Klebsiella virulence typing – part II

Assistant Professor Zoe A. Dyson

London School of Hygiene and Tropical Medicine

zoe.dyson@lshtm.ac.uk

@msmicrobiocode

Today's schedule

Time	Activity
09:00-09:50 (50 mins)	**Kaptive hands on practical (continued)**
09:50-10:00 (10 mins)	**Class discussion**
10:00-10:50 (50 mins)	**Lecture: *Klebsiella* virulence typing - part II** • The colibactin genotoxin • Siderophores • An introduction to Kleborate
10:50-11:00 (10 mins)	**Class discussion**
11:00-11:15	**Break**
11:15-12:00 (45 mins)	**Lecture: *Klebsiella* antimicrobial resistance (AMR) typing** • An introduction to AMR determinant detection • AMR in *Klebsiella pnuemoniae* • AMR detection & score analysis with Kleborate
12:00-12:10 (10 mins)	**Class discussion**
12:10-13:00 (50 mins)	**Kleborate hands on practical**
13:00-14:00 (1 hour)	**Lunch**
14:00-15:15 (1 hour 15 mins)	**Kleborate hands on practical**
15:15-15:30 (15 mins)	**Break**
15:30-16:30 (1 hour)	**Kleborate hands on practical (continued)**

Lecture outline: *Klebsiella* virulence typing – part I

1. The genotoxin Colibactin (*clb/pks*)

2. Siderophores
   - Enterobactin (*ent*)
   - Salmochelin (*iro*)
   - Aerobactin (*iuc*)
   - Yersiniabactin (*ybt*)

3. An introduction to Kleborate

Revision: virulence/pathogenicity determinants

Revision: Different infection types driven by *Klebsiella*

Members of the KpSC can cause a variety of different infection types:

- Those causing **healthcare associated infections (HAI)** in immunocompromised individuals are often referred to as '**classical**' strains
- Those causing **community acquired infections (CAI)** in healthy individuals are often referred to as **hypervirulent** strains
- **Pathogenicity/virulence** factors commonly associated with more severe cases can be detected from whole genome sequencing (WGS) data

	Characteristic(s) for strain type	
Parameter	Classical	Hypervirulent
Common types of infection	Pneumonia, UTI, bacteremia	Pyogenic liver abscess; bacteremia; lung, neck, and kidney abscesses; pneumonia; cellulitis; necrotizing fasciitis; myositis, meningitis; endophthalmitis
Susceptible population(s)	Immunosuppressed (diabetics, patients with malignancies)	Diabetics, healthy people
Capsule type(s)	Capsule serotypes K1–K78	Hypercapsule serotype K1 (93%) or K2
Siderophores (% of strains expressing siderophore)	Enterobactin (100), yersiniabactin (17–46), salmochelin (2–4), aerobactin (6)	Enterobactin (100), yersiniabactin (90), salmochelin (>90), aerobactin (93–100)
Geographical concentration	Worldwide	Primarily Taiwan and Southeast Asia
Primarily acquired infection type	Nosocomial	Community acquired
Frequency of reports of antibiotic resistance	Frequent (ESBL and carbapenemase producing)	Infrequent

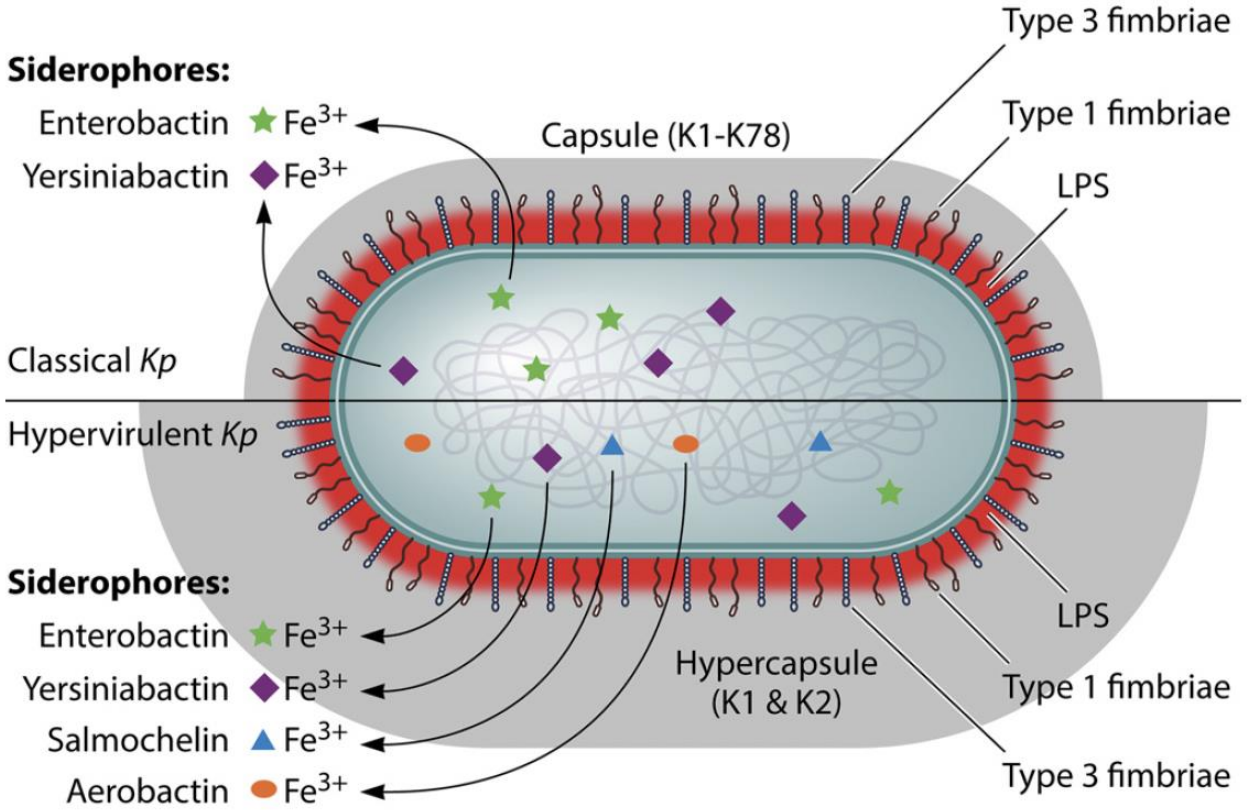
Paczosa and Mecsas 2016, MMBR

Revision: Pathogenicity/virulence factors in *Klebsiella*

- Many factors contribute to the ability of *K. pneumoniae* strains to cause disease and evade host defences

- The most well studied of these are:
  - Siderophores
  - Fimbriae/pilli
  - Capsule
  - Lipopolysaccharide
  - Toxins e.g. colibactin

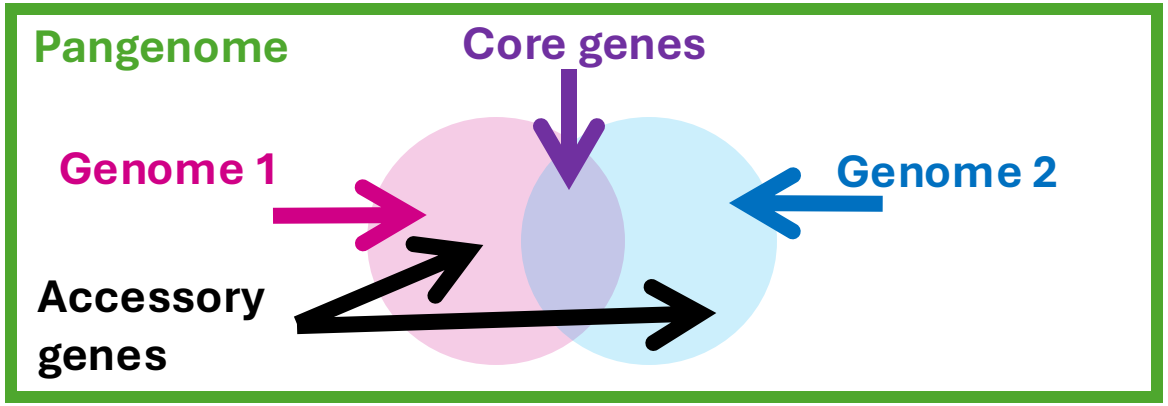
- Other less well studied factors include:
  - Outer membrane proteins (OMPs)
  - Porins
  - Efflux-pumps
  - Iron-transport systems
  - Allantoin metabolism systems
  - Many others…

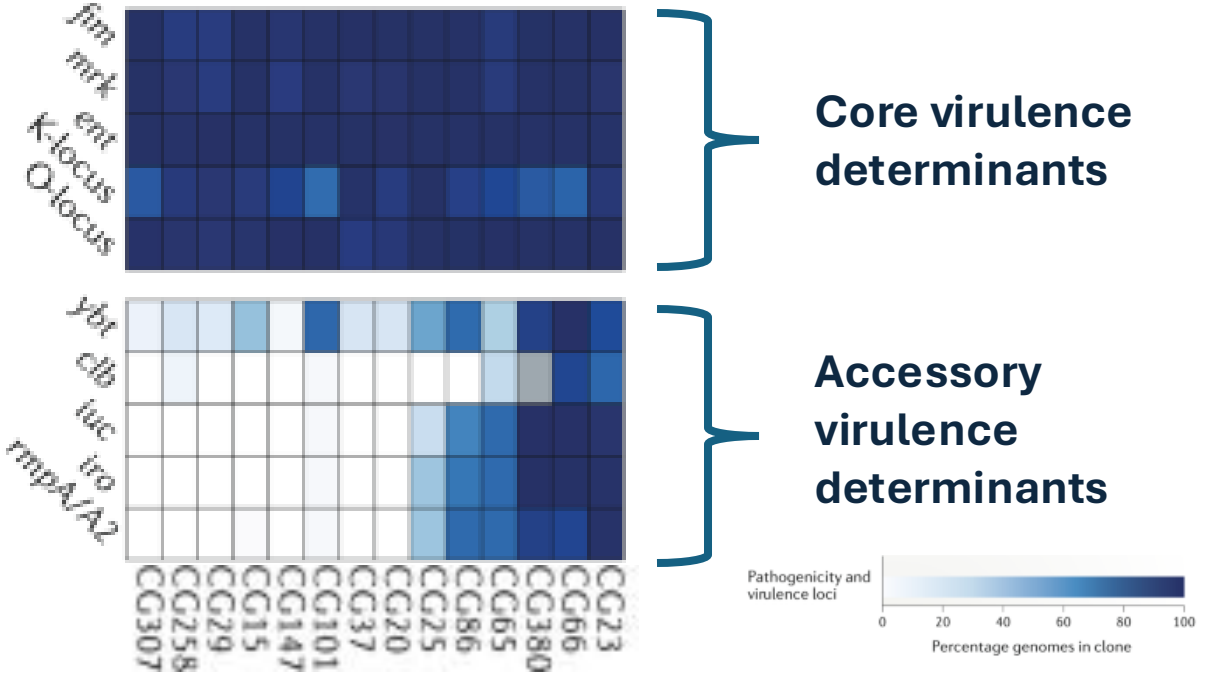


Paczosa and Mecsas 2016, MMBR

Revision: Pathogenicity/virulence factors in *Klebsiella*

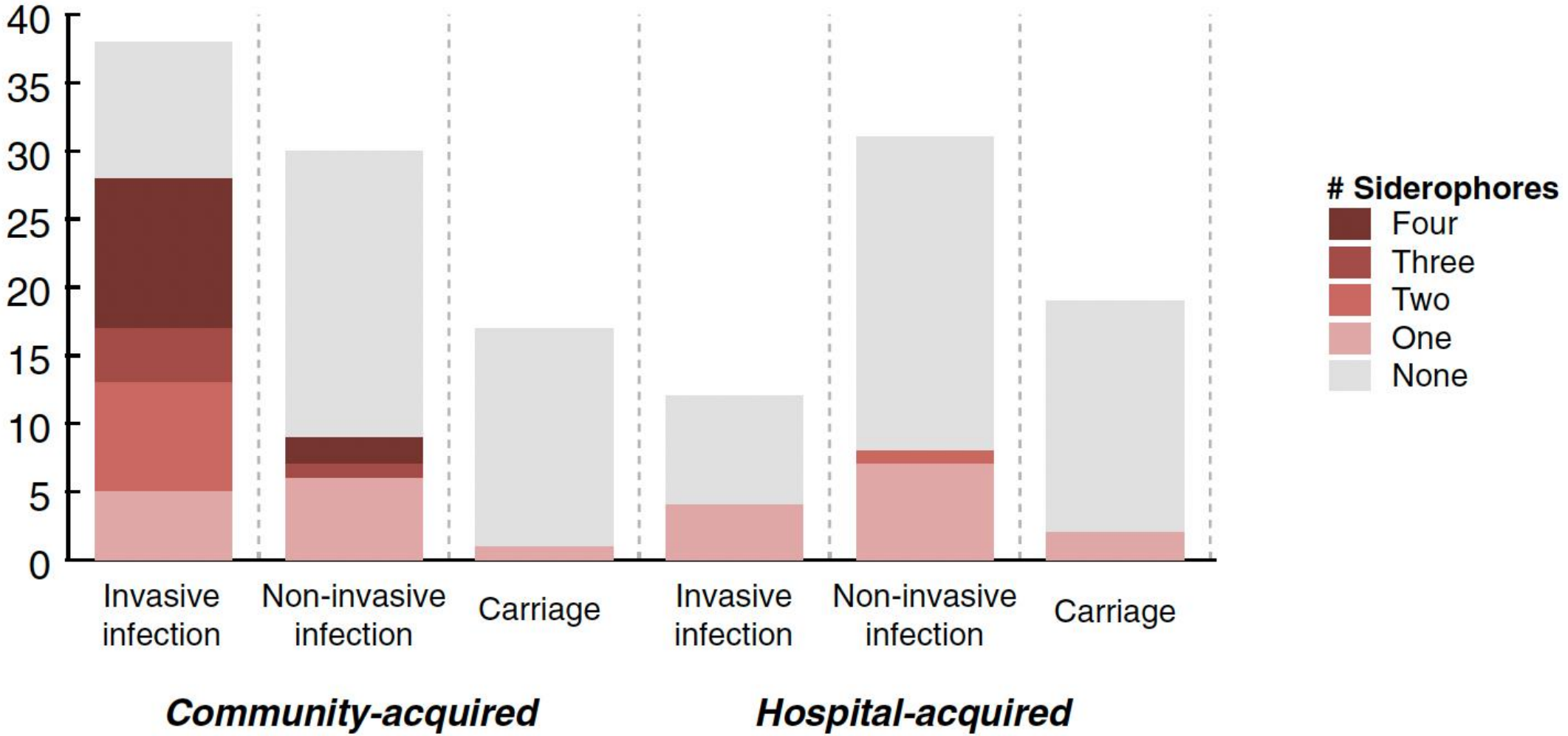
- All *K. pneumoniae* encode a subset of four **core** chromosomally integrated pathogenicity/virulence factors for establishing infections in mammals:
  - *Ent* locus encoding the siderophore enterobactin
  - Types 1and 3 Fimbriae/pilli (*fim* and *mrk* loci)
  - Lipopolysaccharide (O-antigen)
  - Capsular polysaccharide (K-antigen)
- Hypervirulent strains may have:
  - Specific capsule types
  - Other **accessory** siderophores (e.g. yersiniabactin, aerobactin, salmochelin)
  - The genotoxin colibactin
- This lecture focuses on:
  - The genotoxin colibactin
  - Siderophores





Paczosa and Mecsas 2016, MMBR; Wyres et al. 2020, Nat Rev Microbiol

Distribution of accessory virulence determinants among different infection types



Holt et al. PNAS, 2015

The genotoxin colibactin

The genotoxin colibactin (*clb/pks*)

- Colibactin is a genotoxic polykeytide that induces double stranded DNA damage in eukaryotic cells, thus it appears to
  - Promotes mucosal & gut colonisation
  - Promotes dissemination to blood and other organs (e.g. brain)
  - Contribute to colorectal cancer

- First described in *Escherichia coli*

- Present in ~10% of KpSC members

- Commonly associated with known hypervirulent clonal groups
  e.g. CG66, CG23, CG65, CG380

Lam et al. 2018, Microb Genom; Wyres et al. 2020, Nat Rev Microbiol;
Lu et al. 2017, Front Cell Infect Microbiol; Lai et al. 2014, PLOS ONE

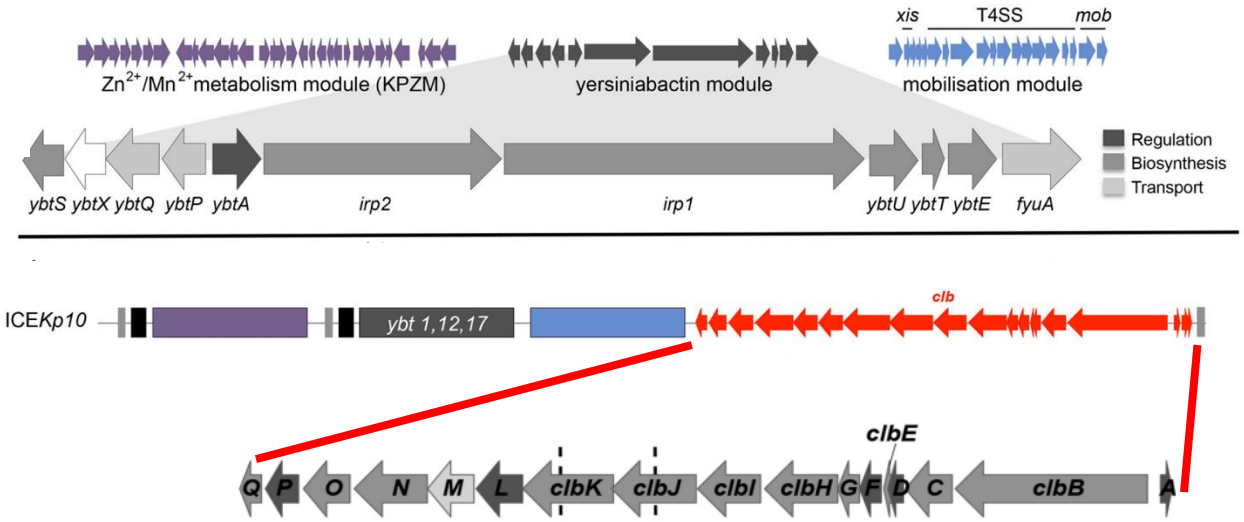
The genotoxin colibactin (*clb/pks*)

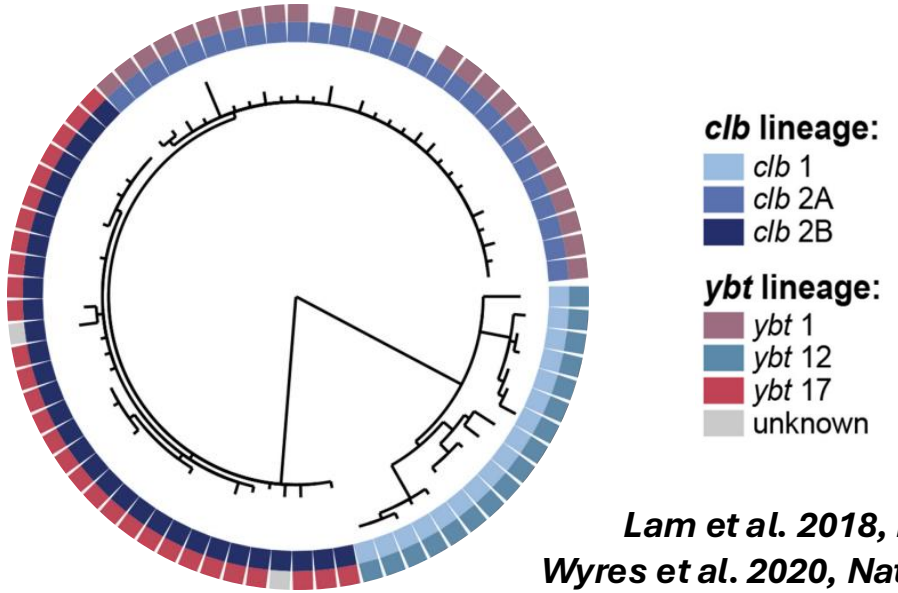
- Synthesised by multi-enzyme complex encoded in the *clb/pks* gene locus (~50 kbp)

- Commonly associated with an Integrative Conjugative Element (ICE*Kp10*)
  - ICE are self-transmissible mobile elements that encode genes for their own excision, circularisation, mobilisation and integration
  - Commonly integrate near tRNA-*Asn*
  - ICE*Kp* are common virulence-associated mobile genetic elements among members of the KpSC
  - Can also carry genes for siderophores,
    e.g. 38% of strains carry colibactin + yersiniabactin

- 3 colibactin lineages
  - Associated with specific yersiniabactin (*ybt*) lineages i.e. *ybt* 1, *ybt* 12, *ybt* 17

- Colibactin Sequence Typing (CbST) scheme
  - Available via PubMLST and BIGSdb
  - Integrated into Kleborate
  - Useful epidemiological markers





Lam et al. 2018, Microb Genom
Wyres et al. 2020, Nat Rev Microbiol
Lai et al. 2014, PLOS ONE

Revision: Multi-locus sequence typing (MLST)

- Defined set of seven core genes for typing (e.g. *rpoB, gapA, mdh, pgi, phoE, infB, tonB* for *Klebsiella*)
- For each gene, every unique allele is assigned a number (e.g. *gapA*-1, *gapA*-2, *gapA*-3)

locus	allele id	sequence
gapA	1	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	2	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	3	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	4	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	5	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	6	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	7	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	8	AATCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	9	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	10	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	11	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC
gapA	12	AACCTGAAGTGGGAC . . . ACCGGTATGGCGTTC

***Klebsiella gapA* alleles in MLST scheme**
- *gapA* currently has 392 unique alleles

Revision: Multilocus sequence typing (MLST)

- Defined set of seven core genes for typing (e.g. *rpoB, gapA, mdh, pgi, phoE, infB, tonB* for *Klebsiella*)
- For each gene, every unique allele is assigned a number (e.g. *gapA*-1, *gapA*-2, *gapA*-3)
- Each unique combination of gene alleles defines a unique sequence type (ST)
- Each genome can then be represented by the set of allele numbers across these genes
- MLST database made up of
  (i) set of all allele sequences
  (ii) lookup table of allele number combinations to ST

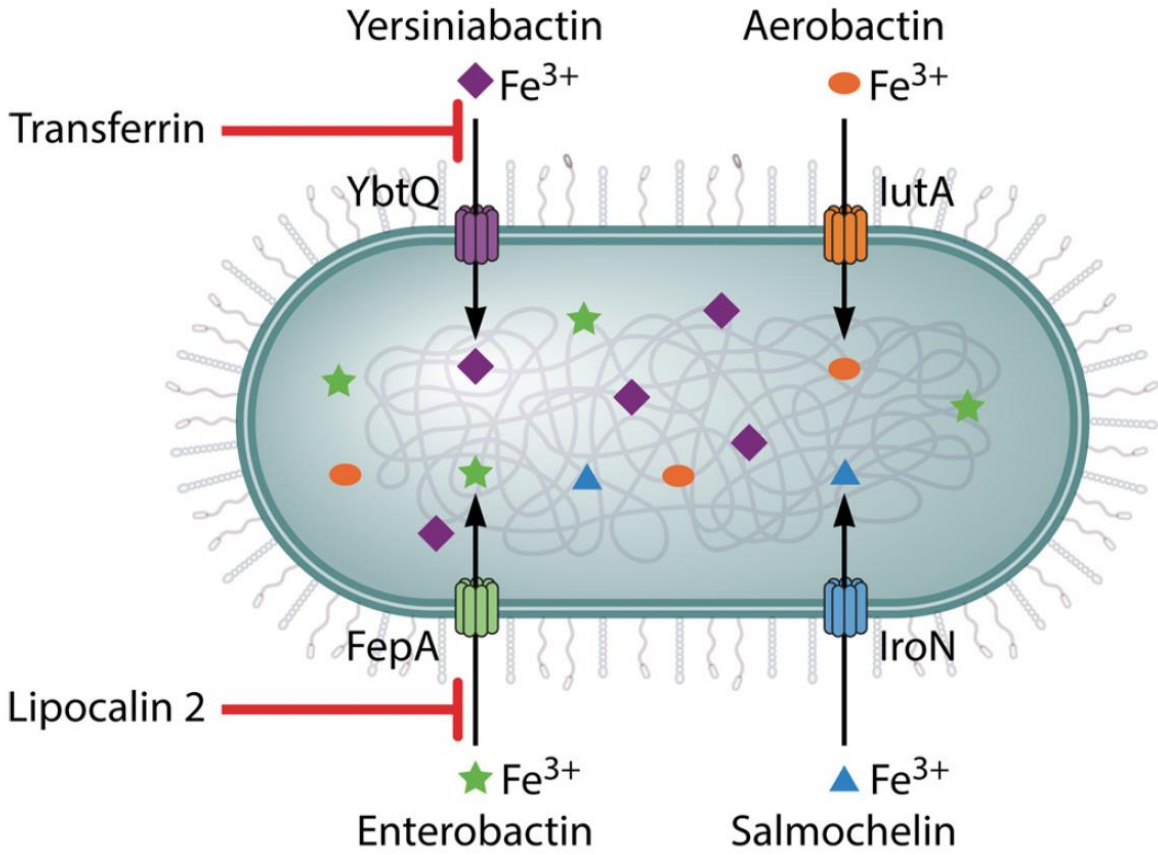
ST	*gapA*	*infB*	*mdh*	*pgi*	*phoE*	*rpoB*	*tonB*
1	4	4	1	1	7	4	10
2	3	4	1	1	9	4	17
3	5	5	1	1	9	6	11
4	3	1	1	1	3	3	1
5	2	2	1	1	3	3	3

***Klebsiella* MLST scheme**
- Currently has >7500 unique allelic profiles

Siderophores

Siderophores

- Iron is a limited resource that is required by members of the KpSC that must be acquired from the environment during an infection
- Iron is not readily available in the host during an infection as it is sequestered by the host as a part of the immune response that restricts pathogen growth
- Siderophores are iron-chelating molecules that can competitively scavenge iron from host iron transport proteins (e.g. transferrin, lactoferrin) as they have a higher affinity to iron
- 4 siderophore systems are commonly found in the KpSC
  - Enterobactin
  - Salmochelin
  - Aerobactin
  - Yersiniabactin
- Enterobactin is a core siderophore
- Salmochelin, Aerobactin & Yersiniabactin are accessory siderophores that can enhance virulence/pathogenicity



Paczosa and Mecsas 2016, MMBR
Wyres et al. 2020, Nat Rev Microbiol

Enterobactin (*ent*)

Core siderophore: enterobactin (*ent*)

- The siderophore enterobactin is considered part of the core genome as it is present in almost all KpSC members

- Highest affinity to iron of the four common siderophores

- Neutralised by human lipocalin-2 (Lcn2) which therefore inhibits KpSC growth and induces an inflammatory response
  - Not reported by Kleborate

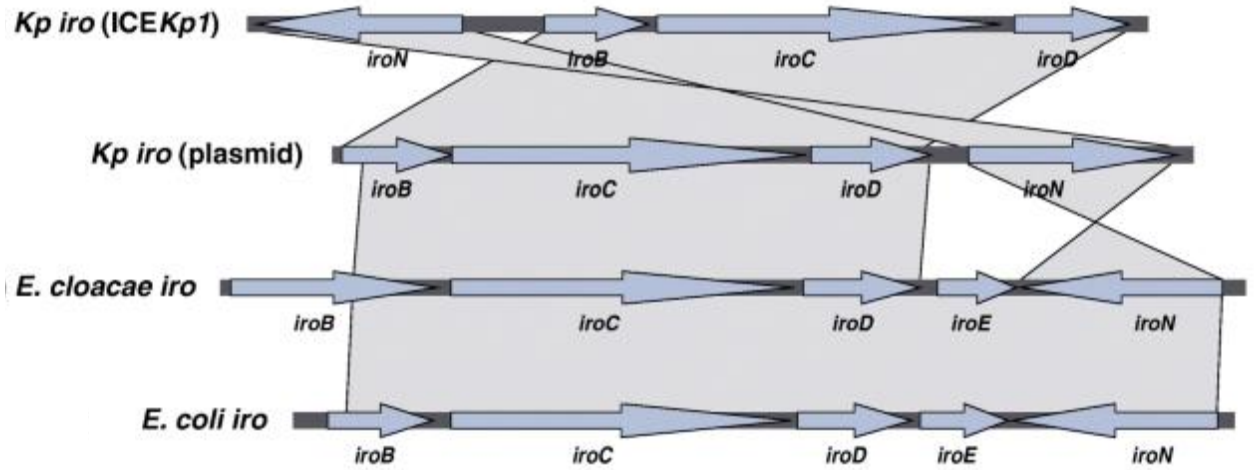
- Enterobactin is encoded by the *ent* locus
  - Encoded by the *entABCDEF* gene locus
  - The *fepABCDG* gene locus encodes proteins that mediate it's transport
    - *fepA* encodes the uptake receptor

Paczosa and Mecsas 2016, MMBR
Wyres et al. 2020, Nat Rev Microbiol

Salmochelin (*iro*)

Accessory siderophore: salmochelin (*iro*)

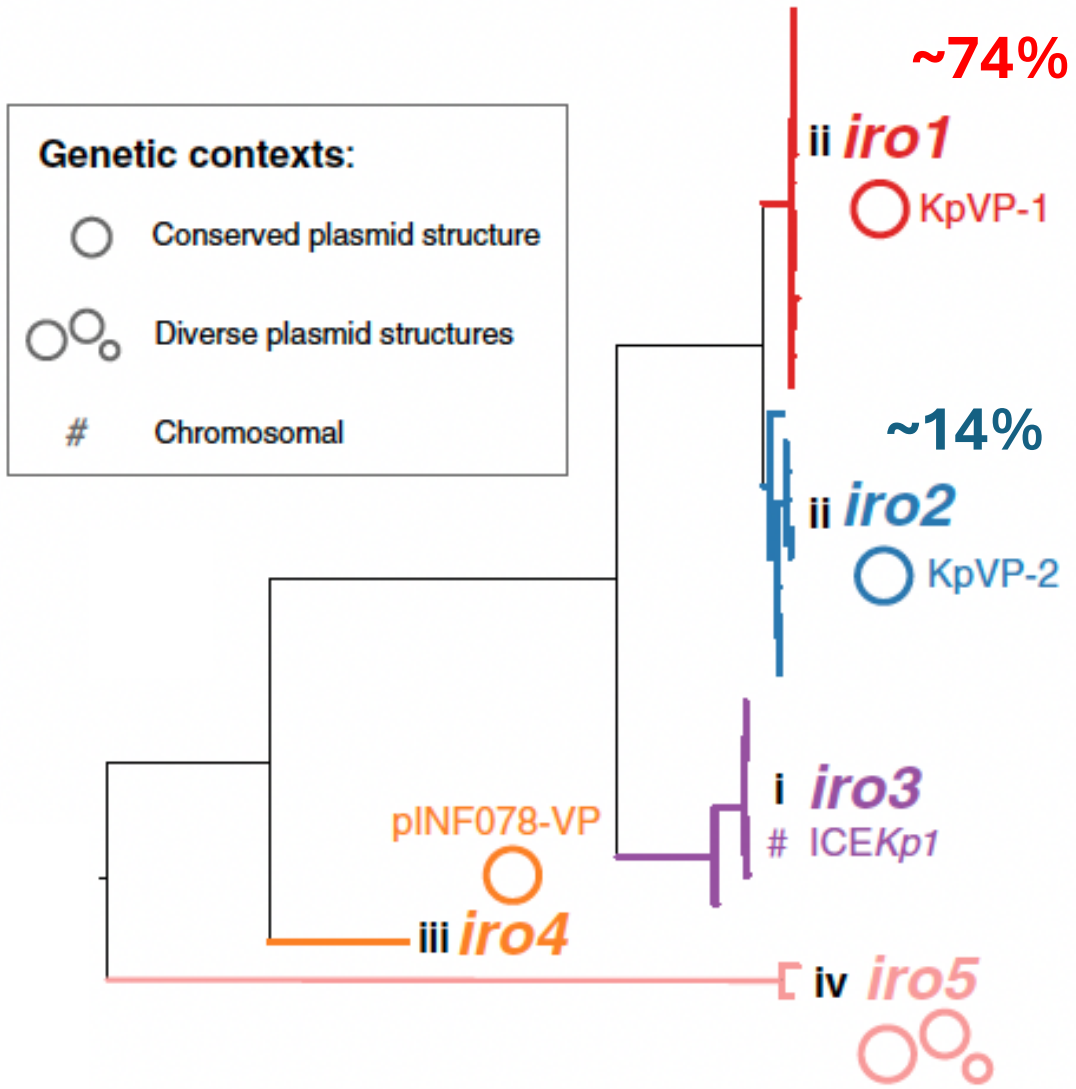
- Salmochelin forms a c-glycosylated form of enterobactin (Gly-Ent)
- Evades lipocalin-2 (Lcn2), and therefore circumvents inflammation while maintaining iron scavenging
- Higher affinity to iron than aerobactin (*iro*) and yersiniabactin (*ybt*)
- Enhances colonisation of the nasopharynx
- Present in <10% of KpSC, but prevalent in hypervirulent strains (~90%)
- Genes responsible for modification of enterobactin (*ent*) are encoded by the genes within the *iro* locus
  - Transport is mediated by IroN
- Associated with multiple mobile genetic elements



Lam et al. 2018, Genome Med
Paczosa and Mecsas 2016, MMBR
Wyres et al. 2020, Nat Rev Microbiol

Accessory siderophore: salmochelin (*iro*)

- Salmochelin sequence typing scheme (SmST)
  - Utilises the same principles of MLST
  - Available via PubMLST and BIGSdb
  - Integrated into Kleborate
- 5 distinct lineages, associated with different mobile genetic elements
  - Lineages 1 & 2 (*iro1* & *iro2*) are associated with *Klebsiella* virulence plasmids (KpVP-1 and KpVP-2)
  - Lineage 4 (*iro4*) is associated with plasmid pINV078-VP
  - Lineage 3 (iro3) is associated with an Integrative Conjugative Element (ICE*Kp1*)
  - Lineage 5 is associated with a diverse range of plasmids
- Useful epidemiological makers for tracking novel acquisitions of siderophores

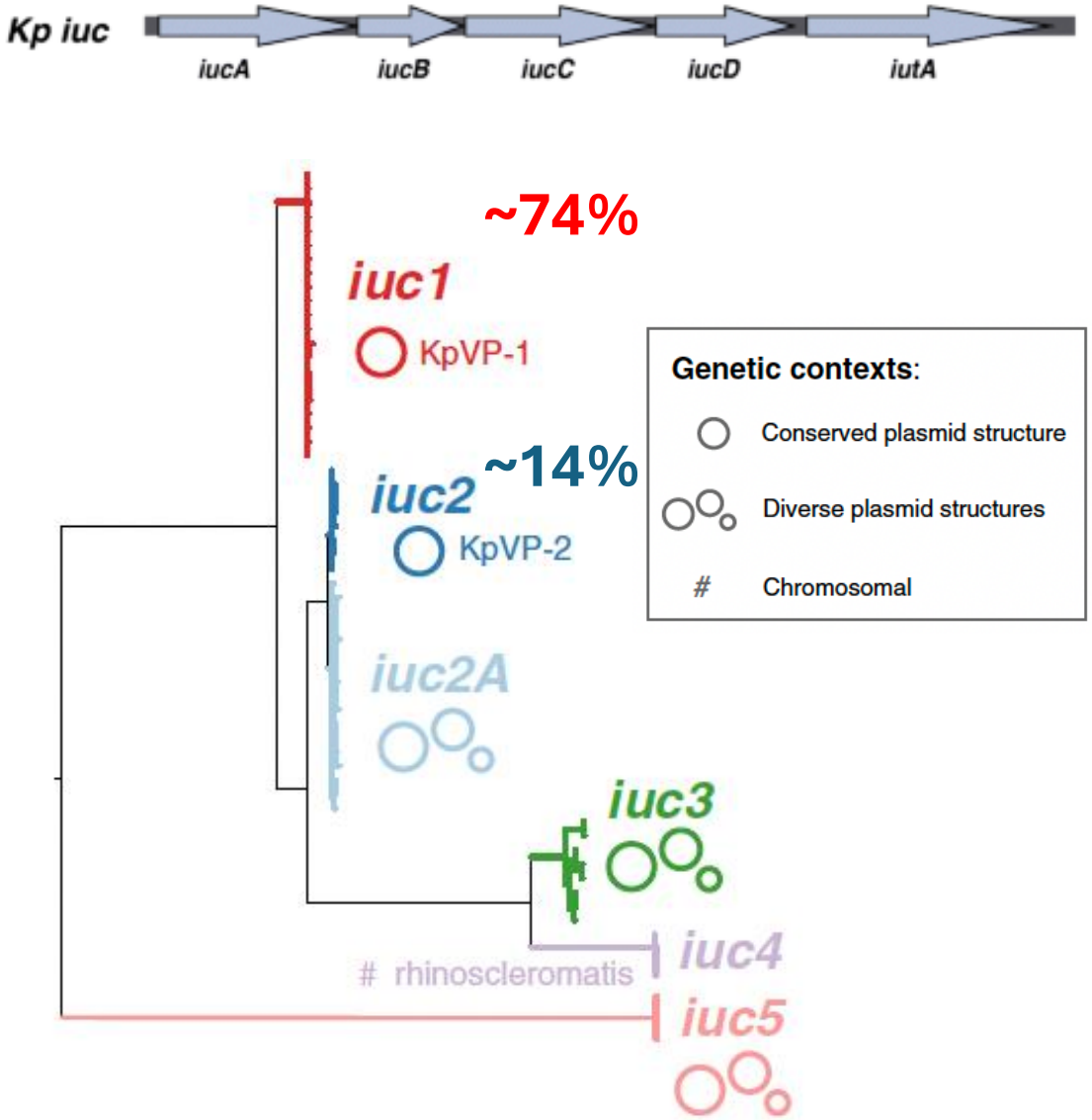


Lam et al. 2018, Genom Med

Aerobactin (*iuc*)

Accessory siderophore: aerobactin (*iuc*)

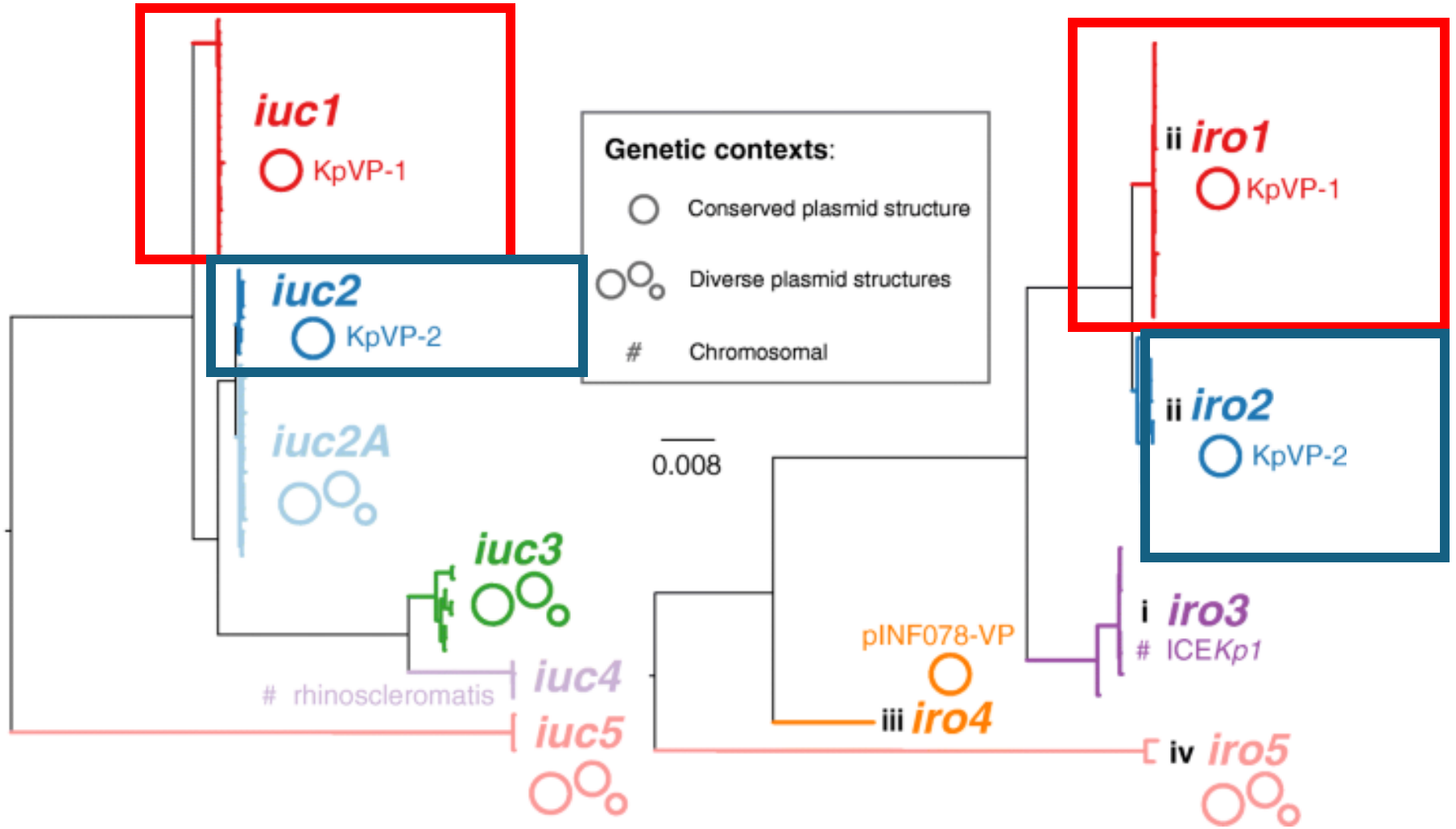
- Citrate-hydroxomate siderophore
- Not inhibited by lipocalin-2 (Lcn2) due to structural differences
- Lowest affinity to iron of all the common KpSC siderophores
- Present in <10% of KpSC, but prevalent among hypervirulent strains (93-100%)
- Encoded by the *iuc* locus
  - Transport medicated by *iutA*
- Aerobactin sequence typing scheme (AbST)
  - Utilises the same principles of MLST
  - Five lineages (iuc1-5) each associated with different mobile genetic elements
  - Available via PubMLST and BIGSdb
  - Integrated into Kleborate



Lam et al. 2018, Genom Med

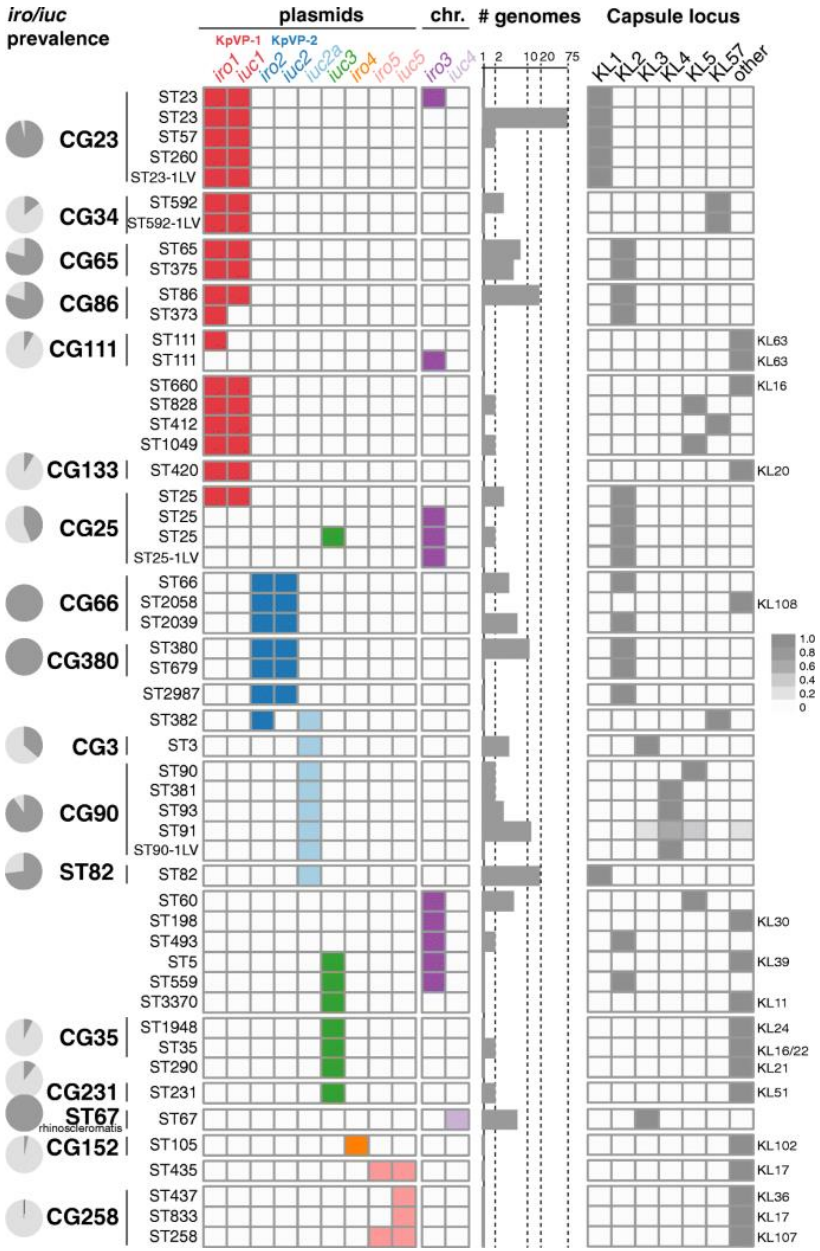
Accessory siderophores: aerobactin (*iuc*) & salmochelin (*iro*)

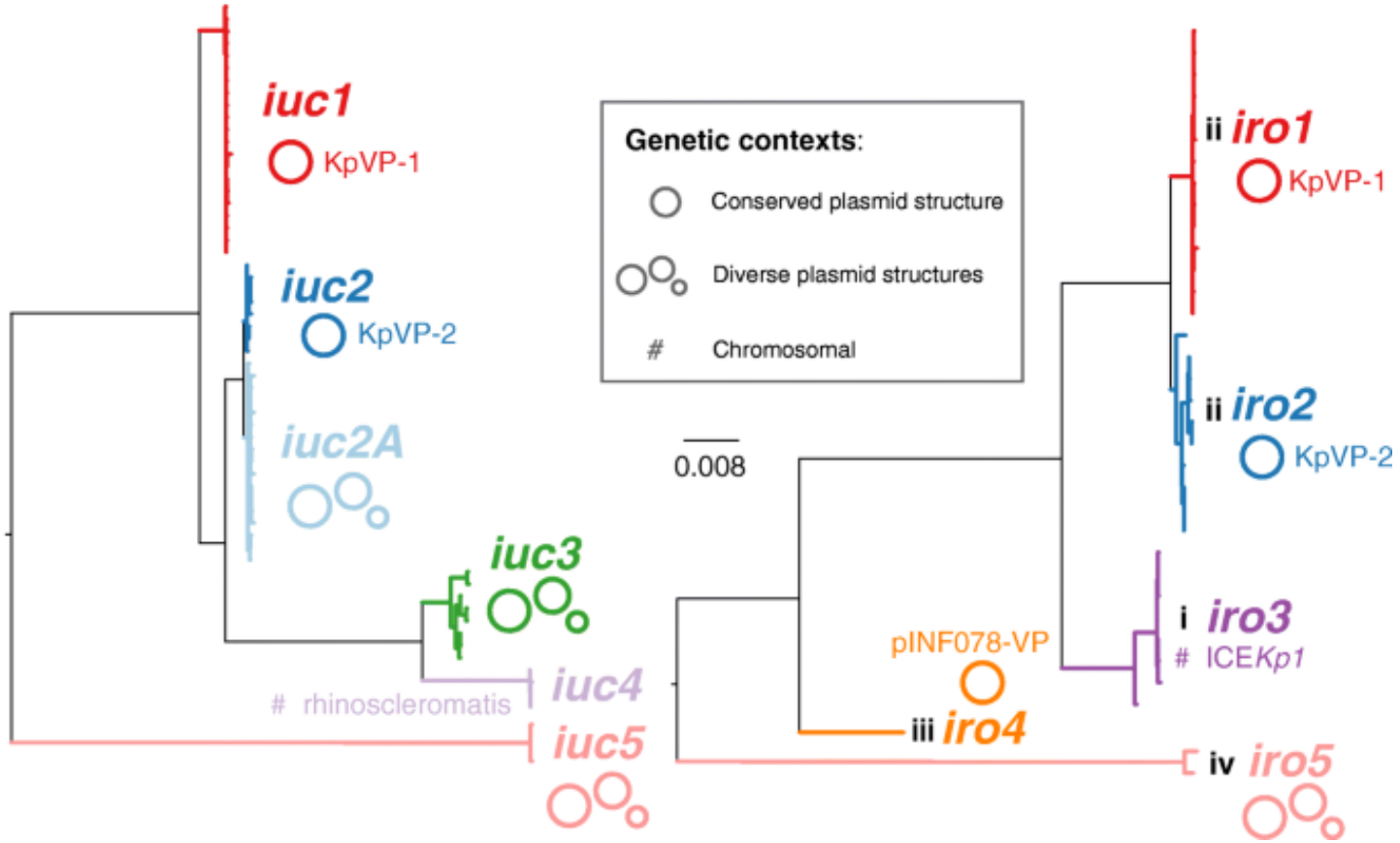
Co-carriage of Salmochelin (iro), Aerobactin (iuc) & rmp genes is common & driven by Klebsiella virulence plasmids (KpVP)



Lam et al. 2018, Genom Med

Accessory siderophores: aerobactin (*iuc*) & salmochelin (*iro*)





Lam et al. 2018, Genom Med

Yersiniabactin (*ybt*)

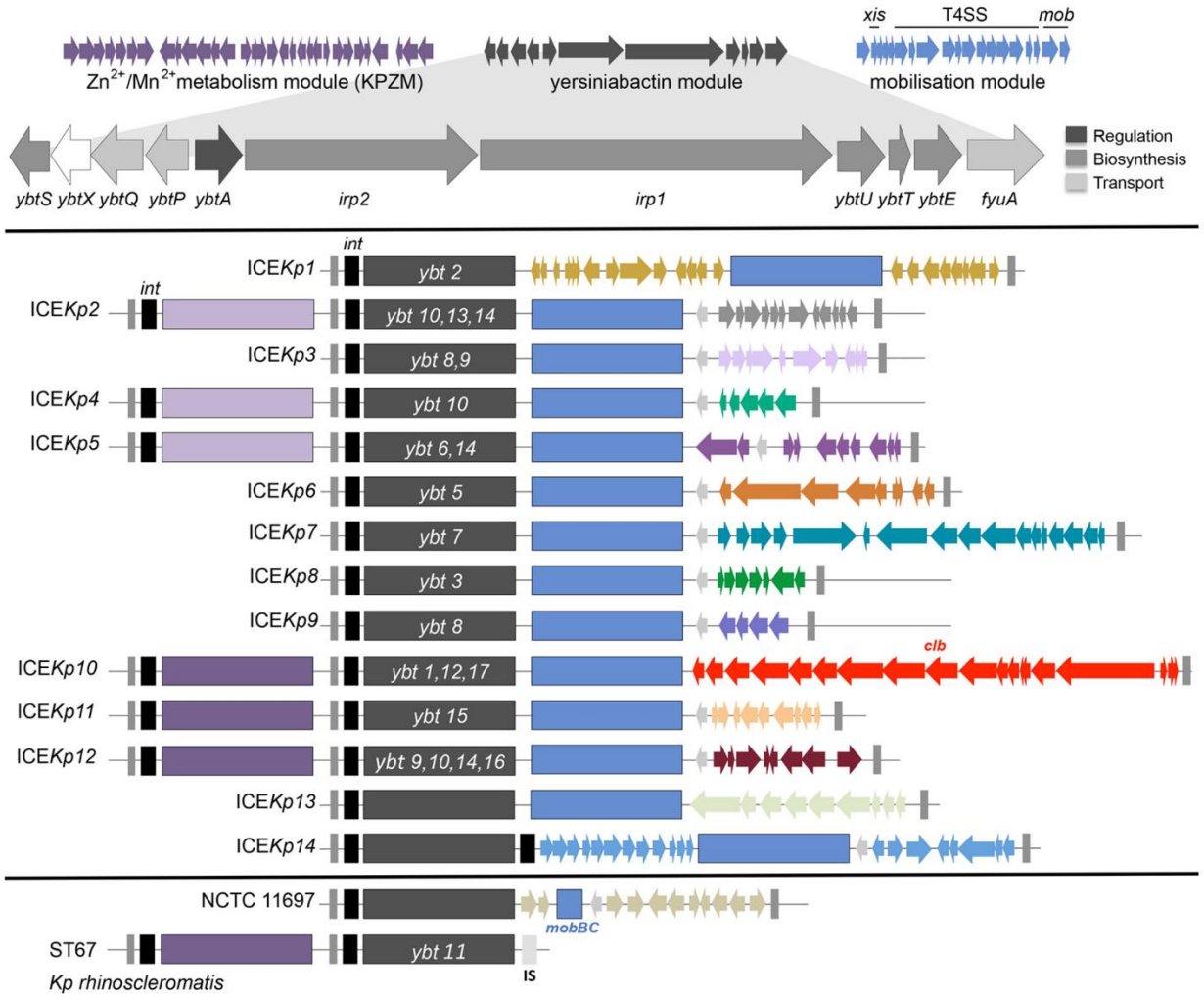
Accessory siderophore: yersiniabactin (*ybt*)

- Yersiniabactin was originally discovered in *Yersinia* spp. as a part of a pathogenicity gene island, but appears to have emerged from the KpSC
- Common among isolates from respiratory tract and in hypervirulent clinical isolates (~78-100%), and in classical MDR isolates (6-80%)
- Expressed during lung infections, allowing for proliferation and high bacterial loads
- Not inhibited by lipocalin-2 (Lcn2) due to structural differences
- Dissemination from the lungs may not be possible without other siderophores as yersinabactin appears unable to acquire iron from transferrin which is concentrated in blood

Wyres et al. 2020, Nat Rev Microbiol; Paczosa and Mecsas 2016, MMBR

Accessory siderophore: yersiniabactin (*ybt*)

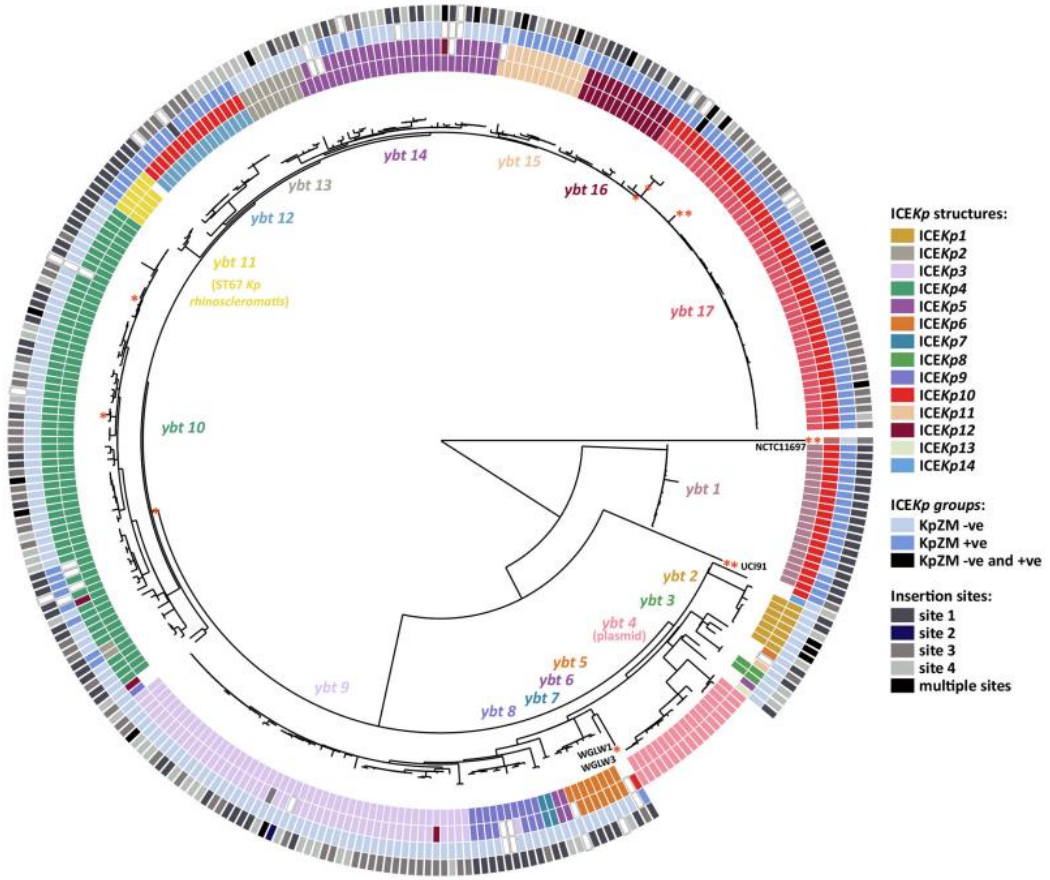
- Commonly associated with an Integrative Conjugative Elements (ICE)
  - Several different structural variants
  - Some co-carriage of colibactin genes (*cbl/pks*)
  - 22 different ICE*Kp* structures
- Synthesis proteins encoded in *irp* genes
  - Transporters mediated by ybt & *fyu* genes



Lam et al. 2018, Microb Genom

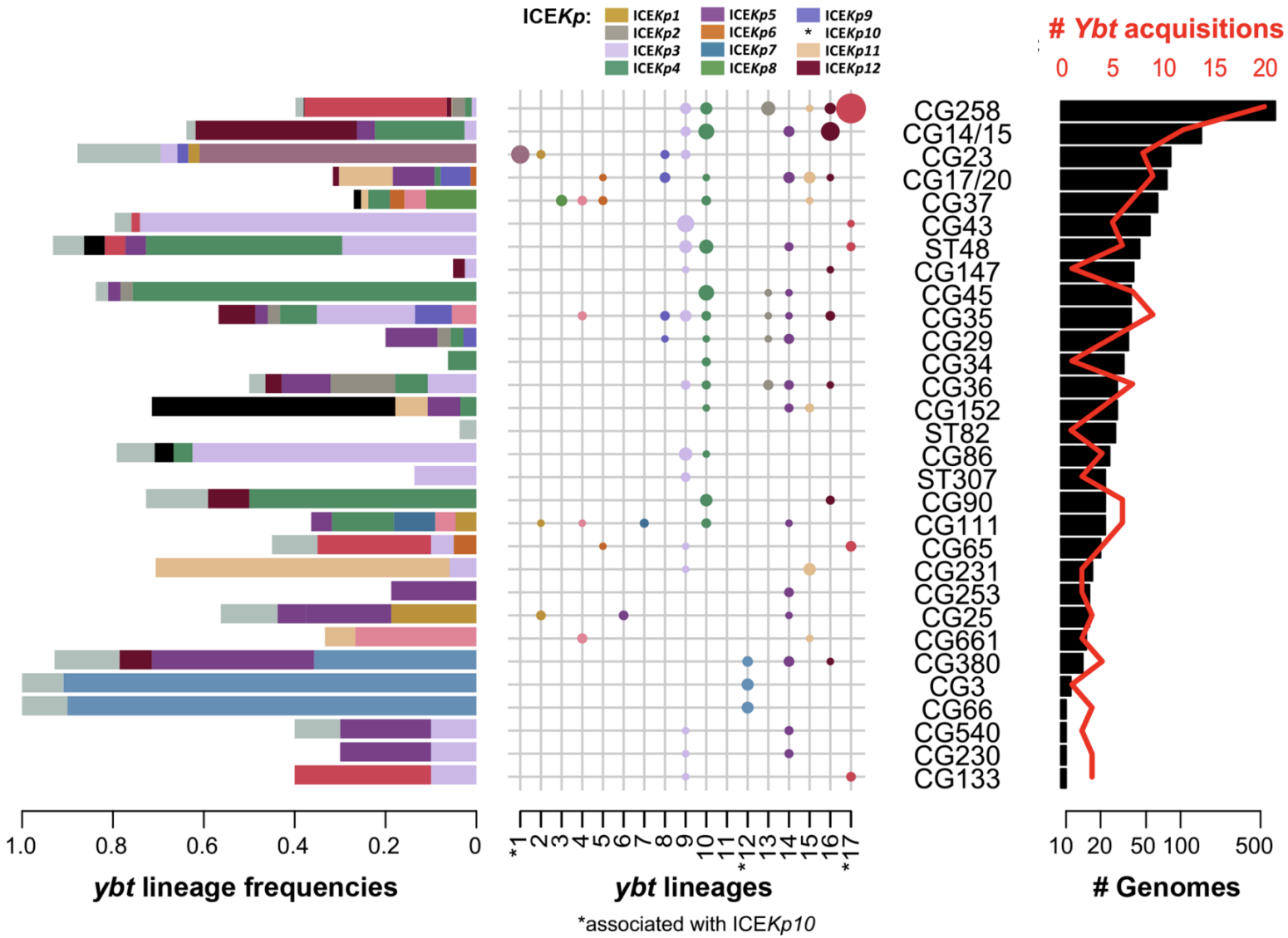
Accessory siderophore: yersiniabactin (*ybt*)

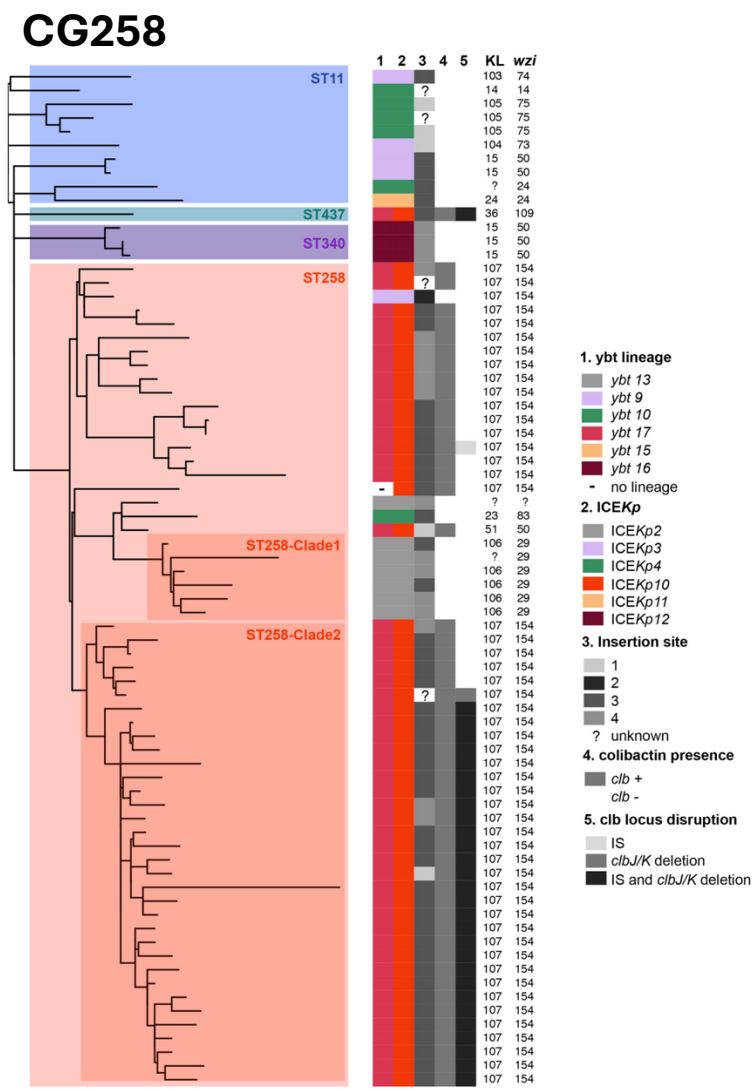
- Yersiniabactin sequence typing (YbST) scheme
  - 28 different lineages
  - Available via PubMLST and BIGSdb
  - Integrated into Kleborate
- Useful epidemiological markers for surveillance of novel acquisitions
- Some associations with specific clonal groups, and other bacterial species



Lam et al. 2018, Microb Genom

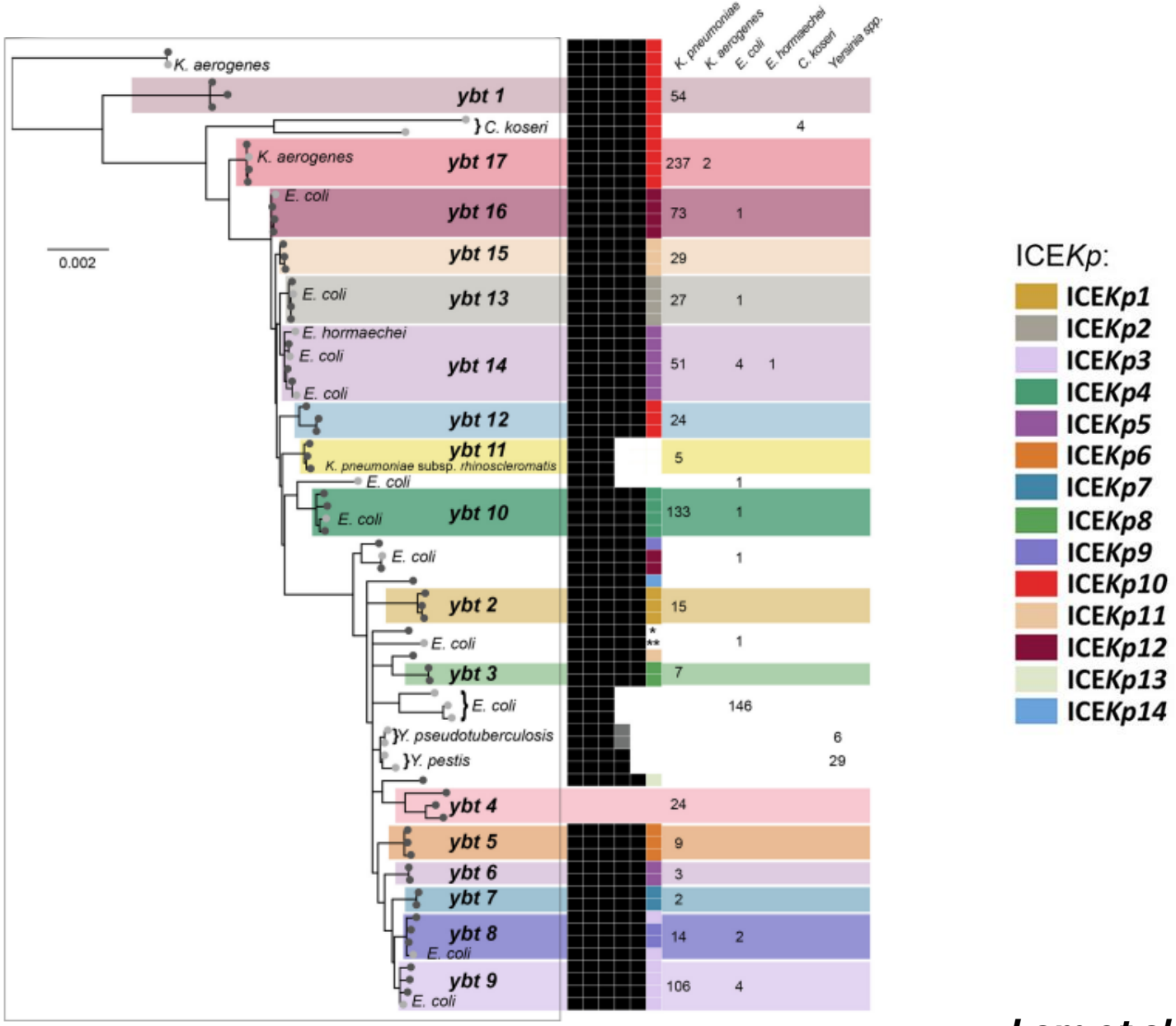
Accessory siderophore: yersiniabactin (*ybt*)





Lam et al. 2018, Microb Genom

Accessory siderophore: yersiniabactin (*ybt*)



Lam et al. 2018, Microb Genom

Typing methods provide useful nomenclature

1. To stratify cases into pathogen subtypes
   - To identify / define those with different genomic / biological traits and assess whether they have distinct epidemiology, so they can be managed in a targeted way
   - May consider phylogenetic relatedness to define groups, or use non-phylogenetic groupings

2. To investigate emergence and spread
   - Of the infectious disease generally, or variants of special clinical interest such as drug resistant or hypervirulent strains
   - Identify sources of infection, track transmission events, investigate outbreaks

An introduction to Kleborate

Lam et al. 2021, Nat Commun

Kleborate: genotyping & surveillance framework

Bioinformatics software for analysing KpSC whole genome sequencing data.

+ additional modules for:

- *Klebsiella oxytoca* species complex (KoSC)
- *Escherichia coli*

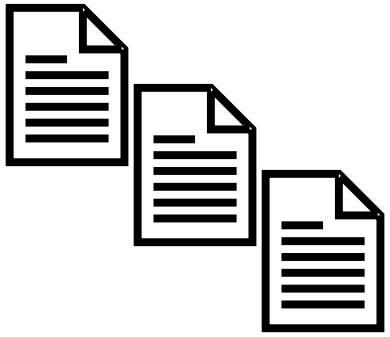
In a single analysis, Kleborate provides data on:

1. Assembly Quality Control Statistics
2. Species typing
3. Multilocus sequence typing (MLST)
4. *In silico* serotyping: K- and O-antigen typing
5. Virulence determinants
6. Antimicrobial Resistance determinants
7. Virulence and AMR scores

Lam et al. 2021, Nat Commun

Kleborate: input & output files

Input



Assembled whole genome sequences (.fasta file)

KLEBORATE

github.com/klebgenomics/Kaptive

Command line tool

Output

Text files summarising genome data (.txt file)

Also available via Galaxy Europe & Pathogenwatch!

Lam et al. 2021, Nat Commun

Kleborate: genotyping & surveillance framework

Installation via:

conda package manager
pip

Modules:

- General Modules
- **Modules for KpSC**
- Modules for KoSC
- Modules for *Escherichia* species complex

```
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

Input/output:
  -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)

Modules:
  --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

Help:
  -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.
```

Example command:

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

Lam et al. 2021, Nat Commun

Kleborate: assembly quality control metrics

The below are provided to help users assess the reliability of genotyping results:

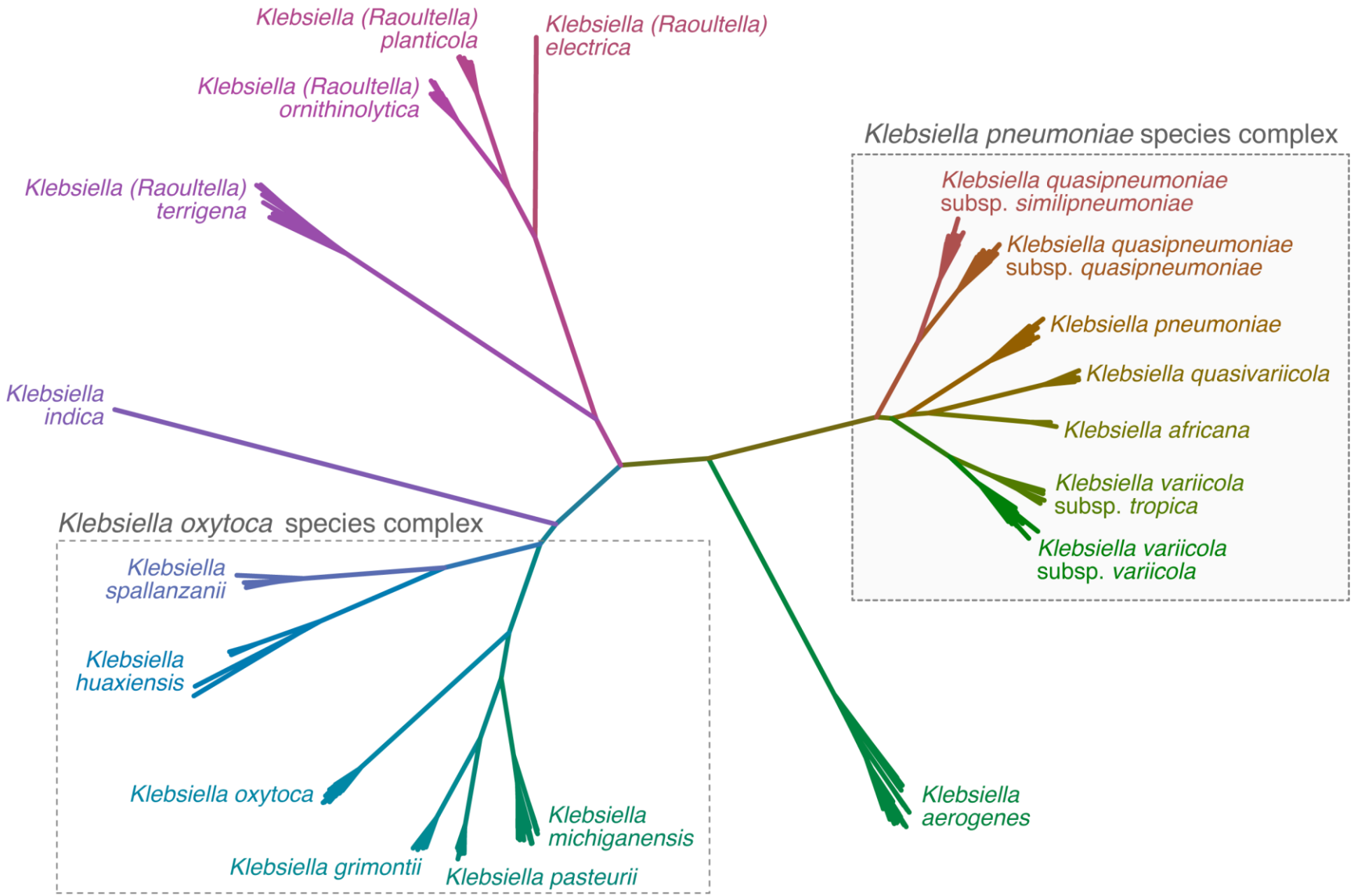
- Contig count
- N50 (sequence length of the shortest contig at 50% of the total assembly length)
- Largest contig size
- Total genome size
- Number of ambiguous bases

- Low-quality warnings triggered by
  - Ambiguous bases
  - Assembly length outside expected range (4.5-7.5 Mbp)
  - N50 < 10,000 bp

- Users should carefully consider the genotyping outputs for low-quality assemblies

Lam et al. 2021, Nat Commun

Kleborate: species typing



Lam et al. 2021, Nat Commun

Kleborate: Multilocus sequence typing (MLST)

- Defined set of **seven core genes** for typing (e.g. *rpoB, gapA, mdh, pgi, phoE, infB, tonB* for *Klebsiella*)
- For each gene, every unique allele is assigned a number (e.g. *gapA*-1, *gapA*-2, *gapA*-3)
- Each unique combination of gene alleles defines a unique sequence type (ST)
- Each genome can then be represented by the set of allele numbers across these genes
- MLST database made up of
  (i) set of all allele sequences
  (ii) lookup table of allele number combinations to ST

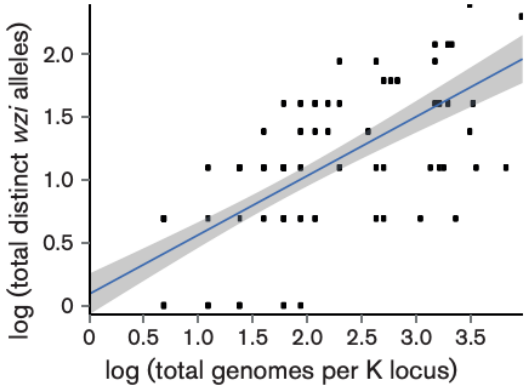
ST	*gapA*	*infB*	*mdh*	*pgi*	*phoE*	*rpoB*	*tonB*
1	4	4	1	1	7	4	10
2	3	4	1	1	9	4	17
3	5	5	1	1	9	6	11
4	3	1	1	1	3	3	1
5	2	2	1	1	3	3	3

KpSC MLST scheme
- Currently has >7500 unique allelic profiles

Lam et al. 2021, Nat Commun

Kleborate: *in silico* serotyping

Option 1: Wzi locus typing (default option)



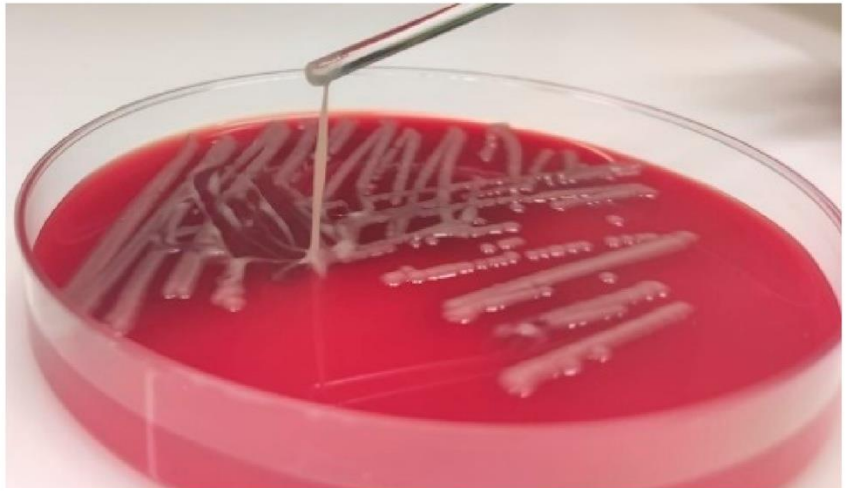
Option 2: Kaptive (must be specifically called)

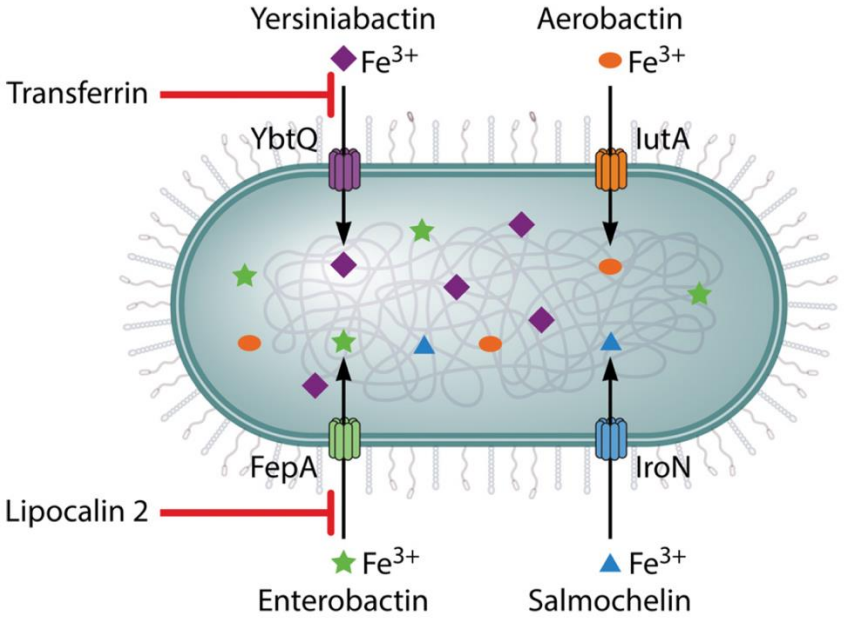
`-m klebsiella_pneumo_complex__kaptive`

Lam et al. 2021, Nat Commun; Brisse et al. 2013, J Clin Microbiol; Wyres et al. 2016, Microb Genom

Kleborate: virulence determinant detection

- Regulators of mucoid phenotype (*rmp*) genes (hypercapsule/hypermucoid capsule phenotypes)
- Colibactin genotoxin
- Siderophores (detection and sequence typing)
  - Enterobactin (*ent*) – not reported – core siderophore inactivated by Lcn2
  - Salmochelin (*iro*) + SmST
  - Aerobactin (*iuc*) + AbST
  - Yersiniabactin (*ybt*) (YbST)





Lam et al. 2021, Nat Commun

Kleborate: virulence scores

Summary of the relative level of acquired virulence/pathogenicity

Low

High

Virulence score	Virulence determinants*
0	No accessory virulence determinants
1	Yersiniabactin (*ybt*) only
2	Colibactin (clb), either with or without yersiniabactin (*ybt*)**
3	Aerobactin (*iuc*), either with or without yersiniabactin + Colibactin
4	Aerobactin (*iuc*) + yersiniabactin (ybt), without Colibactin (*cbl*)
5	Aerobactin (*iuc*) + yersiniabactin (*ybt*) + Colibactin (*cbl*)

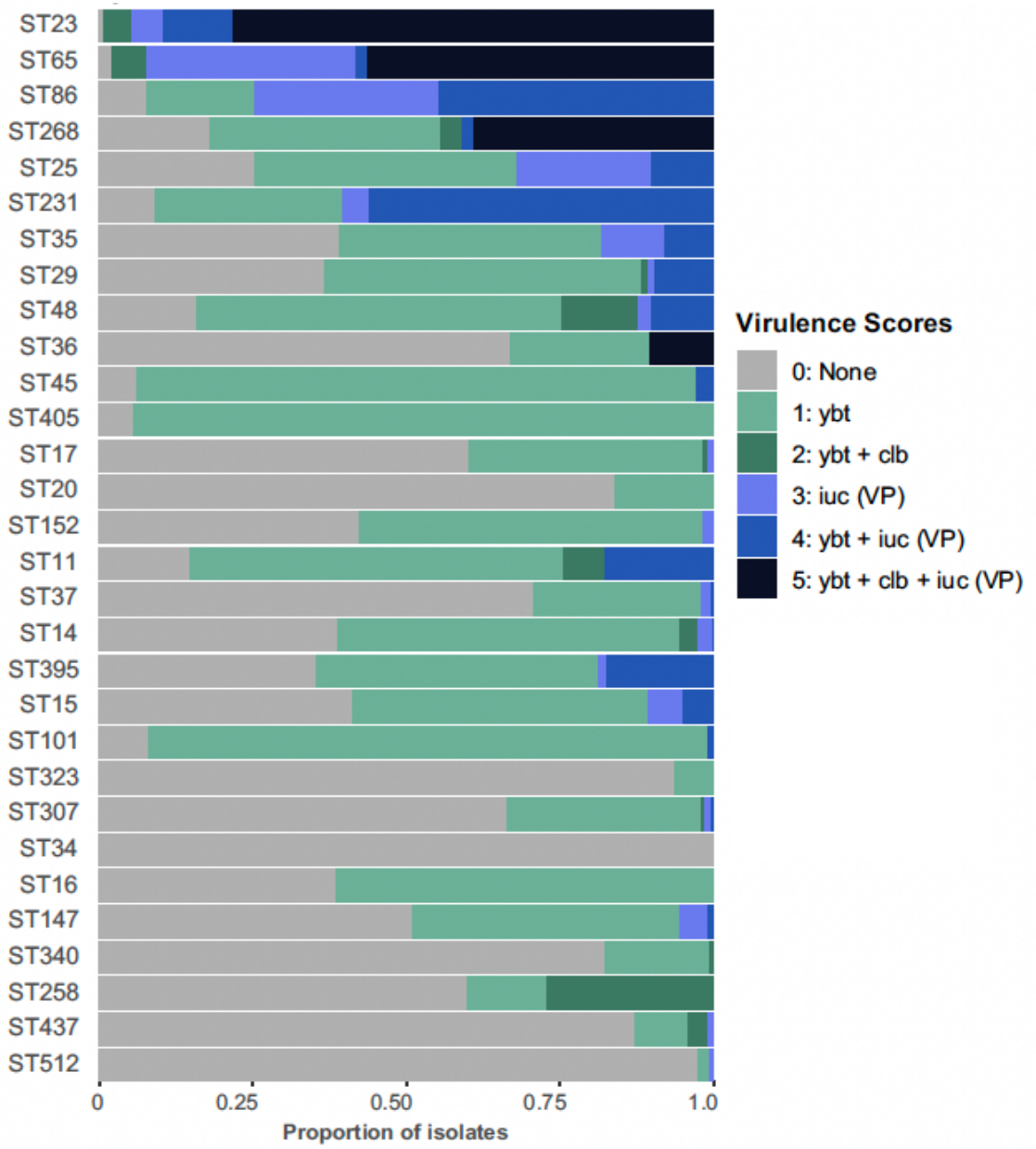
** rmp & Salmochelin (iro) not considered in scoring, but commonly co-carried with aerobactin (iuc) on virulence plasmids (KpVP)*

*** High levels of co-carriage of colibactin and yersiniabactin on ICEKp10*

Lam et al. 2021, Nat Commun

Kleborate: virulence scores

Virulence score	Virulence determinants*
0	No accessory virulence determinants
1	Yersiniabactin (*ybt*) only
2	Colibactin (clb), either with or without yersiniabactin (*ybt*)**
3	Aerobactin (*iuc*), either with or without yersiniabactin + Colibactin
4	Aerobactin (*iuc*) + yersiniabactin (ybt), without Colibactin (*cbl*)
5	Aerobactin (*iuc*) + yersiniabactin (*ybt*) + Colibactin (*cbl*)



Lam et al. 2021, Nat Commun

Kleborate: ONT R9 chemistry

- ONT only assemblies were comparable to Illumina only and Illumina-ONT hybrid assemblies
- Reliable capsule (K) type calls for all strains (100% exact or best matching locus)
- Reliable multi-locus sequence type (MLST) assignment (98.3% exact match or single-locus variants)
- Good detection of acquired AMR genes and mutations (88–100% correct identification across the various drug classes)
- Good detection of virulence determinants e.g. for yersiniabactin - 100% correct identification & correct lineage calls

MICROBIAL GENOMICS

RESEARCH ARTICLE

Foster-Nyarko *et al.*, *Microbial Genomics* 2023;9:000936

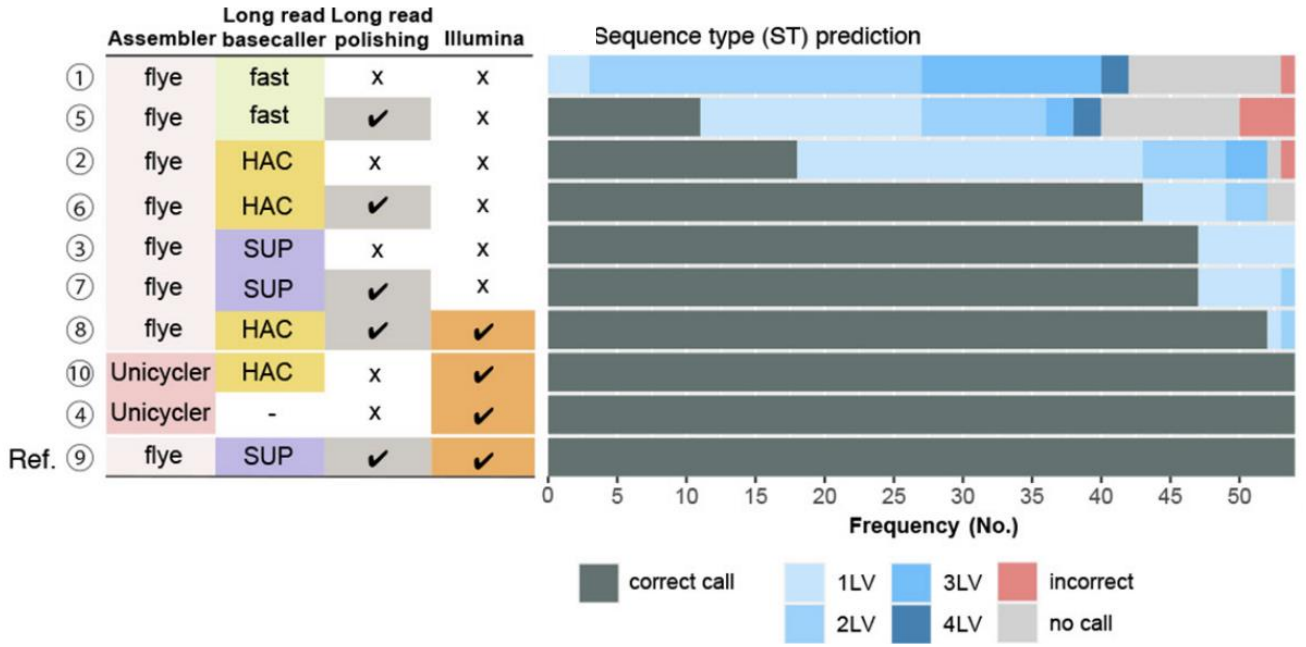
DOI 10.1099/mgen.0.000936

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Nanopore-only assemblies for genomic surveillance of the global priority drug-resistant pathogen, *Klebsiella pneumoniae*

Ebenezer Foster-Nyarko[1,*], Hugh Cottingham[2], Ryan R. Wick[2], Louise M. Judd[2], Margaret M. C. Lam[2], Kelly L. Wyres[2], Thomas D. Stanton[1], Kara K. Tsang[1], Sophia David[3], David M. Aanensen[3], Sylvain Brisse[4] and Kathryn E. Holt[1,2]



Kleborate: genotyping & surveillance framework

Bioinformatics software for analysing KpSC whole genome sequencing data.

In a single analysis, Kleborate provides data on:

1. Assembly Quality Control Statistics
2. Species typing
3. Multilocus sequence typing (MLST)
4. *In silico* serotyping: K- and O-antigen typing
5. Virulence determinants
6. Antimicrobial Resistance determinants
7. Virulence and AMR scores

Discussed next lecture!

Lam et al. 2021, Nat Commun

Any questions or reflections?