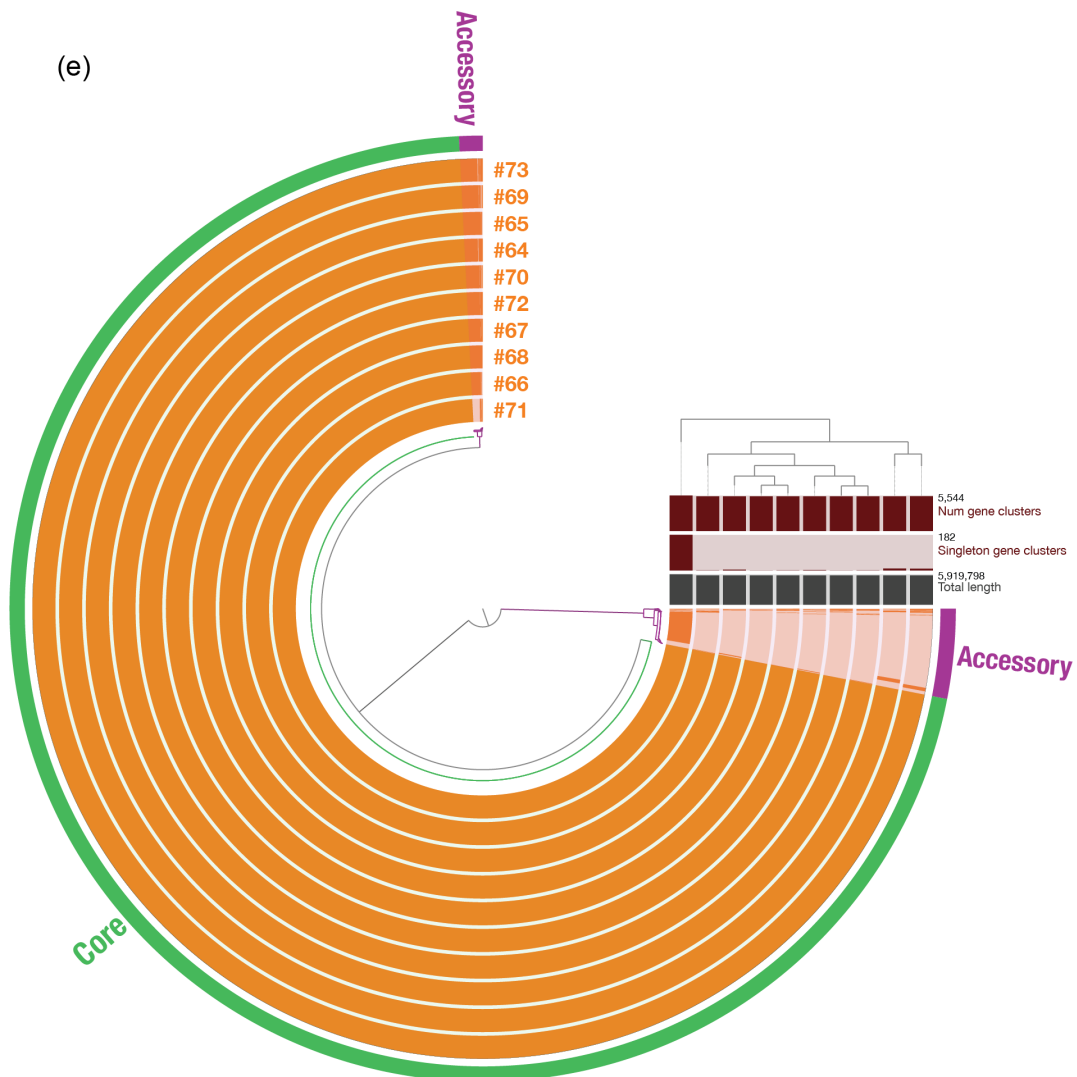
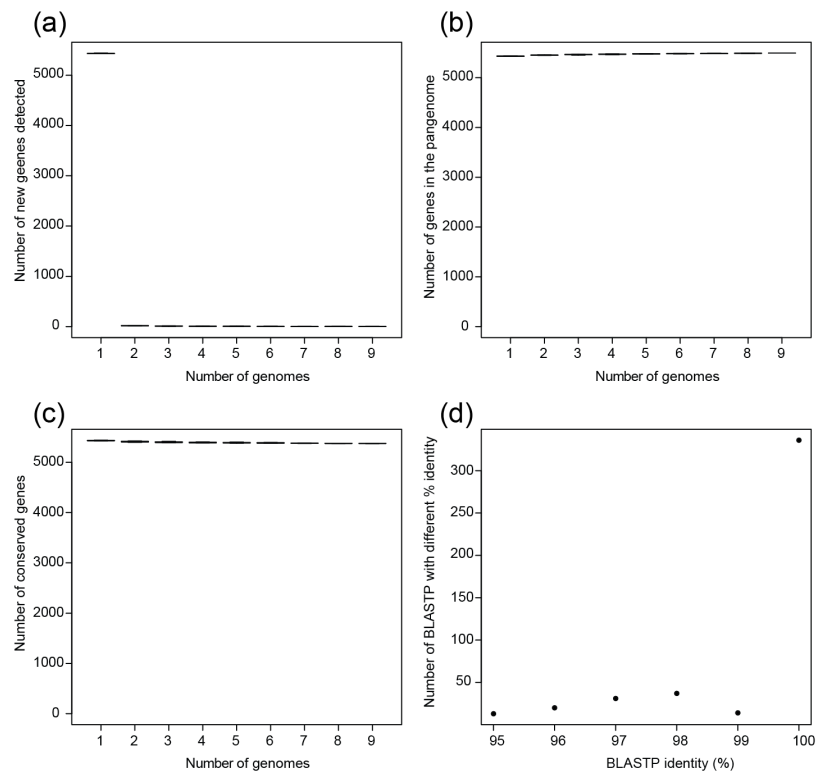
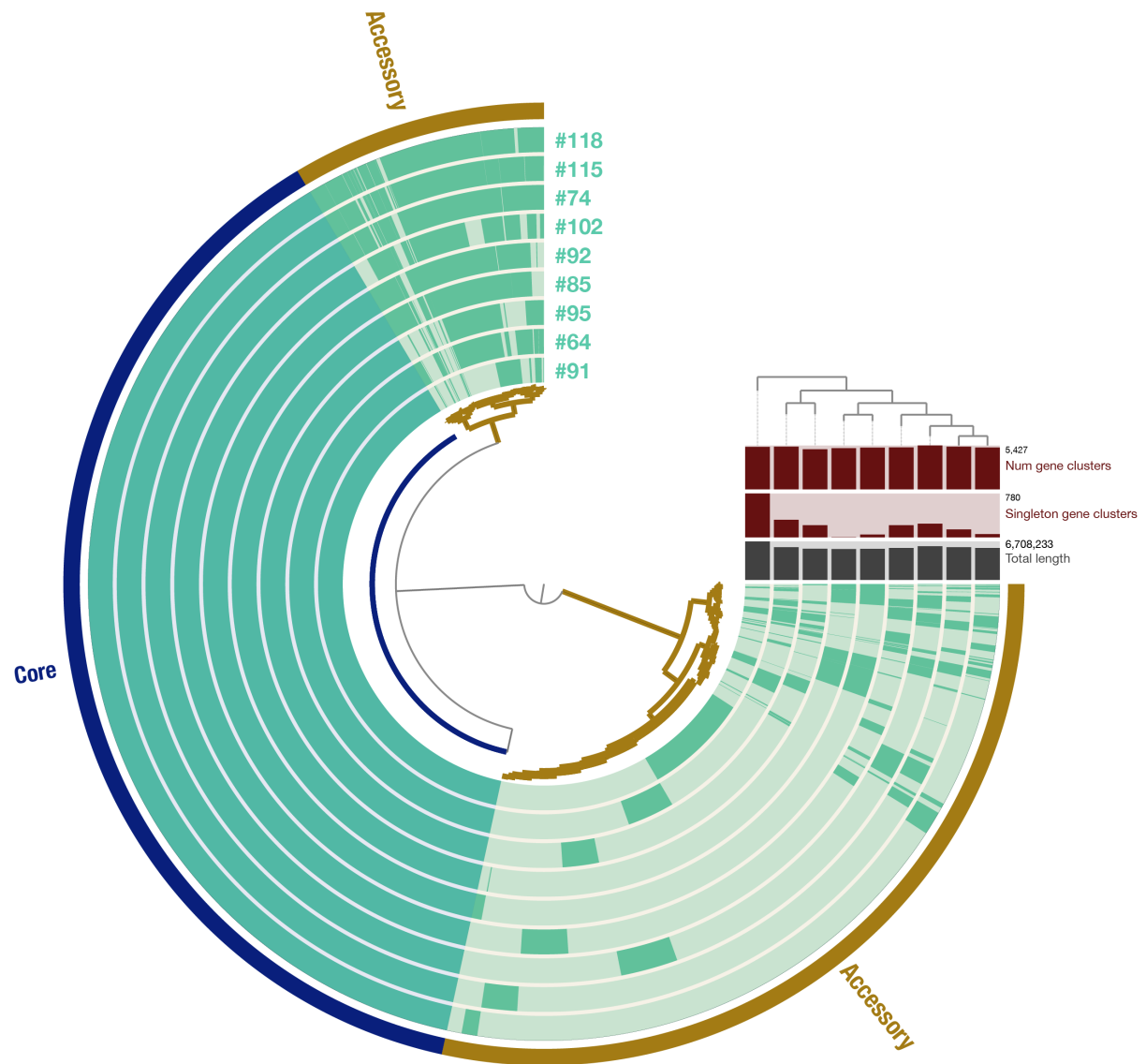


Supplementary Figures associated with Chen *et al.* – Preterm infants harbour diverse *Klebsiella* populations, including atypical species that encode and produce an array of antimicrobial resistance- and virulence-associated factors



Supplementary Figure 1. Genomic characterization of nine *K. pneumoniae* isolates recovered from neonate P008. (a) Change in number of new genes detected as the number of genomes increased. The number of new genes detected fell to almost 0 when a second genomes joined the first, and stayed at around 0 as further genomes were added to the analyses. (b, c) Change in total number of genes and conserved genes in the pangenome. As the number of genomes increased, the number of total genes (b) and conserved genes (c) is almost stable, with only negligible changes observed. (d) Number of BLASTP hits at different percentage identity. Identities of all hits are over 95 %, with the majority of them 100 % identical, providing more evidence to show the isolates from neonate P008 share the same genetic content. (e) Anvi'o representation of genomes of the isolates recovered from neonate P008. Differences between the isolates are due to the presence of single gene clusters in genomes, considered to be due to differences in genome coverage across the isolates.



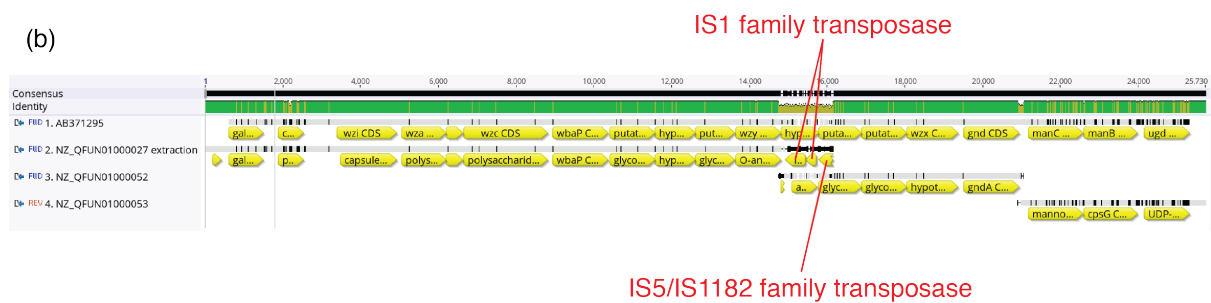
Supplementary Figure 2. Anvi'o representation of genomes of the *K. pneumoniae* isolates recovered from different infants. The genomes are clearly different from one another, indicating each isolate represents a different strain of *K. pneumoniae*.

(a)

Result



(b)

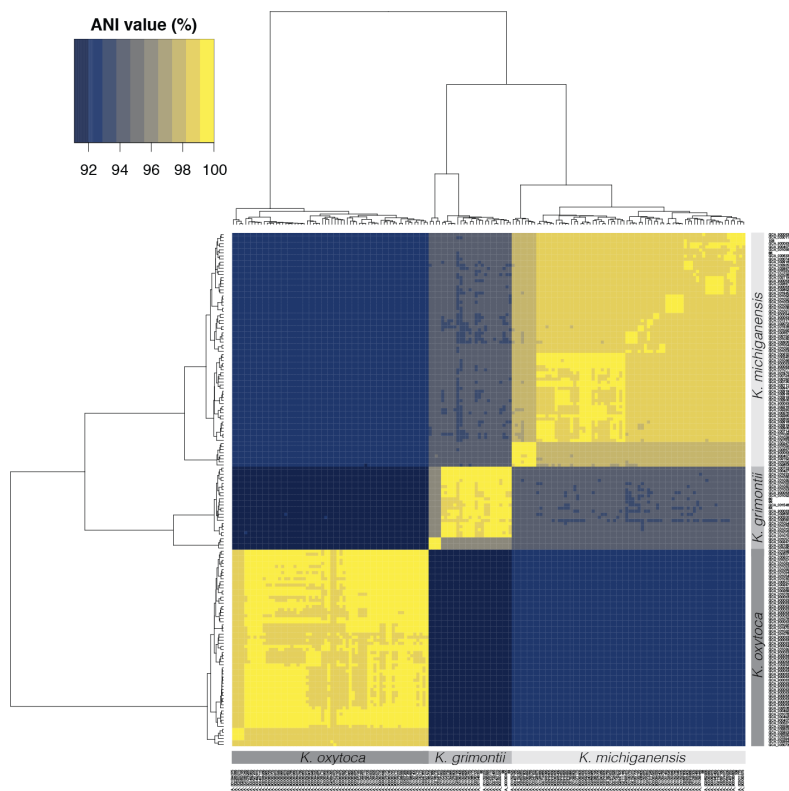


Supplementary Figure 3. Isolate #74 does not have an intact set of K62 capsular genes. (a) Analysis of the isolate's genome using Kaptive showed hits across three contigs and suggested the isolate encoded a variant of the K62 capsule type. (b) Comparison of the Kaptive K62 reference sequence (GenBank accession number AB371295) with the three contigs of #74 showed the K62 genes are non-contiguous. The ClustalW 2.1 alignment was created using Geneious Prime v2019.2.1.

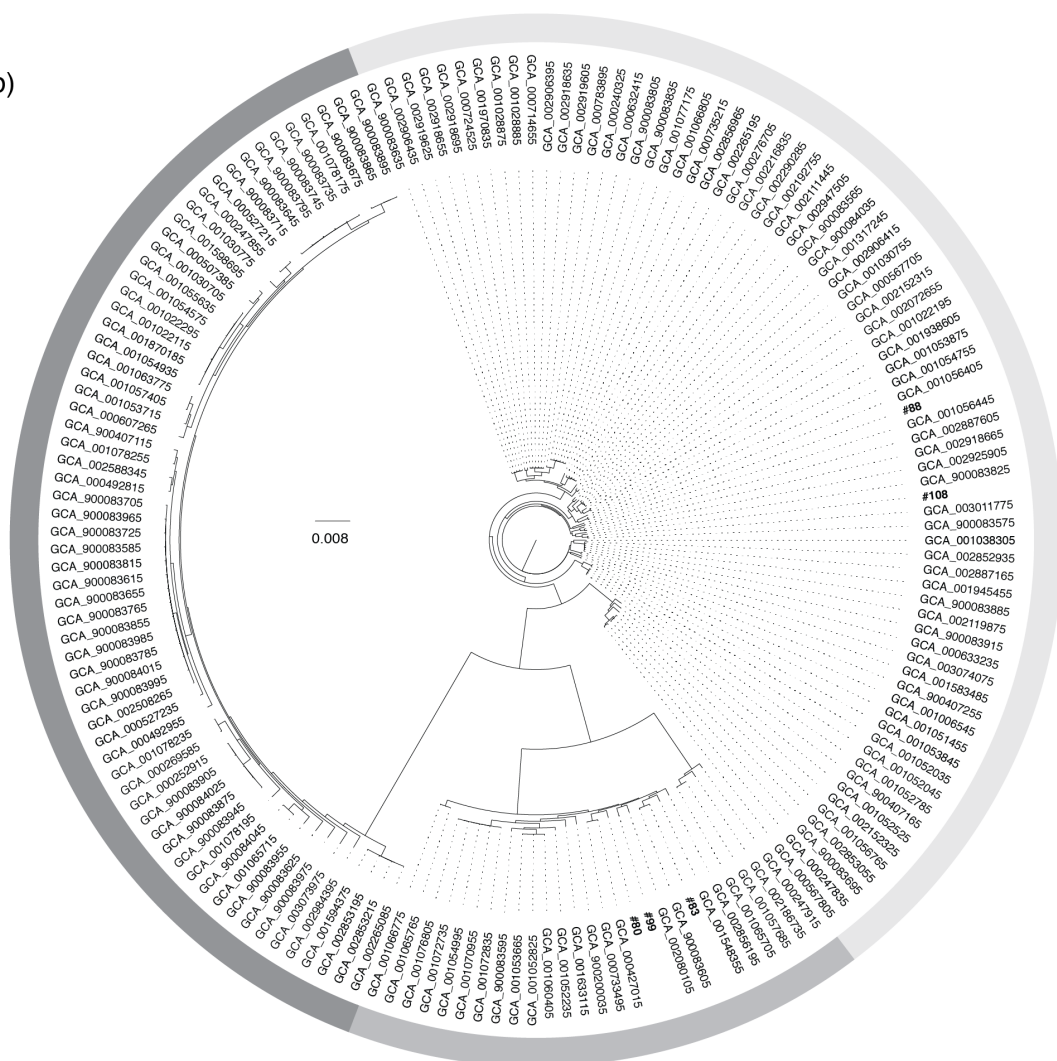
Consensus Identity	1	10	20	30	40	50	
1. NC_012731_Waal.	Met Ser Thr Leu Lys Val Ser Asn Phe Ala Arg Gln Gly Tyr Ser Ile	Val Phe Pro Leu Phe Leu Phe Phe Ser Ala	Ile Phe Cys Met Ser Thr Arg Thr Asn Asn Leu Leu His Leu Ser	Ile Leu Leu Leu Leu Leu Ser Leu Val Arg Gln Gly Asn			
2. AF146532_Waal.	Met Ser Thr Leu Lys Val Ser Asn Phe Ala Arg Gln Gly Tyr Ser Ile	Val Phe Pro Leu Phe Leu Phe Phe Ser Ala	Ile Phe Cys Met Ser Thr Arg Thr Asn Asn Leu Leu His Leu Ser	Ile Leu Leu Leu Leu Leu Ser Leu Val Arg Gln Gly Asn			
3. #91_Waal.	Met Ser Thr Leu Lys Val Ser Asn Phe Ala Arg Gln Gly Tyr Ser Ile	Val Phe Pro Leu Phe Leu Phe Phe Ser Ala	Ile Phe Cys Met Ser Thr Arg Thr Asn Asn Leu Leu His Leu Ser	Ile Leu Leu Leu Leu Leu Ser Leu Val Arg Pro Gly Asn			
Consensus Identity	60	70	80	90	100		
1. NC_012731_Waal.	Arg Gln Ala Leu Ala Gly Val Leu Arg Glu Gln Trp Gln Thr Thr Leu Leu Ala Ala Phe Phe Ile	Tyr Tyr Ala Leu Ser Asn Val Trp Gly His Thr Pro Gln His Ile Asp Ser Pro Ile	Thr His Gly Val Tyr Leu Thr Gly Tyr Leu Leu Leu				
2. AF146532_Waal.	Arg Gln Ala Leu Ala Gly Val Leu Arg Glu Gln Trp Gln Thr Thr Leu Leu Ala Ala Phe Phe Ile	Tyr Tyr Ala Phe Ser Asn Val Trp Gly His Thr Pro Gln His Ile Asp Ser Pro Ile	Thr His Gly Val Tyr Leu Thr Gly Tyr Leu Leu Leu				
3. #91_Waal.	Arg Gln Ala Leu Ala Gly Val Leu Arg Glu Gln Trp Gln Thr Thr Leu Leu Ala Val Phe Phe Ile	Tyr Tyr Ala Leu Ser Asn Leu Trp Gly His Thr Pro Gln His Ile Asp Ser Pro Ile	Thr His Gly Val Tyr Leu Thr Gly Tyr Leu Leu Leu				
Consensus Identity	110	120	130	140	150		
1. NC_012731_Waal.	Met Thr Met Leu Leu Arg Asp Gly Arg Thr Arg Arg Leu Ala Met Leu Ala Val Val Gly Gly Ile	Thr Val Leu Ser Leu Trp Thr Leu Ile	Ile Asp His Thr Leu Val Leu Thr Glu Arg Ala Val Ser Pro Glu Asn Pro Gly Pro Thr Asn Val Ile				
2. AF146532_Waal.	Met Thr Met Leu Leu Arg Asp Gly Arg Thr Arg Arg Leu Ala Met Leu Ala Val Val Gly Gly Ile	Thr Val Leu Ser Leu Trp Thr Leu Ile	Ile Asp His Thr Leu Val Leu Thr Glu Arg Ala Val Ser Pro Glu Asn Pro Gly Pro Thr Asn Val Ile				
3. #91_Waal.	Met Thr Met Leu Leu Arg Asp Gly Arg Thr Arg Arg Leu Ala Met Leu Ala Val Val Gly Gly Ile	Thr Val Leu Ser Leu Trp Thr Leu Met Phe Asp His Thr Leu Val Leu Thr Glu Arg Ala Val Ser Pro Glu Asn Pro Gly Pro Thr Asn Val Ile					
Consensus Identity	160	170	180	190	200	210	
1. NC_012731_Waal.	Asp Leu Ala Gly Tyr Cys Gly Ile Gly Ile Leu Ile Cys Gly Met Leu Leu Lys Glu Lys Ala Ser His Trp Leu Tyr Leu Pro Val Val	Ile Met Leu Val Met Leu Leu Leu Thr Gln Ser Arg Gly Pro Ile	Ile Ala Leu Val Leu Ala Val Gly Cys				
2. AF146532_Waal.	Asp Leu Ala Gly Tyr Cys Gly Ile Gly Ile Leu Ile Cys Gly Met Leu Leu Lys Glu Lys Ala Ser His Trp Leu Tyr Leu Pro Val Val	Ile Met Leu Val Met Leu Leu Leu Thr Gln Ser Arg Gly Pro Ile	Ile Ala Leu Val Leu Ala Val Gly Cys				
3. #91_Waal.	Asp Leu Ala Gly Tyr Cys Gly Ile Gly Ile Leu Ile Cys Gly Met Leu Leu Lys Glu Lys Ala Ser His Trp Leu Tyr Leu Pro Val Val	Ile Met Leu Val Met Met Leu Leu Thr Gln Ser Arg Gly Pro Ile	Ile Ala Leu Val Leu Ala Val Gly Cys				
Consensus Identity	220	230	240	250	260	270	
1. NC_012731_Waal.	Thr Leu His Leu His Val Phe Thr Arg Arg Asn Leu Leu Ile Ala Ala Ala Leu Ala Val Leu Val Ala Leu Leu Leu Val Met Thr Pro Val Gly Asp Met Leu Leu Ala Arg Phe Glu Glu Leu Gly Thr Gln Ser Gly Leu Arg Leu Ser	Ile Trp His					
2. AF146532_Waal.	Thr Leu His Leu His Val Phe Thr Arg Arg Asn Leu Leu Ile Ala Ala Ala Leu Ala Val Leu Val Ala Leu Leu Leu Val Met Thr Pro Val Gly Asp Met Leu Leu Ala Arg Phe Glu Glu Leu Gly Thr Gln Ser Gly Leu Arg Leu Ser	Ile Trp His					
3. #91_Waal.	Thr Leu His Leu His Val Phe Thr Arg Arg Asn Leu Leu Ile Ala Ala Ala Leu Ala Val Leu Val Ala Leu Leu Leu Val Met Thr Pro Val Gly Asp Met Leu Leu Ala Arg Phe Glu Glu Leu Gly Thr Gln Ser Gly Leu Arg Leu Ser	Ile Trp His					
Consensus Identity	280	290	300	310	320		
1. NC_012731_Waal.	His Thr Leu Ser Glu Met Ala Ser Gln Pro Trp Leu Gly Arg Gly Phe Ser Tyr Glu Leu Asp Phe Ile Asn Tyr Ser Gly Glu His Ile Thr Thr Thr His Ser Val Tyr Met Gly Ala Leu Leu Lys Gly Gly Ile	Val Gly Leu Leu Leu Leu Ala					
2. AF146532_Waal.	His Thr Leu Ser Glu Met Ala Ser Gln Pro Trp Leu Gly Arg Gly Phe Ser Tyr Glu Leu Asp Phe Ile Asn Tyr Ser Gly Glu His Ile Thr Thr Thr His Ser Val Tyr Met Gly Ala Leu Leu Lys Gly Gly Ile	Val Gly Leu Leu Leu Leu Ala					
3. #91_Waal.	His Thr Leu Ser Glu Met Ala Ser Gln Pro Trp Leu Gly Arg Gly Phe Ser Tyr Glu Leu Asp Phe Ile Asn Tyr Ser Gly Glu His Ile Thr Thr Thr His Ser Val Tyr Met Gly Ala Leu Leu Lys Gly Gly Ile	Val Gly Leu Leu Leu Leu Ala					
Consensus Identity	330	340	350	360	370		
1. NC_012731_Waal.	Ile Ile Ala Cys Gly Leu Trp Gln Ala Trp Arg Lys Arg His Thr Asp Ser Arg Tyr Ser Leu Ala Ile Leu Phe Tyr Ala Leu Val Phe Met Ala Ser Gln Gly Met Phe Ile	Ile Ser Asn Pro Arg Glu Thr Trp Val Leu Phe Trp Leu Pro Leu Gly					
2. AF146532_Waal.	Ile Ile Ala Cys Gly Leu Trp Gln Ala Trp Arg Lys Arg His Thr Asp Ser Arg Tyr Ser Leu Ala Ile Leu Phe Tyr Ala Leu Val Phe Met Ala Ser Gln Gly Met Phe Ile	Ile Ser Asn Pro Arg Glu Thr Trp Val Leu Phe Trp Leu Pro Leu Gly					
3. #91_Waal.	Val Ile Ala Cys Gly Leu Trp Gln Ala Trp Arg Lys Arg His Thr Asp Ser Arg Tyr Ser Leu Ala Ile Leu Phe Tyr Ala Leu Val Phe Met Ala Ser Gln Gly Met Phe Ile	Ile Ser Asn Pro Arg Glu Thr Trp Val Leu Phe Trp Leu Pro Leu Gly					
Consensus Identity	380	390					
1. NC_012731_Waal.	Ile Ala Leu Ser Lys Gly Val Ala Glu Lys Arg						
2. AF146532_Waal.	Ile Ala Leu Ser Lys Gly Val Ala Glu Lys Arg						
3. #91_Waal.	Ile Ala Leu Ser Lys Gly Val Ala Glu Lys Arg						

Supplementary Figure 4. Demonstration that *K. quasipneumoniae* #91 K11:O3 encodes a type 1 LPS core. The WaaL-encoding sequence of #91 was identified by comparison of the genome of #91 with the *waa* gene cluster described by Regué *et al.* (66) (GenBank accession number AF146532). WaaL of *K. quasipneumoniae* #91 shared 95.6 % pairwise identity with WaaL of *Klebsiella pneumoniae* subsp. *pneumoniae* NTUH-K2044 (GenBank accession number NC_012731), which had previously been shown to belong to LPS core type 1 (67). WaaL of AF146532 shared 98.9 % pairwise identity with that of NC_012731. The ClustalW 2.1 alignment was created in Geneious Prime v2019.2.1 using the BLOSUM 62 matrix.

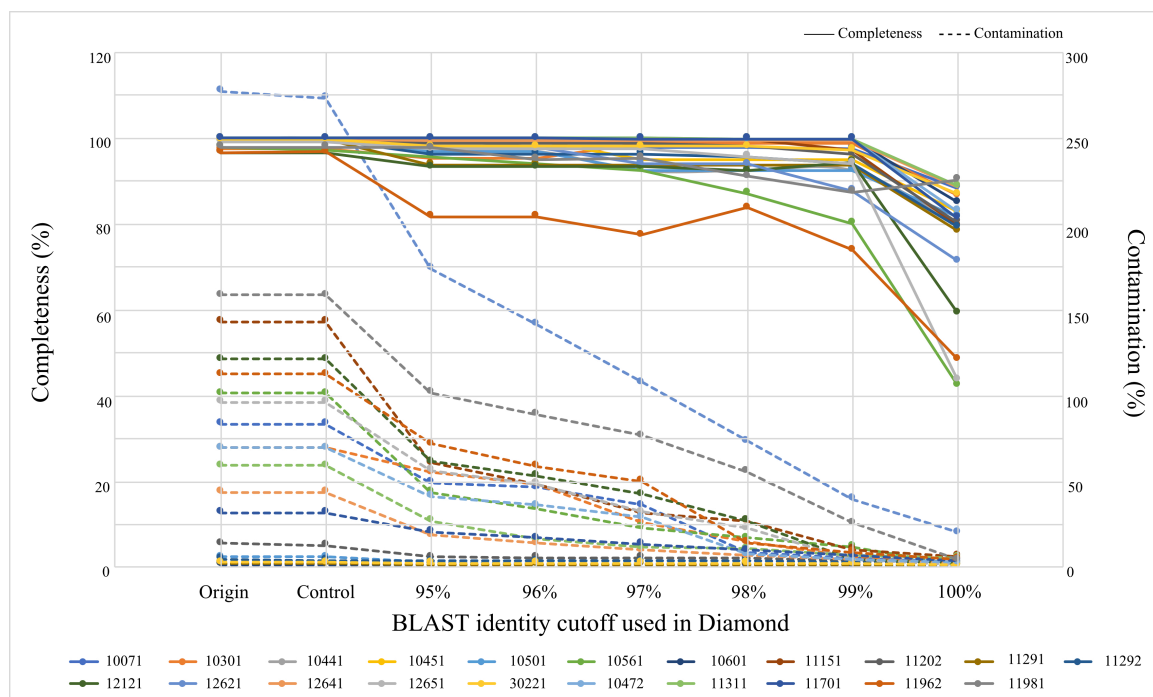
(a)



(b)

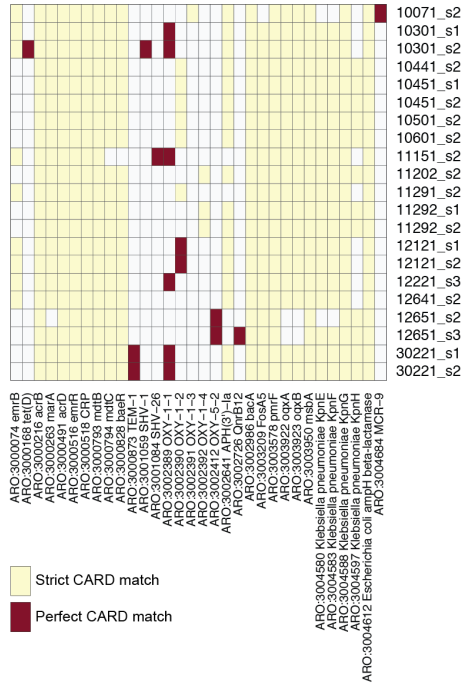


Supplementary Figure 5. Confirmation that the *K. oxytoca*-related isolates recovered from infants were strains of *K. michiganensis* and *K. grimontii*. (a) Heatmap generated with the R package heatmap.2() from FastANI (24) outputs. (b) Phylogenetic tree showing the placement of the genomes of the isolates among confirmed strains of *K. oxytoca*, *K. michiganensis* and *K. grimontii* (36). FastTree v2.1.10 (39) was used to generate the tree from the core gene alignment produced by v3.12.0 (default settings) (38), with the tree visualized using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). In (a, b), affiliation of strains #88 and #108 with *K. michiganensis* is confirmed, while strains #80, #83 and #89 belong to *K. grimontii*.

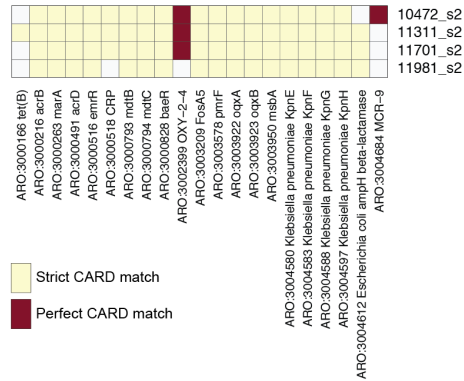


Supplementary Figure 6. Genome quality assessment of MAGs. The change of genome completeness and percentage contaminations when different Diamond BLAST identities were applied to MAGs recovered from stage2 samples. Solid line represents completeness and dotted line represents contamination. An identity of 99 % is generally the most suitable cut-off to do decontamination as genome completeness is kept at a reasonable level but the removal of contaminants is effective.

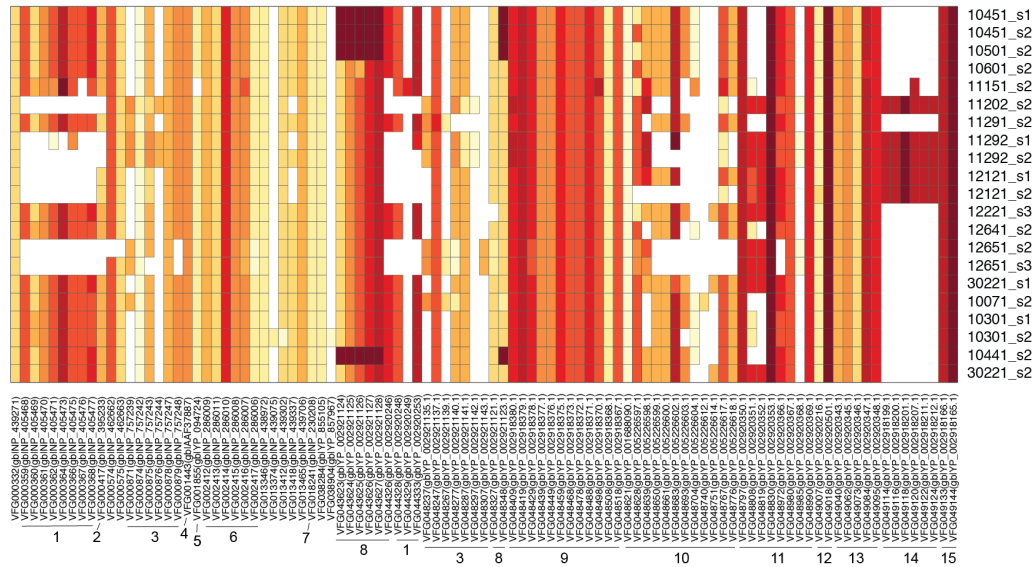
(a) *Klebsiella michiganensis* MAGs – CARD



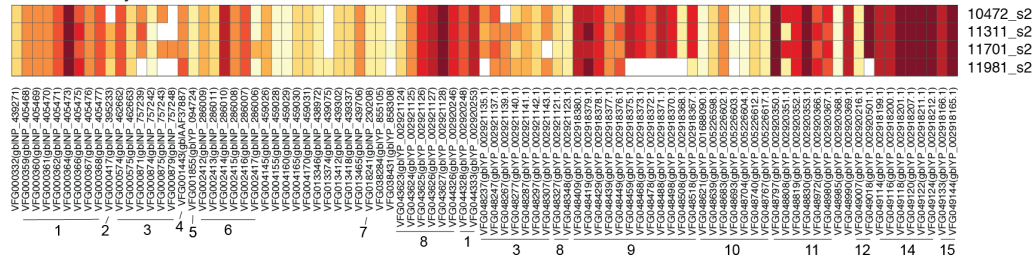
Klebsiella oxytoca MAGs – CARD



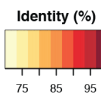
(b) *Klebsiella michiganensis* MAGs – VFDB



Klebsiella oxytoca MAGs – VFDB



1. Yersiniabactin
2. Plia, plasminogen activator
3. Type I fimbriae
4. OmpA (Outer membrane protein A)
5. Hsp60
6. ECP (*E. coli* common pilus)
7. Al-2 (autoinducer-2)
8. Type 3 fimbriae
9. Enterobactin (incomplete)
10. Type VI secretion system (incomplete)
11. Capsule (incomplete)
12. RcsAB (regulation of capsule synthesis)
13. LPS (incomplete)
14. Allantoin utilization
15. AcrAB efflux pump



Supplementary Figure 7. Summary of (a) antibiotic resistance and (b) virulence factor genes found in the MAGs recovered from the data of Ward *et al.* (1) by comparison of protein sequences with those of the CARD and VFDB, respectively. (a) Strict CARD match, not identical but the bit-score of the matched sequence is greater than the curated BLASTP bit-score cut-off; perfect CARD match, 100 % identical to the reference sequence along its entire length. Loose matches are not shown to avoid presenting false positives based on sequences with low homology and bit-scores below CARD BLASTP cut-off recommendations. (b) Identity (%), BLASTP reported only for those proteins sharing >70 % identity and 90 % query coverage with VFDB protein sequences.