Kleborate Tutorial

Contact:
kat.holt@lshtm.ac.uk
mary.gathoni@lshtm.ac.uk
margaret.lam@monash.edu

Overview

In this tutorial we will explore Kleborate v3, a tool for genotyping loci of clinical relevance in *Klebsiella pneumoniae* and its close relatives in the *K. pneumoniae* Species Complex (KpSC).

We will explore Kleborate features:

- Species identification
- 7-locus MLST typing
- Virulence genotyping
- Antimicrobial resistance (AMR) genotyping
- K and O locus typing via Kaptive

We will demonstrate how to install and run Kleborate using the command line, and explore some example Kleborate outputs for genomes published as part of the BARNARDS study of neonatal sepsis (Sands *et al.*, 2021) and European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE) project (Grundmann et al., 2017)

Introduction

Klebsiella pneumoniae is a commensal bacterium that causes opportunistic infections in hospitals. It has six close relatives (species and subspecies), known as the *K. pneumoniae* species complex (KpSC). These related species are often difficult to distinguish from one another in clinical labs using biotyping or MALDI-TOF and consequently can be confused for *K. pneumoniae*.

K. pneumoniae are intrinsically resistant to ampicillin, and resistance to additional antimicrobials frequently arises through horizontal gene transfer and/or chromosomal mutations. Multi-drug resistance (MDR) is increasing globally and MDR strains with resistance to the carbapenems are of particular concern, earning *K. pneumoniae* a top position in the World Health Organization's priority list of drug-resistant pathogens for which novel control strategies are urgently required.

A handful of 'hypervirulent' *K. pneumoniae* clonal groups are also recognised, comprising strains that encode a constellation of acquired virulence factors and which can cause invasive disease outside the hospital setting. Fortunately, most of these hypervirulent strains have so far remained susceptible to the majority of antimicrobials. But evidence is now mounting that other *K. pneumoniae*, including MDR and carbapenem-resistant strains, can acquire the virulence factors – siderophores (yersiniabactin, salmochelin and aerobactin), regulators of hypermucoidy (the *rmpADC* locus and potentially also *rmpA2*) and/or the genotoxin colibactin – resulting in enhanced virulence potential. This so-called 'convergence' of MDR and acquired virulence factors further heightens the public health risk associated with *K. pneumoniae* because the resulting strains have the potential to cause severe infections that are extremely difficult to treat.

Capsule (K) and LPS (O) antigen variation in *K. pneumoniae* is of increasing interest to the research community, due to its importance in host-pathogen and phage interactions, and thus potential relevance to novel disease control measures such as vaccines, immunotherapy and phage therapy.

To learn more about taxonomy and population genomics of *Klebsiella pneumoniae* and the KpSC, and what we know so far about the distribution of AMR, virulence, K and O types in the *K. pneumoniae* population, see Wyres, Lam & Holt, 2020, Nature Reviews Microbiology.

The Kleborate Genotyping Framework

Kleborate was primarily developed to screen genome assemblies of *Klebsiella pneumoniae* and the *Klebsiella pneumoniae* species complex (KpSC) for:

- Species (e.g. *K. pneumoniae*, *K. quasipneumoniae*, *K. variicola*, etc.)
- *K. pneumoniae* species complex MLST
- ICEKp-associated virulence loci: yersiniabactin (*ybt*), colibactin (*clb*), salmochelin (*iro*), hypermucoidy (*rmp*)
- Virulence plasmid associated loci: salmochelin (*iro*), aerobactin (*iuc*), hypermucoidy (*rmp*, *rmpA2*)
- Antimicrobial resistance determinants: acquired genes, SNPs, gene truncations and intrinsic β-lactamases
- K (capsule) and O antigen (LPS) serotype prediction, via *wzi* alleles and Kaptive

Kleborate v3 includes a range of modules for typing bacterial genomes, most of which are specific to a particular species or complex (*Klebsiella pneumoniae SC*, *Klebsiella oxytoca SC*, *Escherichia coli*).

Kleborate v3 modules are divided into:

1. General Modules
2. Modules for *Klebsiella pneumoniae* species complex
3. Modules for *Klebsiella oxytoca* species complex
4. Modules for *Escherichia* species complex

For this tutorial, we will only go through modules relevant to *Klebsiella pneumoniae* species complex

General modules

Species detection

`-m enterobacterales__species`

This module will attempt to identify the species of each input assembly. It does this by comparing the assembly using Mash to a curated set of *Klebsiella* and other *Enterobacteriaceae* assemblies from NCBI and reporting the species of the closest match. Kleborate considers a Mash distance ≤0.02 to be a strong species match. A distance of >0.02 is a weak match and might indicate that your sample is a novel lineage or a hybrid between multiple *Klebsiella* species.

Outputs

The output of the species typing module is the following columns:

Species	Species name (scientific name)
species_match	Strength of the species call indicated as `strong` (Mash distance ≤ 0.02) or `weak` (Mash distance of > 0.02 and ≤ 0.04, may be novel or hybrid species)

Contig stats

`-m general__contig_stats`

The quality and completeness of Kleborate results depend on the quality of the input genome assemblies. In general, you can expect good results from draft genomes assembled with tools like SPAdes from high-depth (>50x) Illumina data, however, it is always possible that key genes subject to genotyping may be split across contigs, which can create problems for detecting and typing them accurately.

This module takes `enterobacterales__species` as a prerequisite and generates some basic assembly statistics to help users understand their typing results in the context of assembly quality, although we recommend users conduct more comprehensive QC themselves before typing genomes (e.g. screen for contamination, etc).

The module reports a standard set of assembly quality metrics (see Outputs below).

It will also flag in the `QC_warnings` column if an assembly size falls outside those specified in the `species_specification.txt` in the module directory, or if N50 <10 kbp or ambiguous bases (Ns) are detected in the sequence.

Outputs

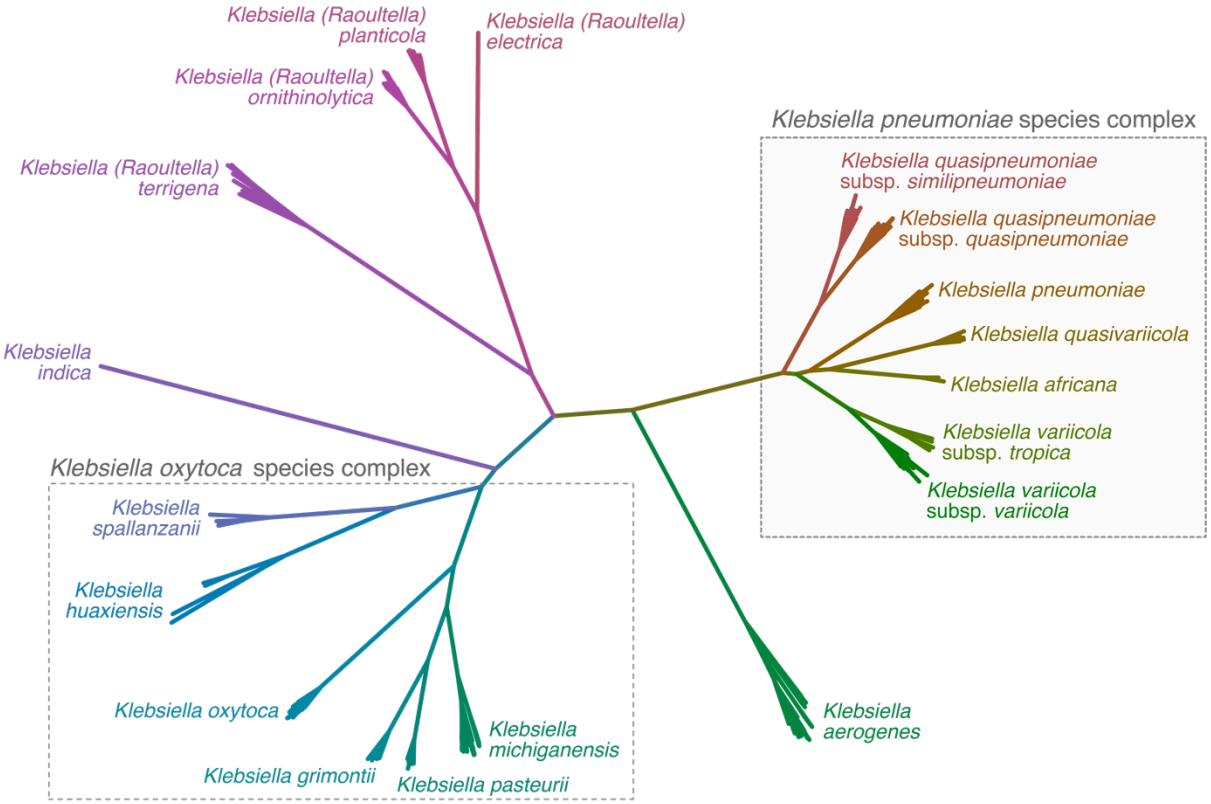
The output of the contig stats module is the following columns:

contig_count	Number of contigs in the input assembly
N50	N50 calculated from the contig sizes
largest_contig	Size of largest contig (in bp)
total_size	Total assembly size (in bp)
ambiguous_bases	Detection of ambiguous bases (yes or no). If yes, the number of ambiguous bases is also provided in brackets.
QC_warnings	List of QC issues detected, including `ambiguous_bases` (ambiguous bases detected) `N50` (N50 < 10 kbp), `total_size` (genome size falls outside expected range).

Modules for *Klebsiella pneumoniae* species complex

`--preset kpsc`

Modules for *K. pneumoniae* will be run if the `enterobacterales__species` module confirms the input assembly as a member of the *K. pneumoniae* species complex (KpSC) labelled in the tree below.



***K. pneumoniae* species complex (KpSC):** Kleborate is designed for detailed genotyping of the well-studied *K. pneumoniae* species complex (KpSC) labelled on the tree, which includes the seven species listed in the table below. These were previously considered as phylogroups within *K. pneumoniae*. We've included the phylogroup numbers in the table below for backwards compatibility with older literature, but these names are not used in the Kleborate output. See this review for an overview of the species complex

Species	Kp phylogroup[a]	Kp phylogroup (alternative)[b]	Reference
K. pneumoniae	Kp1	KpI	Brenner, D.J. 1979 Int J Syst Evol Microbiol 29: 38-41
K. quasipneumoniae subsp *quasipneumoniae*	Kp2	KpIIa	Brisse et al., 2014 Int J Syst Evol Microbiol 64:3146-52
K. quasipneumoniae subsp *similipneumoniae*	Kp4	KpIIb	Brisse et al. 2014 Int J Syst Evol Microbiol 64:3146-52
K. variicola subsp *variicola*	Kp3	KpIII	Rosenblueth et al. 2004 Syst Appl Microbiol 27:27-35
K. variicola subsp *tropica*	Kp5	-	Rodrigues et al., 2019 Res Microbiol S0923-2508:30019-1 (described as subsp *tropicalensis* in paper)
K. quasivariicola	Kp6	-	Long et al. 2017 Genome Announc 5: e01057-17
K. africana	Kp7	-	Rodrigues et al. 2019 Res Microbiol S0923-2508:30019-1 (described as *africanensis* in this paper)

[a] Kp phylogroup numbers as described in Rodrigues et al. 2019
[b] alternative (older) Kp phylogroup numbers as described in Brisse et al. 2001 and Fevre et al. 2005 prior to the identification of *K. variicola* subsp *tropica*, *K. quasivariicola* and *K. africana*.

KpSC MLST

`-m klebsiella_pneumo_complex__mlst`

Genomes identified by Kleborate as belonging to the *K. pneumoniae* species complex are subjected to MLST using the 7-locus scheme described at the *K. pneumoniae* BIGSdb hosted at the Pasteur Institute. Note that this scheme is not specific to *K. pneumoniae* but covers the whole *K. pneumoniae* species complex.

NB: A copy of the MLST alleles and ST definitions is stored in the /data directory of this module.

Rhinoscleromatis and Ozaenae

The *K. pneumoniae* clonal group CG67 is known as *K. pneumoniae* subsp. *rhinoscleromatis* because it causes rhinoscleroma (chronic granulomatous infection of the nose and upper airways), and clonal group CG91 is known as *K. pneumoniae* subsp. *ozaenae* as it can cause ozena (atrophic rhinitis). To alert users to this, when STs belonging to these clonal groups are detected by Kleborate this is flagged in the ST column, e.g. 'ST67 (subsp. rhinoscleromatis)' or 'ST97 (subsp. ozaenae)'.

The relevant STs are:

Species column	**ST**	**MLST column**
K. pneumoniae	67, 68, 69, 3772, 3819	ST67 (subsp. rhinoscleromatis)
K. pneumoniae	90, 91, 92, 93, 95, 96, 97, 381, 777, 3193 3766, 3768, 3771, 3781, 3782, 3784, 3802, 3803	ST91 (subsp. ozaenae)

Outputs

The output of the KpSC MLST module is the following columns:

ST	sequence type
gapA, infB, mdh, pgi, phoE, rpoB, tonB	allele number

- Kleborate reports the closest matching ST if a precise match is not found.
- Imprecise allele matches are indicated with a *.
- Imprecise ST calls are indicated with -nLV, where n indicates the number of loci that disagree with the ST reported. So, 258-1LV indicates a single-locus variant (SLV) of ST258, i.e. 6/7 loci match ST258.

KpSC virulence modules

Typing modules are available for the five key acquired virulence loci that are associated with invasive infections and are found at high prevalence among hypervirulent *K. pneumoniae* strains: the siderophores yersiniabactin (*ybt*), aerobactin (*iuc*) and salmochelin (*iro*), the genotoxin colibactin (*clb*), and the hypermucoidy locus *rmpADC*. Each of these loci comprises multiple genes and will only be reported if >50% of the genes are detected.

There is also a module to screen for the alternative hypermucoidy marker gene *rmpA2*.

For each module, if the target locus is detected, the typer will:

- Call a sequence type using the same logic as for 7-gene MLST
- Report the phylogenetic lineage associated with each sequence type, as outlined below and detailed in the corresponding papers
- Report the structural variant of the mobile genetic element that is usually associated with that phylogenetic lineage (for *ybt* and *rmpADC* only)

The *ybt*, *clb*, *iuc*, *iro* and *rmpADC* locus-specific ST schemes, and *rmpA2* alleles, are defined in the *K. pneumoniae* Bacterial Isolate Genome Sequence Database.

Virulence alleles are treated in the same way as [MLST] alleles:

- To consider a Minimap2 hit, it must exceed both 80% identity and 40% coverage (adjustable via the –min_spurious_identity and –min_spurious_coverage options).

- Hits that fail to meet 90% identity and 80% coverage (adjustable via the --min_identity and --min_coverage options) are reported in the spurious_virulence_hits column but not used for sequence typing.

- Imperfect hits (either <100% identity or <100% coverage) are reported with a *. E.g. 15* means that no perfect match was found but the closest match is allele 15.

- Kleborate will next translate the hit into amino acid sequence and look for truncations (expressed as % amino acid length from the start codon). If the result is less than 90%, it is added to the result (e.g. 15*-42%).

- Virulence locus STs are only reported if >50% of the genes in a locus are detected (e.g. at least 6 of the 11 *ybt* locus genes are required to report a *ybt* ST).
- If <50% of the genes in a locus are detected, Kleborate reports the ST as 0 and the lineage as –.

- If <100% but >50% of the genes in a locus are detected, Kleborate will report the locus as (incomplete), along with the closest matching ST and its corresponding phylogenetic lineage. E.g. if only 7 of the 11 *ybt* genes are detected, this will be reported as ybtX; ICEKpX (incomplete).

- For genomes with multiple copies of a virulence locus (e.g. a strain that carries ICE *Kp1* and the KpVP-1 plasmid will have two copies of *iro* and *rmp*), Kleborate will report

and assign a ST or closest matching ST to each of these virulence loci provided that the locus is relatively intact in the genome (i.e. >50% of the genes in a locus are present on a single contig) and according to the above criteria.

Yersiniabactin and colibactin

`-m klebsiella__ybst, klebsiella__cbst`

We previously explored the diversity of the *K. pneumoniae* integrative conjugative element (ICE *Kp*), which mobilises the yersiniabactin locus *ybt*, using genomic analysis of a diverse set of 2498 *Klebsiella* (see this article). Overall, we found *ybt* in about a third of all *K. pneumoniae* genomes (and *clb* in about 14%). We identified 17 distinct lineages of *ybt* (see figure) embedded within 14 structural variants of ICE *Kp* that can integrate at any of four tRNA-Asn sites in the chromosome. One type was found to be plasmid-borne. Based on this analysis, we developed a MLST-style approach for assigning yersiniabactin sequence types (YbST) and colibactin sequence types (CbST), which is implemented in Kleborate.

Note that while ICE *Kp1* is occasionally found in other species within the KpSC, and even in other genera of Enterobacteriaceae (see original paper), most of the known variation included in the database is derived from *K. pneumoniae*.

Yersiniabactin outputs

The output of the ybst module is the following columns:

Yersiniabactin	Lineage (ICEKp prediction)
YbST	Yersiniabactin sequence type
ybtS, ybtX, ybtQ, ybtP, ybtA, irp2, irp1, ybtU, ybtT, ybtE, fyuA	allele number (ybt locus)

Colibactin outputs

The output of the cbst module is the following columns:

Colibactin	Lineage
CbST	Colibactin sequence type
clbA, clbB, clbC, clbD, clbE, clbF, clbG, clbH, clbI, clbL, clbM, clbN, clbO, clbP, clbQ	allele number (clb / pks locus)

Aerobactin and salmochelin

`-m klebsiella__abst, klebsiella__smst`

We further explored the genetic diversity of the aerobactin (*iuc*) and salmochelin (*iro*) loci among a dataset of 2733 *Klebsiella* genomes (see this publication). We identified five *iro* and six *iuc* lineages, each of which was associated with a specific location within *K. pneumoniae* genomes (primarily virulence plasmids). Based on this analysis, we developed a MLST-style approach for assigning aerobactin sequence types (AbST) and salmochelin sequence types (SmST) which is implemented in Kleborate.

- The most common lineages are *iuc1* and *iro1*, which are found together on the FIBk virulence plasmid KpVP-1 (typified by pK2044 or pLVPK common to the hypervirulent clones ST23, ST86, etc).
- *iuc2* and *iro2* lineages were associated with the alternative FIBk virulence plasmid KpVP-2 (typified by Kp52.145 plasmid II from the K2 ST66 lab strain known as Kp52.145 or CIP 52.145 or B5055).
- *iuc5* and *iro5* originate from *E. coli* and are carried (often together) on *E. coli* FII plasmids that can transfer to *K. pneumoniae*.
- The lineages *iuc2A*, *iuc3* and *iro4* were associated with other novel FIBk plasmids that had not been previously described in *K. pneumoniae*, but sequences for which are included in the paper.
- The salmochelin locus present in ICE *Kp1* constitutes its own lineage *iro3*, and the aerobactin locus present in the chromosome of ST67 *K. pneumoniae* subsp *rhinoscleromatis* strains constitutes its own lineage *iuc4*.

Note on *iucA* sequence update:

In Kleborate version 2.2.0 and earlier, the majority of *iucA* alleles had a sequence length of 1791 bp, with the exception being those associated with lineage *iuc 5* which have a length of 1725 bp. Related to this, *iucA* in genomes with *iuc 3* encoded a premature stop codon resulting in a significantly truncated and presumably non-functional IucA protein (i.e. at 2% length of the intact amino acid sequence), despite experimental evidence showing siderophore activity in *iuc 3*+ isolates. Considering this evidence, the sequences of *iucA* genes with the longer ~1791 bp length were updated to ~1725 bp by removing the first 66 bp. These changes are captured in Kleborate version 2.3.0 onwards and address the truncation issue in *iuc 3*+ genomes. The following *iucA* alleles and AbST profiles have also been retired due to sequence redundancy following the update:

- alleles: iucA48, iucA49, iucA52
- profiles: AbST 70, 82, 83

Aerobactin outputs

The output of the abst module is the following columns:

Aerobactin	Lineage (plasmid prediction)
AbST	Sequence type
iucA, iucB, iucC, iucD, iutA	allele number (iuc locus)

Salmochelin outputs

The output of the smst module is the following columns:

Salmochelin	Lineage (plasmid prediction)
SmST	Sequence type
iroB, iroC, iroD, iroN	allele number (iro locus)

Hypermucoidy loci

`-m klebsiella__rmst, klebsiella__rmpa2`

The *rmpA* locus is associated with the hypermucoidy phenotype that is a virulence feature that is often observed in hypervirulent *K. pneumoniae* strains. Recent work has revealed that *rmpA* serves as a transcriptional regulator for the *rmpD* and *rmpC* genes, and together these genes comprise the *rmpADC* (or *rmp*) locus. *rmpC* is involved in the upregulation of capsule expression while *rmpD* drives hypermucoviscosity (see the paper on rmpC and this one on rmpD for more information.)

In light of this information, we screened and extracted the *rmpA*, *rmpD* and *rmpC* sequences from the 2733 genomes included in the aerobactin and salmochelin study and generated a RmST typing scheme. We observed four distinct *rmp* lineages, which were associated with the KpVP-1 (*rmp 1*), KpVP-2 (*rmp 2*), *iuc2A* virulence plasmids (*rmp 2A*), ICE *Kp1* (rmp 3) and the *rmp4* lineage which is associated with *K. pneumoniae* CG67 Lam et al., 2024 BioRxiv

The `klebsiella__rmst` module screens for *rmpADC* and will report a sequence type, along with the associated lineage and mobile genetic element.

The *rmpA2* gene is homologous to *rmpA*, and the `klebsiella__rmpa2` module screens for alleles of *rmpA2*.

Note:

- Alleles for each gene are sourced from the BIGSdb-pasteur, while additional *rmpA* alleles have also been added to Kleborate.
- The *rmpA* and *rmpA2* genes share ~83% nucleotide identity so is easily distinguished.
- Unique (non-overlapping) nucleotide Minimap2 hits with >95% identity and >50% coverage are reported. Note multiple hits to the same gene are reported if found. E.g. the NTUH-K2044 genome carries *rmpA* in the virulence plasmid and also in ICE *Kp1*, which is reported in the *rmpA* column as rmpA_11(ICEKp1),rmpA_2(KpVP-1).
- As with the other virulence genes, truncations in the *rmpA* and *rmpA2* genes are expressed as a percentage of the amino acid length from the start codon, e.g. rmpA_5-54% indicates the RmpA protein is truncated after 54% length of the intact amino acid sequence. These truncations appear to be common, due to insertions and deletions within a poly-G tract, and almost certainly result in loss of protein function.

Rmp outputs

The output of the rmst module is the following columns:

RmpADC	Lineage
RmST	Sequence type
rmpA, rmpD, rmpC	allele number (rmp locus)

rmpA2 outputs

The output of the rmst module is the following columns:

rmpA2	best matching allele

Virulence score

`-m klebsiella_pneumo_complex__virulence_score`

This module takes klebsiella__abst, klebsiella__cbst, klebsiella__ybst as prerequisites and calculates a virulence score, which ranges from 0 to 5 as outlined below. Note neither the salmochelin (iro) locus nor rmpADC are explicitly considered in the virulence score, for simplicity. The iro and rmpADC loci typically appear alongside the aerobactin (iuc) locus on the Kp virulence plasmids, and so presence of iuc (score of 3-5) generally implies presence of iro and rmpADC. However we prioritise iuc in the calculation of the score, as aerobactin is specifically associated with growth in blood and is a stronger predictor of the hypervirulence phenotype see this review. The iro and rmpADC loci are also occasionally present with ybt, in the ICEKp variant - ICEKp1, but this will still score 1.

0	negative for all of yersiniabactin (ybt), colibactin (clb), aerobactin (iuc)
1	yersiniabactin only
2	yersiniabactin and colibactin (or colibactin only)
3	aerobactin (without yersiniabactin or colibactin)
4	aerobactin with yersiniabactin (without colibactin)
5	yersiniabactin, colibactin and aerobactin

Virulence score outputs

Virulence score is output in the following column:

virulence_score	Score of 0-5, as defined above

Antimicrobial Resistance (KpSC AMR)

`-m klebsiella_pneumo_complex__amr`

Acquired AMR genes

This module screens input genomes against a curated version of the CARD database of acquired resistance gene alleles (see the following spreadsheet for details on curation), and groups these by drug class for reporting purposes. The chromosomal *fosA* and *oqxAB* genes that are intrinsic to all KpSC are not reported and usually do not confer fosfomycin and fluoroquinolone resistance in these species.

Kleborate has logic to choose the best allele hit, annotate that hit with extra information and place it in an appropriate column in the output.

In brief:

- Exact nucleotide matches are preferred, followed by exact amino acid matches, followed by inexact nucleotide matches.
- Annotations indicate aspects of the hit: **`^`** (inexact nucleotide but exact amino acid match), **`*`** (inexact nucleotide and inexact amino acid match), **`?`** (incomplete match), **`-X%`** (truncated amino acid sequence), **`$`** (mutated start codon, translation may be disrupted).
- The column indicates the confidence of the hit: strong hits go in the column for their drug class, truncated hits go in the **`truncated_resistance_hits`** column and low identity/coverage hits go in the **`spurious_resistance_hits`** column.

And here is the logic in more detail:

- In order to consider a Minimap hit, it must exceed both 80% identity and 40% coverage (adjustable via the --min_spurious_identity and --min_spurious_coverage options).
- If the hit is 100% identity and 100% coverage, then it will be reported with no further annotation (e.g. **`TEM-15`**).
- If no exact nucleotide match is found, Kleborate searches for an exact amino acid match, and will report this with a **`^`** symbol. E.g. **`TEM-15^`** indicates an exact match to the **`TEM-15`** protein sequence but with one or more nucleotide differences.
- If no exact amino acid match is found, the closest nucleotide match is reported with a * symbol. E.g. **`TEM-15*`** indicates no precise nucleotide or amino acid match is found, but the closest nucleotide match is to **`TEM-15.`**
- If the hit is less than 100% coverage, **`a ?`** is added to the result E.g. **`TEM-15?`** indicates an incomplete match at 100% identity, and TEM-15*? indicates an incomplete match at <100% identity.

- Kleborate will next translate the hit into amino acid sequence and look for truncations (expressed as % amino acid length from the start codon). If the result is less than 90%, it is added to the result (e.g. **`TEM-15*-42%)`** and the hit is reported in the **`truncated_resistance_hits`** column.

- If the hit is less than 90% identity or 80% nucleotide coverage (adjustable via the **`--min_identity`** and **`--min_coverage`** options), it is reported in the spurious_resistance_hits column. Otherwise, it is reported in the column for its drug class (e.g. Bla_ESBL_acquired).

Note that Kleborate reports resistance results for all antimicrobial classes with confidently attributable resistance mechanisms in KpSC. Not all of these are actually used clinically for treatment of KpSC infections (e.g. MLS, Rif) but they are still reported as the presence of acquired resistance determinants to these classes is of interest to researchers for other reasons (e.g. these genes can be useful markers of MGEs and MGE spread; there is potential for use of these drugs against other organisms to select for KpSC in co-infected patients or in the environment). For an overview of antimicrobial resistance and consensus definitions of multidrug resistance (MDR), extensive drug resistance (XDR) and pan drug resistance in Enterobacteriaceae, see Magiorakos, 2012

SHV beta-lactamases

All KpSC carry a core chromosomal beta-lactamase gene (SHV in *K. pneumoniae*, LEN in *K. variicola*, OKP in *K. quasipneumoniae*) that confers clinically significant resistance to ampicillin. Some KpSC also carry acquired mobile SHV alleles, which can confer additional inhibitor resistance and/or resistance to extended spectrum beta-lactams.

Kleborate will report all of the SHV alleles it detects and separate them into columns based on the resistance phenotype they are predicted to encode:

- SHV alleles associated with ampicillin resistance only, will be reported in the **`Bla_chr`** column because they are assumed to represent the chromosomal allele. These genes are not included in the count of acquired resistance genes or drug classes.

- Other SHV alleles e.g. those predicted to encode ESBLs (extended-spectrum beta-lactamases) or beta-lactamases with inhibitor resistance will be reported in the relevant **`Bla_ESBL_acquired`** or **`Bla_inhR_acquired`** columns etc (see below), because these SHV alleles are almost always carried on plasmids. (However, it is possible to have a mutation in a chromosomal SHV gene that gives a match to an ESBL allele, which would also be reported in the **`Bla_ESBL_acquired`** column and counted as an acquired gene because it is very hard to tell the difference without manual exploration of the genetic context.)

The specific mutations, and assignment of alleles to class, is detailed in this preprint from KlebNET-GSP: Tsang et al, 2024 BioRxiv.

Additional chromosomal mutations associated with AMR

- Fluoroquinolone resistance mutations: GyrA 83 & 87 and ParC 80 & 84. These appear in the **`Flq_mutations`** column.

- Colistin resistance due to truncation or loss of core genes MgrB or PmrB. If these genes are missing or truncated, this information will be reported in the 'Col_mutations' column (truncations are expressed as % amino acid length from the start codon, if there is a mutation in the start codon this is indicated as **`$`** to flag that the gene is present but may not be translated correctly). Note if MgrB and PmrB are present and not truncated then nothing about them will be reported in the 'Col' column.

- OmpK35 and OmpK36 truncations and point mutations shown to result in reduced susceptibility to beta-lactamases (insertions GD or TD in the third loop or synonymous C > T at nucleotide 25 **`ompK36_c25t`**). This information will be reported in the **`Omp_mutations`** column (truncations are expressed as % amino acid length from the start codon). Note that if a gene is fragmented across multiple contigs, Kleborate will attempt to predict the closest matching allele based on the longest fragment. If this longest fragment does not contain the start of the gene, the truncation will be reported as -0%. Additionally, if these core genes are present and not truncated then nothing about them will be reported in the 'Omp' column. The specific effect of OmpK mutations on drug susceptibility depends on multiple factors including what combinations of OmpK35 and OmpK36 alleles are present and what beta-lactamase genes are present (this is why we report them in their own column separate to Bla genes). See e.g. paper and this one for more information on OmpK genes and drug resistance.

Note these do not count towards acquired resistance gene counts but do count towards drug classes (with the exception of Omp mutations, whose spectrum of effects depends on the presence of acquired beta-lactamases and thus their impact on specific beta-lactam drug classes is hard to predict).

AMR outputs

Results of the KpSC AMR module are grouped by drug class (according to the ARG-Annot DB), with beta-lactamases further broken down into Lahey classes (now maintained at BLDB), as follows:

AGly_acquired	aminoglycoside resistance genes
Col_acquired	colistin resistance genes
Fcyn_acquired	fosfomycin resistance genes
Flq_acquired	fluoroquinolone resistance genes

Gly_acquired	glycopeptide resistance genes
MLS_acquired	macrolide resistance genes
Phe_acquired	phenicol resistance genes
Rif_acquired	rifampin resistance genes
Sul_acquired	sulfonamide resistance genes
Tet_acquired	tetracycline resistance genes
Tgc_acquired	tigecycline resistance genes
Tmt_acquired	trimethoprim resistance genes
Bla_acquired	beta-lactamases (other than SHV) that have no known extended-spectrum, carbapenemase, or inhibitor-resistance activity
Bla_ESBL_acquired	extended-spectrum beta-lactamases, including SHV alleles with known ESBL activity
Bla_ESBL_inhR_acquired	extended spectrum beta-lactamases with resistance to beta-lactamase inhibitors, including SHV alleles associated with these traits
Bla_Carb_acquired	carbapenemases
Bla_chr	SHV alleles associated with ampicillin resistance only (assumed core chromosomal genes)
SHV_mutations	mutations in the SHV beta-lactamase known to be associated with expansion of enzyme activity
Omp_mutations	resistance-related mutations in the OmpK35 and OmpK36 osmoporins
Col_mutations	reports if MgrB or PmrB genes are not intact
Flq_mutations	reports mutations found in the quinolone-resistance determining regions of GyrA and ParC
truncated_resistance_hits	list of acquired resistance genes in which the encoded protein is predicted to be truncated (e.g. due to a stop codon or frameshift mutation within the open reading frame)
spurious_resistance_hits	list of acquired resistance genes detected below the identity or coverage thresholds (default <90% identity or <80% nucleotide coverage)

Resistance scores and counts

Running the KpSC AMR module automatically runs additional modules for generating counts of resistance genes and drug classes, and calculating a resistance score. These modules take **`klebsiella_pneumo_complex__amr`** as a prerequisite and can be specified manually as follows:

```
-m klebsiella_pneumo_complex__resistance_score,
klebsiella_pneumo_complex__resistance_gene_count,
klebsiella_pneumo_complex__resistance_class_count
```

This module calculates a resistance score, which ranges from 0 to 3 as follows

0	no ESBL, no carbapenemase (regardless of colistin resistance)
1	ESBL, no carbapenemase (regardless of colistin resistance)
2	Carbapenemase without colistin resistance (regardless of ESBL genes or OmpK mutations)
3	Carbapenemase with colistin resistance (regardless of ESBL genes or OmpK mutations)

This module quantifies how many acquired resistance genes are present and how many drug classes (in *addition* to ampicillin to which KpSC are intrinsically resistant) have at least one resistance determinant detected (i.e. ignoring genes recorded in the Bla_chr and Bla_acquired columns).

A few things to note:

- The presence of resistance *mutations*, and non-ESBL forms of core genes SHV/LEN/OKP, do not contribute to the resistance *gene* count.
- Mutations do contribute to the drug class count, e.g. fluoroquinolone resistance will be counted if a GyrA mutation is encountered regardless of whether or not an acquired quinolone resistance (*qnr*) gene is also present. The exceptions are Omp mutations, which do not contribute to the drug class count as their effect depends on the strain background and the presence of acquired beta-lactamase enzymes; hence this information is provided in a separate column, and interpretation is left to the user (see the Antimicrobial Resistance page).
- Genes reported in the truncated_resistance_genes and spurious_resistance_genes columns do not contribute to the counts.
- Note that since a drug class can have multiple resistance determinants, the gene count is typically higher than the class count.
- most ESBL+ *K. pneumoniae* also carry multiple other resistance genes, associated with multiple drug classes. Therefore, genomes with scores >0 are also typically multi-drug resistant. (See fig below, showing distribution of AMR classes and genes

amongst a non-redundant set of 9705 public genomes, reproduced from the Kleborate paper).

Resistance scores and counts outputs

Resistance scores and counts are output in the following columns:

resistance_score	Score of 0-3, as defined above
num_resistance_genes	Number of acquired resistance genes
num_resistance_classes	Number of drug classes to which resistance determinants have been acquired (in addition to intrinsic ampicillin)

KpSC K and O locus typing with Kaptive

`-m klebsiella_pneumo_complex__kaptive`

The two key surface antigens produced by *K. pneumoniae* are the K antigen (capsular polysaccharide) and O antigen (lipopolysaccharide). Serological typing is not widely available, but we can predict K and O antigens based on identification and typing of their biosynthesis loci.

Nomenclature: Serologically defined K types are named as K1, K2, K3, etc. Each K type is associated with a unique K locus (KL), with a unique set of sugar processing genes that produce a unique capsular polysaccharide structure. These K loci are numbered according to the K type they produce - i.e. KL1 produces K1 antigen, KL2 produces K2 antigen, etc. There are currently 77 serologically defined capsule types but >160 distinct K loci have been defined on the basis of distinct gene content. K locus numbers greater than 100 correspond to loci that differ from those in the serotype reference strains, i.e. they are presumed to encode novel serotypes. A similar nomenclature system is followed for O antigens and O types, except that additional genes outside the O locus can modify the antigen produced. See this paper for details.

This module will run the Kaptive v3 tool to identify capsule (K) and O antigen loci. See the Kaptive documentation for more details of how Kaptive works, tutorials, and citations.

`-t, --threads`

Number of threads for alignment (default: 1)

`--k-db, kpsc_k`

Kaptive database for K-locus typing

`--o-db, kpsc_o`

Kaptive database for o-locus typing

Kaptive output

Column Name	Description
Best match locus	The locus type which most closely matches the assembly.
Best match type	The predicted serotype/phenotype of the assembly.
Match confidence	Typeable or Untypeable

Problems	Characters indicating issues with the locus match.
Identity	Weighted percent identity of the best matching locus to the assembly.
Coverage	Weighted percent coverage of the best matching locus in the assembly.
Length discrepancy	If the locus was found in a single piece, this is the difference between the locus length and the assembly length.
Expected genes in locus	A fraction indicating how many of the genes in the best matching locus were found in the locus part of the assembly.
Expected genes in locus, details	Gene names for the expected genes found in the locus part of the assembly.
Missing expected genes	A string listing the gene names of expected genes that were not found.

KpSC Wzi typing for K antigen prediction

`-m klebsiella_pneumo_complex__wzi`

This module reports the closest match amongst the *wzi* alleles in the BIGSdb. This is a marker of capsule locus (KL) type, which is predictive of capsule (K) serotype. Although there is not a 1-1 relationship between *wzi* allele and KL/K type, there is a strong correlation (see Wyres et al, MGen 2016 and Brisse et al, J Clin Micro 2013). Note the *wzi database* is populated with alleles from the *Klebsiella pneumoniae* species complex and is not reliable for other species.

The *wzi* allele can provide a handy way of spotting the hypervirulence-associated capsule types (wzi=K1, wzi2=K2, wzi5=K5); or spotting capsule switching within clones, e.g. you can tell which ST258 lineage you have from the _wzi_ type (wzi154: the main lineage II; wzi29: recombinant lineage I; others: probably other recombinant lineages). But the K locus predictions from the Kaptive module are more specific and reliable.

Wzi typing results are output in the following columns:

wzi	wzi allele
K_locus	K locus typically associated with this wzi allele

How to install Kleborate 💿

We will demonstrate how to install and run Kleborate using a command line computing environment. You can also consult Kleborate's online documentation.

Kleborate can be installed for use on your personal computer or cluster. Kleborate rely on the use of third-party dependencies including.

- Python v3.9 or later
- Biopython v1.75 or later
- Mash v2.0 or later
- Minimap2
- Kaptive
- DNA Features Viewer

To reduce the chances of dependency-related errors, we highly recommend that these tools be used within a virtual environment with conda 🐍 .

Set up a Conda environment

We will set up a Conda environment **`klebsiella_analysis`** with the `-n/-name` flag, specify the python version as `3.9` and list our dependencies to install.

```
conda create -n klebsiella_analysis -c bioconda python=3.9
minimap2 mash -y
```

Once your environment has installed successfully, we need to 'activate' it so all the binaries are visible to our `$PATH`.

```
conda activate klebsiella_analysis
```

Install Kleborate from PyPI:

```
pip install kleborate
```

How to use Kleborate 🖥️

Command line

To use Kleborate on the command line, use the `kleborate` command.

You can see the full set of usage options by typing

```
$ kleborate --help
```

```
usage: kleborate [-a ASSEMBLIES [ASSEMBLIES ...]] [-o OUTDIR] [-r] [--trim_headers] [--list_modules] [-p PRESET] [-m MODULES] [-h]
                 [--help_all] [--version]

Kleborate: a tool for characterising virulence and resistance in pathogen assemblies

           :
  -a ASSEMBLIES [ASSEMBLIES ...], --assemblies ASSEMBLIES [ASSEMBLIES ...]
                                        FASTA file(s) for assemblies
  -o OUTDIR, --outdir OUTDIR            Directory for storing output files
  -r, --resume                          append the output files (default: False)
  --trim_headers                        Trim headers in the output files (default: False)

       :
  --list_modules                        Print a list of all available modules and then quit (default: False)
  -p PRESET, --preset PRESET            Module presets, choose from: kpsc, kosc, escherichia
  -m MODULES, --modules MODULES         Comma-delimited list of Kleborate modules to use

    :
  -h, --help                            Show this help message and exit
  --help_all                            Show a help message with all module options
  --version                             Show program's version number and exit

If you use Kleborate, please cite the paper:
Lam MMC, et al. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nature
Communications. 2021. doi:10.1038/s41467-021-24448-3.

If you turn on the Kaptive option for full K and O typing, please also cite:
Wyres KL, et al. Identification of Klebsiella capsule synthesis loci from whole genome data. Microbial Genomics. 2016.
doi:10.1099/mgen.0.000102.
```

To view a list of available modules for Kleborate

```
$ kleborate --list_modules
```

For more information, see Kleborate's online documentation

PRACTICAL

Download example data from BARNARDS and EuSCAPE projects

The assemblies used in this workshop, along with the expected output files, can be accessed on Google drive . The data can also be found here (total file size: 8.2 MB).

You can download the folder to your computer using:

Objectives.

1) ***To familiarize yourself with the Kleborate command line options***
2) **Run Kleborate to analyse *Klebsiella pneumoniae* genomic assemblies**

Run Kleborate on example data

```
cd kleborate_workshop_data

kleborate -a *.fasta.gz -o kleborate_results -p kpsc --trim_headers
```

Here we are using:

- **`-a *.fasta.gz`:** Specifies the input files (assemblies) to be analysed (.fasta or.fasta.gz).
- **`-o`:** Specifies the directory where the output files will be saved (one output file per species/complex detected).
- **`-p`:** Specifies the preset modules to run (kpsc, kosc, escherichia).
- **`--trim_headers`:** Trim module names from column headers in the output.

Once Kleborate completes the analysis, it will generate output files in the specified directory **`kleborate_results`.** Each file is named according to the species/complex detected, such as klebsiella_pneumo_complex_output.txt.

Examples - interpreting results

A typical output

A Kleborate run with the `-p kpsc` flag will yield a tab delimited `.txt` output file containing 113 columns of results data.

You can view a copy of the example Kleborate results file in this Google drive.

The first column, labelled ‘strain’, contains the input genome name/s (taken from the input filenames, before the extension. fasta).

Here is the first few columns of the Kleborate output file from our example above, viewed in Microsoft Excel:

strain	species	species_match	contig_count	N50	largest_contig	total_size	ambiguous_bases	QC_warnings	ST
ERR4920436	Klebsiella pneumoniae	strong	109	281841	406372	5501897	no	-	ST218
ERR4920450	Klebsiella pneumoniae	strong	110	192649	419803	5440718	no	-	ST218
ERR4920551	Klebsiella pneumoniae	strong	306	48613	146404	5869852	no	-	ST15

We will now step through the columns in the Kleborate output file, using genome **ERR4920436** as an example.

Output columns: Species identification

strain	species	species_match
ERR4920436	Klebsiella pneumoniae	strong

Species assignment is based on mash comparison to the reference tree, see the documentation.
The match here is strong; weak matches are usually indicative of assembly contamination (mixed sample) or potentially presence of a species hybrid. The genome size and QC metrics can help give an indication if the problem is a mixed sample.

Output columns: Assembly quality

contig_count	N50	largest_contig	total_size	ambiguous_bases	QC_warnings
109	281841	406372	5501897	no	-

These columns should be quite self-explanatory. Ambiguous bases refer to the presence of non-A/G/C/T bases (e.g. ‘N’) in the input assembly. If any are found, the number will be printed here.

QC Warnings will be given here if:

a. ambiguous bases are detected
b. assembly length is <4.5 or >7.5 Mbp
c. N50 is below 10,000 bp

Output columns: genotyping summary

ST	virulence_score	resistance_score	num_resistance_classes	num_resistance_genes
ST218	4	1	7	14

The ST column reports the 7-locus MLST call, which here indicates exact match to ST218. Individual allele calls for the MLST loci are also reported:

ST	gapA	infB	mdh	pgi	phoE	rpoB	tonB
ST218	2	3	1	1	9	4	12

The virulence score of 4 indicates a high level of virulence: specifically, that aerobactin and Yersiniabactin were detected, but not colibactin. See explanation of scores

The resistance score of 1 indicates the presence of an ESBL gene but no carbapenemase. The number of acquired resistance genes, and the number of drug classes they are associated with, are also reported. See Kleborate documentation for details.

Output columns: Virulence detection/genotyping

YbST	Yersiniabactin
581-2LV	ybt 9; ICEKp3

The two columns indicate that the closest matching Yersiniabactin sequence type (YbST) is YbST581, but with allele mismatches for 2 of the *ybt* locus genes ('-2LV'). YbST581 belongs to lineage *ybt* 9, which is typically mobilised by ICE*Kp3* (note Kleborate is not specifically searching for the full-length ICEKp3, it is inferred from the ST, see documentation.

CbST	Colibactin
0	-

No colibactin locus is detected.

AbST	Aerobactin	iucA	iucB	iucC	iucD	iutA
1-1LV	iuc 1 (truncated)	1	1	1*-61%	1	1

Aerobactin is detected, with closest matching AbST being AbST 1, with a single mismatching allele ('-1LV'). Scrolling across to the individual *iuc* locus columns at the end of the file, we can see this is because the *iucC* nucleotide sequence is truncated. AbST1 corresponds to aerobactin lineage *iuc* 1, but the locus is reported as truncated *iucC* gene.

SmST	Salmochelin
1	iro 1

Salmochelin is detected, with an exact match to SmST1, which corresponds to lineage *iro* 1.

The *rmp* locus is detected, with closest match to RmST26, corresponding to lineage *rmp* 1 which is usually located on the virulence plasmid KpVP-1. But, there is a single allele mismatch, and the locus is truncated. Scrolling across to the individual *rmp* locus columns, we can see this is because the *rmpA* protein sequence is truncated and so not expected to be functional.

RmST	RmpADC	rmpA	rmpD	rmpC
26-1LV	rmp 1; KpVP-1 (truncated)	2*-0%	2	2

The closest allelic match for *rmpA2* is rmpA2_8, but the protein sequence is truncated and so not expected to be functional.

rmpA2
rmpA2_8-60%

The exact matches are required to call an ST:

- Yersiniabactin; 6 genes
- Colibactin; 8 genes
- Aerobactin; 3 genes
- Salmochelin; 2 genes
- RmpADC; 2 genes

Output columns: AMR detection/genotyping

AGly_acquired	Phe_acquired	Rif_acquired	Tet_acquired	Tmt_acquired
aac(3)-IIa.v1^;aac(6')-Ib-cr.v2;aadA*;strA.v1;strB.v1	catII.2*;cmlA5	arr-2	tet(A).v1	dfrA14.v2*

Acquired genes are reported in columns labelled by the drug class. Here we can see we have acquired genes associated with resistance to aminoglycosides (AGly), phenicols (Phe), rifampicin (Rif), tetracycline (Tet) and trimethoprim (Tmt).

^ indicates an exact match at the amino acid level, but inexact at nucleotide level.
* Indicates no exact match at amino acid or nucleotide level was found, the reported allele is the closest nucleotide match. Note some genes have multiple nucleotide sequences in the database that share the same allele name. The specific sequence variants are labelled as .v1, .v2 (e.g. strA.v1 above) and are considered to share the same functionality.

Col_acquired	Fcyn_acquired	Flq_acquired	Gly_acquired	MLS_acquired	Sul_acquired	Tgc_acquired
-	-	-	-	-	-	-

These empty columns indicate that no known resistance genes associated with colistin (Col), fosfomycin (Fcyn), fluoroquinolones (Flq), glycopeptides (Gly), macrolides (MLS), sulfonamides (Sul) or tigecycline (Tgc) were detected.

Beta-lactamases

Bla_acquired	Bla_inhR_acquired	Bla_ESBL_acquired	Bla_ESBL_inhR_acquired	Bla_Carb_acquired	Bla_chr	SHV_mutations
OXA-1;OXA-10;TEM-1D.v1^	-	CTX-M-15	-	-	SHV-1^	-

The next lot of columns tell us about the **beta-lactamase enzymes** that were found. These are divided into classes to indicate their expected spectrum of activity. Genes in the `Bla_acquired` column are narrow/broad spectrum, which are not expected to have much impact on phenotype as *K. pneumoniae* are intrinsically resistant to ampicillin anyway due to the presence of chromosomal *bla*SHV. However, this strain carries three additional enzymes which may increase MIC to certain cephalosporins. The acquired ESBL, CTX-M-15, is expected to confer resistance to third-generation cephalosporins, such as ceftriaxone. These are commonly used to treat infections in the hospital setting, which is why the presence of ESBLs elevates the resistance score to 1. No carbapenemase (`Carb`) is detected and no alleles associated with resistance to beta-lactamase inhibitors (`inhR`) are detected. SHV-1 is detected, this is the most common chromosomal SHV allele and so we assume it is chromosomally located in this genome (and report it in the `Bla_chr` column), although we don't specifically check for its genetic context. No mutations are found in the SHV gene that are associated with enhanced spectrum of activity, so the `SHV_mutations` column is blank.

Resistance mutations

Omp_mutations	Col_mutations	Flq_mutations	truncated_resistance_hits	spurious_resistance_hits
-	-	-	CatB4.v1?-81%	-

No resistance mutations were found in chromosomal genes. A truncated copy of the acquired chloramphenicol resistance gene *catB4* was detected. But as this genome also has full hits to chloramphenicol genes *cmlA5* and *catII* this doesn't make much difference to our interpretation, which would be that the strain is likely chloramphenicol resistant.

Output columns: K and O loci

wzi	K_locus	K_type	K_locus_confidence	K_locus_problems	K_locus_identity	K_Missing_expected_genes
wzi77	KL57	K57	Typeable	!	99.52%	-

The first column indicates a precise match to the *wzi77* allele. The K_locus column reports the best matching K locus matched to *wzi77.*

Kleborate also calls the Kaptive module for serotype prediction. The results indicate the best matching K locus was KL57, which shared 99.52% nucleotide identity with the reference

KL57 sequence and carried all expected KL57 genes: leading to a ‘Typeable’ confidence in the call, and corresponding prediction of K type K57.

O_locus	O_type	O_locus_confidence	O_locus_problems	O_locus_identity	O_Missing_expected_genes
O1/O2v2	O2afg	Typeable	-	99.27%	-

The O locus results were also generated by Kaptive. This indicates the best matching O locus was O1/O2v2, leading to a ‘Typeable’ confidence in the call and corresponding prediction of O2 subtype O2afg (note subtype is based on identification of additional modifying genes outside the O locus, see this paper).

Example 2

Resistance mutations

We will now explore resistance mutations reported in **ERR1415571** assembly

strain	Flq_acquired	Bla_ESBL_acquired	Bla_Carb_acquired	SHV_mutations	Omp_mutations	Flq_mutations
ERR1415571	-	CTX-M-15;SHV-12	OXA-232	238S;240K;35Q	OmpK35-69%;OmpK36GD	GyrA-83Y;GyrA-87G;ParC-80I

This genome carries a non-wildtype form of the **SHV beta-lactamase**. The allele detected is `SHV-12`, which is a known ESBL allele and so is reported in the `Bla_ESBL_acquired` column. The `SHV_mutations` column shows us this SHV gene carries three common mutations, two of which (238S, 240K) are known to extend the enzyme's spectrum of activity resulting in resistance to third-generation cephalosporins.

This strain also carries **porin mutations** reported in the `'Omp_mutations'` column, both a truncation of OmpK35 and an insertion (GD) in the beta-loop strand of OmpK36. These mutations restrict the transfer of small molecules into the cell, reducing the accumulation of certain drugs and impacting susceptibility. It is likely that the OXA-232 carbapenemase detected in this genome cannot on its own confer full clinical resistance, however combined with the porin mutations the effect is likely to raise the MIC above the threshold for clinical resistance.

The genome lacks acquired fluoroquinolone resistance genes; however, it harbours three mutations related to fluoroquinolone resistance, reported in the `'Flq_mutations'` column. This combination of three mutations, in specific codons of the fluoroquinolone drug targets GyrA and ParC, most likely confer clinical resistance to fluoroquinolones such as ciprofloxacin.

AMR and hypervirulence convergence

Given the public health risks posed by strains that possess both AMR and hypervirulence determinants, there is much interest around the detection of such convergent strains. As mentioned in the introduction of this tutorial, Kleborate enables straightforward detection of these strains, via the virulence and resistance scores. We define convergence as virulence score ≥3 (indicating presence of iuc, which is also a marker for the virulence plasmid) and resistance score ≥1 (indicating presence of an ESBL and/or carbapenemase).

Two assemblies included in the test data repository, **ERR4920551** and **ERR4920450** from the BARNARDS neonatal dataset, show evidence of convergence. When we check the acquired AMR columns, we observe `CTX-M-15` in the `Bla_ESBL_acquired` column for both assemblies and `NDM-1` in the `Bla_Carb_acquired` column for ERR4920551. The latter also has a OmpK35 truncation, which is associated with increased resistance to carbapenems.

strain	Bla_acquired	Bla_ESBL_acquired	Bla_Carb_acquired	Bla_chr	SHV_mutations	Omp_mutations
ERR4920450	OXA-1;OXA-10;TEM-1D.v1^	CTX-M-15	-	SHV-1^	-	-
ERR4920551	CMY-6;TEM-1D.v1^	CTX-M-15	NDM-1	SHV-28^	-	OmpK35-46%

Based on the relevant virulence columns in the Kleborate output (see below), both strains probably carry deletion variants of the virulence plasmid (i.e. missing *iro*, *rmpADC* and *rmpA2* in ERR4920551; missing *rmpADC* in ERR4920450).

strain	YbST	Yersiniabactin	AbST	Aerobactin	SmST	Salmochelin	RmST	RmpADC	rmpA2	Bla_ESBL_acquired	Bla_Carb_acquired	resistance_score
ERR4920450	581-2LV	ybt 9; ICEKp3	1-1LV	iuc 1 (truncated)	1	iro 1	0	-	rmpA2_8*-0%	CTX-M-15	-	1
ERR4920551	277	ybt 16; ICEKp12	1	iuc 1	0	-	0	-	-	CTX-M-15	NDM-1	2

References

Tools and data

Argimón, S. *et al.* (2021) 'Rapid Genomic Characterization and Global Surveillance of *Klebsiella* Using Pathogenwatch', *Clinical Infectious Diseases*, 73(Supplement_4), pp. S325–S335. Available at: https://doi.org/10.1093/cid/ciab784.

Brisse, S., Passet, V., Haugaard A.B., *et al*. (2013) '*wzi* Gene Sequencing, a Rapid Method for Determination of Capsular Type for *Klebsiella* Strains', *Journal of Clinical Microbiology*, 51(12), p. 4073. Available at: https://doi.org/10.1128/JCM.01924-13

David, S., Reuter, S., Harris, S.R., *et al.* (2019) 'Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread', *Nature Microbiology,* 4(11), p. 1919. Available at: https://doi.org/10.1038/s41564-019-0492-8.

Grundmann, HajoKoraqi, Andi et al. (2017). Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study 17(2), 153 - 163

Lam, M.M.C., Wick, R.R., Judd, L.M., *et al.* (2022) Kaptive 2.0: updated capsule and LPS locus typing for the *Klebsiella pneumoniae* species complex. *Microbial Genomics,* 8(3). Available at: https://doi.org/10.1099/mgen.0.000800.

Lam, M.M.C., Wick, R.R., Watts, S.C., *et al.* (2021) 'A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex', *Nature Communications*, 12(1), p. 4188. Available at: https://doi.org/10.1038/s41467-021-24448-3.

Lam, M.M.C., Wyres, K.L., Judd, L.M., *et al.* (2018) 'Tracking key virulence loci encoding aerobactin and salmochelin siderophore synthesis in *Klebsiella pneumoniae*'. Genome Medicine, 10:77. Available at: https://doi.org/10.1186/s13073-018-0587-5.

Lam, M.M.C., Wick, R.R., Wyres, K.L., *et al.* (2018) 'Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICE*Kp* in *Klebsiella pneumoniae* populations'. *Microbial Genomics,* 4(9). Available at: https://doi.org/10.1099/mgen.0.000196.

Sands *et al.* (2021) 'Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries', *Nature Microbiology*, 6(4), pp. 512–523. Available at: https://doi.org/10.1038/s41564-021-00870-7.

Wyres, K.L., Lam, M.M.C. and Holt, K.E. (2020) 'Population genomics of *Klebsiella pneumoniae*', *Nature Reviews Microbiology* 18:344. Available at: https://rdcu.be/b1Fbb.

Wyres, K.L. *et al.* (2016) 'Identification of *Klebsiella* capsule synthesis loci from whole genome data', *Microbial Genomics*, 2(12). Available at: https://doi.org/10.1099/mgen.0.000102.

Klebsiella virulence

Bachman, M.A., Oyler, J.E., Burns, S.H., *et al.* (2011) '*Klebsiella pneumoniae* Yersiniabactin Promotes Respiratory Tract Infection through Evasion of Lipocalin 2'. *Infection and Immunity* 79(8). Available at: https://doi.org/10.1128/IAI.05114-11

Fajardo-Lubián, A, Ben Zakour, N.L., Agyekum, A., *et al.* (2019) 'Host adaptation and convergent evolution increases antibiotic resistance without loss of virulence in a major human pathogen'. *PLoS Pathogens.* Available at: https://doi.org/10.1371/journal.ppat.1007218

Lam, M.M.C. and Wyres, K.L., *et al.* (2018) 'Population genomics of hypervirulent Klebsiella pneumoniae clonal-group 23 reveals early emergence and rapid global dissemination'. *Nature Communications* 9, Article 2703. Available at: https://doi.org/10.1038/s41467-018-05114-7

Nougayrède, J.P., Homburg, S., Taier, F., *et al.* (2006) '*Escherichia coli* Induces DNA Double-Strand Breaks in Eukaryotic Cells'. *Science* 313(5788). Available at: https://doi.org/10.1126/science.1127059

Russo, T.A., Olson, R., MacDonald, U., *et al.* (2015) 'Aerobactin, but Not Yersiniabactin, Salmochelin, or Enterobactin, Enables the Growth/Survival of Hypervirulent (Hypermucoviscous) *Klebsiella pneumoniae Ex Vivo* and *In Vivo*'. *Infection and Immunity* 83(8). Available at: https://doi.org/10.1128/IAI.00430-15

Russo, T.A., Olson, R., Fang, C.T., *et al.* (2018) 'Identification of Biomarkers for Differentiation of Hypervirulent *Klebsiella pneumoniae* from Classical *K. pneumoniae*'. *Journal of Clinical Microbiology* 56(9): e00776-18. Available at: https://doi.org/10.1128/JCM.00776-18

Walker, K.A., Treat, L.P., Sepulveda, W.E., *et al.* (2020) 'The Small Protein RmpD Drives Hypermucoviscosity in *Klebsiella pneumoniae*'. *mBio* 11(5). Available at: https://doi.org/10.1128/mbio.01750-20

Walker, K.A., Miner, T.A., Palacios, M., *et al.* (2019) 'A Klebsiella pneumoniae Regulatory Mutant Has Reduced Capsule Expression but Retains Hypermucoviscosity'. *mBio* 10(2). Available at: https://doi.org/10.1128/mbio.00089-19

Klebsiella resistance mutations

Fajardo-Lubián, A., Ben Zakour, N.L., Agyekum A., *et al*. (2019) 'Host adaptation and convergent evolution increases antibiotic resistance without loss of virulence in a major human pathogen'. *PLoS Pathogens*, 15(3): e1007218. Available at: https://doi.org/10.1371/journal.ppat.1007218.

Neubauer, S. Madzgalla, S., Marquet, M., *et al.* (2020) 'Genotype-Phenotype Correlation Study of SHV β-Lactamases Offers New Insight into SHV Resistance Profiles'. *Antimicrobial Agents and Chemotherapy*, 64(7). Available at: https://doi.org/10.1128/AAC.02293-19

Wong, J.L.C., Romano, M., Kerry, L.E., *et al.* (2019) 'OmpK36-mediated Carbapenem resistance attenuates ST258 *Klebsiella pneumoniae* in vivo'. *Nature Communications* 10, Article 3957. Available at: https://doi.org/10.1038/s41467-019-11756-y