fig 1A) and 1,931 STs (S1 Data). Pan-genome analysis [59] identified 61,595 gene clusters, 11,184 (18.2%) of which were matched to a total of 2,601 unique KEGG Orthologs [61] with associated substrate-reaction data. This represented a lower bound estimate of the proportion of ‘metabolic genes’ in the *KpSC* pan-genome, as 38.9% of gene clusters could not be assigned functional annotations (hypothetical proteins). Metabolism-associated KEGG Orthologs (hereafter ‘metabolic orthologs’) were bimodally distributed (Fig 1B and S4 Data), with 1,499 orthologs (57.4%) present in $\geq 95\%$ of genomes and 899 rare orthologs present in $< 15\%$ of genomes. This bimodal distribution is consistent with the general distribution of accessory genes in the *KpSC* [4], and other bacterial species with similarly large effective population sizes [64,89].

To analyze the phenotypic diversity across the *KpSC*, a strain-specific metabolic model was generated for each genome using Bactabolize [65], which was previously shown to produce models with overall 93.2% and 78.1% predictive accuracies in aerobic and anaerobic conditions, respectively (data for 37 diverse *KpSC* each tested for growth using 124 distinct carbon substrates via the Omnilog microarray; range 5.4%–100.0%, median 100.0% accuracy per substrate in aerobic conditions; range 0.0%–100.0%, median 88.9% accuracy per substrate in anaerobic conditions [66]). Growth phenotypes were predicted for 345 carbon, 192 nitrogen, 68 phosphorus, and 34 sulfur substrates as sole sources in minimal media under aerobic and anaerobic conditions (a total of 1,278 conditions). Individual strain-specific models comprised a mean of $1677.0 \text{ genes} \pm 17.3$ (standard deviation, SD) and $3330.1 \text{ reactions} \pm 13.5$ (SD). Individual isolates were predicted to grow in a mean of 678.2 ± 17.9 (SD) conditions. Seventeen isolates were identified as outliers with fewer positive growth phenotypes in anaerobic conditions, likely due to sequencing coverage issues rather than genuine biological deficiencies [90] (S5 Data, S1A Fig, and S1 Text).

The distribution of positive growth predictions mirrored that for metabolic gene content. At the population-level, 643 conditions (47.6%) were predicted to support growth for $\geq 95\%$ isolates (‘core’, Fig 1C and S5 Data) and were comprised of 349 aerobic and 294 anaerobic conditions, which systematically confirmed the facultative anaerobe status of the *KpSC*. These 643 core conditions consisted of 207 total substrates (note that some substrates can be tested as sole sources of multiple elements, e.g., carbon and nitrogen), with at least 204 of these being known human metabolites, 134 diet-related compounds, and 203 gut microbiome metabolites (S5 Data). In contrast, 509 modeled growth conditions (43.6%) did not support growth of any isolates (224 aerobic and 285 anaerobic conditions). Carbon substrate usage displayed the greatest variability, with 41 aerobic and 37 anaerobic conditions predicted to support growth in $< 95\%$ isolates (Figs 1C and S1B).

Metabolic variation is structured within and between species

The set of core growth conditions for each *KpSC* taxon was greater than the species complex as a whole (11–39 additional conditions each). Each taxon was associated with a distinct core growth profile mirrored by distinct core metabolic

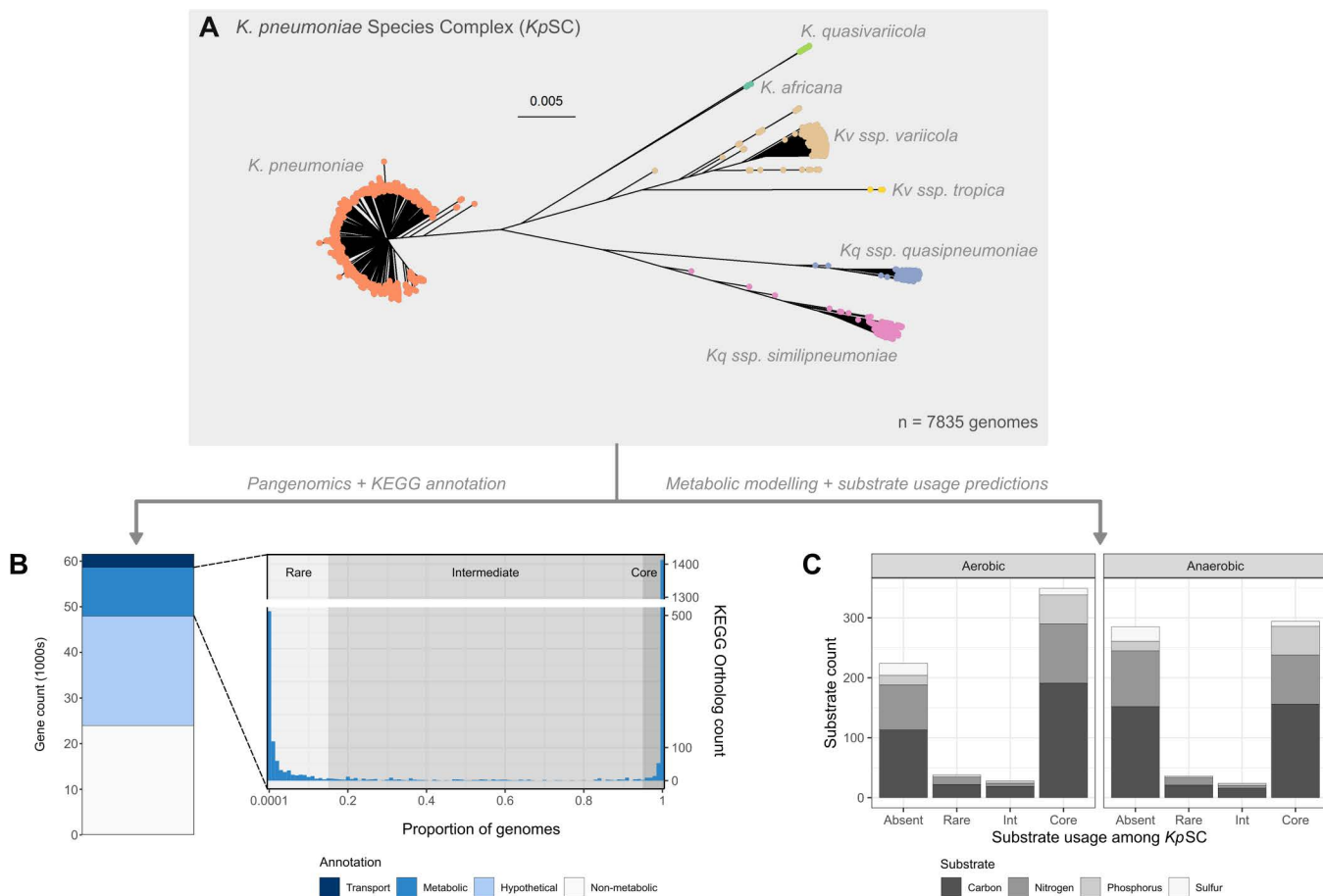


Fig 1. Population-level metabolic diversity in the *K. pneumoniae* species complex. A: Neighbor-joining tree of the 7,835 KpSC genomes used in this study. Tip colors indicate species as labeled. An interactive version found at <https://microreact.org/project/pcTMwQZAGCsqqhbEZzaj11-a-metabolic-atlas-of-the-klebsiella-pneumoniae-species-complex>. B: Stacked bars show pan-genome content by metabolic gene (KEGG) category as indicated in legend (left). Population distribution of KEGG orthologs associated with metabolic genes ('metabolic orthologs', right). The light gray background shows rare orthologs (>0–15%), the medium gray shows intermediate orthologs (>15–<95%) while the dark gray shows core orthologs (≥95%). C: Stacked bars show population frequency categories for predicted substrate usage, coloured by substrate type as indicated in legend. The data and phylogeny underlying this Figure can be found at Figshare (<https://doi.org/10.6084/m9.figshare.24503737>) and [S4 Data](#) and [S5 Data](#).

<https://doi.org/10.1371/journal.pbio.3003559.g001>

ortholog profiles as expected [70,91] (S2–S4A Figs, S4 Data, S5 Data, and S1 Text), and there was evidence of specialization at the level of macromolecular functions as determined by COG analysis (S5 Fig, S1 Data, and S1 Text). These predicted growth capabilities were generally consistent with biochemical tests described in formal species definitions [92–95]. However, there were a small number of notable discrepancies that indicate that species identification protocols should be revisited (see S1 Text and S6 Data for details).

While individual KpSC taxa are clearly defined by distinct metabolic profiles, there was also clear evidence of intra-taxon variation (S2 and S3 Figs). Four taxa were represented by sufficient genomes for comparisons ($n > 200$ each); *K. pneumoniae* (hereafter Kp, $n = 6,652$), *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* (Kqq, $n = 201$), *K. quasipneumoniae* subsp. *similipneumoniae* (Kqs, $n = 285$), and *Klebsiella variicola* subsp. *variicola* (Kvv, $n = 672$). In Kp, 938 metabolic orthologs were variably present in <95% genomes (544, 586, and 738 orthologs variable among Kqq, Kqs, and Kvv, respectively) and 79 conditions were predicted to support variable growth (81, 78, and 67 for Kqq, Kqs, and Kvv). We

hypothesized that this variation is not randomly distributed in the population, and our analysis of all-vs-all genome pairs within taxa showed statistically significant positive relationships between pairwise Average Nucleotide Identity (ANI, generally indicative of ancestral relatedness) and pairwise Jaccard similarity across metabolic genes (general linear hypothesis test, $p < 2 \times 10^{-16}$ for each of the four well-sampled taxa, **s 2A**). Comparison of the regression gradients indicated significant differences between taxa; however, the R^2 values, which indicate how well the models fit the data, were low (range 0.04–0.48), suggesting a large amount of variation in metabolic gene Jaccard similarities was not explained by the relationship with ANI.

Next, we used the recently established core-genome Life Identifier Number (LIN) code scheme to assign genomes to SLs [54]. These SLs represent deep-branching phylogenetic clades, separated by hundreds of years of evolutionary divergence (with notable exceptions discussed below). Of the 7,835 genomes, 7,833 were assigned to 1,418 distinct SLs and two were unassigned due to insufficient alleles detected.

Pairwise Jaccard similarities for metabolic ortholog profiles were significantly higher within SLs (median 0.956, IQR 0.947–0.965) than between SLs within the same species (median 0.923, IQR 0.914–0.932), $p < 0.0001$ (Kruskal–Wallis test). Mirroring this, pairwise Jaccard similarities for predicted substrate usage profiles were significantly higher for pairs within SLs (median 0.990, IQR 0.983–0.997) than between SLs (median 0.97, IQR 0.96–0.98) for both aerobic and anaerobic substrate usage ($p < 0.0001$ for all comparisons, Kruskal–Wallis test with Holm correction, followed by Dunn’s post-hoc, Fig 2B).

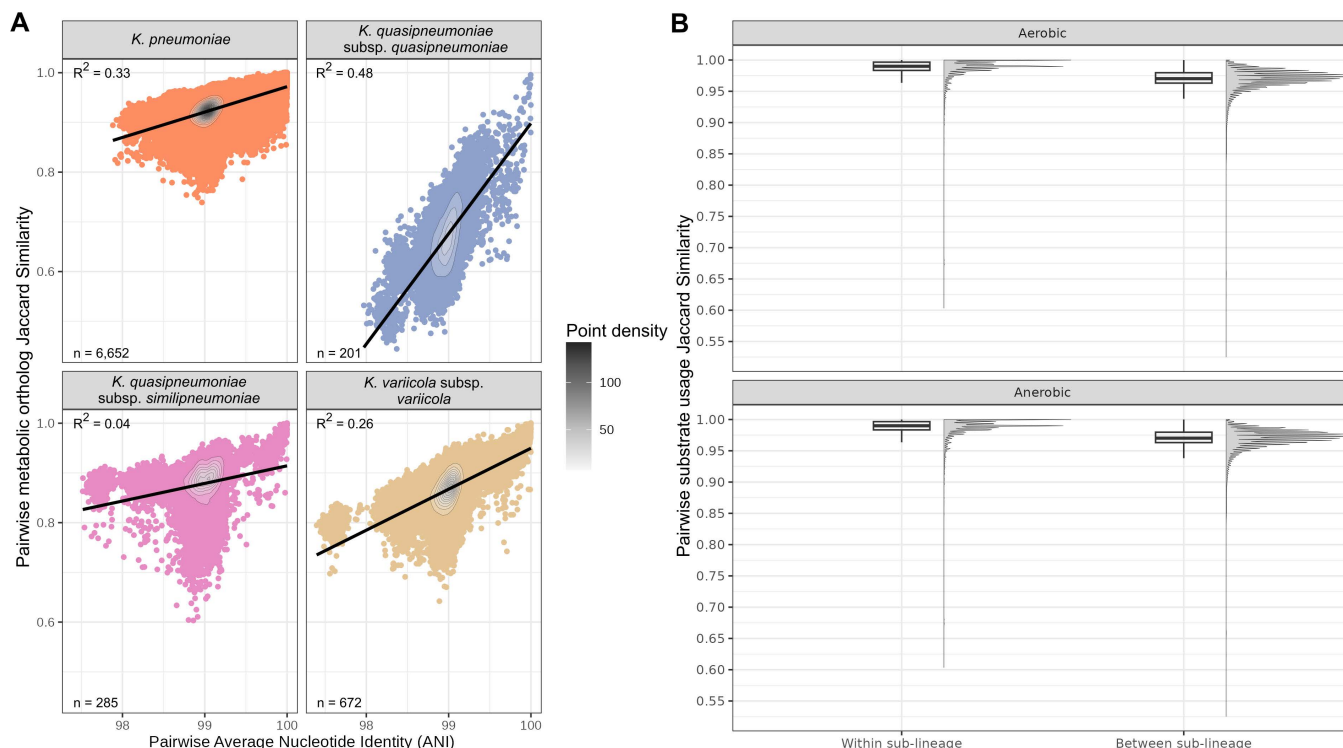


Fig 2. Metabolic heterogeneity within taxa is nonrandomly distributed. A: Pairwise Average Nucleotide Identity (ANI) vs. pairwise Jaccard similarity of metabolic orthologs within each of the well-sampled taxa ($n > 200$ genomes). B: Distributions of pairwise (all vs. all) Jaccard Similarity of simulated substrate usage for pairs of genomes within vs. between SLs in the same species. Significance asterisks excluded as each group was significantly different from every other group ($p < 2 \times 10^{-16}$), calculated using a nonparametric Kruskal–Wallis test with Holm correction, followed by Dunn’s post-hoc test. The data underlying this Figure can be found at Figshare (<https://doi.org/10.6084/m9.figshare.24503737>).

<https://doi.org/10.1371/journal.pbio.3003559.g002>

Finally, we sought to explore structuring among accessory metabolic traits, using a co-occurrence analysis. Twelve pairs of co-occurring substrate-usages were identified and 11 of these pairs were mechanistically-linked, e.g., the same substrate tested as different element sources ($n=5$ pairs) or substrates for which the catabolic pathways are overlapping ($n=4$ pairs, see [S1 Text](#)). Just one co-occurring phenotype pair (formamide and Fe(III)dicitrate as sources of carbon) was not obviously mechanistically-linked, indicating that they may be co-selected or encoded by genes that are physically linked in the genome or on plasmids. Accordingly, we have identified the genes enabling use of formamide and Fe(III)dicitrate as sources of carbon co-located on plasmids in the completed genomes of multiple unrelated strains, e.g., INF008 (SL221), INF044 (SL258), and INF119 (SL17).

Major SLs are defined by unique metabolic profiles

To explore the association between metabolism and population structure at higher-resolution, we calculated the relative prevalence of predicted growth phenotypes within 48 *Kp* SLs that were each represented by genomes from broad geographies and time-spans (≥ 15 genomes, 3–12 geographic regions, 5–86 years). These included SLs corresponding to globally-distributed multi-drug resistant (SL14, SL15, SL17, SL29, SL37, SL101, SL147, SL258, and SL307) and hypervirulent clones (SL23, SL25, SL65, and SL86) [3]. The data showed that each SL is associated with a distinct, core metabolic fingerprint, indicating metabolic specialization ([Figs 3](#), [S4B](#), [S6](#), and [S5 Data](#)), with 253–316 core aerobic and 198–223 core anaerobic conditions each. Note that in some cases the SL-specific core was smaller than the *Kp* species core reported above using the $\geq 95\%$ genome threshold (363 aerobic, 307 anaerobic conditions), because most SLs represent far less than 5% of the total *Kp* genome collection. This contradiction highlights the influence of sample size and diversity on core definitions, which should be considered as indicative of the general trends in the population. We tested for the impact of sampling biases on our SL comparisons. The data show no correlation between the number of SL core traits and either of sample size or temporal range ([S7A](#) and [S7B Fig](#), Pearson's product moment correlation, $r=0.20$ and 0.04 , $p=0.1751$ and 0.7774 , respectively). There was a moderate positive correlation between the number of SL core traits and number of geographic regions represented ([S7C Fig](#), $r=0.49$, $p=0.0004$). However, the latter is counter intuitive and seems to be driven by a small number of SLs that have a low number of core traits, rather than a genuine trend ([S7C Fig](#)). We are therefore confident that the trends we describe are not the result of sampling biases. Nonetheless, we acknowledge that the precise sets of SL-specific core traits listed here should be considered an approximation, because the addition of further diverse genomes for any given SL could result in the removal of a minority of traits from its core list.

No pair of SLs shared the same core growth profile, even pairs that are thought to share relatively recent common ancestry, such as SL14 and SL15 (common ancestor ~ 45 years ago [4]) and SL65 and SL25 (common ancestor yet to be dated) [96] ([Figs 3](#) and [S6](#)). However, there was also clear evidence of diversity within SLs, with 3–102 variable growth phenotypes each, a pattern that was replicated by metabolic ortholog distributions ([S3](#), [S6 Figs](#), and [S1 Text](#)). After excluding growth conditions predicted to support growth among $\geq 95\%$ (46/48) *K. pneumoniae* SLs ($n=615$ conditions, 340 aerobic and 275 anaerobic, corresponding to 202 total substrates) and those that did not support any growth among the 48 SLs ($n=525$ conditions), the data support a broad classification of phenotypes as follows: i) 89 that were core in some SLs (range = 1–45 SLs) but variably present in others (range = 1–17) which we call 'SL-specific core traits'; ii) 13 that were not core in any SL (accessory in 3–46 SLs, including the only pair of co-occurring phenotypes that was not mechanistically-linked) ([Figs 3](#) and [S6](#)). The SL-specific core traits could be further divided into those that were core to the majority ($\geq 66\%$) of SLs ($n=57$ phenotypes that we call 'majority SL-specific core traits'), those that were core to an intermediate number of SLs ($\geq 25\%$ – $<66\%$, $n=13$ phenotypes, corresponding to six substrates, that we call 'common SL-specific core traits') and those that were core in only very few SLs ($<25\%$, $n=13$ 'rare SL-specific core traits', eight aerobic and five anaerobic, corresponding to seven total substrates).

We have previously tested the predictive accuracy of 16 of 68 variable aerobic growth phenotypes, 14 of which were estimated $\geq 70\%$, including 12 estimated $\geq 90\%$ (see [Fig 3](#)), while the accuracy for 11 of 16 tested anaerobic

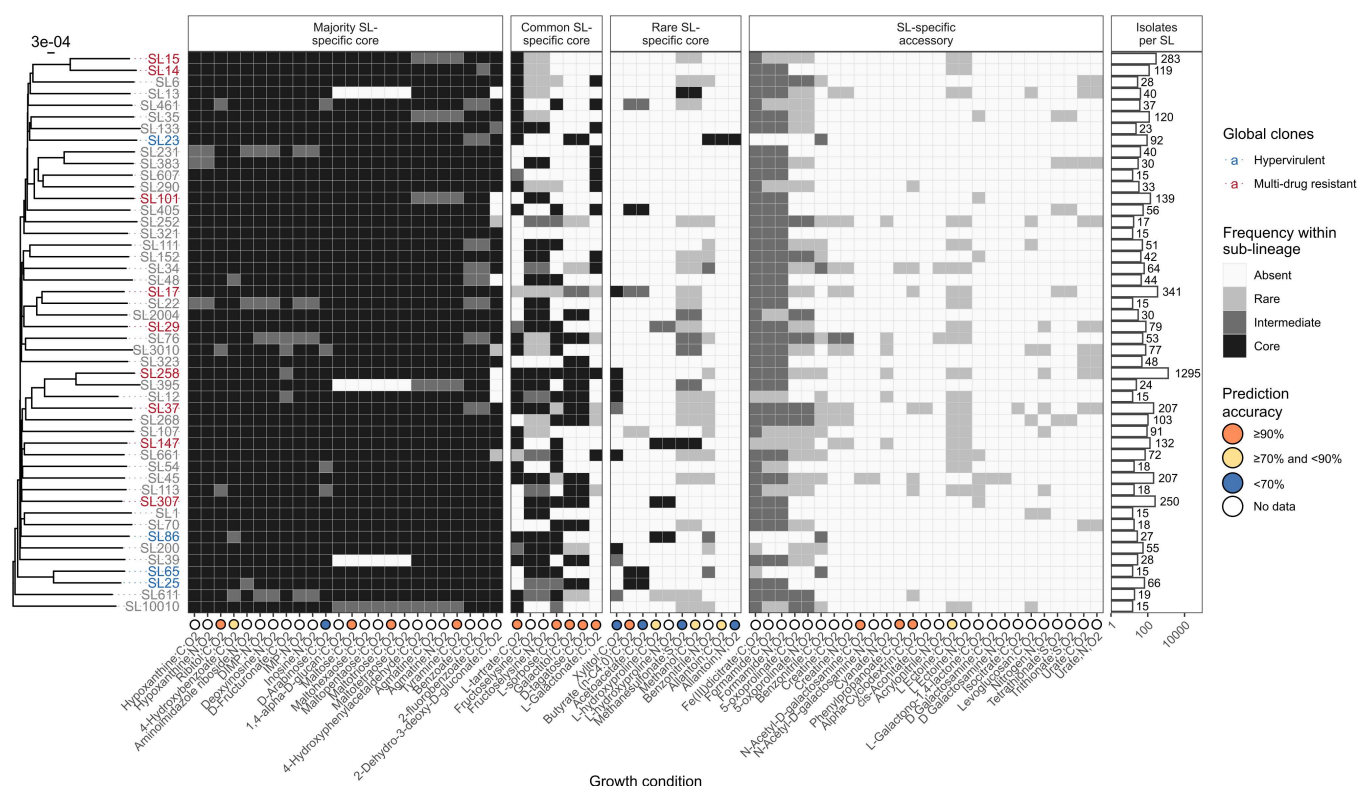


Fig 3. Distinct metabolic fingerprints of SLs within species. Heatmap showing frequency of variable aerobic substrate usage across 48 global *K. pneumoniae* sub-lineages (SL). Rows are ordered by phylogeny. SL labels are coloured to indicate the globally-distributed clones described in [3]: Blue shows hypervirulent while red shows multi-drug resistant. Substrate shown along X-axis. The element source of each substrate is semi-colon separated and abbreviated for brevity: C=Carbon, N=Nitrogen, and S=Sulfur. O2 indicates aerobic conditions. Frequency of substrate usage indicated by shading as shown in legend. Where available, substrate-specific prediction accuracies are indicated in coloured circles below the X-axis. Number of isolates per each SL shown in bars. Anaerobic growth predictions shown in S6 Fig. The data underlying this Figure can be found in S5 Data.

<https://doi.org/10.1371/journal.pbio.3003559.g003>

growth phenotypes was estimated $\geq 70\%$, including three estimated $\geq 90\%$ (see S6 Fig). We propose that the majority SL-specific core traits and the accessory traits are unlikely to play a major role in defining SL-specific evolutionary trajectories or ecological interactions, and hence we focused our follow-up investigations here on the common and rare SL-specific core traits. Our prior work had confirmed predictive accuracies $\geq 97\%$ for five of the six common SL-specific core substrates; L-sorbose, L-tartrate, galactitol, D-tagatose, and L-galactonate as sole sources of carbon in aerobic conditions [66] (tested using the Omnilog phenotyping microarray, for 37 diverse *KpSC* isolates). The sixth, fructose-lysine, is not present in the Omnilog array and could not be tested here due to its high cost. One of the seven rare SL-specific core substrates, allantoin, was previously tested as a sole source of carbon in aerobic conditions, for which the predictive accuracy was 89.2% (37 isolates tested [66]). Here, we tested growth using allantoin as a sole source of nitrogen and found the model accuracy was much lower (15% for 13 isolates tested). We also tested the six other rare SL-specific core growth substrates (13 isolates tested for each), confirming high accuracy (85%) for L-hydroxyproline as a carbon source, 77% accuracy for each of butyrate and methanol as sources of carbon, 46% for xylitol as a source of carbon, 31% for methanesulfonate as source of sulphur, and 15% for acetoacetate as a source of carbon (S2 Data, see S1 Text for more details, Fig 3).

SLs co-operate through cross-feeding metabolites

We have shown substrate usage varies across the *KpSC* and some of this variation is SL-dependent. We hypothesized that such diversity promotes the maintenance of distinct SLs by limiting inter-lineage competition and/or promoting commensalism. We tested this hypothesis using in vitro co-culture experiments.

We selected four carbon substrates (galactitol, L-tartrate, L-sorbose, and L-hydroxyproline) for use in three independent substrate combinations and corresponding SL pairs for which predicted substrate-usages were mutually exclusive SL-specific core traits. One isolate from each SL was selected for co-culture aerobically in each of two conditions (minimal media plus each of the SL-specific core substrates), such that each experiment contained one isolate that was able to utilize the substrate as a sole carbon source (the prototroph) and one that was not (the auxotroph). In short, Strain A as an auxotroph and Strain B as a prototroph, after which we then swapped their roles via use of another substrate, so that A would be a prototroph and B would be an auxotroph. This was done within each pair to evaluate reciprocal cross-feeding between SLs under different conditions. Growth of the auxotroph was measured using OD_{600} and compared to the same isolate growing in monoculture in the same media. Co-cultured isolates were physically separated by a Transwell membrane but had shared access to the growth medium (Fig 4A). A unique aspect of this experiment was the use of independent isolates from the same SLs as biological replicates rather than replicates of the same isolate. This allowed us to limit the impact of strain-specific growth biases while demonstrating conserved behavior across a greater number of independent isolates. SLs representing clinically-relevant clones were selected (multi-drug resistant SL37 versus hypervirulent SL29, hypervirulent SL86 versus hypervirulent SL23, and multi-drug resistant SL17 versus SL29).

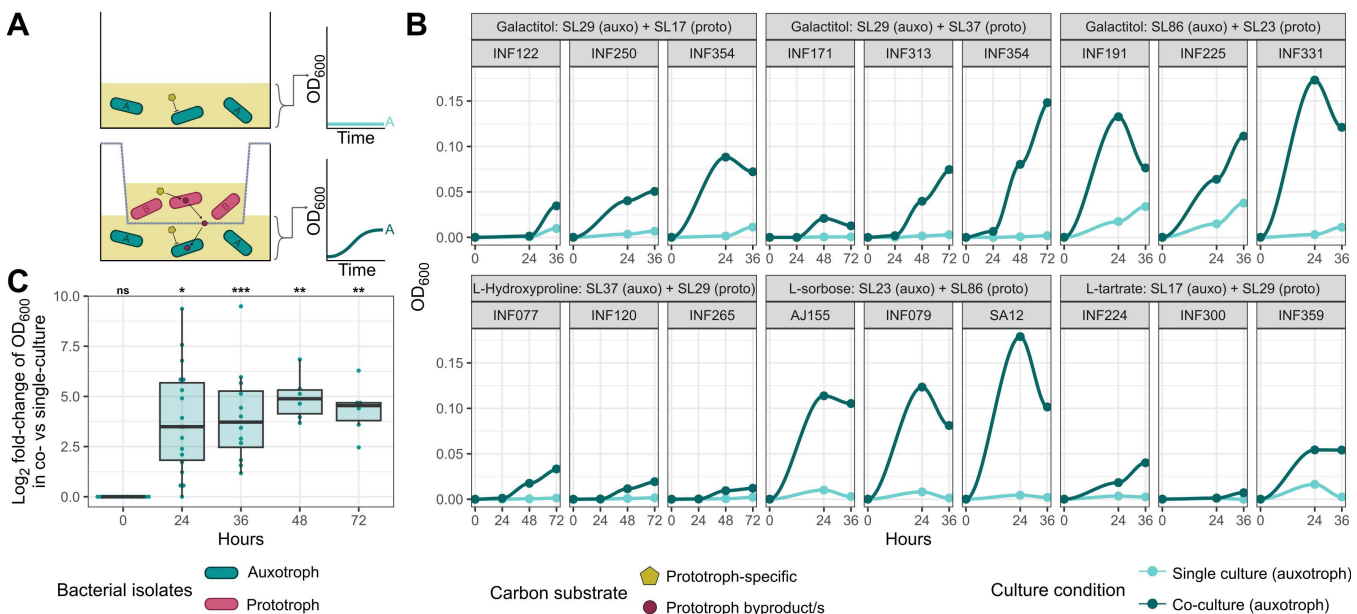


Fig 4. Nutrient prototrophic lineages support growth of auxotrophic lineages. Co-culture growth assays between prototrophs and auxotrophs under various substrate usage conditions. Time points captured were at 0, 24, 36, 48, and 72 hours depending upon the substrate pair (Materials and methods). A: Graphic demonstrating the experimental design measuring commensalism between an auxotrophic strain unable to utilize a carbon source and a prototrophic strain which can. The growth of the auxotroph was compared in single vs. co-culture conditions. B: Growth results (OD_{600}) over time for each auxotrophic strain in the bottom well ($n=3$ strains per SL as biological triplicates) in single and co-culture with a prototroph for the substrate. C: \log_2 -fold-change distribution of growth results (OD_{600}) at each time point under different culture conditions. Statistical significance comparing the OD_{600} values of co-culture to single culture is shown above plot (Dunn's post-hoc with Holm correction, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). The data underlying this Figure can be found in [S3 Data](#).

<https://doi.org/10.1371/journal.pbio.3003559.g004>

Our data showed that co-culturing auxotrophs with prototrophs in the presence of a SL-specific core substrate significantly increased the OD₆₀₀ of the auxotrophs (Fig 4B) with mean log₂-fold-change 3.78 ± 2.64 SD at 24 hours, $p=0.0128$; 4.02 ± 2.34 SD at 36 hours, $p=0.0001$; 4.94 ± 1.14 SD at 48 hours, $p=0.0039$; and 4.35 ± 1.27 at 72 hours $p=0.0039$; Kruskal–Wallis + Dunn’s post-hoc with Holm correction (Fig 4C). To confirm this phenomenon only occurred in the presence of a cross-feeding partner and eliminate the possibility of contamination or horizontal transfer of the genes conferring the prototroph phenotype, aliquots from each co-culture well were plated onto LB agar to confirm a single, distinct colony morphology. Colonies were then inoculated into M9 minimal media plus the auxotrophic substrate as a sole carbon source, confirming the original growth behavior of each isolate (S3 Data).

To further explore these pairwise strain interactions, we performed in silico co-culture simulations, which indicated that prototrophs were likely cross-feeding the auxotrophs by catabolising the sole carbon substrates, then exporting metabolites that were capable of supporting auxotroph growth (see S1 Text). Subsequent metabolomics analysis of representative prototroph culture supernatants supported the presence of many of the implicated metabolites (S6 Data, S8 Data, and S8 Fig): Between two and 13 metabolites predicted to be involved in cross-feeding were detected with significantly higher peak intensity in the prototroph culture supernatant compared to the no-isolate controls ($p<0.05$ by two-tailed Wilcoxon Rank Sum Test with Benjamini-Hochberg multiple testing correction). Among these, between two and 11 metabolites were each detected with >1 log₂ fold intensity increase, indicating a substantial increase in the quantity of the metabolite present in the prototroph supernatant as compared to the matched control (S6 Data). Metabolites with substantial intensity changes included those involved in central carbon metabolism and/or that have been previously shown to support growth for all or almost all *KpSC*, e.g., alanine, aspartate, glycerate, glutamate, lactate, malate, and pyruvate [70] (labeled in S8 Fig).

We cannot rule out the possibility that the putative cross-feeding metabolites were released following cell lysis; however, we note that the associated peak intensity shifts were among the most prominent of all metabolites tested, including ubiquitous intracellular metabolites, e.g., other amino acids (see S8 Fig). Therefore, it is unlikely that cell lysis alone explains the metabolite quantities we detected. Furthermore, each of the substrates named above is known to be actively exported by the closely related organism *E. coli* [97–102], and a tBLASTn search showed that all six of the representative prototrophs carried orthologs of genes encoding the relevant substrate exporter proteins characterized from *E. coli* (alanine, glutamate, malate, lactate, and pyruvate) or *Tetragenococcus halophila* (aspartate). (We were unable to identify any known exporters of glycerate. Also see S1 Text.)

Discussion

Here, we present a bacterial population metabolism study of the *K. pneumoniae* species complex, in which we used quantitative metabolic modeling to predict and compare growth phenotypes for $>7,000$ members of a single species complex, and contextualized the results against an established population genomics framework [3]. This approach allowed us to identify key metabolic differences that are indicative of metabolic fingerprints at both the species and sub-species (sub-lineage) levels. Our subsequent in vitro co-culture experiments (Fig 4) provide evidence that these metabolic differences can facilitate co-operation between SLs.

Our data shows that metabolic capabilities are structured in the population: we observed significant positive correlations between pairwise ANI and metabolic gene Jaccard similarity among each of the four well-sampled *KpSC* taxa (Fig 2A), indicating that closely related strains (high ANI) generally share a greater overlap of metabolic genes than more distantly related strains (low ANI). However, the proportion of variation explained was low (≤ 0.48 for all taxa). While there is likely a small impact of unavoidable assembly and/or annotation artifacts, we hypothesize the dominant driver is horizontal gene transfer; i) chromosomal recombination, which can introduce nucleotide variation into core genes and inflate ANI values for otherwise closely related genomes that share common metabolic genes; and ii) movement of mobile genetic elements, driving acquisition or loss of genes among closely related genomes (high ANI) and/or acquisition of common

sets of genes by distantly related genomes (low ANI). This is further supported by the comparison of pairwise growth prediction Jaccard similarities for isolate pairs within versus between SLs (Fig 2B, wherein the distributions were overlapping but shifted towards higher values for within SL pairs) and by our exploration of 48 common SLs (Figs 3 and S6). The latter showed that each SL was associated with a distinct set of core traits (metabolic orthologs and growth phenotypes), alongside clear evidence of variation within SLs, mirroring variations in total gene content that have been reported previously [4,96].

SL-specific core traits likely represent those vertically inherited from the SL common ancestor, whereas variable traits within SLs reflect acquisitions and/or losses through evolutionary time. Interestingly, several predicted growth phenotypes were observed only as accessory traits, i.e., they were not core to any single SL, most likely due to acquisition on mobile genetic elements rather than vertical inheritance followed by repeated losses in diverse strain backgrounds. Accordingly, we identified the genes associated with the two co-occurring accessory growth phenotypes, formamide and Fe(III)dicitrate, co-located on plasmids in genomes representing multiple distinct SLs. A plasmid born lactose utilization operon has been reported previously among *KpSC* [4], and *E. coli* plasmids have been shown to carry a variety of metabolic genes [103], but the potential role for plasmids in driving metabolic diversity within species has not been fully appreciated.

Our co-culture experiments indicated that metabolic variation, including SL-specific core growth capabilities, can enable co-operation between *KpSC* wherein an isolate which cannot metabolize a given substrate (an auxotroph), is supported to grow by the breakdown of the substrate by another isolate (a prototroph) (Fig 4). These findings were replicated for each of three distinct isolate pairs, representing three distinct and distantly related SL pair combinations, each tested for two different substrates. In each case, substrates predicted to support growth of the auxotroph were identified within the culture supernatant of one of the prototrophs, and include those ubiquitously metabolized by *KpSC*. We therefore predict that this capacity for ‘cross feeding’ is a common phenomenon among *KpSC*, although we cannot comment on its true frequency in natural environments nor in the context of the broader microbial communities within which *KpSC* exist. Further work leveraging microbial community models in representative conditions, and/or using in vivo models will be required to explore these intra-species dynamics.

Intriguingly, our data may shed light on the question of how so many SLs can be maintained within the *KpSC* population. In the most intuitive scenario, the ability to consume different substrates may enable *KpSC* SLs to inhabit distinct metabolic niches within spatially overlapping environments, thereby reducing the impact of nutrient competition. Supporting this hypothesis, it has been shown previously that metabolic niche differentiation can support the co-existence of multiple strains of *E. coli* in an in vivo model of mouse gut colonization [104]. In a more complex scenario, the ability to consume a substrate that cannot be consumed by a competitor can provide a selective advantage resulting in a relative frequency increase. This in turn can result in a depletion of the food resource and shift in selective pressure to favor the competitor. Although comparatively understudied, commensal interactions such as the metabolite cross-feeding implicated by our co-culture experiments (Fig 4), can also drive similar frequency-dependent selective shifts [106]. Over time, such fluctuations may prevent the exclusion of SLs and facilitate a stable equilibrium of co-existence. Notably, a similar model of negative frequency-dependent selection has been proposed previously as a driver for the maintenance of diverse SLs (entities) within bacterial populations [105,106]. However, this model is notoriously difficult to substantiate [106], requiring a long-term dataset from the same location over time and/or laboratory protocols that are beyond the scope of the current work, and therefore it remains to be confirmed.

Niche range no doubt further complicates the eco-evolutionary dynamics within *Kp* populations, e.g., where SLs are exposed to fluctuating selective pressures due to transient migration between niches, and/or where SLs undergo niche-adaptation. Recent analyses of contemporaneous *Kp* collections have suggested that some SLs may not be equally distributed between different hosts and environmental sources, which may indicate some level of niche-adaptation [107–109]. However, the same studies also identify near-identical strains from multiple sources and a subset of SLs that appear to occur frequently in diverse environments, which is indicative of niche migration. Importantly, the available

evidence suggests that wherever *KpSC* are sampled there exists substantial SL diversity, and the SLs identified in distinct niches are intermingled throughout the species phylogeny [4,33,43,108,109]. Therefore, while our convenience sample of genomes comprised primarily human-derived isolates (6,465/7,835 genomes), we anticipate that the trends we report regarding SL metabolic differentiation are applicable to the broader population. Notably, the limited genomes from nonhuman niches included here suggest that metabolic traits cluster isolates by SL rather than by niche (S4B and S4C Fig). However, we acknowledge that there may be rarer traits and/or SLs associated with nonhuman niches that we have not captured in our sample, and studies encompassing more diverse isolates are required to robustly test for niche-metabolism associations.

While genome-scale metabolic models are powerful approaches for exploring cellular metabolism, they are not without caveats. Firstly, model construction is reliant on curated databases that are limited by current biochemical knowledge [61], and while our pan-metabolic reference model was built from a large diverse collection of strains [66], it cannot possibly capture all metabolism in the *KpSC* population. Hence, we are likely underestimating the metabolic diversity within the highly variable *KpSC*. Secondly, growth prediction accuracies can vary by substrate, and are generally lower for anaerobic conditions for which the metabolic processes are less well understood [66]. While many of the core substrates and common SL-specific core traits have been confirmed with high accuracy in previous work, we have not tested all substrates and our phenotypic testing in this work showed variable accuracy. This is not unexpected, particularly for rarer substrate catabolism, which is comparatively understudied and therefore difficult to model, and/or requires unknown regulatory stimuli. Of note, our assays showed that allantoin nitrogen usage, a trait previously associated with hypervirulent infections [68] and predicted by our models as a SL-specific core trait in the dominant hypervirulent SL23, may in fact be core to *KpSC* (see S1 Text). These and other inaccurate predictions highlight gaps in our current knowledge that are noted for future investigations, and in some cases may have resulted in an overestimation of the population metabolic diversity. On the other hand, given the limitations to modeling traits involving pathways that are not well studied and/or for which the reaction stoichiometries are not defined; and that metabolic models do not consider the impact of sequence variations within metabolic genes nor variations in gene regulation; our analyses are very likely to represent an underestimate of the true metabolic diversity within and between *KpSC* species and SLs. Regulatory variations likely play an important role in determining metabolic phenotypes *in vivo*, and in driving competitive interactions and ecological adaptation, and our metabolic models provide a platform on to which transcriptomic, proteomic and metabolomic data can be integrated and contextualized in the future to further probe diversity.

Notwithstanding the caveats of *in silico* modeling, our analyses clearly challenge the assumption that *Kp* performs a defined and homogenous metabolic function, and this has important implications for how we understand and model *KpSC* pathogenicity, and/or microbial communities containing this organism. There is increasing interest in gut microbiota manipulation for prevention of pathogen colonization and subsequent opportunistic infection [110,111], and studies on *Kp* have thus far implicated competitive exclusion via nutrient competition as the major mechanistic driver [111,112]. Notably, the *Kp* SL-specific core traits we identified include those implicated in driving competitive interactions and colonization resistance in the mammalian gut, e.g., fructoselysine and galactitol usage [113,114]. Our work therefore provides further evidence for the importance of including diverse strains in experimental studies of *KpSC* competitive interactions, colonization dynamics, pathogenicity, and virulence (including hypervirulence), and highlights the need to consider population structure when designing and interpreting metabolic association studies. A comprehensive understanding of the population metabolic diversity will greatly benefit the design of novel control strategies targeting *Kp*, particularly those harnessing microbial competitive interactions or targeting metabolic processes.

Supporting information

S1 Text. Description of supplementary results. (DOCX)

S1 Data. List of all 7,835 genomes used for this study along with accessions, metadata, per-genome summary of simulated growth phenotypes, and proportional Clusters of Orthologous Genes (COG) data.

(XLSX)

S2 Data. Outcomes of in vitro growth experiments for metabolic model validation.

(XLSX)

S3 Data. Raw data and statistical analysis of co-culture experiments.

(XLSX)

S4 Data. Metabolic gene (metabolic KEGG Ortholog) presence/absence table for entire 7,835 genome dataset.

(XLSX)

S5 Data. Binarised growth predictions from metabolic modeling for entire 7,835 genome dataset.

(XLSX)

S6 Data. Metabolites detected by LC-MS of prototroph culture supernatants and comparison of peak intensities to matched no-isolate controls.

(XLSX)

S7 Data. Biochemical tests described in formal species definitions and consistency with model predictions.

(DOCX)

S8 Data. Substrates predicted to be involved in cross-feeding and LC-MS detection.

(DOCX)

S1 Fig. Substrate usage frequencies across *KpSC* dataset.

(PDF)

S2 Fig. Taxon-specific substrate usage.

(PDF)

S3 Fig. Taxa are associated with unique metabolic ortholog profiles.

(PDF)

S4 Fig. PHATE analysis of metabolic orthologs.

(PDF)

S5 Fig. Taxon-specific specialization of macromolecular functions.

(PDF)

S6 Fig. Distinct anerobic metabolic fingerprints of sub-lineages within species.

(PDF)

S7 Fig. Relationships between the number of core aerobic phenotypes predicted for each common sub-lineage and sample characteristics.

(PDF)

S8 Fig. Metabolite peak intensities in isolate culture supernatants as compared to no-isolate controls in M9 minimal media supplemented with various carbon sources as specified.

(PDF)

Acknowledgments

This study used BPA-enabled (Bioplatforms Australia)/NCRIS-enabled (National Collaborative Research Infrastructure Strategy) infrastructure located at the Monash Proteomics and Metabolomics Platform. This work was further supported by Monash eResearch capabilities, including Research Data Storage, Nectar Research Cloud and the MASSIVE M3 HPC facility (www.massive.org.au).

Author contributions

Conceptualization: Ben Vezina, Sylvain Brisse, Jonathan M. Monk, Kathryn E. Holt, Kelly L. Wyres.

Data curation: Ben Vezina, Martin Rethoret-Pasty.

Formal analysis: Ben Vezina, Helena B. Cooper, Christopher K. Barlow, Kelly L. Wyres.

Funding acquisition: Sylvain Brisse, Jonathan M. Monk, Kathryn E. Holt, Kelly L. Wyres.

Investigation: Ben Vezina, Helena B. Cooper.

Methodology: Ben Vezina, Christopher K. Barlow, Sylvain Brisse, Jonathan M. Monk, Kathryn E. Holt, Kelly L. Wyres.

Project administration: Kelly L. Wyres.

Resources: Ben Vezina, Sylvain Brisse, Jonathan M. Monk, Kathryn E. Holt, Kelly L. Wyres.

Supervision: Sylvain Brisse, Jonathan M. Monk, Kelly L. Wyres.

Validation: Ben Vezina.

Visualization: Ben Vezina, Kelly L. Wyres.

Writing – original draft: Ben Vezina, Kelly L. Wyres.

Writing – review & editing: Ben Vezina, Helena B. Cooper, Christopher K. Barlow, Martin Rethoret-Pasty, Sylvain Brisse, Jonathan M. Monk, Kathryn E. Holt, Kelly L. Wyres.

References

1. GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet*. 2024;404(10459):1199–226. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1) PMID: [39299261](https://pubmed.ncbi.nlm.nih.gov/39299261/)
2. Sati H, Carrara E, Savoldi A, Hansen P, Garlasco J, Campagnaro E, et al. The WHO Bacterial Priority Pathogens List 2024: a prioritisation study to guide research, development, and public health strategies against antimicrobial resistance. *Lancet Infect Dis*. 2025;25(9):1033–43. [https://doi.org/10.1016/S1473-3099\(25\)00118-5](https://doi.org/10.1016/S1473-3099(25)00118-5) PMID: [40245910](https://pubmed.ncbi.nlm.nih.gov/40245910/)
3. Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. *Nat Rev Microbiol*. 2020;18(6):344–59. <https://doi.org/10.1038/s41579-019-0315-1> PMID: [32055025](https://pubmed.ncbi.nlm.nih.gov/32055025/)
4. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci U S A*. 2015;112(27):E3574–81. <https://doi.org/10.1073/pnas.1501049112> PMID: [26100894](https://pubmed.ncbi.nlm.nih.gov/26100894/)
5. Watkins ER, Maiden MC, Gupta S. Metabolic competition as a driver of bacterial population structure. *Future Microbiol*. 2016;11:1339–57. <https://doi.org/10.2217/fmb-2016-0079> PMID: [27660887](https://pubmed.ncbi.nlm.nih.gov/27660887/)
6. Cui Y, Yang X, Didelot X, Guo C, Li D, Yan Y, et al. Epidemic clones, oceanic gene pools, and Eco-LD in the free living marine pathogen *Vibrio parahaemolyticus*. *Mol Biol Evol*. 2015;32(6):1396–410. <https://doi.org/10.1093/molbev/msv009> PMID: [25605790](https://pubmed.ncbi.nlm.nih.gov/25605790/)
7. Reuter S, Corander J, de Been M, Harris S, Cheng L, Hall M, et al. Directional gene flow and ecological separation in *Yersinia enterocolitica*. *Microb Genom*. 2015;1(3):e000030. <https://doi.org/10.1099/mgen.0.000030> PMID: [28348815](https://pubmed.ncbi.nlm.nih.gov/28348815/)
8. Sheppard SK, Cheng L, Méric G, de Haan CPA, Llaena A-K, Martinen P, et al. Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. *Mol Ecol*. 2014;23(10):2442–51. <https://doi.org/10.1111/mec.12742> PMID: [24689900](https://pubmed.ncbi.nlm.nih.gov/24689900/)
9. Top J, Baan J, Bisschop A, Arredondo-Alonso S, van Schaik W, Willems RJL. Functional characterization of a gene cluster responsible for inositol catabolism associated with hospital-adapted isolates of *Enterococcus faecium*. *Microbiology (Reading)*. 2021;167(9). <https://doi.org/10.1099/mic.0.001085> PMID: [34491894](https://pubmed.ncbi.nlm.nih.gov/34491894/)

10. Connor CH, Zucoloto AZ, Munnoch JT, Yu I-L, Corander J, Hoskisson PA, et al. Multidrug-resistant *E. coli* encoding high genetic diversity in carbohydrate metabolism genes displace commensal *E. coli* from the intestinal tract. *PLoS Biol.* 2023;21(10):e3002329. <https://doi.org/10.1371/journal.pbio.3002329> PMID: 37847672
11. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet.* 2014;15(2):107–20. <https://doi.org/10.1038/nrg3643> PMID: 24430943
12. Omid A, Patrick P, Omkar SM, Lars KN, Bernhard OP. Pangenome reconstruction of *Lactobacillaceae* metabolism predicts species-specific metabolic traits. *bioRxiv.* 2023;2023.09.18.558222.
13. Seif Y, Kavvas E, Lachance J-C, Yurkovich JT, Nuccio S-P, Fang X, et al. Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nat Commun.* 2018;9(1):3771. <https://doi.org/10.1038/s41467-018-06112-5> PMID: 30218022
14. Blázquez B, San León D, Rojas A, Tortajada M, Nogales J. New insights on metabolic features of *Bacillus subtilis* based on multistrain genome-scale metabolic modeling. *Int J Mol Sci.* 2023;24(8):7091. <https://doi.org/10.3390/ijms24087091> PMID: 37108252
15. Norsigian CJ, Attia H, Szubin R, Yassin AS, Palsson BØ, Aziz RK, et al. Comparative genome-scale metabolic modeling of metallo-beta-lactamase-producing multidrug-resistant *Klebsiella pneumoniae* clinical isolates. *Front Cell Infect Microbiol.* 2019;9:161. <https://doi.org/10.3389/fcimb.2019.00161> PMID: 31179245
16. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A.* 2013;110(50):20338–43. <https://doi.org/10.1073/pnas.1307797110> PMID: 24277855
17. Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BØ. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc Natl Acad Sci U S A.* 2016;113(26):E3801–9. <https://doi.org/10.1073/pnas.1523199113> PMID: 27286824
18. Arzamasov AA, Rodionov DA, Hibberd MC, Guruge JL, Kent JE, Kazanov MD, et al. Integrative genomic reconstruction reveals heterogeneity in carbohydrate utilization across human gut bifidobacteria. *Nat Microbiol.* 2025;10(8):2031–47. <https://doi.org/10.1038/s41564-025-02056-x> PMID: 40670725
19. Ardalani O, Phaneuf PV, Krishnan J, Li G, Burrows J, Nielsen LK, et al. A diverse genetic basis for metabolic reactions is revealed through pangenome analysis. *bioRxiv.* 2025;2025.01.23.634541.
20. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard A-S, et al. Genomic definition of hypervirulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. *Emerg Infect Dis.* 2014;20(11):1812–20. <https://doi.org/10.3201/eid2011.140206> PMID: 25341126
21. Wang M, Earley M, Chen L, Hanson BM, Yu Y, Liu Z, et al. Clinical outcomes and bacterial characteristics of carbapenem-resistant *Klebsiella pneumoniae* complex among patients from different global regions (CRACKLE-2): a prospective, multicentre, cohort study. *Lancet Infect Dis.* 2021.
22. Davis GS, Waits K, Nordstrom L, Weaver B, Aziz M, Gauld L, et al. Intermingled *Klebsiella pneumoniae* populations between retail meats and human urinary tract infections. *Clin Infect Dis.* 2015;61(6):892–9. <https://doi.org/10.1093/cid/civ428> PMID: 26206847
23. Ellington MJ, Heinz E, Wailan AM, Dorman MJ, de Goffau M, Cain AK, et al. Contrasting patterns of longitudinal population dynamics and antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single center. *Genome Biol.* 2019;20(1):184.
24. Grundmann H, Glasner C, Albiger B, Aanensen DM, Tomlinson CT, Andrasević AT, et al. Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. *Lancet Infect Dis.* 2017;17(2):153–63. [https://doi.org/10.1016/S1473-3099\(16\)30257-2](https://doi.org/10.1016/S1473-3099(16)30257-2) PMID: 27866944
25. Franklin-Alming FV, Kaspersen H, Hetland MAK, Bakksjø R-J, Nesse LL, Leangapichart T, et al. Exploring *Klebsiella pneumoniae* in healthy poultry reveals high genetic diversity, good biofilm-forming abilities and higher prevalence in Turkeys than broilers. *Front Microbiol.* 2021;12(2609).
26. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clin Infect Dis.* 2017;65(2):208–15. <https://doi.org/10.1093/cid/cix270> PMID: 28369261
27. Hawkey J, Vezina B, Monk JM, Judd LM, Harshgyi T, López-Fernández S, et al. A curated collection of *Klebsiella* metabolic models reveals variable substrate usage and gene essentiality. *Genome Res.* 2022;32(5):1004–14. <https://doi.org/10.1101/gr.276289.121> PMID: 35277433
28. Henson SP, Boinett CJ, Ellington MJ, Kagia N, Mwarumba S, Nyongesa S, et al. Molecular epidemiology of *Klebsiella pneumoniae* invasive infections over a decade at Kilifi County Hospital in Kenya. *Int J Med Microbiol.* 2017;307(7):422–9. <https://doi.org/10.1016/j.ijmm.2017.07.006> PMID: 28789913
29. Huynh B-T, Passet V, Rakotondrasoa A, Diallo T, Kerleguer A, Hennart M, et al. *Klebsiella pneumoniae* carriage in low-income countries: antimicrobial resistance, genomic diversity and risk factors. *Gut Microbes.* 2020;11(5):1287–99. <https://doi.org/10.1080/19490976.2020.1748257> PMID: 32404021
30. Leangapichart T, Lunha K, Jiwakanon J, Angkititrakul S, Järhult JD, Magnusson U, et al. Characterization of *Klebsiella pneumoniae* complex isolates from pigs and humans in farms in Thailand: population genomic structure, antibiotic resistance and virulence genes. *J Antimicrob Chemother.* 2021;76(8):2012–6. <https://doi.org/10.1093/jac/dkab118> PMID: 33829268
31. Lee IR, Molton JS, Wyres KL, Gorrie C, Wong J, Hoh CH, et al. Differential host susceptibility and bacterial virulence factors driving *Klebsiella* liver abscess in an ethnically diverse population. *Sci Rep.* 2016;6:29316. <https://doi.org/10.1038/srep29316> PMID: 27406977
32. Lewis JM, Mphasa M, Banda R, Beale MA, Mallewa J, Heinz E, et al. Genomic and antigenic diversity of colonizing *Klebsiella pneumoniae* isolates mirrors that of invasive isolates in Blantyre, Malawi. *Microb Genom.* 2022;8(3):000778. <https://doi.org/10.1099/mgen.0.000778> PMID: 35302438

33. Ludden C, Moradigaravand D, Jamroz D, Gouliouris T, Blane B, Naydenova P, et al. A One Health Study of the genetic relatedness of *Klebsiella pneumoniae* and their mobile elements in the east of England. *Clin Infect Dis*. 2020;70(2):219–26. <https://doi.org/10.1093/cid/ciz174> PMID: 30840764
34. McDougall FK, Wyres KL, Judd LM, Boardman WSJ, Holt KE, Power ML. Novel strains of *Klebsiella africana* and *Klebsiella pneumoniae* in Australian fruit bats (*Pteropus poliocephalus*). *Res Microbiol*. 2021;172(7–8):103879. <https://doi.org/10.1016/j.resmic.2021.103879> PMID: 34506927
35. Musicha P, Msefula CL, Mather AE, Chaguza C, Cain AK, Peno C, et al. Genomic analysis of *Klebsiella pneumoniae* isolates from Malawi reveals acquisition of multiple ESBL determinants across diverse lineages. *J Antimicrob Chemother*. 2019;74(5):1223–32. <https://doi.org/10.1093/jac/dkz032> PMID: 30778540
36. Potter RF, Lainhart W, Twentyman J, Wallace MA, Wang B, Burnham CA, et al. Population structure, antibiotic resistance, and uropathogenicity of *Klebsiella variicola*. *mBio*. 2018;9(6):e02481–18. <https://doi.org/10.1128/mBio.02481-18> PMID: 30563902
37. Raffelsberger N, Hetland MAK, Svendsen K, Småbrekke L, Löhr IH, Andreassen LLE, et al. Gastrointestinal carriage of *Klebsiella pneumoniae* in a general adult population: a cross-sectional study of risk factors and bacterial genomic diversity. *Gut Microbes*. 2021;13(1):1939599. <https://doi.org/10.1080/19490976.2021.1939599> PMID: 34182896
38. Rodrigues C, Hauser K, Cahill N, Ligowska-Marzeta M, Centorotola G, Cornacchia A, et al. High prevalence of *Klebsiella pneumoniae* in European food products: a multicentric study comparing culture and molecular detection methods. *Microbiol Spectr*. 2022;10(1):e0237621. <https://doi.org/10.1128/spectrum.02376-21> PMID: 35196810
39. Smit PW, Stoesser N, Pol S, van Kleef E, Oonsivilai M, Tan P, et al. Transmission dynamics of hyper-endemic multi-drug resistant *Klebsiella pneumoniae* in a Southeast Asian neonatal unit: a longitudinal study with whole genome sequencing. *Front Microbiol*. 2018;9.
40. Thorpe H, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, et al. One health or three? Transmission modelling of *Klebsiella* isolates reveals ecological barriers to transmission between humans, animals and the environment. *bioRxiv*. 2021:2021.08.05.455249.
41. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, et al. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J Antimicrob Chemother*. 2013;68(10):2234–44. <https://doi.org/10.1093/jac/dkt180> PMID: 23722448
42. Stoesser N, Giess A, Batty EM, Sheppard AE, Walker AS, Wilson DJ, et al. Genome sequencing of an extended series of NDM-producing *Klebsiella pneumoniae* isolates from neonatal infections in a Nepali hospital characterizes the extent of community- versus hospital-associated transmission in an endemic setting. *Antimicrob Agents Chemother*. 2014;58(12):7347–57. <https://doi.org/10.1128/AAC.03900-14> PMID: 25267672
43. Vezina B, Judd LM, McDougall FK, Boardman WSJ, Power ML, Hawkey J, et al. Transmission of *Klebsiella* strains and plasmids within and between grey-headed flying fox colonies. *Environ Microbiol*. 2022;24(9):4425–36. <https://doi.org/10.1111/1462-2920.16047> PMID: 35590448
44. Wand ME, Baker KS, Benthall G, McGregor H, McCowen JWJ, Deheer-Graham A, et al. Characterization of pre-antibiotic era *Klebsiella pneumoniae* isolates with respect to antibiotic/disinfectant susceptibility and virulence in *Galleria mellonella*. *Antimicrob Agents Chemother*. 2015;59(7):3966–72. <https://doi.org/10.1128/AAC.05009-14> PMID: 25896708
45. Wyres KL, Nguyen TNT, Lam MMC, Judd LM, van Vinh Chau N, Dance DAB, et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive *Klebsiella pneumoniae* from South and Southeast Asia. *Genome Med*. 2020;12(1):11. <https://doi.org/10.1186/s13073-019-0706-y> PMID: 31948471
46. Zheng Z, Gorden PJ, Xia X, Zheng Y, Li G. Whole-genome analysis of *Klebsiella pneumoniae* from bovine mastitis milk in the U.S. *Environ Microbiol*. 2021.
47. European-Nucleotide-Archive. enaBrowserTools. 1.6 ed. Available from: <https://github.com/2020>
48. Krueger F. Trim Galore GitHub. 2012. Available from: <https://github.com/FelixKrueger/TrimGalore>
49. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017;13(6):e1005595. <https://doi.org/10.1371/journal.pcbi.1005595> PMID: 28594827
50. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom*. 2021;7(11):000685. <https://doi.org/10.1099/mgen.0.000685> PMID: 34739369
51. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat Commun*. 2021;12(1):4188. <https://doi.org/10.1038/s41467-021-24448-3> PMID: 34234121
52. Argimón S, David S, Underwood A, Abrudan M, Wheeler NE, Kekre M, et al. Rapid genomic characterization and global surveillance of *Klebsiella* using Pathogenwatch. *Clin Infect Dis*. 2021;73(Suppl_4):S325–35.
53. Jolley KA, Maiden MCJ. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 2010;11:595. <https://doi.org/10.1186/1471-2105-11-595> PMID: 21143983
54. Hennart M, Guglielmini J, Bridel S, Maiden MCJ, Jolley KA, Criscuolo A, et al. A dual barcoding approach to bacterial strain nomenclature: genomic taxonomy of *Klebsiella pneumoniae* strains. *Mol Biol Evol*. 2022;39(7):msac135. <https://doi.org/10.1093/molbev/msac135> PMID: 35700230
55. Palma F, Hennart M, Jolley KA, Crestani C, Wyres KL, Bridel S, et al. Bacterial strain nomenclature in the genomic era: Life Identification Numbers using a gene-by-gene approach. *bioRxiv*. 2024:2024.03.11.584534.
56. Jain C, Rodriguez-R LM, Philipp AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114. <https://doi.org/10.1038/s41467-018-07641-9> PMID: 30504855

57. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res.* 2019;29(2):304–16. <https://doi.org/10.1101/gr.241455.118> PMID: 30679308
58. Wick RR, Holt KE. Assembly Dereplicator. 0.1.0 ed. 2019. Available from: <https://github.com/rwrick/Assembly-Dereplicator>
59. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 2020;21(1):180. <https://doi.org/10.1186/s13059-020-02090-4> PMID: 32698896
60. Rice P, Longden I, Bleasby A. EMBOSS: the european molecular biology open software suite. *Trend Genet.* 2000;16(6):276–7.
61. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27> PMID: 10592173
62. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics.* 2020;36(7):2251–2. <https://doi.org/10.1093/bioinformatics/btz859> PMID: 31742321
63. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: community ecology package. 2.5-7 ed. 2020.
64. Horesh G, Taylor-Brown A, McGimpsey S, Lassalle F, Corander J, Heinz E, et al. Different evolutionary trends form the twilight zone of the bacterial pan-genome. *Microb Genom.* 2021;7(9):000670. <https://doi.org/10.1099/mgen.0.000670> PMID: 34559043
65. Vezina B, Watts SC, Hawkey J, Cooper HB, Judd LM, Jenney AWJ, et al. Bactabolize is a tool for high-throughput generation of bacterial strain-specific metabolic models. *Elife.* 2023;12:RP87406. <https://doi.org/10.7554/eLife.87406> PMID: 37815531
66. Cooper HB, Vezina B, Hawkey J, Passet V, López-Fernández S, Monk JM, et al. A validated pangenome-scale metabolic model for the *Klebsiella pneumoniae* species complex. *Microb Genom.* 2024;10(2):001206. <https://doi.org/10.1099/mgen.0.001206> PMID: 38376382
67. Whelan FJ, Rusilowicz M, McInerney JO. Coinfinder: detecting significant associations and dissociations in pangenomes. *Microb Genom.* 2020;6(3):e000338. <https://doi.org/10.1099/mgen.0.000338> PMID: 32100706
68. Chou H-C, Lee C-Z, Ma L-C, Fang C-T, Chang S-C, Wang J-T. Isolation of a chromosomal region of *Klebsiella pneumoniae* associated with allantoin metabolism and liver infection. *Infect Immun.* 2004;72(7):3783–92. <https://doi.org/10.1128/IAI.72.7.3783-3792.2004> PMID: 15213119
69. Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature.* 2002;420(6912):186–9. <https://doi.org/10.1038/nature01149> PMID: 12432395
70. Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. Metabolic diversity of the emerging pathogenic lineages of *Klebsiella pneumoniae*. *Environ Microbiol.* 2017;19(5):1881–98. <https://doi.org/10.1111/1462-2920.13689> PMID: 28181409
71. Diener C, Gibbons SM, Resendis-Antonio O. MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota. *mSystems.* 2020;5(1):e00606–19. <https://doi.org/10.1128/mSystems.00606-19> PMID: 31964767
72. Creek DJ, Jankevics A, Breitling R, Watson DG, Barrett MP, Burgess KEV. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal Chem.* 2011;83(22):8703–10. <https://doi.org/10.1021/ac2021823> PMID: 21928819
73. Creek DJ, Jankevics A, Burgess KEV, Breitling R, Barrett MP. IDEOM: an excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics.* 2012;28(7):1048–9. <https://doi.org/10.1093/bioinformatics/bts069> PMID: 22308147
74. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol.* 2012;30(10):918–20. <https://doi.org/10.1038/nbt.2377> PMID: 23051804
75. Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics.* 2008;9:504. <https://doi.org/10.1186/1471-2105-9-504> PMID: 19040729
76. Scheltema RA, Jankevics A, Jansen RC, Swertz MA, Breitling R. PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal Chem.* 2011;83(7):2786–93.
77. R-Core-Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2020.
78. RStudio-Team. RStudio: integrated development for R. Boston, MA: RStudio; 2020.
79. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. *JOSS.* 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>
80. Garnier S. viridis: Default Color Maps from 'matplotlib'. 2018.
81. Neuwirth E. RColorBrewer: ColorBrewer palettes. 1.1-2 ed. 2014.
82. Kassambara A. ggpubr: 'ggplot2' based publication ready plots. 0.5.0 ed. 2022.
83. Aphalo PJ, Slowikowski K, Mouksassi S. ggpmisc: miscellaneous Extensions to 'ggplot2'. 0.4.1 ed. 2021.
84. Yu G. aplot: Decorate a 'ggplot' with Associated Information. 0.0.6 ed. 2020.
85. Zeileis A, Fisher JC, Hornik K, Ihaka R, McWhite CD, Murrell P, et al. Colorspace: a toolbox for manipulating and assessing colors and palettes. *J Stat Softw.* 2020;96(1):1–49.
86. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 2016;8(1):28–36. <https://doi.org/10.1111/2041-210x.12628>
87. Kassambara A. rstatix: pipe-friendly framework for basic statistical tests. 2023.

88. Campitelli E. ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'. 0.4.5 ed. 2021.
89. McInerney JO. Prokaryotic pangenomes act as evolving ecosystems. *Mol Biol Evol.* 2023;40(1):msac232. <https://doi.org/10.1093/molbev/msac232> PMID: 36288801
90. Gunasekera S, Abraham S, Stegger M, Pang S, Wang P, Sahibzada S, et al. Evaluating coverage bias in next-generation sequencing of *Escherichia coli*. *PLoS One.* 2021;16(6):e0253440. <https://doi.org/10.1371/journal.pone.0253440> PMID: 34166413
91. Hawkey J, Wyres KL, Judd LM, Harshegyi T, Blakeway L, Wick RR, et al. ESBL plasmids in *Klebsiella pneumoniae*: diversity, transmission and contribution to infection burden in the hospital setting. *Genome Med.* 2022;14(1):97. <https://doi.org/10.1186/s13073-022-01103-0> PMID: 35999578
92. Imhoff J. Bergey's manual of systematic bacteriology. 2005. p. 587–850.
93. Rosenblueth M, Martínez L, Silva J, Martínez-Romero E. *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. *Syst Appl Microbiol.* 2004;27(1):27–35. <https://doi.org/10.1078/0723-2020-00261> PMID: 15053318
94. Brisse S, Passet V, Grimont PAD. Description of *Klebsiella quasipneumoniae* sp. nov., isolated from human infections, with two subspecies, *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., and demonstration that *Klebsiella singaporensis* is a junior heterotypic synonym of *Klebsiella variicola*. *Int J Syst Evol Microbiol.* 2014;64(Pt 9):3146–52. <https://doi.org/10.1099/ijs.0.062737-0> PMID: 24958762
95. Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S. Description of *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and *Klebsiella variicola* subsp. *variicola* subsp. nov. *Res Microbiol.* 2019;170(3):165–70.
96. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. *PLoS Genet.* 2019;15(4):e1008114. <https://doi.org/10.1371/journal.pgen.1008114> PMID: 30986243
97. Kurgan G, Kurgan L, Schneider A, Onyeabor M, Rodriguez-Sanchez Y, Taylor E, et al. Identification of major malate export systems in an engineered malate-producing *Escherichia coli* aided by substrate similarity search. *Appl Microbiol Biotechnol.* 2019;103(21–22):9001–11. <https://doi.org/10.1007/s00253-019-10164-y> PMID: 31641813
98. Yasid NA, Rolfe MD, Green J, Williamson MP. Homeostasis of metabolites in *Escherichia coli* on transition from anaerobic to aerobic conditions and the transient secretion of pyruvate. *R Soc Open Sci.* 2016;3(8):160187. <https://doi.org/10.1098/rsos.160187> PMID: 27853594
99. Kim S, Jin SH, Lim HG, Lee B, Kim J, Yang J, et al. Synthetic cellular communication-based screening for strains with improved 3-hydroxypropionic acid secretion. *Lab Chip.* 2021;21(22):4455–63. <https://doi.org/10.1039/d1lc00676b> PMID: 34651155
100. Halpern YS, Barash H, Druck K. Glutamate transport in *Escherichia coli* K-12: nonidentity of carriers mediating entry and exit. *J Bacteriol.* 1973;113(1):51–7. <https://doi.org/10.1128/jb.113.1.51-57.1973> PMID: 4567140
101. Abe K, Ohnishi F, Yagi K, Nakajima T, Higuchi T, Sano M, et al. Plasmid-encoded asp operon confers a proton motive metabolic cycle catalyzed by an aspartate-alanine exchange reaction. *J Bacteriol.* 2002;184(11):2906–13. <https://doi.org/10.1128/JB.184.11.2906-2913.2002> PMID: 12003930
102. Hori H, Yoneyama H, Tobe R, Ando T, Isogai E, Katsumata R. Inducible L-alanine exporter encoded by the novel gene *ygaW* (*alaE*) in *Escherichia coli*. *Appl Environ Microbiol.* 2011;77(12):4027–34. <https://doi.org/10.1128/AEM.00003-11> PMID: 21531828
103. Palomino A, Gewurz D, DeVine L, Zajmi U, Morales J, Abu-Rumman F, et al. Metabolic genes on conjugative plasmids are highly prevalent in *Escherichia coli* and can protect against antibiotic treatment. *ISME J.* 2023;17(1):151–62. <https://doi.org/10.1038/s41396-022-01329-1> PMID: 36261510
104. Meador JP, Caldwell ME, Cohen PS, Conway T. *Escherichia coli* pathotypes occupy distinct niches in the mouse intestine. *Infect Immun.* 2014;82(5):1931–8. <https://doi.org/10.1128/IAI.01435-13> PMID: 24566621
105. Harrow GL, Lees JA, Hanage WP, Lipsitch M, Corander J, Colijn C, et al. Negative frequency-dependent selection and asymmetrical transformation stabilise multi-strain bacterial population structures. *ISME J.* 2021;15(5):1523–38. <https://doi.org/10.1038/s41396-020-00867-w> PMID: 33408365
106. Christie MR, McNickle GG. Negative frequency dependent selection unites ecology and evolution. *Ecol Evol.* 2023;13(7):e10327. <https://doi.org/10.1002/ece3.10327> PMID: 37484931
107. Calland JK, Haukka K, Kpordze SW, Brusah A, Corbella M, Merla C, et al. Population structure and antimicrobial resistance among *Klebsiella* isolates sampled from human, animal, and environmental sources in Ghana: a cross-sectional genomic One Health study. *Lancet Microbe.* 2023;4(11):e943–52. [https://doi.org/10.1016/S2666-5247\(23\)00208-2](https://doi.org/10.1016/S2666-5247(23)00208-2) PMID: 37858320
108. Thorpe HA, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, et al. A large-scale genomic snapshot of *Klebsiella* spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings. *Nat Microbiol.* 2022;7(12):2054–67. <https://doi.org/10.1038/s41564-022-01263-0> PMID: 36411354
109. Hetland MAK, Winkler MA, Kaspersen HP, Håkonsholm F, Bakksjø R-J, Bernhoff E, et al. A genome-wide One Health study of *Klebsiella pneumoniae* in Norway reveals overlapping populations but few recent transmission events across reservoirs. *Genome Med.* 2025;17(1):42. <https://doi.org/10.1186/s13073-025-01466-0> PMID: 40296028
110. Spragge F, Bakkeren E, Jahn MT, B N Araujo E, Pearson CF, Wang X, et al. Microbiome diversity protects against pathogens by nutrient blocking. *Science.* 2023;382(6676):eadj3502. <https://doi.org/10.1126/science.adj3502> PMID: 38096285

111. Furuichi M, Kawaguchi T, Pust M-M, Yasuma-Mitobe K, Plichta DR, Hasegawa N, et al. Commensal consortia decolonize Enterobacteriaceae via ecological control. *Nature*. 2024;633(8031):878–86. <https://doi.org/10.1038/s41586-024-07960-6> PMID: [39294375](#)
112. Osbelt L, Wende M, Almási É, Derksen E, Muthukumarasamy U, Lesker TR, et al. *Klebsiella oxytoca* causes colonization resistance against multidrug-resistant *K. pneumoniae* in the gut via cooperative carbohydrate competition. *Cell Host Microbe*. 2021;29(11):1663–79.e7.
113. Eberl C, Weiss AS, Jochum LM, Durai Raj AC, Ring D, Hussain S, et al. *E. coli* enhance colonization resistance against *Salmonella* Typhimurium by competing for galactitol, a context-dependent limiting carbon source. *Cell Host Microbe*. 2021;29(11):1680–92.e7.
114. Gül E, Abi Younes A, Huuskonen J, Diawara C, Nguyen BD, Maurer L, et al. Differences in carbon metabolic capacity fuel co-existence and plasmid transfer between *Salmonella* strains in the mouse gut. *Cell Host Microbe*. 2023;31(7):1140–1153.e3. <https://doi.org/10.1016/j.chom.2023.05.029> PMID: [37348498](#)