# A metabolic atlas of the *Klebsiella pneumoniae* species complex reveals lineage-specific metabolism that supports co-existence of diverse lineages

## Short title: *K. pneumoniae* lineage specific metabolism

## Authors

Ben Vezina[1,2+], Helena B. Cooper[1,2], Martin Rethoret-Pasty[3], Sylvain Brisse[3], Jonathan M. Monk[4], Kathryn E. Holt[1,5] and Kelly L. Wyres[1,2+]

[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] Institut Pasteur, Université Paris Cité, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, France

[4] Department of Medicine, University of California, San Diego School of Medicine, San Diego, CA, United States of America

[5] Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, UK

+ Corresponding authors Ben Vezina (benjamin.vezina@monash.edu); Kelly L. Wyres (Kelly.wyres@monash.edu)

## Competing interests

The authors declare no competing interests

# 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/subspecies and hundreds of distinct co-circulating sub-lineages 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 as a driver of population structure. 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 modelling 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 sub-lineages exhibit unique metabolic profiles. *In vitro* co-culture experiments demonstrated reciprocal commensalistic cross-feeding between sub-lineages, effectively extending the range of conditions supporting individual growth. We propose that these substrate specialisations promote the existence and persistence of co-circulating sub-lineages by reducing nutrient competition and facilitating commensal interactions via negative frequency-dependent selection.

Our findings have implications for understanding the eco-evolutionary dynamics of *K. pneumoniae* 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 (*Kp*SC) (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-lineages 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 sub-lineage epidemiologies i.e. where a subset are recognised 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 non-randomly distributed between sub-lineages and that up to 37% of the total *Kp*SC gene pool encodes proteins contributing to cellular metabolism (4).

We therefore hypothesised that sub-lineages differ in terms of metabolic capabilities, which likely contributes to their epidemiological behaviours and facilitates the maintenance of diverse sub-lineages 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 non-random 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 modelling analyses can be utilised to probe diversity within bacterial species populations, and recent larger-scale studies (n = 50-2,446 isolates) have highlighted the power of these approaches to identify industrially relevant, ecology-, drug-resistance- and pathogenicity-associated metabolic traits (12-17). While a recent study of 2,773 metabolic models confirmed that *E. coli* phylogroups (which are approximately as diverse as *Kp*SC species) are associated with unique metabolic reactions (18). 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 sub-lineages. Such information is crucial to understanding the interaction between metabolism, ecology and evolution within species, and for design of broad-acting control strategies that seek to exploit metabolic vulnerabilities.

Here, we describe a population-level analysis of 7,835 *Kp*SC genomes to evaluate the distribution and diversity of metabolism in the context of the population structure, first at the level of species, then sub-lineages 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 *Kp*SC 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 non-randomly distributed across sub-lineages. Notably, we show for the first time that *Kp*SC sub-lineages, including multi-drug resistant and hypervirulent clones, are associated

with distinct substrate usage profiles that facilitate commensal interactions and likely promote co-existence.

# Materials and Methods

## Genome acquisition, quality filtering and genotyping

Genomes were included from 29 medium-large-scale studies (4, 19-45) that comprised diverse KpSC collections (**Table S1**). Outbreak investigations were explicitly excluded to reduce clonal oversampling and genomes were de-replicated to minimise sampling biases for overrepresented sub-lineages (see **Supplemental Methods**). Reads were acquired from the European Nucleotide Archive using the enaDataGet and enaGroupGet scripts from enaBrowserTools v1.6 (46), trimmed with Trim Galore v0.5.0 using -q 20 option (47) then assembled with Unicycler v0.4.7 with the --keep 0 option (48). 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 analysed to count GFA dead ends using getUnicyclerGraphStats.py from Unicycler (48)

Assemblies were annotated with Bakta v1.1.1 (49) with the '--gram –' option and Bakta database (date accessed: 1/09/2021), while Kleborate v2.0.4 (50) was used to identify species and Sequence Types (STs). Pathogenwatch (https://pathogen.watch/) and BIGSdb (https://bigsdb.pasteur.fr/) (51, 52) were used to assign LIN codes (53) and sub-lineages. Pairwise ANI was calculated using FastANI v1.33 (54) using --fragLen 3000.

## Pangenome analysis

The 7,835 filtered and dereplicated *Kp*SC genomes were split into species, then Panaroo v1.2.8 (55) 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 pangenome.

## Metabolic gene identification and analysis

To analyse metabolic pathways, the pan_genome_reference.fa file from Panaroo was translated using the transeq tool from EMBOSS v6.6.0 (56) with the following options: '-table 11 -frame 1'. Translated sequences were analysed using kofam_scan v1.3.0 (https://github.com/takaram/kofam_scan), the command line version of KEGG's (57) kofamKOALA (58). KEGG release 94.0 ver. 2020-04-02.

Genes with an e-value of ≤0.001 were retained and analysed 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 non-metabolic. 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 (59).

Twilight (60) was used to classify distribution of metabolic orthologs among *Kp*SC taxa and 48 common *K. pneumoniae* sub-lineages (n≥15 genomes each, defined by LIN codes) using the -s 5 and -s 20 commands, respectively. 'Core': ≥95% prevalence. 'Intermediate': <95% and >15% prevalence. 'Rare': ≤15% and >0% prevalence. 'Absent': 0% prevalence.

## Metabolic modelling

Draft metabolic models were generated via Bactabolize v1.0.1 (61) with the *Kp*SC-pan v2.0 model (62) as the input reference (see **Supplemental Methods**). Flux balance analysis as implemented in Bactabolize v1.0.1 was used to predict binary growth phenotypes for all possible carbon, nitrogen, phosphorus and sulfur sources supported by the reference model as sole sources in M9 minimal media, aerobic and anaerobic conditions (**Supplemental Methods**). 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 sub-lineage (n=2,427 genomes), then predicted aerobic growth phenotypes analysed using Coinfinder v1.2.0 (63) with the --associate option.

## *In vitro* growth experiments

To validate substrate usage predictions from the metabolic models, 13 isolates (**Table S2**) were selected from our in-house collections (4, 25, 64) and tested in triplicate on each of nine substrates in minimal media, aerobic conditions. Two experiments were performed: i) endpoint $OD_{600}$ at 24- and 48-hours; ii) 3-day substrate coaxing to induce gene expression (65) for predicted false-negative growth results in the endpoint experiment. Isolates were

coaxed onto the testing substrate via subculturing 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). See **Supplemental Methods** for details.

## Co-culture experiments

We selected three pairs of sub-lineages and three corresponding pairs of substrates (four total substrates) for which one was predicted as a sub-lineage specific core trait in sub-lineage A and the other predicted as a sub-lineage specific core trait in sub-lineage B. For each sub-lineage, three isolates were selected from our in-house collections (4, 25, 66) and treated as biological replicates (**Table S3**).

Isolate pairs were grown in single and co-culture using 0.4 µm pore Transwells (Sigma) and 24-well plates (Costar, Corning). See **Supplemental Methods** for further details. Co-culture results were compared to single cultures and $\log_2$-fold change $OD_{600}$ values calculated.

## Co-culture simulations

*In silico* co-culture experiments were performed using MICOM v0.32.2 (67) 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 was then analysed to determine metabolite sharing from prototrophs to auxotrophs.

## Data visualisation and code availability

Statistical analysis and graphical visualisation were performed using R v4.0.3 (68), RStudio v1.3.1093 (69), with the following software packages: tidyverse v1.3.1 (70), viridis v0.5.1 (71), RColorBrewer v1.1-2 (72), ggpubr v0.4.0 (73) ggpmisc v0.4.4 (74), aplot v0.1.6 (75), colorspace v2.0-2 (76), ggtree v2.4.1 (77), vegan v2.5-7 (59), rstatix v0.7.0 (78) and ggnewscale v0.4.5 (79).

All code used to generate figures, simulations and perform statistical analysis can be found at (https://dx.doi.org/10.6084/m9.figshare.24503737).

# Results

## Metabolic traits are bimodally distributed across the population

The study dataset included 7,835 high-quality *Kp*SC genomes covering all seven taxa in the complex (**Fig. 1A**) and 1,931 sequence types (STs) (**Table S1**). Pan-genome analysis (55) identified 61,595 gene clusters, 11,184 (18.2%) of which were matched to a total of 2,601 unique KEGG Orthologs (57) with associated substrate-reaction data. This represented a lower bound estimate of the proportion of 'metabolic genes' in the *Kp*SC 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, Table S4**), with 1,499 orthologs (57.4%) present in ≥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 *Kp*SC (4), and other bacterial species with similarly large effective population sizes (60, 80).

To analyse the phenotypic diversity across the *Kp*SC, a strain-specific metabolic model was generated for each genome using Bactabolize (61), which was previously shown to produce models with 93.2% and 78.1% predictive accuracy for aerobic and anaerobic growth phenotypes, respectively (62). 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 (total 1,278 conditions). Individual strain-specific models comprised a mean of 1677.0 genes ± 17.3 (standard deviation, SD) and 3330.1 reactions ± 13.5 (SD). Individual isolates were predicted to grow in a mean of 678.2 ± 17.9 (SD) conditions (**Table S5**, **Fig. S1A, Supplemental Results**).

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 ≥95% isolates ('core', **Fig. 1C**, **Table S5**) and were comprised of 349 aerobic and 294 anaerobic conditions, which systematically confirmed the facultative anaerobe status of the *Kp*SC. 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 (**Table S5**). In contrast, 509 modelled 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 anerobic conditions predicted to support growth in <95% isolates (**Fig. 1C, Fig. S1B**).

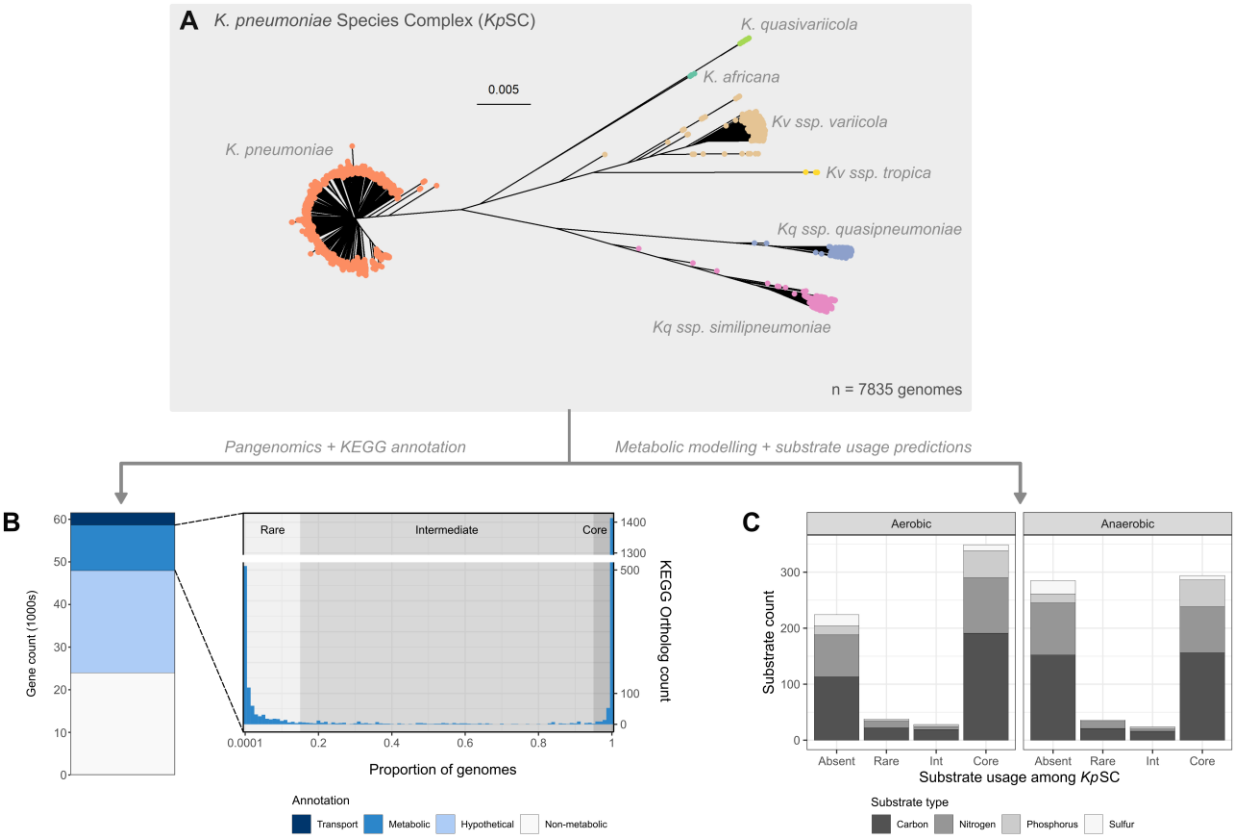



**Fig. 1: Population level metabolic diversity in the *K. pneumoniae* Species Complex**

**A:** Neighbour-joining tree of the 7,835 *Kp*SC genomes used in this study. Tip colours indicate species as labelled. An interactive version can be found at https://microreact.org/project/pcTMwQZAGCsqqhbEZzaj11-a-metabolic-atlas-of-the-klebsiella-pneumoniae-species-complex. **B:** Stacked bars show pangenome 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 grey background shows rare orthologs (>0 to 15%), the medium grey shows intermediate orthologs (>15% to <95%) while the dark grey shows core orthologs (≥95%). **C:** Stacked bars show population frequency categories for predicted substrate usage, coloured by substrate type as indicated in legend.

## Metabolic variation is structured within and between species

The set of core growth conditions for each *Kp*SC 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 ortholog profiles as expected (66, 81) (**Fig. S2 and S3, Tables S4 and S5, Supplemental Results**), and there was evidence of specialisation at the level of macromolecular functions as determined by COG

analysis (**Fig. S4, Table S1, Supplemental Results**). These predicted growth capabilities were generally consistent with biochemical tests described in formal species definitions (82-85). However, there were a small number of notable discrepancies that indicate that species identification protocols should be revisited (see **Supplemental Results** and **Table S6** for details).

While individual *Kp*SC taxa are clearly defined by distinct metabolic profiles, there was also clear evidence of intra-taxon variation (**Fig. S2 and S3**). 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 hypothesised 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, $p<2x10^{-16}$ for each of the four well sampled taxa, **Fig. 2A**). Comparison of the regression gradients indicated significant differences between taxa; however, the $R^2$ values were low (range 0.04-0.48), indicating a large amount of unexplained variation.

Next we used the recently established core-genome Life Identifier Number (LIN) code scheme to assign genomes to sub-lineages (SL) (53). These sub-lineages 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 sub-lineages and two were unassigned due to insufficient alleles detected.

Pairwise Jaccard similarities for metabolic ortholog profiles were significantly higher within sub-lineages (median 0.956, IQR 0.947-0.965) than between sub-lineages 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 sub-lineages (median 0.990, IQR 0.983-0.997) than between sub-lineages (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**).

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 **Supplemental Results**). 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).

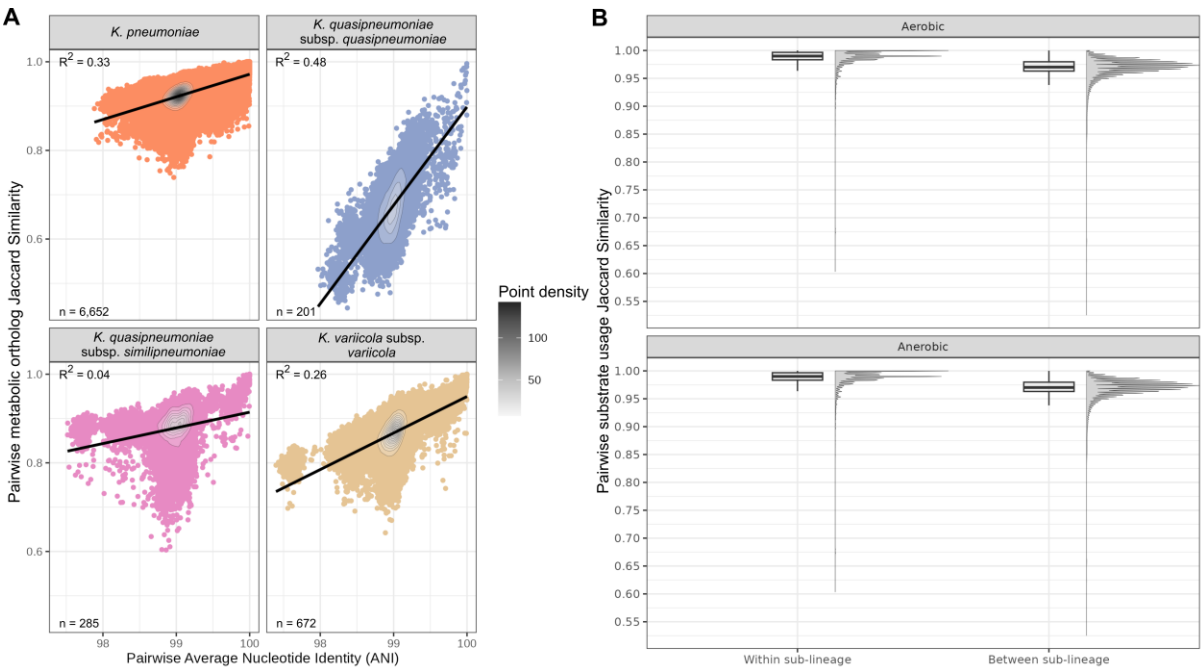



**Fig. 2: Metabolic heterogeneity within taxa is non-randomly distributed**

**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 sub-lineages in the same species. Significance asterisks excluded as each group was significantly different from every other group ($p < 2x10^{-16}$), calculated using a non-parametric Kruskal–Wallis test with Holm correction, followed by Dunn's *post-hoc* test.

## Major sub-lineages 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* sub-lineages that were each represented by genomes from broad geographies and time-spans (≥15 genomes, 3-12 geographic regions, 5-86 years). These included sub-lineages corresponding to globally distributed multi-drug resistant (SL14, SL15, SL17, SL29, SL37, SL101, SL147, SL258, SL307) and hypervirulent clones (SL23, SL25, SL65, SL86) (3). The data showed that each sub-lineage is associated with a distinct, core metabolic fingerprint, indicating metabolic specialisation (**Fig. 3**, **Fig. S5**, **Table S5**), with 253-316 core aerobic and 198-223 core anaerobic conditions each. Note that in some cases the sub-lineage

specific core was smaller than the *Kp* species core reported above using the ≥95% genome threshold (363 aerobic, 307 anaerobic conditions), because most sub-lineages 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. No pair of sub-lineages shared the same core growth profile, even pairs that are thought to share relatively recent common ancestry such as SL14 and SL15 (common ancestor approximately 45 years ago (4)) and SL65 and SL25 (common ancestor yet to be dated) (86) (**Fig. 3**, **Fig. S5**). However, there was also clear evidence of diversity within sub-lineages, with 3-102 variable growth phenotypes each, a pattern that was replicated by metabolic ortholog distributions (**Fig. S3, Fig. S5, Supplemental Results**).

After excluding growth conditions predicted to support growth among ≥95% (46/48) *K. pneumoniae* sub-lineages (n=615 conditions, 340 aerobic and 275 anerobic, corresponding to 202 total substrates) and those that did not support any growth among the 48 sub-lineages (n=525 conditions), the data support a broad classification of phenotypes as follows: i) 89 that were core in some sub-lineages (range=1-45 sub-lineages) but variably present in others (range=1-17) which we call 'sub-lineage-specific core traits'; ii) 13 that were not core in any sub-lineage (accessory in 3-46 sub-lineages, including the only pair of co-occurring phenotypes that was not mechanistically linked) (**Fig. 3**, **Fig. S5**). The sub-lineage-specific core traits could be further divided into those that were core to the majority (≥66%) of sub-lineages (n=57 phenotypes that we call 'majority sub-lineage-specific core traits'), those that were core to an intermediate number of sub-lineages (≥25% to <66%, n=13 phenotypes, corresponding to six substrates, that we call 'common sub-lineage-specific core traits') and those that were core in only very few sub-lineages (<25%, n=13 'rare sub-lineage-specific core traits', eight aerobic and five anerobic, corresponding to seven total substrates). We propose that the majority sub-lineage specific core traits are unlikely to play

a major role in defining sub-lineage specific evolutionary trajectories or ecological interactions, and hence we focussed our follow-up investigations on the common and rare sub-lineage specific core traits.

We have previously confirmed predictive accuracies ≥97% for five of the six common sub-lineage-specific core substrates; L-sorbose, L-tartrate, galactitol, D-tagatose and L-galactonate as sole sources of carbon in aerobic conditions (62) (tested using the Omnilog phenotyping microarray). The sixth, fructoselysine, is not present in the Omnilog array and could not be tested here due to its high cost. However, we tested seven of the rare sub-lineage-specific core growth substrates under aerobic conditions, confirming high accuracy for L-hydroxyproline as a carbon source (84.6%) and variable accuracies for the remainder (**Table S2, Supplemental Results**), which was not unexpected as rare substrate usage is comparatively understudied and therefore difficult to predict.

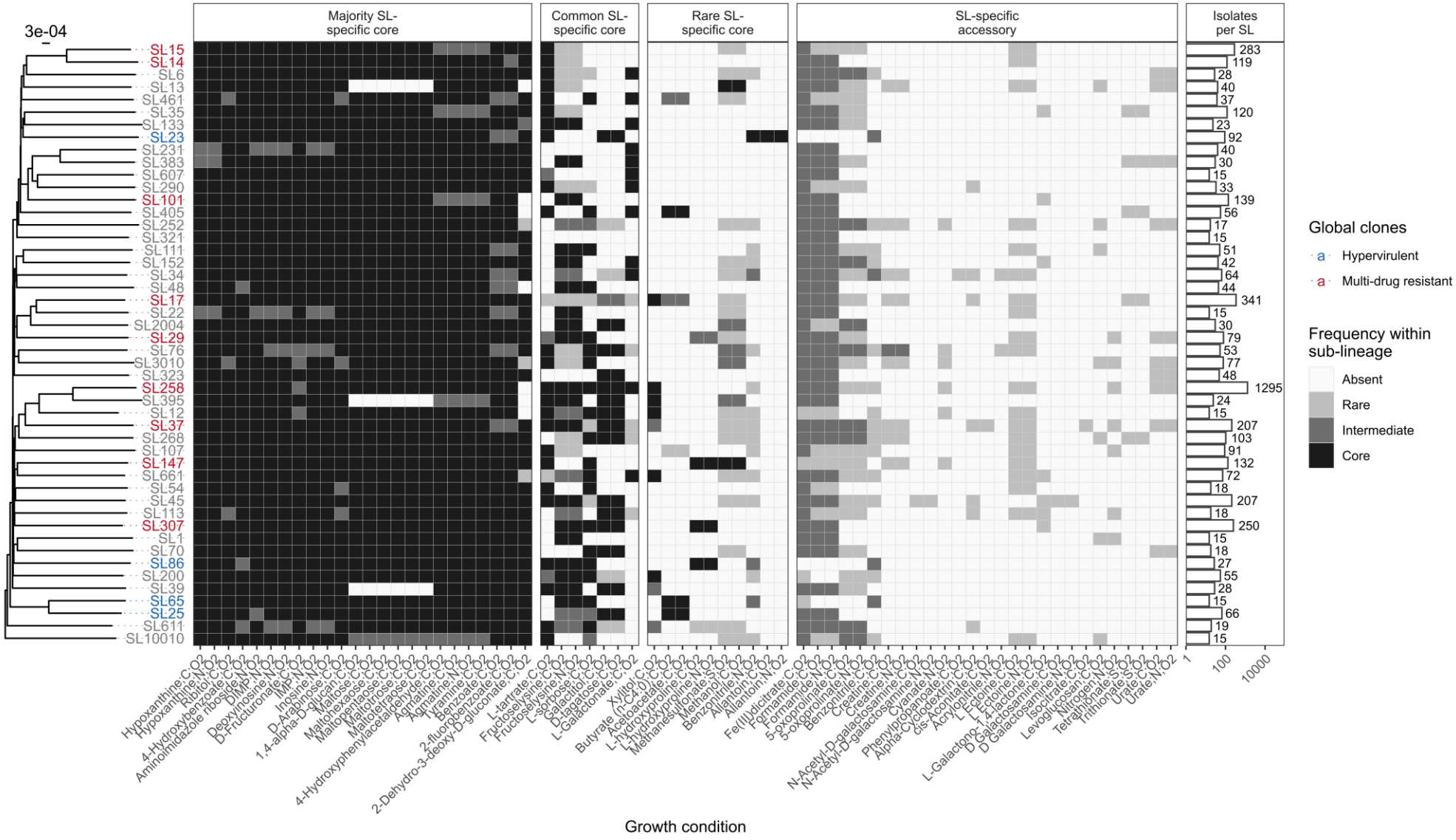

3e-04
Majority SL-specific core
Common SL-specific core
Rare SL-specific core
SL-specific accessory
Isolates per SL
SL15
SL14
SL6
SL13
SL461
SL35
SL133
SL23
SL231
SL383
SL607
SL290
SL101
SL405
SL252
SL321
SL111
SL152
SL34
SL48
SL17
SL22
SL2004
SL29
SL76
SL3010
SL323
SL258
SL395
SL12
SL37
SL268
SL107
SL147
SL661
SL54
SL45
SL113
SL307
SL1
SL70
SL86
SL200
SL39
SL65
SL25
SL611
SL10010
283
119
28
40
37
120
23
92
40
30
15
33
139
56
17
15
51
42
64
44
341
15
30
79
53
77
48
1295
24
15
207
103
91
132
72
18
207
18
250
15
18
27
55
28
15
66
19
15
1
100
10000
Growth condition
Global clones
a Hypervirulent
a Multi-drug resistant
Frequency within sub-lineage
Absent
Rare
Intermediate
Core

**Fig. 3: Distinct metabolic fingerprints of sub-lineages within species**

Heatmap showing frequency of variable aerobic substrate usage across 48 global *K. pneumoniae* sub-lineages (SL). Rows are ordered by phylogeny. Sub-lineage labels are coloured to indicate the globally-distributed clones described in (3): Blue shows hypervirulent while red shows multidrug 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. Number of isolates per each sub-lineage shown in bars. Anaerobic growth predictions shown in **Fig. S5.**

## Sub-lineages co-operate through cross-feeding metabolites

We have shown substrate usage varies across the *Kp*SC and some of this variation is sub-lineage-dependent. We hypothesised that such diversity promotes the maintenance of distinct sub-lineages 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, L-hydroxyproline) for use in three independent substrate combinations and corresponding sub-lineage pairs for which predicted substrate-usages were mutually exclusive sub-lineage-specific core traits. One isolate from each sub-lineage was selected for co-culture aerobically in each of two conditions (minimal media plus each of the sub-lineage-specific core substrates), such that each experiment contained one isolate that was able to utilise 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 sub-lineages 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 sub-lineages 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 behaviour across a greater number of independent isolates. Sub-lineages representing clinically-relevant clones were selected (multi-drug resistant SL37 vs hypervirulent SL29, hypervirulent SL86 vs hypervirulent SL23 and multi-drug resistant SL17 vs SL29).

Our data showed that co-culturing auxotrophs with prototrophs in the presence of a sub-lineage-specific core substrate significantly increased the $OD_{600}$ of the auxotrophs (**Fig. 4B**) with mean $\log_2$-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**). Corresponding *in silico* co-culture simulations 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 **Supplemental Results**).

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 behaviour of each isolate (**Table S3**). Together these results indicated that cross-feeding was occurring from the prototrophs to the auxotrophs across: i) multiple genetic backgrounds (sub-lineages); ii) to and from each strain partner within each pair under different conditions; iii) across independent isolates within sub-lineages, confirming that commensal interactions can occur between co-habiting *Kp* sub-lineages.

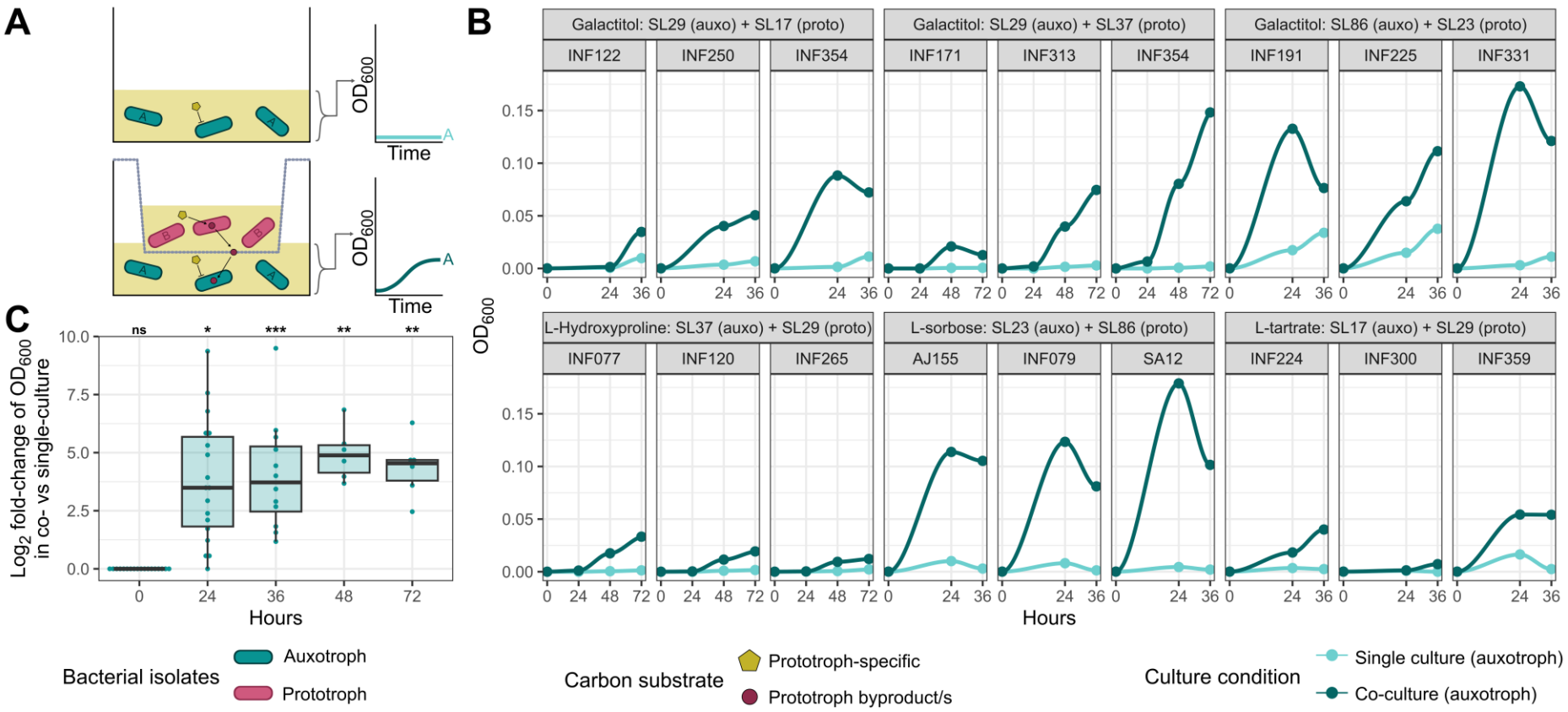



**Fig. 4: Nutrient prototrophic lineages support growth of auxotrophic lineages**

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 (**Methods**). **A:** Graphic demonstrating the experimental design measuring commensalism between an auxotrophic strain unable to utilise 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 sub-lineage 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 timepoint 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, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Raw data can be found in **Table S3**.

# Discussion

Here we present a bacterial population metabolism study of the *K. pneumoniae* species complex, in which we used quantitative metabolic modelling to predict and compare growth phenotypes for >7,000 members of a single species complex, and contextualised 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 using multiple independent isolates from multiple sub-lineages and multiple substrate combinations (**Fig. 4**) provide evidence that these metabolic differences can facilitate commensal interactions between sub-lineages.

Our data shows that growth capabilities are structured in the population: while there was clearly variation within and between sub-lineages, each of the 48 sub-lineages we examined was associated with a distinct set of core traits (metabolic orthologs and growth phenotypes), which likely represents vertical inheritance from the sub-lineage common ancestor (**Fig. 3**, **Fig S5**). This finding is consistent with a model of negative frequency-dependent selection as a driver for the maintenance of diverse sub-lineages (entities) in the population (87, 88). Where two sub-lineages compete for limited resources, e.g. food, the ability to consume a substrate that cannot be consumed by a competitor provides a selective advantage resulting in a relative frequency increase. This in turn results in a depletion of the food resource and shift in selective pressure to favour the competitor. Although comparatively understudied, commensal interactions such as the metabolite cross-feeding implicated by our co-culture experiments (**Fig. 4**), can drive similar frequency-dependent selective shifts (88). Together, these fluctuations prevent the exclusion of sub-lineages and drive the population to a stable equilibrium of co-existence.

Niche range no doubt further complicates the eco-evolutionary dynamics within *Kp* populations e.g. where sub-lineages are exposed to fluctuating selective pressures due to transient migration between niches, and/or where sub-lineages undergo niche adaptation. Recent analyses of contemporaneous *Kp* collections have shown that sub-lineages are not equally distributed between different hosts and environmental sources, which may indicate some level of niche-adaptation (89, 90). However, the same studies also highlighted significant diversity within niches and near-identical strains from multiple sources, which is indicative of niche migration. Unfortunately, we were unable to directly test for associations between sub-lineage, metabolism and ecological niche with our convenience sample of genomes comprising primarily human-derived isolates (6,465/7,835 genomes). Future studies will be required to probe these complex multi-way interactions and to systematically test the model of negative-frequency dependent selection, which is notoriously difficult to detect (88), requiring a long-term dataset from the same location over time, and/or laboratory protocols that are beyond the scope of the current work.

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 (57), and while our pan-metabolic reference model was built from a large diverse collection of strains (62), it cannot possibly capture all metabolism in the *Kp*SC population. Hence, we are likely underestimating the metabolic diversity within the highly variable *Kp*SC. Secondly, growth prediction accuracies can vary by substrate, and are generally lower for anaerobic conditions for which the metabolic processes are less well understood (62). While many of the core substrates and common sub-lineage specific core traits have been confirmed with high accuracy in previous work, our phenotypic testing showed variable accuracy for rarer substrates. Of note, our assays showed that allantoin usage, a trait previously associated with hypervirulent infections (64) and predicted by our models as a sub-lineage specific core

trait in the dominant hypervirulent SL23, may in fact be core to *Kp*SC (see **Supplemental Results**). These and other inaccurate predictions highlight gaps in our current knowledge that are noted for future investigations. Finally, metabolic models do not consider the impact of sequence variations within metabolic genes nor variations in gene regulation that will also play an important role in determining metabolic phenotypes and ecological adaptation.

Notwithstanding the caveats of *in silico* modelling, 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 microbial communities containing this organism. There is increasing interest in gut microbiota manipulation for prevention of pathogen colonisation and subsequent opportunistic infection (91, 92), and studies on *Kp* have thus far implicated competitive exclusion via nutrient competition as the major mechanistic driver (92, 93). Notably, the *Kp* sub-lineage-specific core traits we identified include those implicated in driving competitive interactions and colonisation resistance in the mammalian gut e.g. fructoselysine and galactitol usage (94, 95). We argue that a comprehensive understanding of the population distribution of these and other metabolic traits, will greatly benefit the design of novel control strategies targeting *Kp*, particularly those harnessing microbial competitive interactions or targeting metabolic processes.

# Author Contributions

Conceptualization: BV, SB, JMM, KEH, KLW.

Methodology: BV, SB, JMM, KEH, KLW

Validation: BV.

Formal analysis: BV, HBC, KLW.

Investigation: BV, HBC.

Resources: KLW, SB, KEH.

Data Curation: BV, MR-P

Writing: BV, HBC, MR-P, SB, JMM, KEH, KLW.

Visualization: BV, KLW.

Supervision: SB, KLW.

Project administration: KLW.

Funding acquisition: SB, JMM, KEH, KLW.

# Data availability

All genomic data used is publicly available - accessions and genomes used can be found in **Table S1**. All raw data can be found in **Supplemental Tables**. R and MICOM code can be found at Figshare (https://dx.doi.org/10.6084/m9.figshare.24503737). Interactive phylogeny can be found at https://microreact.org/project/pcTMwQZAGCsqghbEZzaj11-a-metabolic-atlas-of-the-klebsiella-pneumoniae-species-complex.

# Funding

This work was funded by the Australian Research Council Discovery Project (DP200103364, awarded to KLW, KEH, SB and JMM). KLW is supported by a National Health and Medical Research Council of Australia Investigator Grant (APP1176192).

# Acknowledgements

This work was supported by the MASSIVE HPC facility (www.massive.org.au).

# Supplemental Materials

Supplementary Methods, Supplementary Results, Table S6 and Figures S1-S5 are available in a combined PDF (Vezina_Metabolic_Atlas_Supplementary_Materials.pdf). Tables S1-S5 are available in independent Excel files.

**Table S1 (Excel file):** 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.

**Table S2:** Outcomes of *in vitro* growth experiments for metabolic model validation.

**Table S3:** Raw data and statistical analysis of co-culture experiments.

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

**Table S5:** Binarised growth predictions from metabolic modelling for entire 7,835 genome dataset.

**Table S6:** Biochemical tests described in formal species definitions and consistency with model predictions.

**Fig. S1:** Substrate usage frequencies across *Kp*SC dataset.

**Fig. S2:** Taxon-specific substrate usage.

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

**Fig. S4:** Taxon-specific specialisation of macromolecular functions.

**Fig. S5:** Distinct anerobic metabolic fingerprints of sub-lineages within species.

# References


1. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. The Lancet. 2024;404(10459):1199-226.
2. WHO. WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. In: Organisation WH, editor. 2024.
3. Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. Nature Reviews Microbiology. 2020;18(6):344-59.
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 *Klebsiella pneumoniae*, an urgent threat to public health. Proceedings of the National Academy of Sciences. 2015;112(27):E3574.
5. Watkins ER, Maiden MCJ, Gupta S. Metabolic competition as a driver of bacterial population structure. Future Microbiology. 2016;11(10):1339-57.
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. Molecular Biology and Evolution. 2015;32(6):1396-410.
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.
8. Sheppard SK, Cheng L, Méric G, de Haan CP, Llarena AK, Marttinen P, et al. Cryptic ecology among host generalist Campylobacter jejuni in domestic animals. Mol Ecol. 2014;23(10):2442-51.
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. 2021;167(9).
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 Biology. 2023;21(10):e3002329.
11. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics. 2014;15(2):107-20.
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. Nature Communications. 2018;9(1):3771.
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. International Journal of Molecular Sciences [Internet]. 2023; 24(8).
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. Frontiers in Cellular and Infection Microbiology. 2019;9(161).
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. Proceedings of the National Academy of Sciences. 2013;110(50):20338-43.
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. Proceedings of the National Academy of Sciences. 2016;113(26):E3801-E9.

18. 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.
19. 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.
20. 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. The Lancet Infectious Diseases. 2021.
21. 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. Clinical Infectious Diseases. 2015;61(6):892-9.
22. 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 Biology. 2019;20(1):184.
23. 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. The Lancet Infectious Diseases. 2017;17(2):153-63.
24. 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. Frontiers in Microbiology. 2021;12(2609).
25. Gorrie CL, Mirčeta 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. Clinical Infectious Diseases. 2017;65(2):208-15.
26. Hawkey J, Vezina B, Monk JM, Judd LM, Harshegyi T, López-Fernández S, et al. A curated collection of *Klebsiella* metabolic models reveals variable substrate usage and gene essentiality. Genome Research. 2022.
27. 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. International Journal of Medical Microbiology. 2017;307(7):422-9.
28. 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.
29. 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. Journal of Antimicrobial Chemotherapy. 2021.
30. 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. Scientific Reports. 2016;6(1):29316.
31. 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. Microbial Genomics. 2022;8(3).
32. Ludden C, Moradigaravand D, Jamrozy 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. Clinical Infectious Diseases. 2020;70(2):219-26.
33. 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. 2021:103879.
34. 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. Journal of Antimicrobial Chemotherapy. 2019;74(5):1223-32.
35. 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.
36. 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.
37. Rodrigues C, Hauser K, Cahill N, Ligowska-Marzęta 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. Microbiology Spectrum. 2022;10(1):e02376-21.
38. 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. Frontiers in Microbiology. 2018;9.
39. 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.
40. 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. Journal of Antimicrobial Chemotherapy. 2013;68(10):2234-44.
41. 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. Antimicrobial Agents and Chemotherapy. 2014;58(12):7347-57.
42. 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. Environmental Microbiology. 2022;24(9):4425-36.
43. Wand Matthew E, Baker Kate S, Benthall G, McGregor H, McCowen James WI, Deheer-Graham A, et al. Characterization of Pre-Antibiotic Era *Klebsiella pneumoniae* Isolates with Respect to Antibiotic/Disinfectant Susceptibility and Virulence in *Galleria mellonella*. Antimicrobial Agents and Chemotherapy. 2015;59(7):3966-72.
44. 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 Medicine. 2020;12(1):11.
45. Zheng Z, Gorden PJ, Xia X, Zheng Y, Li G. Whole-genome analysis of *Klebsiella pneumoniae* from bovine mastitis milk in the U.S. Environmental Microbiology. 2021;n/a(n/a).
46. European-Nucleotide-Archive. enaBrowserTools. 1.6 ed. https://github.com/2020.
47. Krueger F. Trim Galore GitHub2012 [Available from: https://github.com/FelixKrueger/TrimGalore.
48. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology. 2017;13(6):e1005595.
49. 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. Microbial Genomics. 2021;7(11).
50. 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. Nature Communications. 2021;12(1):4188.

51. 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-s35.
52. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11(1):595.
53. 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. Molecular Biology and Evolution. 2022;39(7):msac135.
54. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications. 2018;9(1):5114.
55. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome biology [Internet]. 2020 2020/07//; 21(1):[180 p.]. Available from: http://europepmc.org/abstract/MED/32698896

https://doi.org/10.1186/s13059-020-02090-4

https://europepmc.org/articles/PMC7376924

https://europepmc.org/articles/PMC7376924?pdf=render.
56. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics. 2000;16(6):276-7.
57. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27-30.
58. 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. 2019;36(7):2251-2.
59. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2.5-7 ed2020.
60. 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. Microbial Genomics. 2021;7(9).
61. 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.
62. 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. Microbial Genomics. 2024;10(2).
63. Whelan FJ, Rusilowicz M, McInerney JO. Coinfinder: detecting significant associations and dissociations in pangenomes. Microbial Genomics. 2020;6(3).
64. Chou HC, Lee CZ, Ma LC, Fang CT, Chang SC, Wang JT. Isolation of a chromosomal region of Klebsiella pneumoniae associated with allantoin metabolism and liver infection. Infect Immun. 2004;72(7):3783-92.
65. 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.
66. Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. Metabolic diversity of the emerging pathogenic lineages of *Klebsiella pneumoniae*. Environmental Microbiology. 2017;19(5):1881-98.
67. Diener C, Gibbons Sean M, Resendis-Antonio O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. mSystems. 2020;5(1):e00606-19.
68. R-Core-Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2020.
69. RStudio-Team. RStudio: Integrated Development for R. Boston, MA RStudio; 2020.

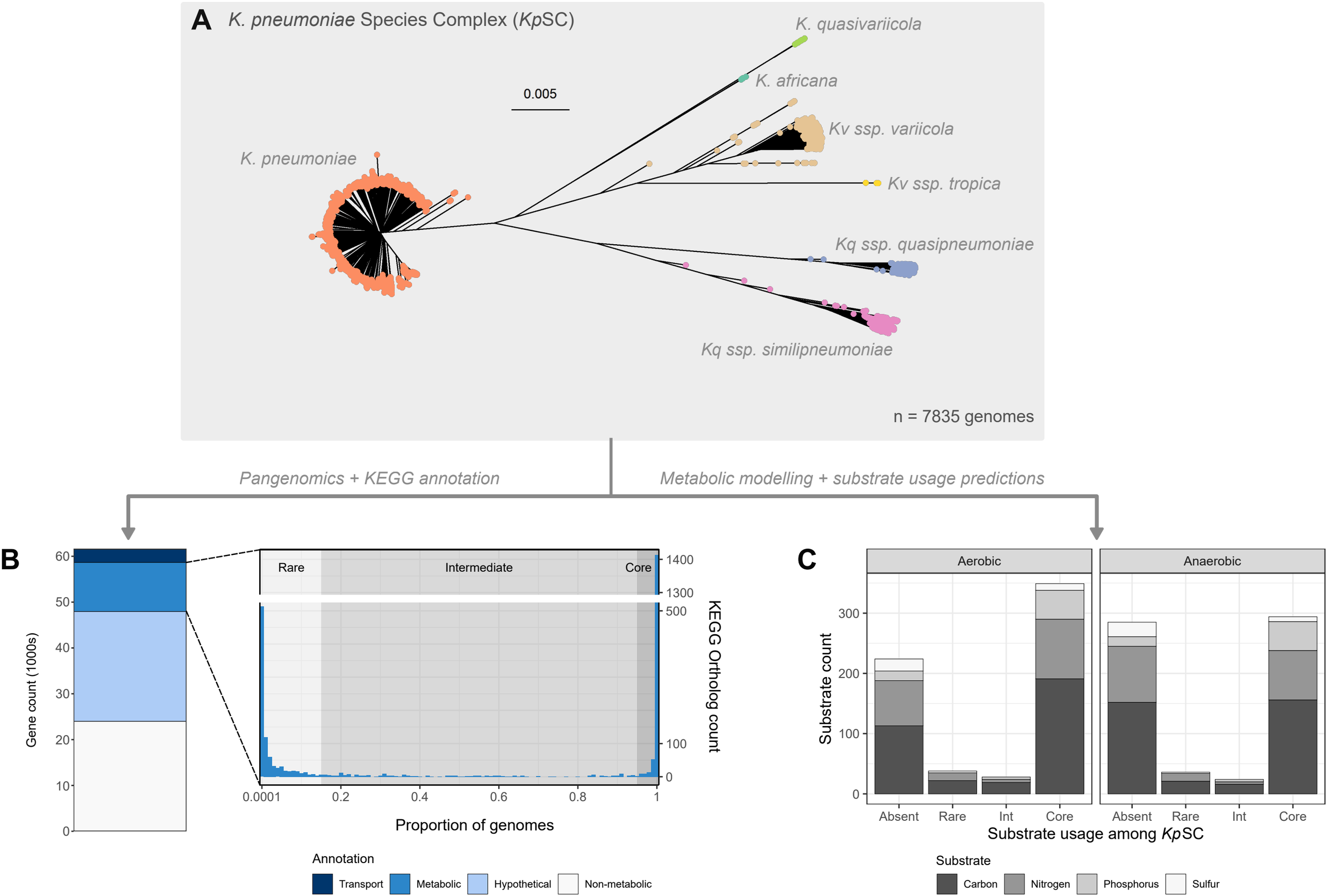
70. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4(43):1686.
71. Garnier S. viridis: Default Color Maps from 'matplotlib'. 2018.
72. Neuwirth E. RColorBrewer: ColorBrewer Palettes. 1.1-2 ed2014.
73. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. 0.5.0 ed2022.
74. Aphalo PJ, Slowikowski K, Mouksassi S. ggpmisc: Miscellaneous Extensions to 'ggplot2'. 0.4.1 ed2021.
75. Yu G. aplot: Decorate a 'ggplot' with Associated Information. 0.0.6 ed2020.
76. Zeileis A, Fisher JC, Hornik K, Ihaka R, McWhite CD, Murrell P, et al. colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes. Journal of Statistical Software. 2020;96(1):1 - 49.
77. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017;8(1):28-36.
78. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. 2023.
79. Campitelli E. ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'. 0.4.5 ed2021.
80. McInerney JO. Prokaryotic Pangenomes Act as Evolving Ecosystems. Molecular Biology and Evolution. 2023;40(1):msac232.
81. 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 Medicine. 2022;14(1):97.
82. Imhoff J. Bergey's Manual® of Systematic Bacteriology. 2005. p. 587-850.
83. Rosenblueth M, Martínez L, Silva J, Martínez-Romero E. *Klebsiella variicola*, A Novel Species with Clinical and Plant-Associated Isolates. Systematic and Applied Microbiology. 2004;27(1):27-35.
84. 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.
85. 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.
86. 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 Genetics. 2019;15(4):e1008114.
87. 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. The ISME Journal. 2021;15(5):1523-38.
88. Christie MR, McNickle GG. Negative frequency dependent selection unites ecology and evolution. Ecol Evol. 2023;13(7):e10327.
89. 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. The Lancet Microbe. 2023;4(11):e943-e52.
90. 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. Nature Microbiology. 2022;7(12):2054-67.
91. 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.382(6676):eadj3502.

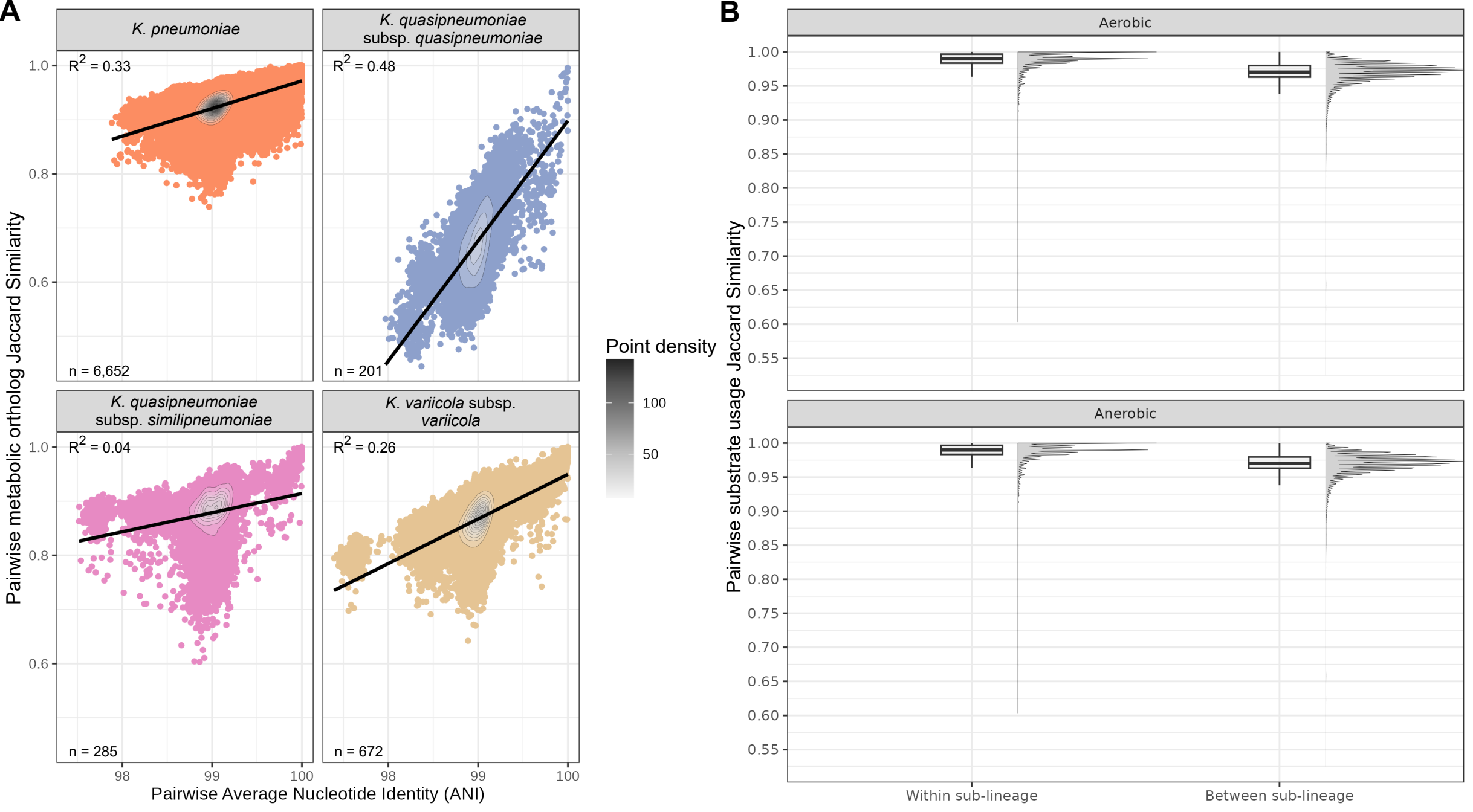
92. Furuichi M, Kawaguchi T, Pust M-M, Yasuma-Mitobe K, Plichta DR, Hasegawa N, et al. Commensal consortia decolonize Enterobacteriaceae via ecological control. Nature. 2024;633(8031):878-86.

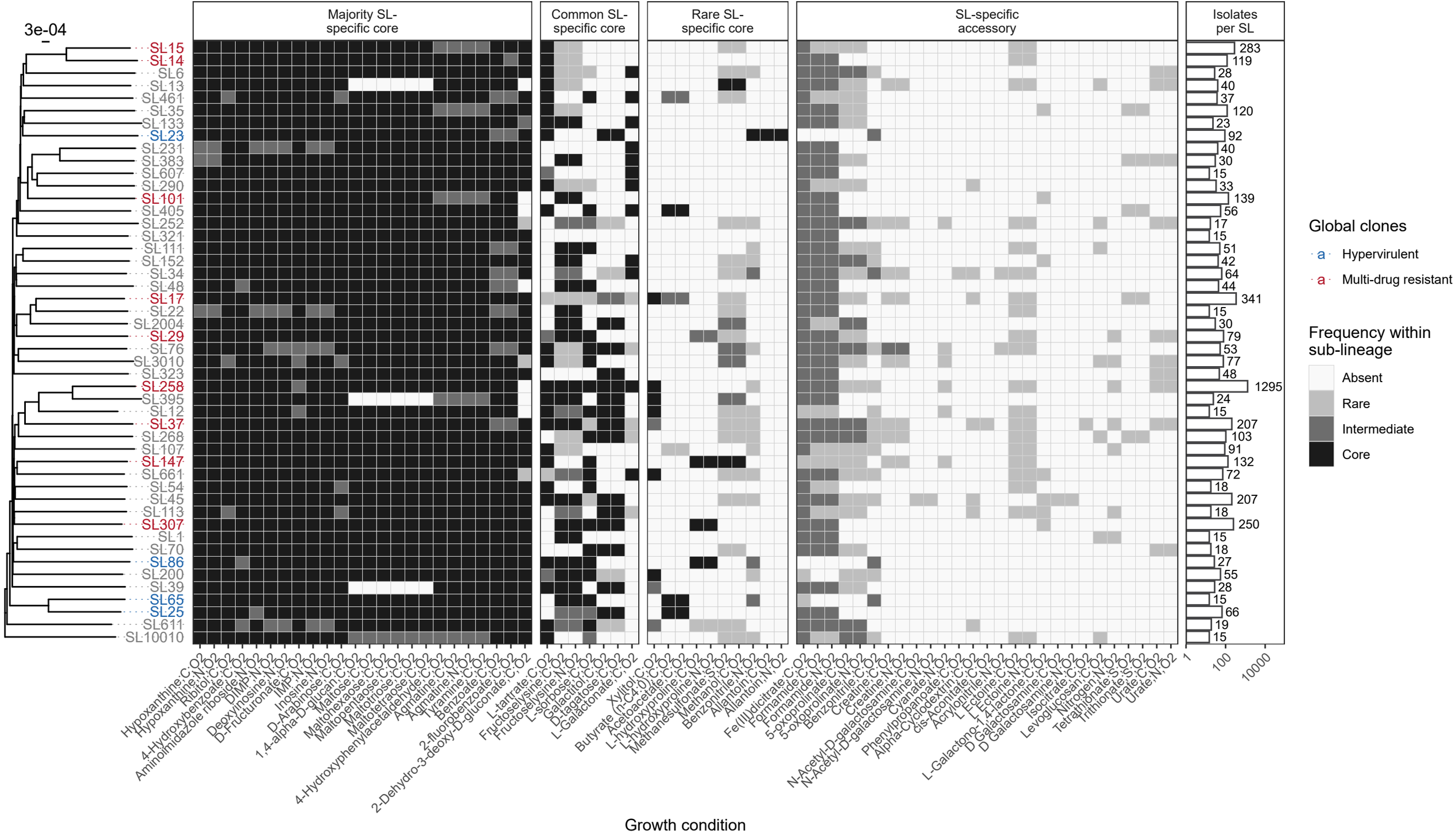
93. Osbelt L, Wende M, Almási É, Derksen E, Muthukumarasamy U, Lesker TR, et al. Klebsiella oxytoca causes colonization resistance against multidrug-resistant K. pneumoniae in the gut via cooperative carbohydrate competition. Cell Host & Microbe. 2021;29(11):1663-79.e7.

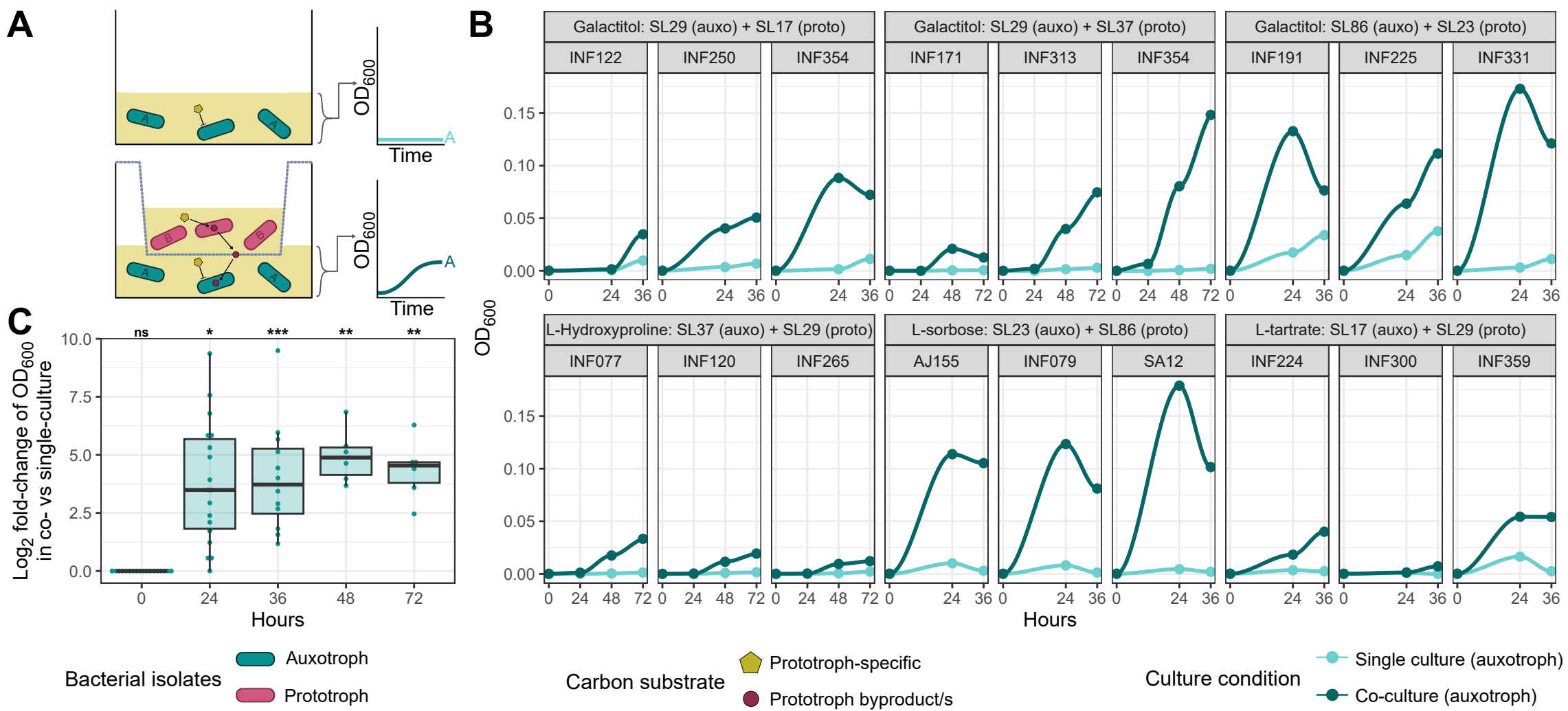
94. Eberl C, Weiss AS, Jochum LM, Durai Raj AC, Ring D, Hussain S, et al. *E. coli* enhance colonization resistance against *Salmonella* Typhimurium by competing for galactitol, a context-dependent limiting carbon source. Cell Host & Microbe. 2021;29(11):1680-92.e7.

95. 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):1140-53.e3.


A
K. pneumoniae Species Complex (KpSC)
K. quasivariicola
K. africana
0.005
Kv ssp. variicola
K. pneumoniae
Kv ssp. tropica
Kq ssp. quasipneumoniae
Kq ssp. similipneumoniae
n = 7835 genomes
Pangenomics + KEGG annotation
Metabolic modelling + substrate usage predictions
B
Gene count (1000s)
Rare
Intermediate
Core
KEGG Ortholog count
Proportion of genomes
Annotation
Transport
Metabolic
Hypothetical
Non-metabolic
C
Aerobic
Anaerobic
Substrate count
Absent
Rare
Int
Core
Substrate usage among KpSC
Substrate
Carbon
Nitrogen
Phosphorus
Sulfur


A
K. pneumoniae
R2 = 0.33
n = 6,652
K. quasipneumoniae subsp. quasipneumoniae
R2 = 0.48
n = 201
K. quasipneumoniae subsp. similipneumoniae
R2 = 0.04
n = 285
K. variicola subsp. variicola
R2 = 0.26
n = 672
Pairwise metabolic ortholog Jaccard Similarity
Pairwise Average Nucleotide Identity (ANI)
Point density
100
50
B
Aerobic
Anerobic
Pairwise substrate usage Jaccard Similarity
Within sub-lineage
Between sub-lineage


3e-04
Majority SL-specific core
Common SL-specific core
Rare SL-specific core
SL-specific accessory
Isolates per SL
SL15
SL14
SL6
SL13
SL461
SL35
SL133
SL23
SL231
SL383
SL607
SL290
SL101
SL405
SL252
SL321
SL111
SL152
SL34
SL48
SL17
SL22
SL2004
SL29
SL76
SL3010
SL323
SL258
SL395
SL12
SL37
SL268
SL107
SL147
SL661
SL54
SL45
SL113
SL307
SL1
SL70
SL86
SL200
SL39
SL65
SL25
SL611
SL10010
283
119
28
40
37
120
23
92
40
30
15
33
139
56
17
15
51
42
64
44
341
15
30
79
53
77
48
1295
24
15
207
103
91
132
72
18
207
18
250
15
18
27
55
28
15
66
19
15
Hypoxanthine;C;O2
Hypoxanthine;N;O2
Ribitol;C;O2
4-Hydroxybenzoate;C;O2
Aminoimidazole riboside;N;O2
DIMP;N;O2
Deoxyinosine;C;O2
D-Fructuronate;C;O2
IMP;N;O2
Inosine;N;O2
D-Arabinose;C;O2
1,4-alpha-D-glucan;C;O2
Maltose;C;O2
Maltohexaose;C;O2
Maltopentaose;C;O2
Maltotriose;C;O2
Maltotetraose;C;O2
4-Hydroxyphenylacetaldehyde;C;O2
Agmatine;C;O2
Agmatine;N;O2
Tyramine;C;O2
Benzoate;C;O2
2-fluorobenzoate;C;O2
2-Dehydro-3-deoxy-D-gluconate;C;O2
L-tartrate;C;O2
Fructoselysine;C;O2
Fructoselysine;N;O2
L-sorbose;C;O2
Galactitol;C;O2
D-tagatose;C;O2
L-Galactonate;C;O2
Xylitol;C;O2
Butyrate (n-C4:0);C;O2
Acetoacetate;C;O2
L-hydroxyproline;C;O2
L-hydroxyproline;N;O2
Methanesulfonate;S;O2
Methanol;C;O2
Benzonitrile;N;O2
Allantoin;C;O2
Allantoin;N;O2
Fe(III)dicitrate;C;O2
Formamide;C;O2
Formamide;N;O2
5-oxoprolinate;C;O2
5-oxoprolinate;N;O2
Benzonitrile;C;O2
Creatine;C;O2
Creatine;N;O2
N-Acetyl-D-galactosamine;C;O2
N-Acetyl-D-galactosamine;N;O2
Cyanate;N;O2
Phenylpropanoate;C;O2
Alpha-Cyclodextrin;C;O2
cis-Aconitate;C;O2
Acrylonitrile;N;O2
L-Ectoine;C;O2
L-Ectoine;N;O2
L-Galactono-1,4-lactone;C;O2
D-Galactosamine;C;O2
D-Galactosamine;N;O2
Isocitrate;C;O2
Levoglucosan;C;O2
Nitrogen;N;O2
Tetrathionate;S;O2
Trithionate;S;O2
Urate;C;O2
Urate;N;O2
1
100
10000
Growth condition
Global clones
a Hypervirulent
a Multi-drug resistant
Frequency within sub-lineage
Absent
Rare
Intermediate
Core


A
OD600
Time
A
B
C
Log2 fold-change of OD600 in co- vs single-culture
ns
*
***
**
**
Hours
Galactitol: SL29 (auxo) + SL17 (proto)
Galactitol: SL29 (auxo) + SL37 (proto)
Galactitol: SL86 (auxo) + SL23 (proto)
INF122
INF250
INF354
INF171
INF313
INF354
INF191
INF225
INF331
L-Hydroxyproline: SL37 (auxo) + SL29 (proto)
L-sorbose: SL23 (auxo) + SL86 (proto)
L-tartrate: SL17 (auxo) + SL29 (proto)
INF077
INF120
INF265
AJ155
INF079
SA12
INF224
INF300
INF359
Bacterial isolates
Auxotroph
Prototroph
Carbon substrate
Prototroph-specific
Prototroph byproduct/s
Culture condition
Single culture (auxotroph)
Co-culture (auxotroph)