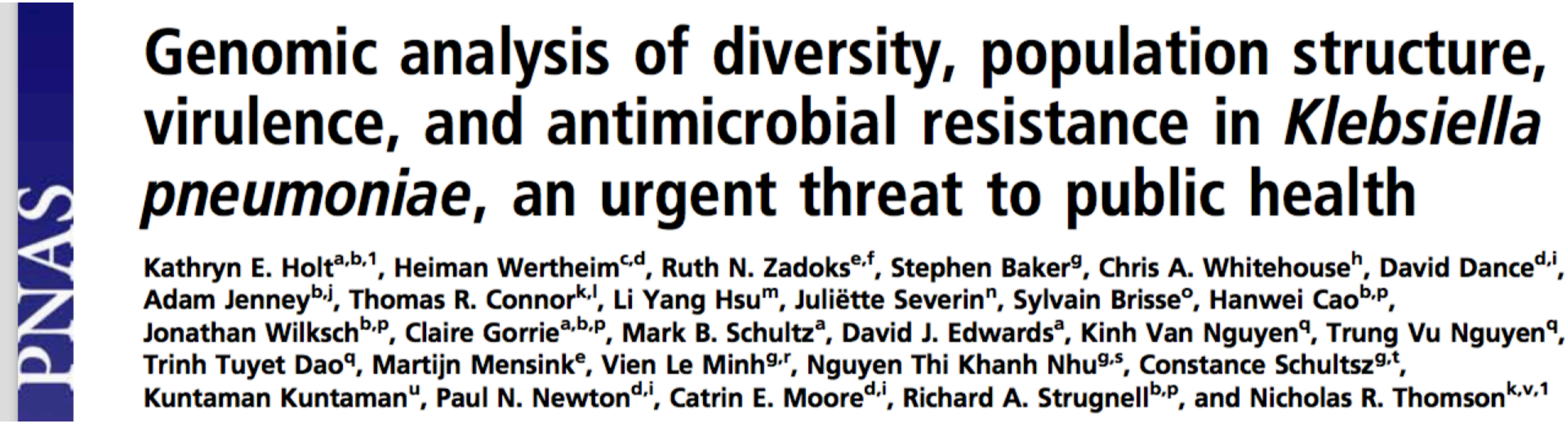


Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health

Kathryn E. Holt[a,b,1], Heiman Wertheim[c,d], Ruth N. Zadoks[e,f], Stephen Baker[g], Chris A. Whitehouse[h], David Dance[d,i], Adam Jenney[b,j], Thomas R. Connor[k,l], Li Yang Hsu[m], Juliëtte Severin[n], Sylvain Brisse[o], Hanwei Cao[b,p], Jonathan Wilksch[b,p], Claire Gorrie[a,b,p], Mark B. Schultz[a], David J. Edwards[a], Kinh Van Nguyen[q], Trung Vu Nguyen[q], Trinh Tuyet Dao[q], Martijn Mensink[e], Vien Le Minh[g,r], Nguyen Thi Khanh Nhu[g,s], Constance Schultsz[g,t], Kuntaman Kuntaman[u], Paul N. Newton[d,i], Catrin E. Moore[d,i], Richard A. Strugnell[b,p], and Nicholas R. Thomson[k,v,1]

Journal Club

14.09.2016

Maido Remm

Eesmärk

- *Klebsiella pneumoniae* is rapidly becoming untreatable using last-line antibiotics. It is especially problematic in hospitals, where it causes a range of acute infections.

- **To approach controlling such a bacterium, we first must define what it is and how it varies genetically.**

Mida tehti

- Sekveneeriti 288 erinevat *K. pneumoniae* isolaati (tüve), mis pärinesid 4 kontinendilt. Lisaks 40 tüve NCBI andmebaasist

- Proove võeti nii haigetelt kui tervetelt inimestelt, lisaks ka lehmadelt ja keskkonnast.

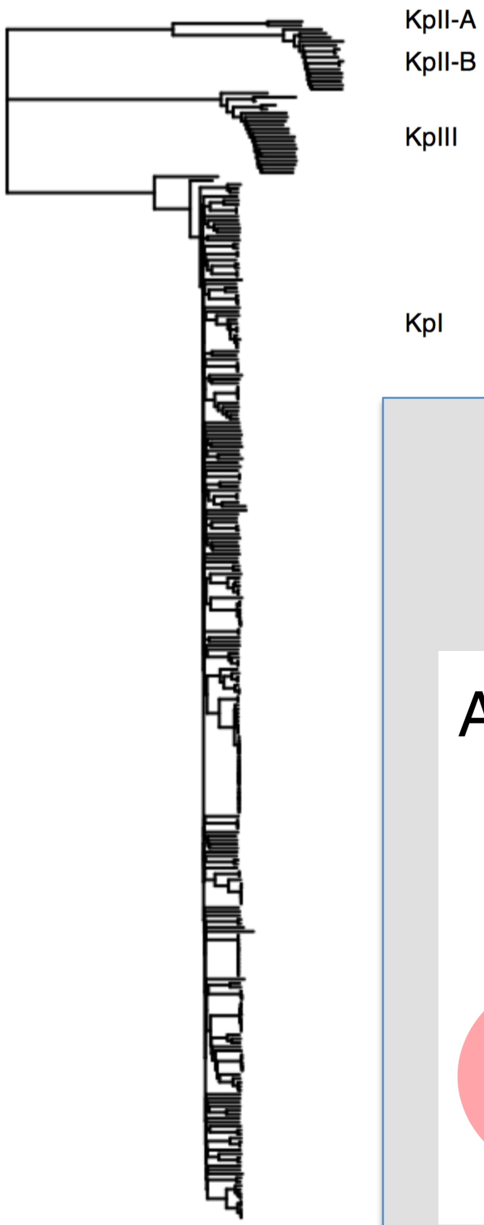
- Analüüsid:
  - Populatsiooni fülogeneetiline struktuur
  - Geenide arv genoomides
  - Geenide ja virulentsuse seoste leidmine

DNA eraldamine ja sekveneerimine:

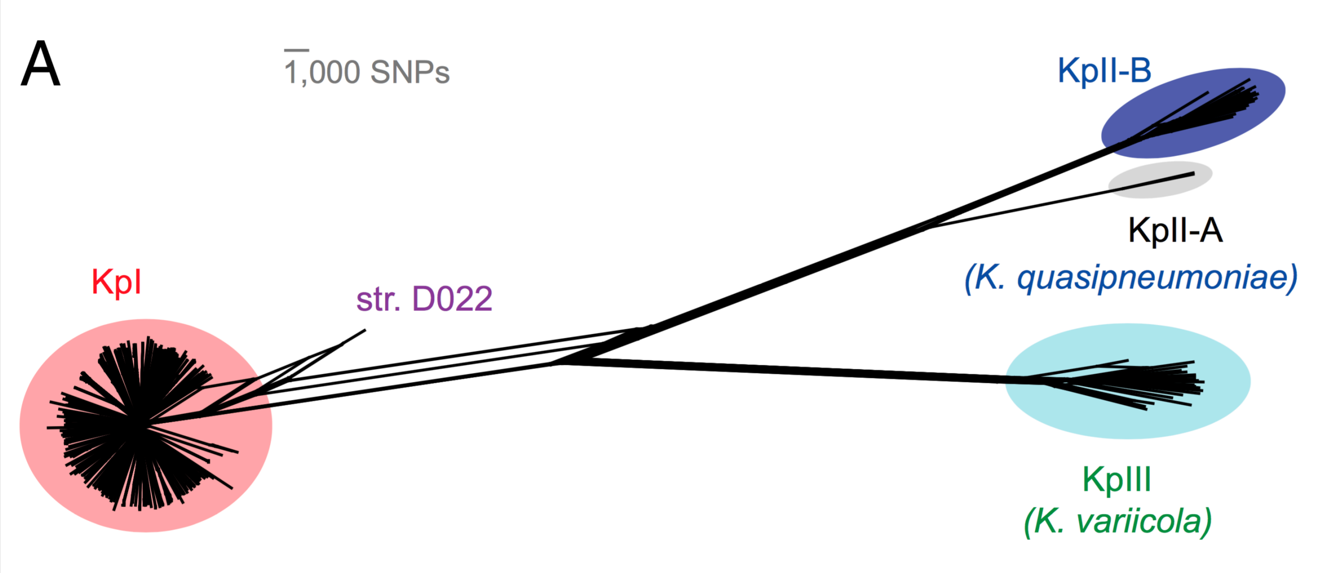
- Genomic DNA was extracted at each site using either
  phenol-chloroform extraction,
  QIAmp (Qiagen),
  High Pure or
  MagNA Pure (Roche).

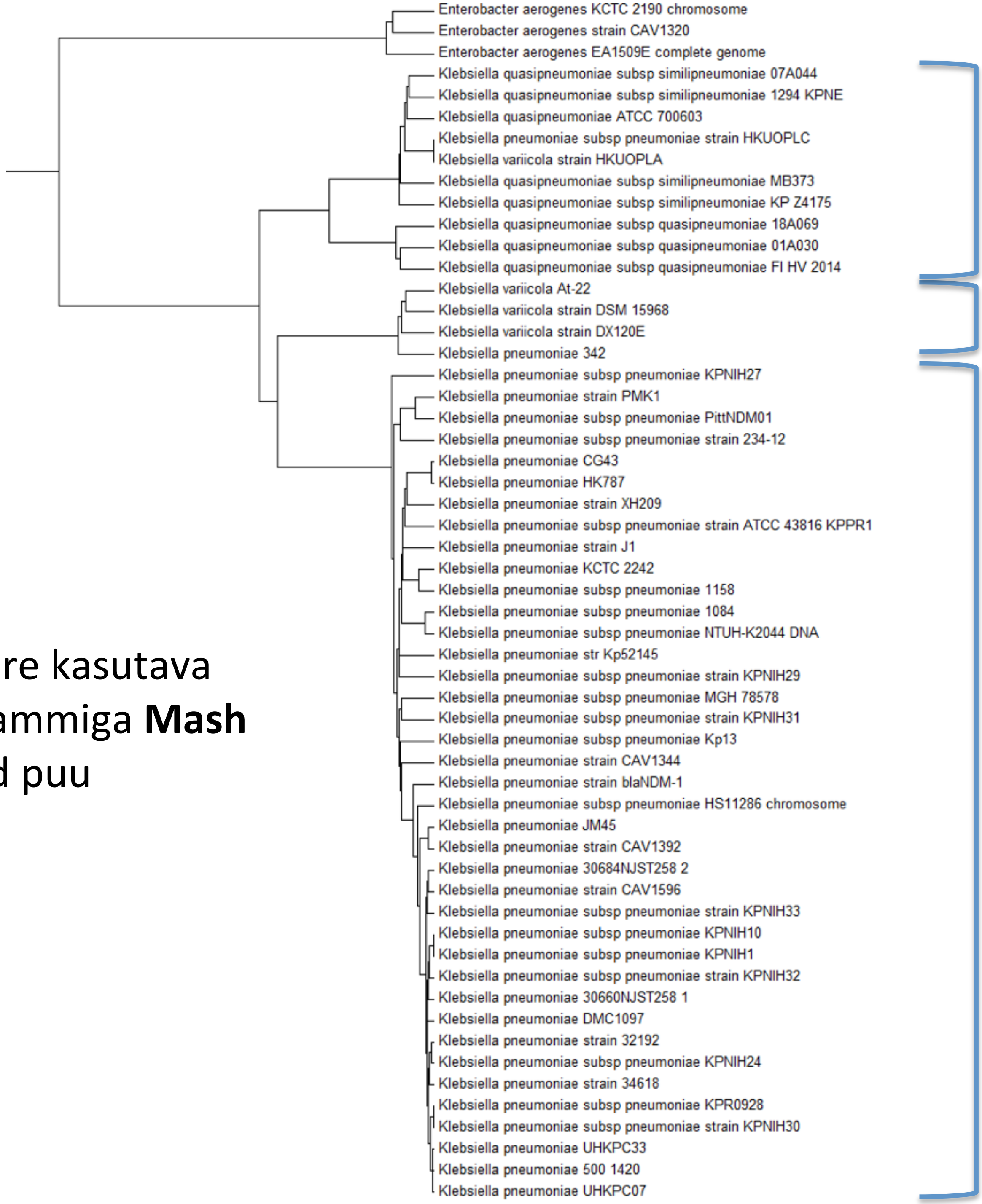
- Index-tagged paired end Illumina sequencing libraries were prepared using one of 12 unique indexing tags (9), combined into pools of uniquely tagged libraries and sequenced on the Illumina Genome Analyzer GAII at the Wellcome Trust Sanger Institute to generate tagged
  76 bp paired-end reads.

Suurima tõepära (ML) sugupuu



Split-network sugupuu




Enterobacter aerogenes KCTC 2190 chromosome
Enterobacter aerogenes strain CAV1320
Enterobacter aerogenes EA1509E complete genome
Klebsiella quasipneumoniae subsp similipneumoniae 07A044
Klebsiella quasipneumoniae subsp similipneumoniae 1294 KPNE
Klebsiella quasipneumoniae ATCC 700603
Klebsiella pneumoniae subsp pneumoniae strain HKUOPLC
Klebsiella variicola strain HKUOPLA
Klebsiella quasipneumoniae subsp similipneumoniae MB373
Klebsiella quasipneumoniae subsp similipneumoniae KP Z4175
Klebsiella quasipneumoniae subsp quasipneumoniae 18A069
Klebsiella quasipneumoniae subsp quasipneumoniae 01A030
Klebsiella quasipneumoniae subsp quasipneumoniae FI HV 2014
Klebsiella variicola At-22
Klebsiella variicola strain DSM 15968
Klebsiella variicola strain DX120E
Klebsiella pneumoniae 342
Klebsiella pneumoniae subsp pneumoniae KPNIH27
Klebsiella pneumoniae strain PMK1
Klebsiella pneumoniae subsp pneumoniae PittNDM01
Klebsiella pneumoniae subsp pneumoniae strain 234-12
Klebsiella pneumoniae CG43
Klebsiella pneumoniae HK787
Klebsiella pneumoniae strain XH209
Klebsiella pneumoniae subsp pneumoniae strain ATCC 43816 KPPR1
Klebsiella pneumoniae strain J1
Klebsiella pneumoniae KCTC 2242
Klebsiella pneumoniae subsp pneumoniae 1158
Klebsiella pneumoniae subsp pneumoniae 1084
Klebsiella pneumoniae subsp pneumoniae NTUH-K2044 DNA
Klebsiella pneumoniae str Kp52145
Klebsiella pneumoniae subsp pneumoniae strain KPNIH29
Klebsiella pneumoniae subsp pneumoniae MGH 78578
Klebsiella pneumoniae subsp pneumoniae strain KPNIH31
Klebsiella pneumoniae subsp pneumoniae Kp13
Klebsiella pneumoniae strain CAV1344
Klebsiella pneumoniae strain blaNDM-1
Klebsiella pneumoniae subsp pneumoniae HS11286 chromosome
Klebsiella pneumoniae JM45
Klebsiella pneumoniae strain CAV1392
Klebsiella pneumoniae 30684NJST258 2
Klebsiella pneumoniae strain CAV1596
Klebsiella pneumoniae subsp pneumoniae strain KPNIH33
Klebsiella pneumoniae subsp pneumoniae KPNIH10
Klebsiella pneumoniae subsp pneumoniae KPNIH1
Klebsiella pneumoniae subsp pneumoniae strain KPNIH32
Klebsiella pneumoniae 30660NJST258 1
Klebsiella pneumoniae DMC1097
Klebsiella pneumoniae strain 32192
Klebsiella pneumoniae subsp pneumoniae KPNIH24
Klebsiella pneumoniae strain 34618
Klebsiella pneumoniae subsp pneumoniae KPR0928
Klebsiella pneumoniae subsp pneumoniae strain KPNIH30
Klebsiella pneumoniae UHKPC33
Klebsiella pneumoniae 500 1420
Klebsiella pneumoniae UHKPC07

Kpn II

Klebsiella quasipneumoniae

Kpn III

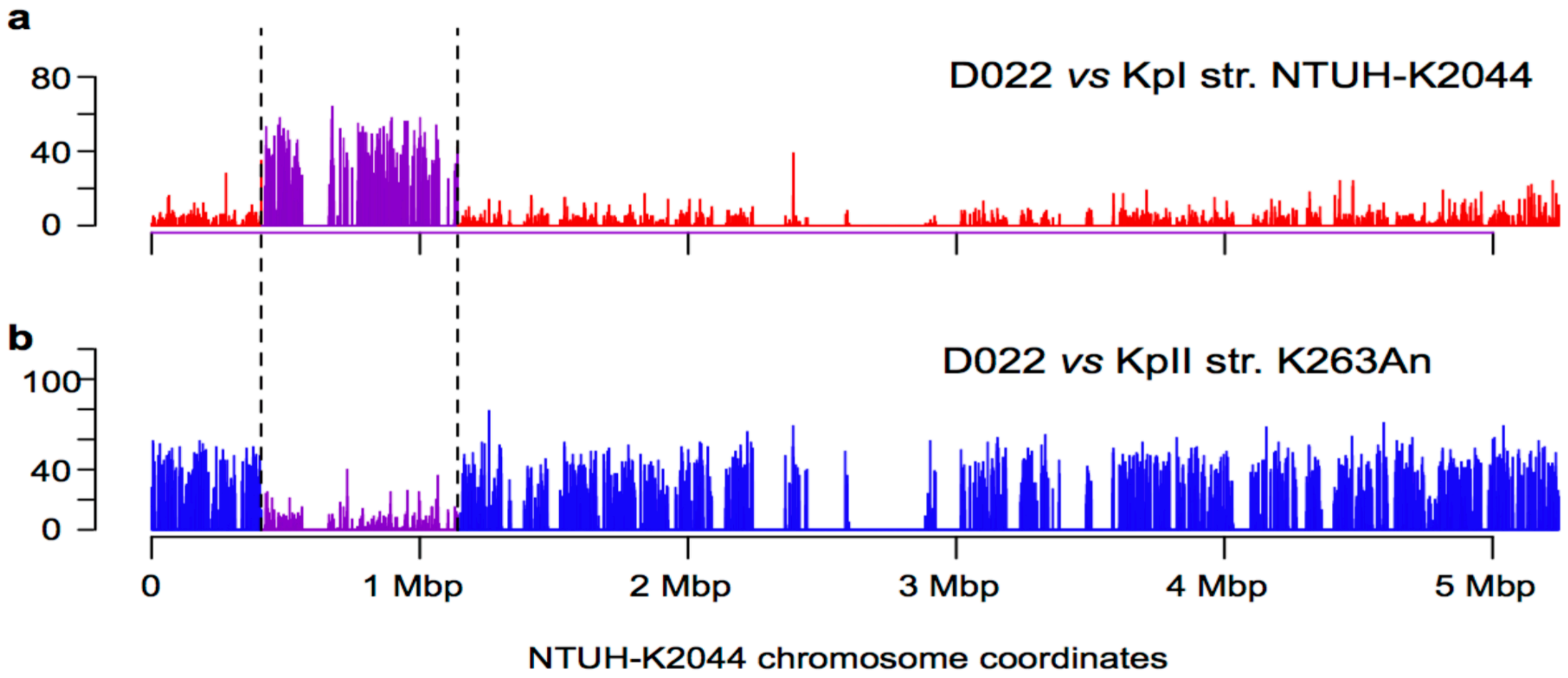
Klebsiella variicola

Kpn I

Klebsiella pneumoniae

k-meere kasutava programmiga **Mash** tehtud puu

Kas need 3 liiki ristuvad omavahel?

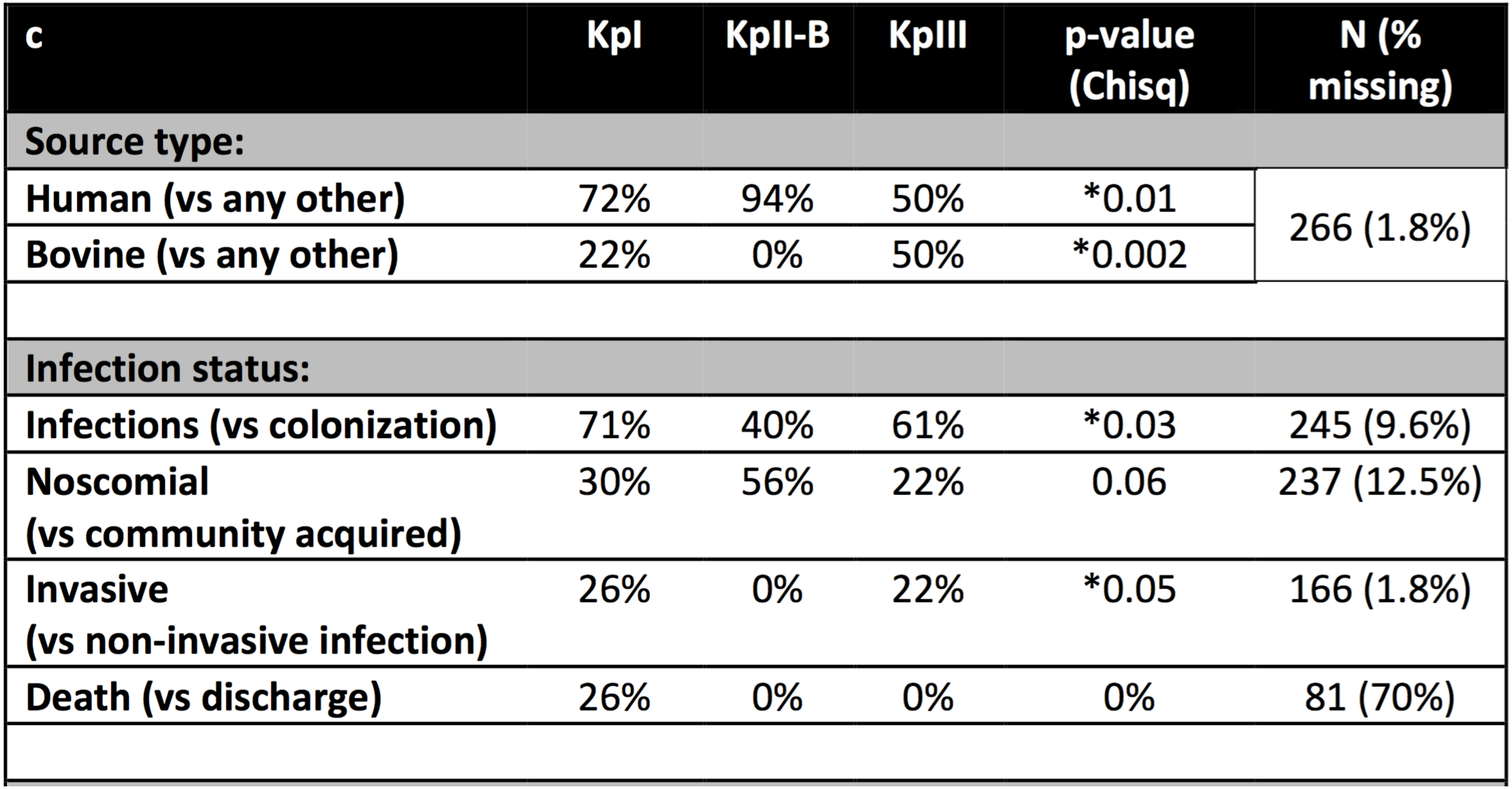


Density of SNPs per kilobase along the genome of Vietnamese human gut carriage isolate D022 compared to reference genomes from KpI (**a**) and KpII (**b**). The 735 kbp region of proposed homologous recombination between D022 and a KpII-like genome is shown in purple and bounded by dashed lines.

=> There are no obvious mechanistic barriers to homologous recombination between KpI, KpII, and KpIII; indeed the observation of a large recombination between KpI and KpII shows that homologous recombination is possible, although the rarity of this event (1 out of >300 genomes) suggests there could be selection against such hybrids.

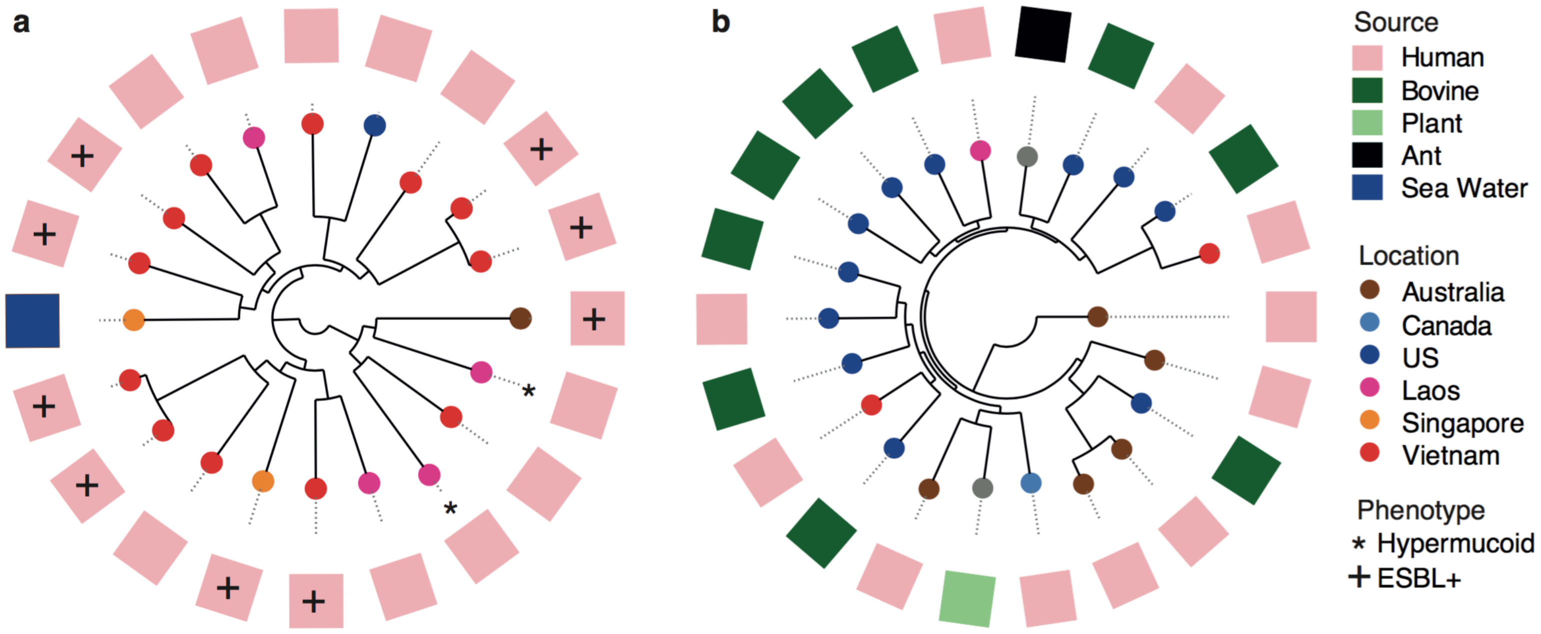
=> The speciation of KpII and KpIII is likely driven by long-term separation in distinct ecological niches.

Klebsiella liigid on erinevate omadustega



c	KpI	KpII-B	KpIII	p-value (Chisq)	N (% missing)
Source type:					
Human (vs any other)	72%	94%	50%	*0.01	266 (1.8%)
Bovine (vs any other)	22%	0%	50%	*0.002	
Infection status:					
Infections (vs colonization)	71%	40%	61%	*0.03	245 (9.6%)
Noscomial (vs community acquired)	30%	56%	22%	0.06	237 (12.5%)
Invasive (vs non-invasive infection)	26%	0%	22%	*0.05	166 (1.8%)
Death (vs discharge)	26%	0%	0%	0%	81 (70%)

Klebsiella liigid on erinevate omadustega



KpII-B

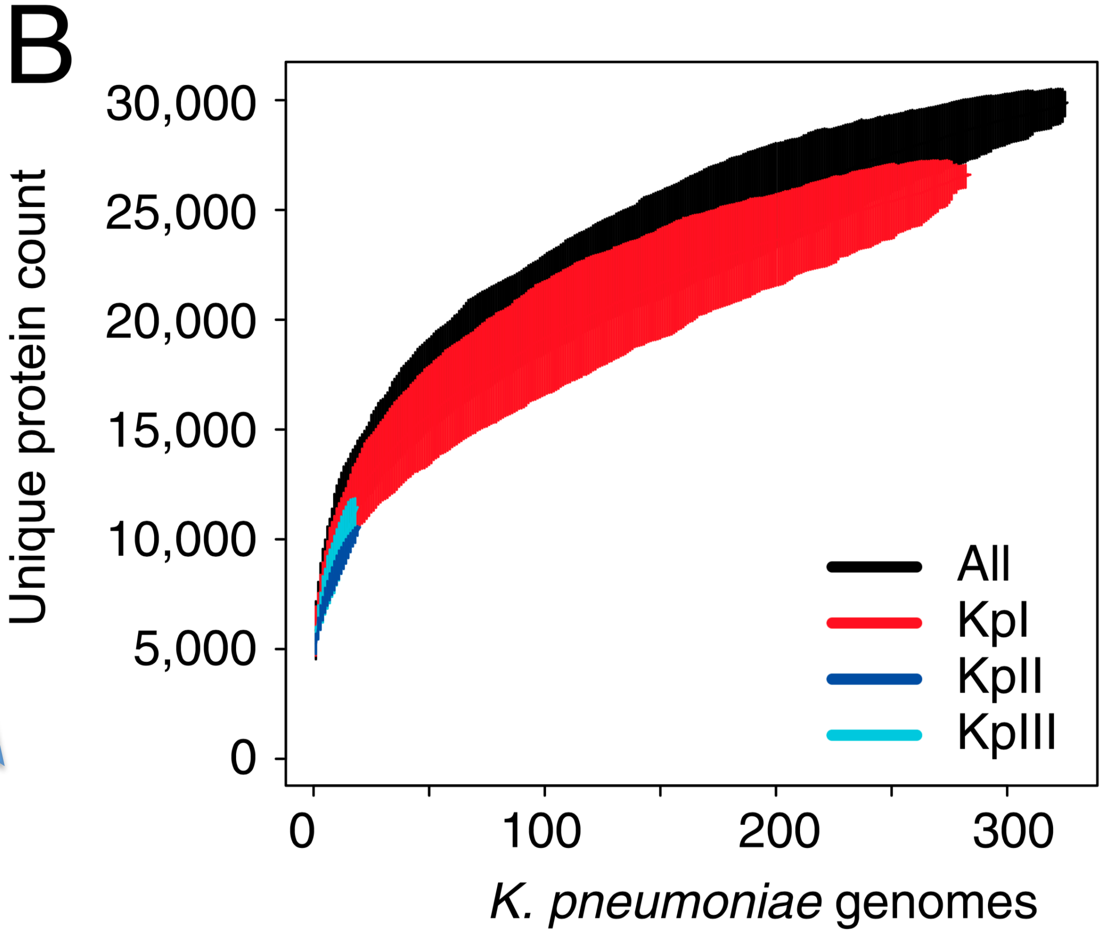
50% Infection
- 39% Invasive
- 50% Nosocomial
- 0 Deaths

KpIII
(*K. variicola*)

63% Infection
- 40% Invasive
- 42% Nosocomial
- 0 Deaths

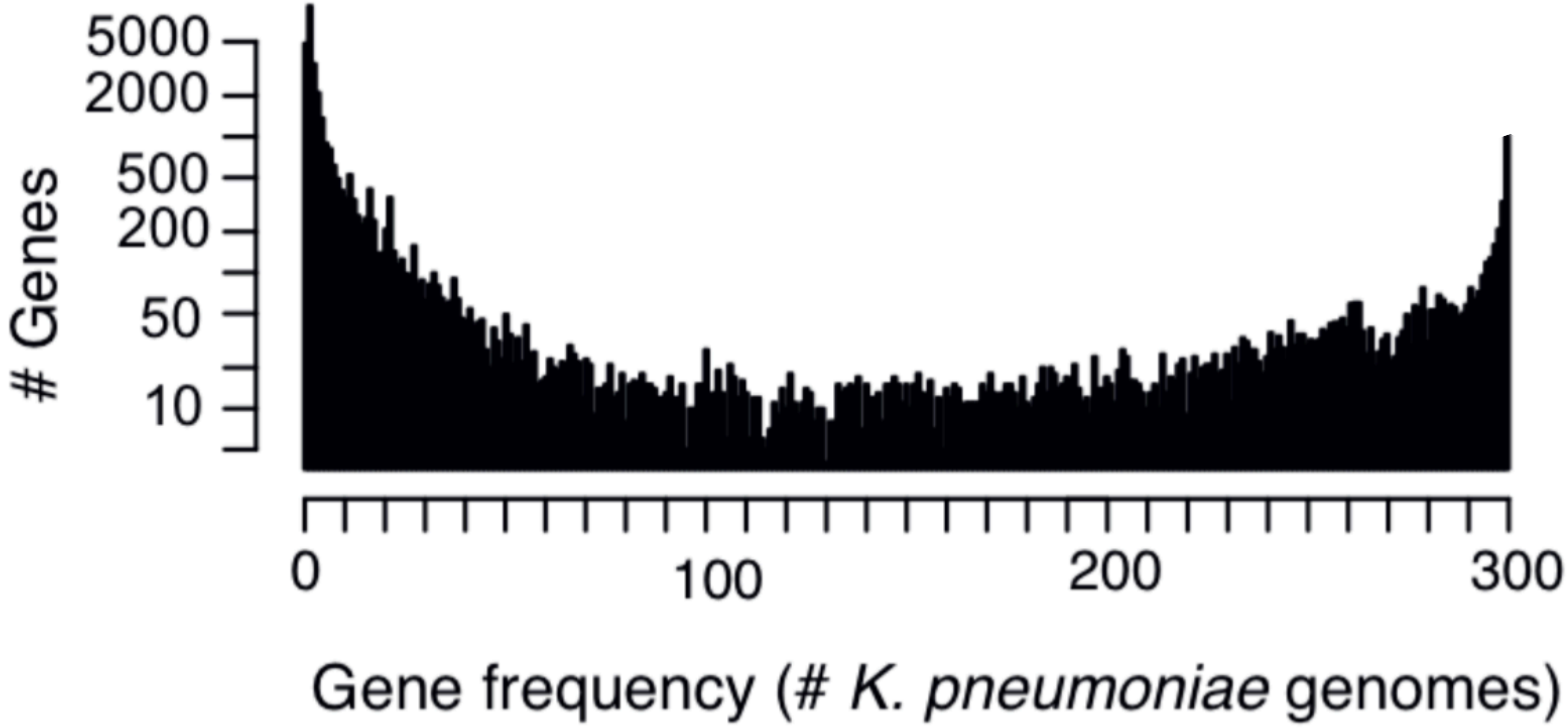
KpII (K. quasipneumoniae) and KpIII (K. variicola)

Pan-genoom



Unique protein -
clustered at the ≤30% amino acid
homology level using CD-HIT

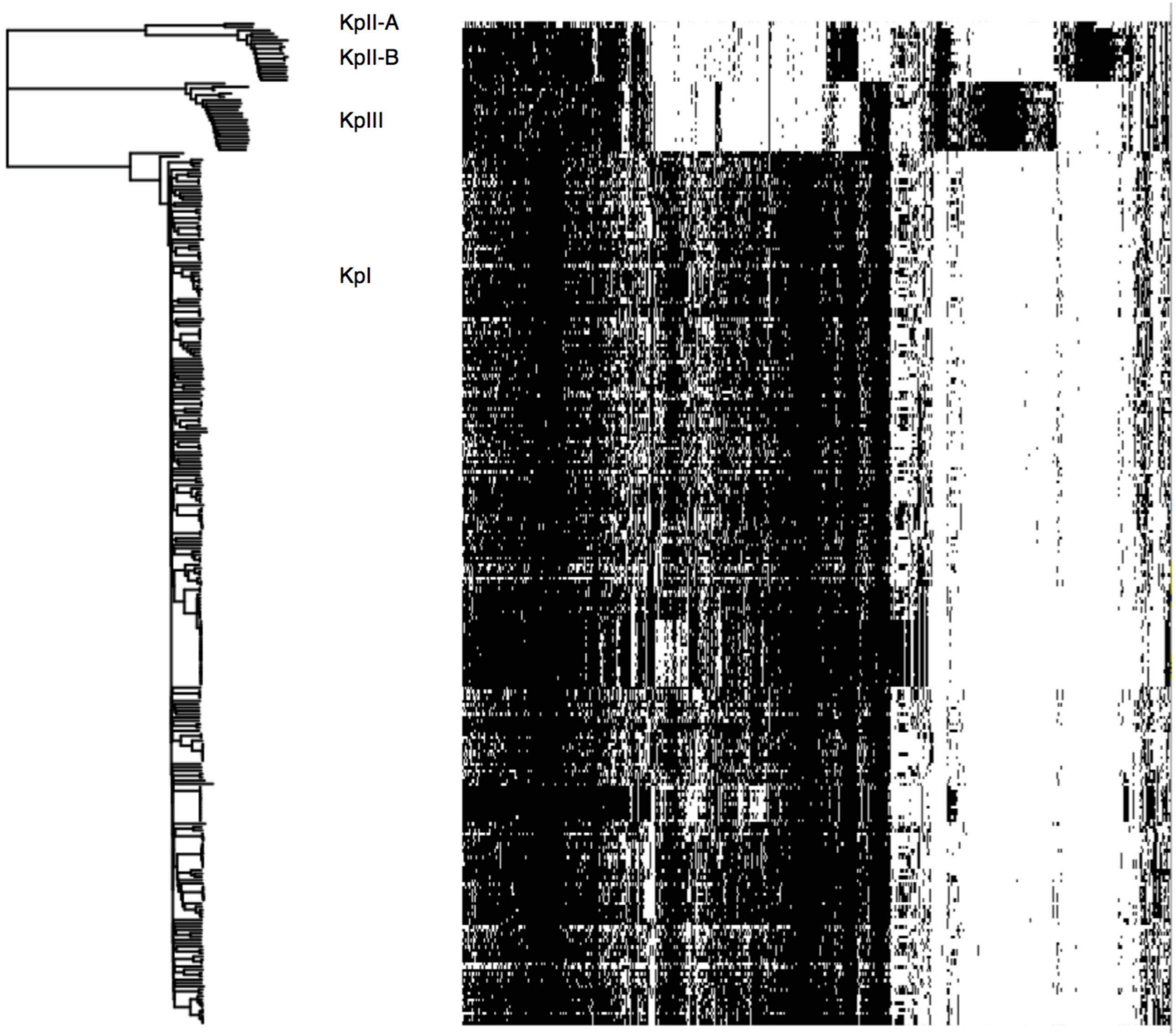
Geenide esinemissagedus eri tüvede genoomides



Pan-genome: Kokku 29,886 erinevat geeni
ca 5000 geeni esineb vaid ühes tüves
1743 geeni esineb kõigis tüvedes
Accessory genome: 17 700 geeni esineb >5% ja <95% tüvedes

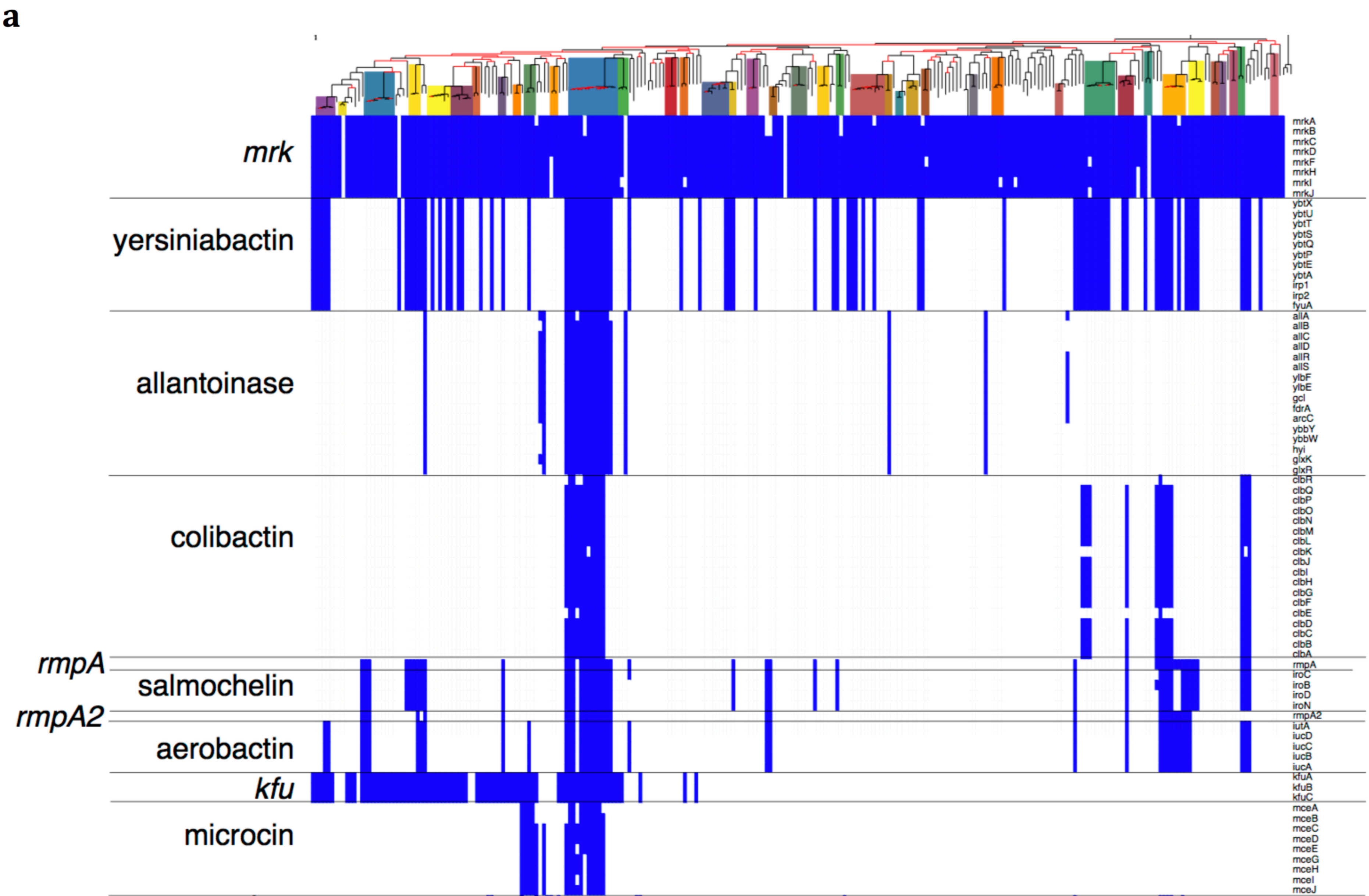
Geenide jaotumine sugupuu järgi

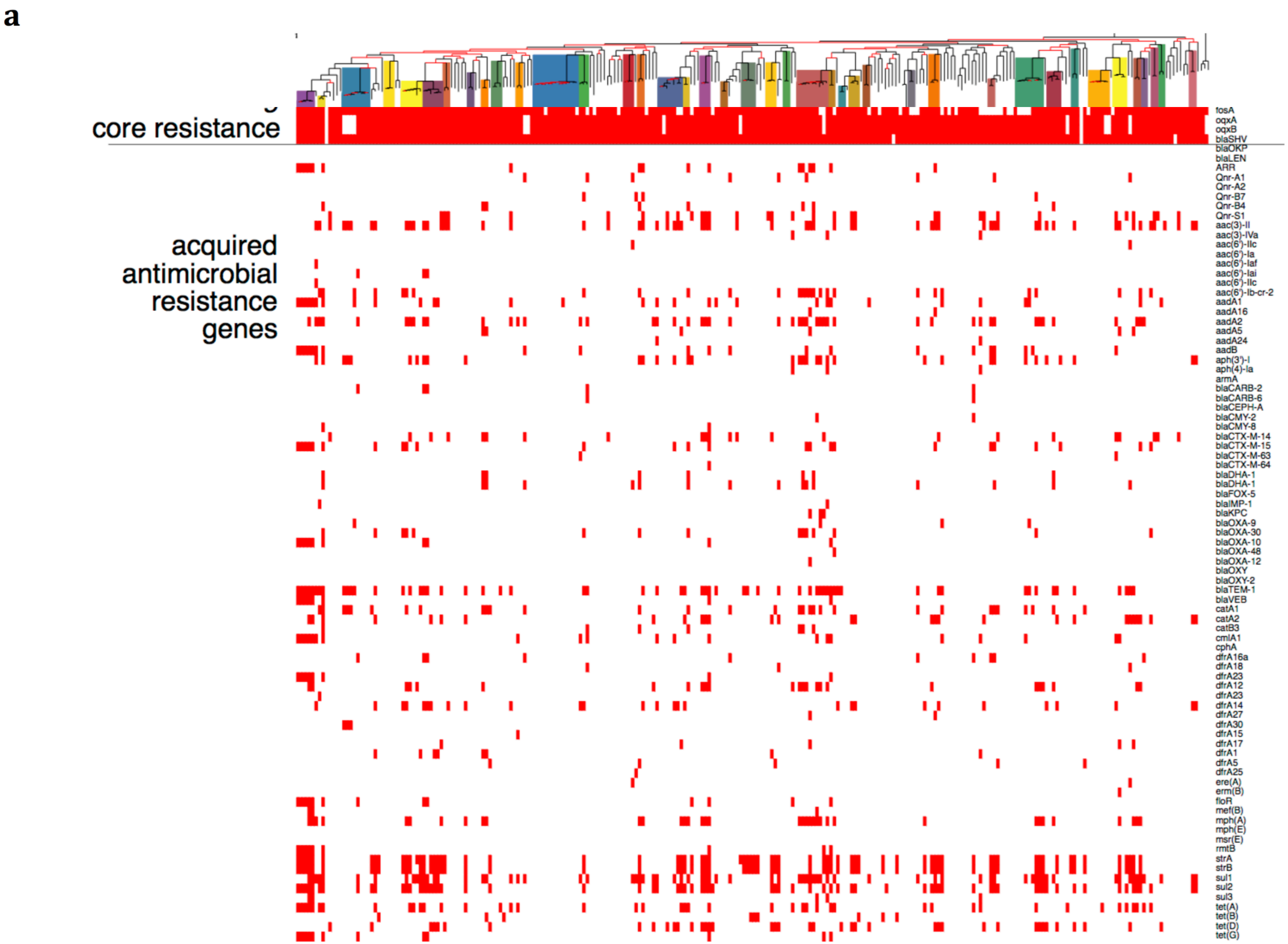
Figure S2 – Accessory genes in *K. pneumoniae*



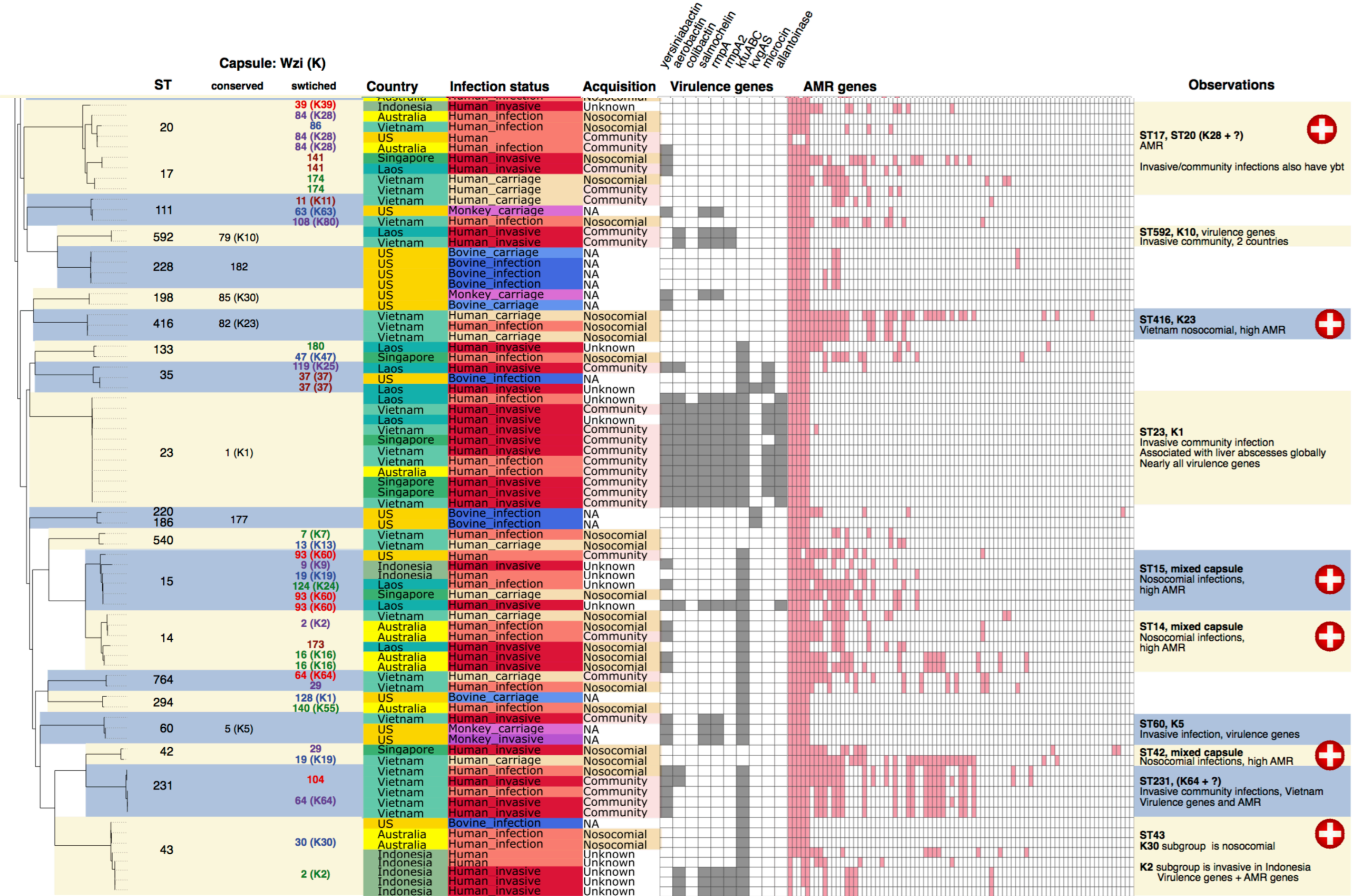
Maximum likelihood tree of 328 *K. pneumoniae* isolates (left), with heatmap indicating presence (black) or absence (white) of 7,771 accessory genes that are present in 5%-95% of genomes (right).

Virulentsusgeenide jaotumine tüvedes




a
core resistance
acquired antimicrobial resistance genes
fosA
oqxA
oqxB
blaSHV
blaOKP
blaLEN
ARR
Qnr-A1
Qnr-A2
Qnr-B7
Qnr-B4
Qnr-S1
aac(3)-II
aac(3)-IVa
aac(6')-IIc
aac(6')-Ia
aac(6')-Iaf
aac(6')-Iai
aac(6')-IIc
aac(6')-Ib-cr-2
aadA1
aadA16
aadA2
aadA5
aadA24
aadB
aph(3')-I
aph(4)-Ia
armA
blaCARB-2
blaCARB-6
blaCEPH-A
blaCMY-2
blaCMY-8
blaCTX-M-14
blaCTX-M-15
blaCTX-M-63
blaCTX-M-64
blaDHA-1
blaDHA-1
blaFOX-5
blaIMP-1
blaKPC
blaOXA-9
blaOXA-30
blaOXA-10
blaOXA-48
blaOXA-12
blaOXY
blaOXY-2
blaTEM-1
blaVEB
catA1
catA2
catB3
cmlA1
cphA
dfrA16a
dfrA18
dfrA23
dfrA12
dfrA23
dfrA14
dfrA27
dfrA30
dfrA15
dfrA17
dfrA1
dfrA5
dfrA25
ere(A)
erm(B)
floR
mef(B)
mph(A)
mph(E)
msr(E)
rmtB
strA
strB
sul1
sul2
sul3
tet(A)
tet(B)
tet(D)
tet(G)

Kogu info ühes tabelis:



ST	Capsule: Wzi (K) conserved	Capsule: Wzi (K) swtiched	Country	Infection status	Acquisition
			Australia	Human_infection	Nosocomial
		39 (K39)	Indonesia	Human_invasive	Unknown
		84 (K28)	Australia	Human_infection	Nosocomial
20		86	Vietnam	Human_infection	Nosocomial
		84 (K28)	US	Human	Community
		84 (K28)	Australia	Human_infection	Community
		141	Singapore	Human_invasive	Nosocomial
		141	Laos	Human_invasive	Community
17		174	Vietnam	Human_carriage	Nosocomial
		174	Vietnam	Human_carriage	Community
111		11 (K11)	Vietnam	Human_carriage	Community
		63 (K63)	US	Monkey_carriage	NA
		108 (K80)	Vietnam	Human_infection	Nosocomial
592	79 (K10)		Laos	Human_invasive	Community
			Vietnam	Human_invasive	Community
228	182		US	Bovine_carriage	NA
			US	Bovine_infection	NA
			US	Bovine_infection	NA
			US	Bovine_infection	NA
198	85 (K30)		US	Monkey_carriage	NA
			US	Bovine_carriage	NA
416	82 (K23)		Vietnam	Human_carriage	Nosocomial
			Vietnam	Human_infection	Nosocomial
			Vietnam	Human_carriage	Nosocomial
133		180	Laos	Human_invasive	Unknown
		47 (K47)	Singapore	Human_infection	Nosocomial
35		119 (K25)	Laos	Human_invasive	Community
		37 (37)	US	Bovine_infection	NA
		37 (37)	Laos	Human_invasive	Unknown
			Laos	Human_infection	Unknown
			Vietnam	Human_invasive	Community
			Laos	Human_invasive	Unknown
			Vietnam	Human_invasive	Community
			Singapore	Human_invasive	Community
23	1 (K1)		Vietnam	Human_invasive	Community
			Vietnam	Human_infection	Community
			Australia	Human_infection	Community
			Singapore	Human_invasive	Community
			Singapore	Human_invasive	Community
			Vietnam	Human_invasive	Community
220			US	Bovine_infection	NA
186	177		US	Bovine_infection	NA
540		7 (K7)	Vietnam	Human_infection	Nosocomial
		13 (K13)	Vietnam	Human_carriage	Nosocomial
		93 (K60)	US	Human	Community
		9 (K9)	Indonesia	Human_invasive	Unknown
15		19 (K19)	Indonesia	Human	Unknown
		124 (K24)	Laos	Human_infection	Unknown
		93 (K60)	Singapore	Human_carriage	Nosocomial
		93 (K60)	Laos	Human_invasive	Unknown
			Vietnam	Human_carriage	Nosocomial
		2 (K2)	Australia	Human_infection	Nosocomial
14			Australia	Human_infection	Community
		173	Laos	Human_invasive	Nosocomial
		16 (K16)	Australia	Human_invasive	Nosocomial
		16 (K16)	Australia	Human_invasive	Nosocomial
764		64 (K64)	Vietnam	Human_carriage	Community
		29	Vietnam	Human_infection	Nosocomial
294		128 (K1)	US	Bovine_carriage	NA
		140 (K55)	Australia	Human_infection	Nosocomial
60	5 (K5)		Vietnam	Human_invasive	Community
			US	Monkey_carriage	NA
			US	Monkey_invasive	NA
42		29	Singapore	Human_invasive	Nosocomial
		19 (K19)	Vietnam	Human_carriage	Nosocomial
			Vietnam	Human_infection	Nosocomial
231		104	Vietnam	Human_invasive	Community
			Vietnam	Human_infection	Community
		64 (K64)	Vietnam	Human_invasive	Community
			Vietnam	Human_invasive	Community
			US	Bovine_infection	NA
			Australia	Human_infection	Nosocomial
43		30 (K30)	Australia	Human_infection	Nosocomial
			Indonesia	Human	Unknown
			Indonesia	Human	Unknown
			Indonesia	Human_invasive	Unknown
		2 (K2)	Indonesia	Human_invasive	Unknown
			Indonesia	Human_invasive	Unknown

Observations

ST17, ST20 (K28 + ?)
AMR

Invasive/community infections also have ybt

ST592, K10, virulence genes
Invasive community, 2 countries

ST416, K23
Vietnam nosocomial, high AMR

ST23, K1
Invasive community infection
Associated with liver abscesses globally
Nearly all virulence genes

ST15, mixed capsule
Nosocomial infections,
high AMR

ST14, mixed capsule
Nosocomial infections,
high AMR

ST60, K5
Invasive infection, virulence genes

ST42, mixed capsule
Nosocomial infections, high AMR

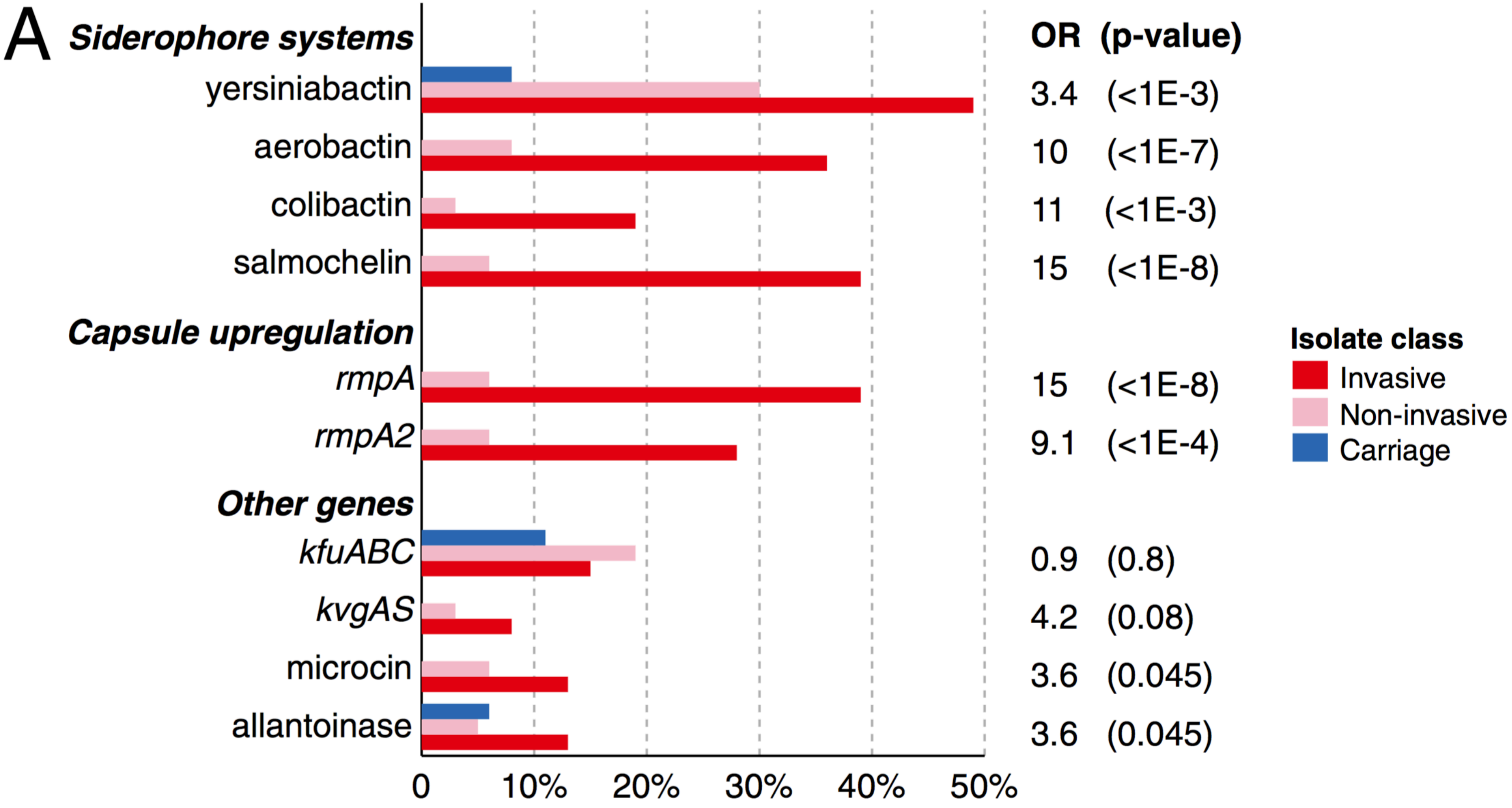
ST231, (K64 + ?)
Invasive community infections, Vietnam
Virulence genes and AMR

ST43
K30 subgroup is nosocomial

K2 subgroup is invasive in Indonesia
Virulence genes + AMR genes

Kas virulentsusgeenide suhteline sagedus on invasiivse fenotüübiga tüvedes erinev?

Odd ratio invasive vs others:

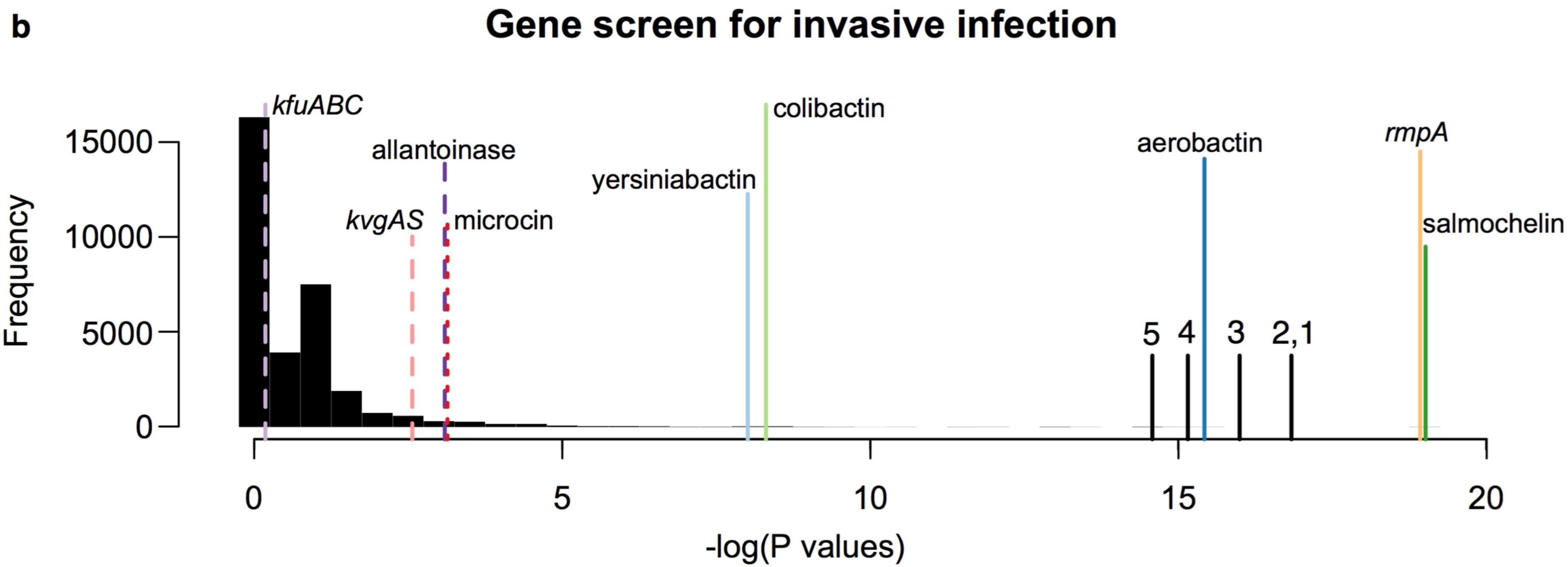


	OR	(p-value)
yersiniabactin	3.4	(<1E-3)
aerobactin	10	(<1E-7)
colibactin	11	(<1E-3)
salmochelin	15	(<1E-8)
rmpA	15	(<1E-8)
rmpA2	9.1	(<1E-4)
kfuABC	0.9	(0.8)
kvgAS	4.2	(0.08)
microcin	3.6	(0.045)
allantoinase	3.6	(0.045)

Invasiivne on tüvi, mis on isoleeritud kudedest, mis on tavaliselt steriilsed

PGWAS

- To extend these observations, we performed a PGWAS, screening each gene in the KpI pangenome for **association with infection in humans**. The strongest associations, reaching pan-genome-wide significance after correcting for multiple testing, were the rmpA/2 and siderophore genes and **five additional predicted iron-metabolism genes** (OR 8–10; 95% CI 3–35), also present on the virulence plasmid pK2044.



Each gene was screened for association with (**a**) infection (vs. carriage) and (**b**) invasive infection (vs. non-invasive infection or carriage), amongst human KpI isolates, using Fisher's exact test. Black bars indicate the distribution of the resulting p-values (note P-values on the x-axis are shown on the natural log scale). P-values for known virulence gene clusters are indicated with labelled vertical bars, coloured as in Figure 4; solid lines indicate significant associations ($P \leq 0.05$, i.e. $-\log(P) \geq 3$); dashed lines indicate non-significant associations ($P > 0.05$, i.e. $-\log(P) < 3$). P-values for novel virulence genes are indicated with black lines: 1, pK2044_00025 (FepB domain); 2, pK2044_00325 (CobW domain); 3, pK2044_00355 (Fur domain); 4, pK2044_00335 (FeoB domain); 5, pK2044_00030 (FepC domain).

Fülogeneetilise tausta võimaliku mõju taandamiseks kasutati logistilist regressiooni

- **Yersiniabactin**, whose synthesis is encoded by the ybt, irp1, irp2, and fyuA genes, was the most prevalent virulence-associated locus, present in one third of the KpI human isolates.

- Despite this high prevalence it was a strong predictor of infection vs. carriage in humans, with an OR of 7.4 [95% confidence interval (CI), 2.2–40; P = 0.0001; Fisher's exact test] and a positive predictive value of 95%.

- **This effect was not dependent on chromosomal background, because yersiniabactin was significantly associated with infection in a logistic regression model that included phylogenetic lineage (OR 1.3; P = 0.003).**

- **Pangenome analysis**
- Illumina reads were assembled using the *de novo* short read assembler Velvet and Velvet Optimiser (19). Publicly available finished or draft genomes were also included. Contigs less than 100 bp in size were excluded from further analysis. We then used an iterative mapping approach as described previously (20), to generate a pangenome representing the non-redundant set of all **>5% divergent sequences of length ≥100 bp** among the 328 *K. pneumoniae* (using the nucmer algorithm in MUMmer).

- Open reading frames (ORFs) were identified and annotated using Prokka (23) with primary reference to NTUH-K2044 protein sequences. **A total of 84,175 ORFs were identified**, which were unique at the 5% DNA homology level. All *K. pneumoniae* read sets were aligned to this pangenome sequence.

- The ORFs were translated into protein using EMBOSS and **clustered at the ≤30% amino acid homology level** using CD-HIT (24), resulting in **29,886 protein/gene clusters**.

- Alternative clustering at ≤10%, ≤20% or ≤40% homology resulted in 46718, 37609 or 29779 clusters, respectively; hence 30% was taken as the point of inflection.

- These data were then processed to generate a binary gene content matrix in which the presence of a gene is defined as >90% coverage of at least one ORF belonging to the corresponding protein cluster. Hence genomes that encode alleles with ≥30% amino acid homology across the length of the sequence are considered to encode the same functional gene in our analysis.