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*.

/	×
10	s١
10	<b>1</b>
· · -	· /

		Results			Download raw results table (TXT)	Download raw results (JSON)
GCF_008120605.1_ASM812060	v1_genomic	Best locus: KL62	Match confidence 10 : Good	Cov (): 100.00%	ID 🚯 : 99.33%	Genes: 18 / 19
KL62 reference 0:						
100.00% 10: 100.00%	97.14% 400.00% 10:100.00%	0% 10: 100.00% 10: 100.00%	99.72%	D: 99.42% D: 100.00% ID: 100.00% N: 100.00% ID: 99.49%	ev: 100.00% 10: 99.17% 100.00% 10: 99.74% 100.00% 10: 100.00% 10: 100.00% 10: 100.00%	0.00% 0.00% 10:99.79% 100.00% 10:99.49% 100.00% 10:99.49%
partie Colv. oppenCiP U	Cov: 12 wzs Cov: wzb C	Varia Contra	WP07 COL KIG2_00 KIG2_00 KIG2_00 KIG2_00	VEL COV. 14162-12 12 12 12 12 13	YL92_14 verx Coll. and Coll. man	c cut mane cut upd cov. T
Other genes found in locus	19:0 ▼ Other	genes found outside	locus <b>@</b> : 4 ▼			
Allelic type 🔁 :	wzc: 62 wzi: 14	9				
Assembly pieces () :	Download as FAST.	A	KL62 reference size (): 24656			
Contig name Start	position End position	Length	Length discrepancy 0: n/a			
NZ_QFUN01000027.1 51419	65716	14298				
NZ_QFUN01000052.1 135	6157	6023				
NZ_QFUN01000053.1 420	4763	4344				
				IS1 ·	family transposase	
b)				IS1	family transposase	
b)	1 2,000	4,000	6.000 8.000 10,000	IS1 -	family transposase	20,000 22,000 24,000
b) ensus ity	1 2,000	4,000	6.000 8.000 10.000	IS1 -	family transposase	20,000 22,000 24,000
b) aty 10 1.A8371295	1 2,000 2010 2010	4.000 wzi CDS	οφο ε.οο 10,000 20.00 wtoP C	12,000 - 12,	family transposase	23,000 22,000 24,000
<b>b)</b> ity <sup>10</sup> 1. AB371295 <sup>10</sup> 2. NZ_QFUN01000027 extraction	2.000 2010 2010 2010 2010 2010	4,000 WZICDS WA	sopo sopo topo www.ccD5 wbaPC www.ccD5 wbaPC wbw.ccD5 wbaPC	IS1	family transposase	22,000 22,000 24,000
b) tity 10 1. AB371295 10 2. NZ_QFUN01000027 extraction 10 3. NZ_QFUN01000052	1 2000 2000 2000 2000 2000 2000 2000 200	4.000 wzi CDS capsule	s.çoo s.çoo to,coo a wxx CD5 wtbaP C iys polyascharid wtbaP C	IS1	family transposase	22,000 22,000 24,000
b) by 10 1. AB371295 10 2. NZ_QFUN01000027 extraction 10 3. NZ_QFUN01000052 (V 4. NZ_QFUN01000053	2,000 galas Car galas Da	4000 wziCDS yw capsulepg	sopo acto tapoo www.cDS websP.C. itypolyraecharidwebsP.C.	IS1 12000 140 14	family transposase	23,000 22,000 24,000 grad CD5 markmmark
b) ty to 1. A8371295 <sup>10</sup> 2. NZ_QFUN01000027 extraction <sup>10</sup> 3. NZ_QFUN01000052 <sup>10</sup> 4. NZ_QFUN01000053	2,00 pale	4000 WZICDS WW Copsule pp	s.opa a.coo to.opa www.cDS whaP C. ity	IS1	family transposase	23,000 22,000 24,000 grad CDS manCmanBUgg grad ACk grad ACk
D) Insus 9 1. A8371295 <sup>10</sup> 2. NZ_QFUN01000027 extraction <sup>10</sup> 3. NZ_QFUN01000052 <sup>14</sup> 4. NZ_QFUN01000053	gal p	4,00 wziCDS ww capsulepr	s.opo 8.000 10.000 wz.CDS whaP C. lys. polysaccharid wbaP C.	IS1	family transposase	20,000 22,000 24,000 grid CDS manC manB ugg grid A C grid A C manno cpsG C UDP
ensus   ify   ify   ifv   ifv	2,00 gal gal gal gal	4000 WZICDS WW Capsule pc	s.çoo 8.coo 10.coo w.z.CDS withaP C lys oolysaccharid withaP C	IST 12,000	family transposase	23,000 22,000 24,000 gpd CDS manCman8ugd gpd A C gpd A C manno cpsG C UDP_

**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.

	1	10	20	30	40	50
Consensus Identity	Met Ser Thr Leu Lys Val Ser As	snPhe Ala Arg Gin Gly Tyr Ser	lie Val Phe Pro Leu Phe Leu Phe Phe	Ser Ala lle PheCysMetSer Thr A	rg Thr Asn Asn Leu Leu His Leu Ser 1	le Leu Leu Leu Leu Ser Leu Val Arg Gin Glu Asn
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	Met Ser Thr Leu Lys Val Ser As	nPhe Ala Arg Gin Giy Tyr Ser	lle Val Phe Pro Leu Phe Leu Phe Phe	Ser Ala lle Phe Cys Met Ser Thr A Met Ser Thr A Met Ser Thr A Met Ser Thr A	rg Thr Asn Asn Leu Leu His Leu Ser   rg Thr Asn Asn Leu Leu His Leu Ser   rg Thr Asn Asn Leu Leu His Leu Ser	le LeuLeuLeuLeuSer Leu Val Arg Gin Giu Asn le LeuLeuLeuLeuSer Leu Val Arg Gin Giu Asn le LeuLeuLeuLeuSer Leu Val Arg Pro Giu Asn
Consensus Identity	Arg Gln Ala Leu Ala Gly Val Le	u Arg Glu Gln Trp Gln Thr Trp	<sup>70</sup> Thr Leu Leu Ala Ala PhePhe Ile Tyr	Tyr Ala Leu Ser Asn Val Trp Gly H	is Thr Pro Gin His Ile Asp Ser Pro I	le Thr His Gly Val Tyr Leu Thr Gly Tyr Leu Leu Leu
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	Arg Gln Ala Leu Ala Gly Val Le Arg Gln Ala Leu Ala Gly Val Le Arg Gln Ala Leu Ala Gly Val Le	eu Arg Glu His Trp Gln Thr Trp eu Arg Glu Gln Trp Gln Thr Trp eu Arg Glu Gln Trp Gln Thr Trp 120	Thr Leu Leu Ala Ala PhePhe Ile Tyr Thr Leu Leu Ala Ala PhePhe Ile Tyr Ala Leu Leu Ala Val PhePhe Ile Tyr 130	Tyr Ala Leu Ser Asn Val Trp Gly H Tyr Ala Phe Ser Asn Val Trp Gly H Tyr Ala Leu Ser Asn Leu Trp Gly H	is Thr Pro Gln His lle Asp Ser Pro l is Thr Pro Gln His lle Asp Ser Pro l is Thr Pro Gln His lle Asp Ser Pro l	le Thr His Gly Val Tyr Leu Thr Gly Tyr Leu Leu Leu le Thr His Gly Val Tyr Leu Thr Gly Tyr Leu Leu Leu le Thr His Gly Val Tyr Leu Thr Gly Tyr Leu Leu Leu 50
Consensus Identity	Met Thr Met Leu Leu Arg Asp G	y Arg Thr Arg Arg Leu Ala Met	eu Ala Val Val Gly Gly Ile Thr Va	Leu Ser Leu Trp Thr Leu Ile Ile A	sp His Thr Leu Val Leu Thr Glu Arg /	la Val Ser Pro Glu Asn Pro Gly Pro Thr Asn Val Ile
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	Met Thr Met Leu Leu Arg Asp G Met Thr Met Leu Leu Arg Asp G Met Ala Met Leu Leu Ser Asp G	ly Arg Thr Arg Arg Leu Ala Met I ly Arg Thr Arg Arg Leu Ala Met I lu Arg Thr Arg Arg Leu Ala Met I	eu Ala Val Val Gly Gly Ile Thr Va eu Ala Val Val Gly Gly Ile Thr Va eu Ala Val Val Gly Gly Ile Thr Va	Leu Ser Leu Trp Thr Leu IIe IIe As Leu Ser Leu Arg Thr Leu IIe IIe As Leu Ser Leu Trp Thr Leu Met Phe As	sp His Thr Leu Val Leu Thr Glu Arg / sp His Thr Leu Val Leu Thr Glu Arg / sp His Thr Leu Val Leu Thr Glu Arg /	la Val Ser Pro Glu Asn Pro Gly Pro Thr Asn Val Ile la Val Ser Pro Glu Asn Pro Gly Pro Thr Asn Val Ile la Val Ser Pro Glu Asn Pro Gly Pro Thr Asn Val Ile
Consensus Identity	AspLeu Ala Gly Tyr Cys Gly II	e Gly lle Leu lle Cys Gly Met l	eu Leu Lys Glu Lys Ala Ser His Trp	Leu Tyr Leu Pro Val Val Ile MetLe	eu Val Met Leu Leu Leu Thr Gin Ser A	rg Gly Pro Ile Ile Ala Leu Val Leu Ala Val Gly Cys
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	AspLeu Ala Gly Tyr Cys Gly II AspLeu Ala Gly Tyr Cys Gly II AspLeu Ala Gly Tyr Cys Gly II	e Gly Ile Leu Ile Cys GlyMett e Gly Ile Leu Ile Cys GlyMett e Gly Ile Leu Ile Cys GlyMett	eu Leu Lys Glu Lys Ala Ser His Trp eu Leu Lys Glu Lys Ala Ser His Trp eu Leu Lys Glu Lys Ala Ser His Trp	Leu Tyr Leu Pro Val Val IIe MetLe Leu Tyr Leu Pro Val Val IIe MetLe Leu Tyr Leu Pro Val Ala IIe MetLe	eu Val MetLeu Leu Leu Thr Gin Ser A eu Val MetLeu Leu Leu Thr Gin Ser A eu Val MetMetLeu Leu Thr Gin Ser A	rg Gly Pro IIe IIe Ala Leu Val Leu Ala Val Gly Cys rg Gly Pro IIe IIe Ala Leu Val Leu Ala Val Gly Cys rg Gly Pro IIe IIe Ala Leu Val Leu Ala Val Gly Cys
Consensus Identity	Thr Leu His Leu His Val Phe Th	متر Arg Arg Asn Leu Leu Ile Ala .	Ala Ala Leu Ala Val Leu Val Ala Leu	Leu Leu Val Met Thr Pro Val Gly A	spMetLeuLeu Ala ArgPhe Glu Glu L	eu Gly Thr Gln Ser Gly Leu Arg Leu Ser Ile Trp His
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	Thr Leu His Leu His Val Phe Th Thr Leu His Leu His Val Phe Th Thr Leu His Leu His Val Phe Th	nr Arg Arg Asn Leu Leu IIe Ala nr Arg Arg Asn Leu Leu IIe Ala nr Arg Arg Asn Leu Leu IIe Ala	Ala Ala Leu Ala Val Leu Val Ala Leu Ala Ala Leu Ala Val Leu Val Ala Leu Ala Ala Leu Ala Val Leu Val Ala Leu	Leu Leu Val Met Thr Pro Val Gly A Leu Leu Val Met Thr Pro Val Gly A Leu Leu Val Met Thr Pro Val Gly A Leu Leu Val Met Thr Pro Val Gly A	spMetLeuLeu Ala ArgPhe Glu Glu L spMetLeuLeu Ala ArgPhe Glu Glu L spMetLeuLeu Ala ArgPhe Glu Glu L	eu Gly Thr Gln Ser Gly Leu Arg Leu Ser lle Trp His eu Gly Thr Gln Ser Gly Leu Arg Leu Ser lle Trp His eu Gly Thr Gln Ser Gly Leu Arg Leu Ser lle Trp His
Consensus Identity	His Thr Leu Ser Glu Met Ala Se	er Gln Pro Trp Leu Gly Arg Gly F	he Ser Tyr Glu Leu AspPhe Ile Asr	n Tyr Ser Gly Glu His 👖 Thr Thr Tl	hr His Ser Val Tyr Met Gly Ala Leu L	eu Lys Gly Gly Ile Val Gly Leu Leu Leu Leu Ala
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	His Thr Leu Ser Glu Met Ala Se His Thr Leu Ser Gln Met Ala Se His Thr Leu Ser Glu Met Ala Se	er Gin Pro Trp Leu Giy Arg Giy F er Gin Pro Trp Leu Giy Arg Giy F er Gin Pro Trp Leu Giy Arg Giy F	he Ser Tyr Glu Leu AspPhe lle Asr he Ser Tyr Glu Leu AspPhe lle Asr he Ser Tyr Glu Leu AspPhe lle Asr	Tyr Ser Gly Glu His lle Thr Thr Ti Tyr Ser Gly Glu His lle Thr Thr Ti Tyr Ser Gly Glu His lle Thr Thr Ti Ser Gly Glu His lle Thr Thr Ti	hr His Ser Val Tyr Met Gly Ala Leu L hr His Ser Val Tyr Met Gly Ala Leu L hr His Ser Val Tyr Met Gly Ala Leu L	eu Lys Gly Gly lle Val Gly Leu Leu Leu Leu Leu Ala eu Lys Gly Gly lle Val Gly Leu Leu Leu Leu Leu Ala eu Lys Gly Gly lle Val Gly Leu Leu Leu Leu Leu Ala
Consensus Identity	lle lle Ala Cys Gly Leu Trp G	In Ala Trp Arg Lys Arg His Thr	sep Ser Arg Tyr Ser Leu Ala IIe Leu	Phe Tyr Ala Leu Val PheMet Ala Se	er Gin Gly MetPhe Ile Ile Ser Asn F	ro Arg Glu Thr Trp Val LeuPhe Trp LeuPro Leu Gly
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	lle lle Ala Cys Gly Leu Trp G lle lle Ala Cys Gly Leu Trp G Val lle Ala Cys Gly Leu Trp G	In Ala Trp Arg Lys Arg His Thr In Ala Trp Arg Lys Arg His Thr In Ala Trp Arg Lys Arg Arg Ser	vsp Ser Arg Tyr Ser Leu Ala IIe Leu vsp Ser Arg Tyr Ser Leu Ala IIe Leu vsp Ser Arg Tyr Ser Leu Ala IIe Leu	Phe Tyr Ala Leu Val PheMet Ala Si Phe Tyr Ala Leu Val PheMet Ala Si Phe Tyr Ala Leu Val PheMet Ala Si	er Gin Gly MetPhe Ile – Ile – Ser Asn F er Gin Gly MetPhe Ile – Ile – Ser Asn F er Gin Gly MetPhe Ile – Ile – Ser Asn F	ro Arg Glu Thr Trp Val LeuPhe Trp Leu Pro Leu Gly ro Arg Glu Thr Trp Val LeuPhe Trp Leu Pro Leu Gly ro Arg Glu Thr Trp Val LeuPhe Trp LeuPro Leu Gly
Consensus	Ile Ala Leu Ser Lys Gly Val A	la Glu Lys Arg				
Identity	_					
1. NC_012731_WaaL 2. AF146532_WaaL 3. #91_WaaL	Ile Ala Leu Ser Lys Gly Val A Ile Ala Leu Ser Lys Gly Val A Ile Ala Leu Ser Lys Gly Leu A	a Glu Lys Arg a Glu Lys Arg a 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.



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.



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.