Transmission of *Klebsiella* strains and plasmids within and between grey-headed flying fox colonies

Ben Vezina ,[1*] Louise M. Judd ,[1]
Fiona K. McDougall ,[2] Wayne S. J. Boardman ,[3]
Michelle L. Power ,[2] Jane Hawkey ,[1]
Sylvain Brisse ,[4] Jonathan M. Monk ,[5]
Kathryn E. Holt [1,6] and Kelly L. Wyres [1]

[1]*Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Vic., Australia.*
[2]*Department of Biological Sciences, Macquarie University, NSW, 2109, Australia.*
[3]*School of Animal and Veterinary Sciences, University of Adelaide, SA, 5371, Australia.*
[4]*Institut Pasteur, Université de Paris, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, France.*
[5]*Department of Bioengineering, University of California, San Diego, CA.*
[6]*Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, UK.*

Summary

The grey-headed flying fox (*Pteropus poliocephalus*) is an endemic Australian fruit bat, known to carry zoonotic pathogens. We recently showed they harbour bacterial pathogen *Klebsiella pneumoniae* and closely related species in the *K. pneumoniae* species complex (*Kp*SC); however, the dynamics of *Kp*SC transmission and gene flow within flying fox colonies are poorly understood. High-resolution genome comparisons of 39 *Kp*SC isolates from grey-headed flying foxes identified five putative strain transmission clusters (four intra- and one inter-colony). The instance of inter-colony strain transmission of *K. africana* was found between two flying fox populations within flying distance, indicating either direct or indirect transmission through a common food/water source. All 11 plasmids identified within the *Kp*SC isolates showed 73% coverage (mean) and ≥95% identity to human-associated *Kp*SC plasmids, indicating gene flow between human clinical and grey-headed flying fox isolates. Along with strain transmission, interspecies horizontal plasmid transmission between *K. pneumoniae* and *Klebsiella africana* was also identified within a flying fox colony. Finally, genome-scale metabolic models were generated to predict and compare substrate usage to previously published *Kp*SC models, from human and environmental sources. These models indicated no distinction on the basis of metabolic capabilities. Instead, metabolic capabilities were consistent with population structure and ST/lineage.

Introduction

The *Klebsiella pneumoniae* species complex (*Kp*SC) (Wyres *et al*., 2020) is a group of closely related *Klebsiella* species which includes *Klebsiella pneumoniae*, *Klebsiella quasipneumoniae* subsp. *quasipneumoniae*, *Klebsiella variicola* subsp. *variicola*, *Klebsiella quasipneumoniae* subsp. *similipneumoniae*, *Klebsiella variicola* subsp. *tropica*, *Klebsiella quasivariicola* and *Klebsiella africana*. *Kp*SC are problematic as they function as cosmopolitan opportunistic pathogens (Knittel *et al*., 1977; Navon-Venezia *et al*., 2017), responsible for a worrying proportion of community and hospital-acquired infections (Pendleton *et al*., 2013). Prevalence of multi-drug resistance and acquired virulence factors associated with invasive infections is increasing over time (Lam *et al*., 2021), hence surveillance of potential reservoirs of problematic *Kp*SC sequence types (STs), along with detection of mobile antimicrobial resistance and virulence determinants, is essential for elucidating the potential routes and/or modes of transmission into the human population.

Aside from humans, *Kp*SC have also been isolated from a wide variety of environments including soil (Melo-Nascimento *et al*., 2018), plants (Knittel *et al*., 1977; Rosenblueth *et al*., 2004), freshwater (Hladicz *et al*., 2017; Melo-Nascimento *et al*., 2018), marine environments and organisms (Håkonsholm *et al*., 2020), waste dumps (Awasthi *et al*., 2017), animals including cats and dogs (Marques *et al*., 2019), migratory and domesticated birds (Liao *et al*., 2019; Ahmed *et al*., 2021), agricultural animals (Hamza *et al*., 2016; Massé *et al*., 2020) and grey-headed flying foxes (McDougall *et al*., 2021b). Flying foxes, or fruit bats, have

long been considered important vectors of zoonotic viruses such as Australian bat lyssavirus (Fraser *et al.*, 1996), Menangle virus (Philbey *et al.*, 2008) and Hendra virus (though not as the primary reservoir) (Halpin *et al.*, 2000). There is growing evidence that grey-headed flying foxes also harbour human pathogenic bacteria such as antibiotic-resistant and pathogenic *Escherichia coli* lineages (McDougall *et al.*, 2021a). Carbapenem-resistant *K. pneumoniae* have also previously been isolated from insectivorous bat guano in Algeria (Gharout-Sait *et al.*, 2019), though at very low frequency (two positive out of 110 samples).

Our previous surveillance study sampled *Kp*SC from four wild grey-headed flying fox colonies in 2017–2018 and represented a snapshot of these populations at the time (McDougall *et al.*, 2021b). Briefly, composite faecal matter was swabbed from drop sheets placed beneath the colonies. Swabs were spatially separated to maximize the likelihood that each represented faecal matter originating from distinct flying fox individuals. *Kp*SC was present in 4/101 samplings from Blackalls Park (flying fox population of 500–2249 on 12/12/2017 and 2500–9999 on 10/4/2018), 18/50 from Camellia Gardens (population of 500–2249), 1/52 from Adelaide (population of ~10 000) and 0/52 from Centennial Park wild colonies (population of ~10 000–15 999) (Fig. 1) (Australian Government Department of Agriculture). Additionally, 20 individual captive animals at the Mylor rehabilitation colony were swabbed: *Kp*SC was present in 15/20 animals (total population of ~40). In 2021, grey-headed flying fox populations of Adelaide Botanic Park colony were 30 000–36 000 (Boardman, W, pers. comm.). Population estimates for the remaining colonies do not yet have data available. Notably, this was the first non-human sighting of *K. africana*, a species which has been reported in just two previous studies globally (Rodrigues *et al.*, 2019; Lam *et al.*, 2021). Our initial genomic analyses of these *Kp*SC strains from flying foxes indicated low ST diversity with few STs previously reported among human clinical isolates, as well as low rates of antimicrobial resistance (1.1%) and virulence factors, suggesting that these grey-headed flying fox colonies are not a high-risk environmental reservoir for problematic *Kp*SC isolates. However, this analysis did not explore detailed strain relationships, or the broader genetic content of these *Kp*SC.

Here, we use long-read DNA sequencing to complete the genomes of 13 *Kp*SC from flying foxes and leverage these data alongside the previously reported draft genomes to perform the first high-resolution analysis of *Kp*SC strain and plasmid transmission dynamics within/ between grey-headed flying fox colonies. To better understand the potential for strain and plasmid transfer into the human population, we compare completed plasmid sequences from flying fox isolates to those identified among human-associated *Kp*SC, and utilize the latest genome-scale metabolic modelling approaches to compare the metabolic capabilities of flying fox-derived isolates to *Kp*SC isolates from other sources.

Experimental procedures

Bacterial isolate collection and whole-genome sequencing

Thirty-nine *Kp*SC isolates including 30 *K. pneumoniae*, one *K. variicola* subsp. *variicola* and eight *K. africana* were previously isolated from four different grey-headed flying fox colonies (McDougall *et al.*, 2021b); Blackalls Park, sampled on 12/12/2017 and 10/4/2018; Camellia Gardens, 4/1/2018; Adelaide Botanic Park, 11/2/2018, Mylor (captive colony), 12/2/2018; Centennial Park, 9/3/2018. One isolate (*K. pneumoniae* FF996) was not utilized in this study due to DNA contamination issues identified during high-resolution analysis. See Table S1 for details of all isolates, genotyping results and genome accessions.

Illumina (short-read) whole-genome sequencing data were generated and reported previously (McDougall *et al.*, 2021b). DNA was extracted for long-read sequencing of 13 isolates representing distinct STs (see Table S1), using the GenFind v3 kit (Beckman Coulter), followed by library preparation using the Ligation library kit (SQK-LSK109, Oxford Nanopore Technologies, ONT) with Native Barcode Expansion (EXP-NBD196, ONT) to allow for multiplexing of libraries. DNA was sequenced on MinION flow cells (R9.4.1, ONT) and basecalled using Guppy version 4.0.14 + 8d3226e with the dna_r9.4.1_450bps_hac model (ONT).

Illumina-only genome assemblies were generated as reported previously (McDougall *et al.*, 2021b). ONT reads were assembled using Flye version 2.8.1-b1676 (Kolmogorov *et al.*, 2019) (options: --plasmids -g 5250000 --asm-coverage 60) then polished twice with medaka version 1.0.3 (nanoporetech, 2020) using the medaka_consensus module (options: -m r941_min_-fast_g303). Finally, assemblies were short-read error corrected using Pilon version 1.23 (Walker *et al.*, 2014) (two rounds per genome).

Assemblies were annotated using Prokka version 1.14.6 (Seemann, 2014), and assembly statistics were generated using the statswrapper.sh module from bbmap version 38.81 (Bushnell, 2014). Species, multi-locus STs, antimicrobial resistance and virulence determinants were identified using Kleborate version v2.0.1 (Lam *et al.*, 2021). Suspected hybrid species were analysed in further detail using GenomePainter version 0.0.8 (https://github.com/scwatts/genome_painter) (Watts, 2018).
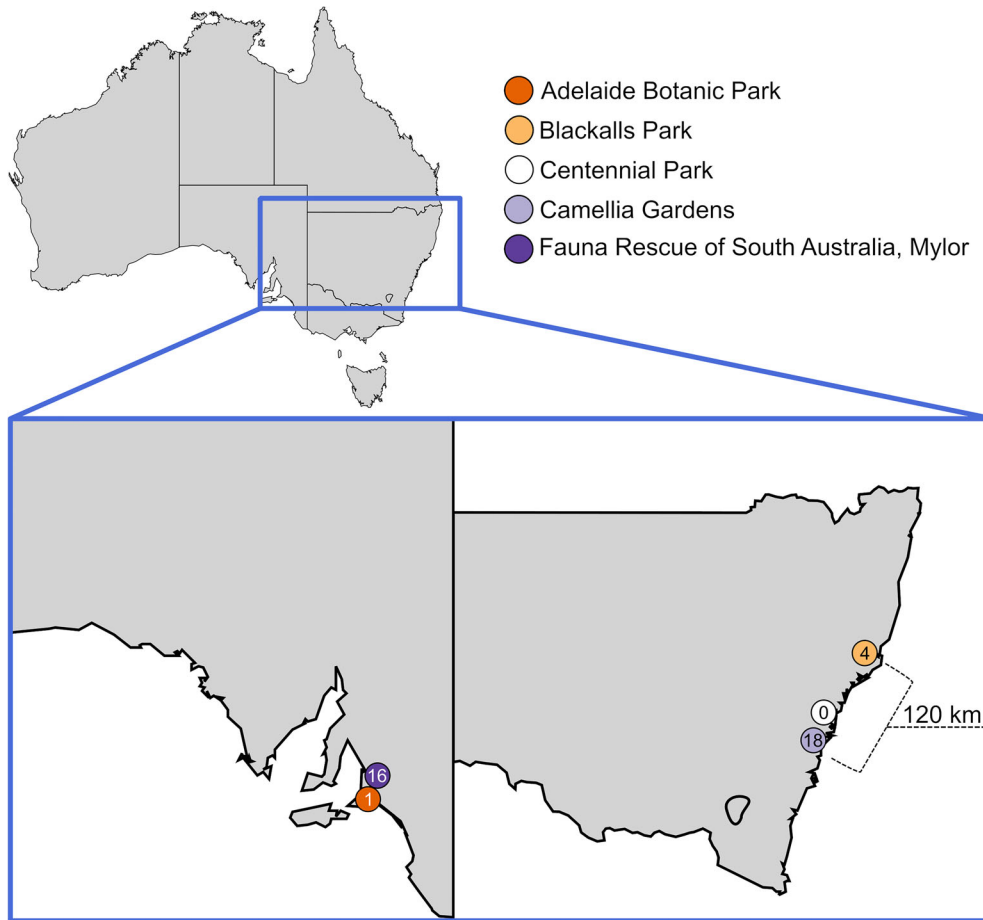
Fig. 1. Sampled flying fox colonies across Australia by McDougall *et al*. (2021b). Coloured circles represent grey-headed flying fox colonies and the number within represents the number of *Kp*SC isolates obtained. *Kp*SC were obtained from all colonies except for Centennial Park. Map of Australia generated by Lokal_Profil from https://upload.wikimedia.org/wikipedia/commons/b/bf/Australia_map%2C_States.svg used as per the Creative Commons Attribution-Share Alike 2.5 Generic licence.

Single nucleotide variant analyses

Single nucleotide variants (SNVs) were identified using RedDog version 1b.10.4 (Edwards *et al*., 2016). Isolates were grouped by ST, and the corresponding Illumina reads mapped to the respective completed chromosome and plasmid reference sequences (Table S1). As a control, the short-read data from the ST reference isolate were also mapped to its completed assembly. Pairwise SNV distances were calculated for each replicon using the snp2dist script at Figshare (https://doi.org/10.6084/m9.figshare.16609054) (Vezina *et al*., 2021). Pairwise chromosomal SNV distances ≤25 were considered indicative of recent strain transmission events, based on previous analyses of *Kp*SC transmission in clinical settings (Gorrie *et al*., 2017; David *et al*., 2019; Sherry *et al*., 2021). An SNV maximum likelihood phylogenetic tree was also constructed from the SNV alignment using RAxML with a GTR substitution model (Vezina *et al*., 2021).

Plasmid analyses

Completed plasmid sequences were analysed using the mob_typer command from Mob-suite version 3.0.0 (Robertson and Nash, 2018). The closest matching (best query coverage and identity) plasmid sequences from human-derived *Klebsiella* isolates were identified using NCBI BLASTn search against the public NCBI nucleotide database as of 01/04/21 (Johnson *et al*., 2008). A MASH distance tree was generated using mash version 2.1.1 (Ondov *et al*., 2016), along with R packages ape version 5.4–1 (Paradis *et al*., 2004) and phangorn version 2.5.5 (Schliep, 2011). The tree was inferred using the FastME algorithm (Lefort *et al*., 2015; Vezina *et al*., 2021).

Visualizations of pairwise plasmid comparisons were generated via EasyFig version 2.2.5 (Sullivan *et al*., 2011) and the BLASTn command, with the following BLAST parameters: Min. length – 1000. Max. *e*-value – 0.0001, to look at gene content and synteny.

Genome-scale metabolic reconstructions and in silico *growth predictions*

Metabolic diversity of *Kp*SC from flying foxes (*n* = 13 completed genomes only) was explored, in comparison to *Kp*SC from other host and/or environmental sources for which metabolic reconstructions have been published previously [*n* = 37 from Hawkey *et al*., 2022 and *n* = 22 from Norsigian *et al*., 2019].

Novel metabolic network reconstructions were generated using cobrapy version 0.20.0 (Ebrahim *et al*., 2013) in a conda environment version 4.9.2 (Anaconda-Software-Distribution, 2016), running Python 3.6.12 (Rossum and Python-Software-Foundation, 1995). The *Kp*SC pan metabolic model (Hawkey *et al*., 2022) was used as a reference as previously described (Thiele and Palsson, 2010; Norsigian *et al*., 2020). The specific code used to generate models was adapted from Norsigian *et al*. (2020) and can be found at Figshare (https://doi.org/10.6084/m9.figshare.16609054) (Vezina *et al*., 2021). The same approach was performed to update the 22 *K. pneumoniae* reconstructions described in Norsigian *et al*. (2019). Model annotations were improved using models iYS1720 (Seif *et al*., 2018), iML1515 (Monk *et al*., 2017) and iWFL_1372 (Monk *et al*., 2013) with a custom script (Vezina *et al*., 2021). MEMOTE reports (Lieven *et al*., 2020) were then generated for each model using the report snapshot option. *In silico* growth capabilities were simulated in minimal media with each of 268 sole carbon, 154 nitrogen, 25 sulfur and 59 phosphorus sources using the flux balance analysis model.optimize method from cobrapy as described previously (Hawkey *et al*., 2022). All genome accessions and model identifiers are shown in Table S1.

To contextualize the metabolic modelling results with regard to population structure, a core genome phylogeny was inferred from a core gene alignment generated with Panaroo version 1.1.2 (Tonkin-Hill *et al*., 2020) (options: --mode relaxed -a core --aligner mafft --core_threshold 0.98 -f 0.90 --merge_paralogs). The resulting alignment of 574 339 variant sites represented 3319 genes shared by ≥71/72 isolates. RAxML-NG version 1.1.0 (Kozlov *et al*., 2019) was used to infer a maximum likelihood phylogeny with 1000 bootstrap replicates (options: --all --model GTR), best of five runs. RAxML was also used to generate an unscaled, daylight phylogenetic tree in the context of plasmid presence (Vezina *et al*., 2021).

Data visualization

Data were visualized using R version 4.0.3 (R Core Team, 2020) and RStudio version 1.3.1093 (RStudio-Team, 2020) with the following software packages: RColorBrewer version 1.1-2 (Neuwirth, 2014), ggplot2 version 3.3.3 (Wickham, 2016), reshape2 version 1.1.4 (Wickham, 2007), aplot version 0.0.6 (Yu, 2020) and ggnewscale version 0.4.5 (Campitelli, 2021). igraph version 1.2.6 (Csardi and Nepusz, 2005), ggraph version 2.0.5 (Pedersen, 2021b) and ggforce version 0.3.3 (Pedersen, 2021a) were used for pairwise SNV analyses only. ggtree version 2.4.1 (Yu *et al*., 2017) was used for phylogenetic tree visualization only. pheatmap version 1.0.12 (Kolde, 2019), grid version 4.0.3 and gridExtra version 2.3 (Auguie and Antonov, 2017) were used for Flux Balance Analysis visualization.

Results

High-resolution SNV analysis identified multiple putative strain transmissions

Among the 13 distinct *Kp*SC STs previously reported from our flying fox isolate collection, five were associated with >1 isolate each (Fig. 2), four of which comprised pairs or groups of isolates that differed by ≤25 SNVs, indicative of recent strain transmissions (Gorrie *et al*., 2017; David *et al*., 2019; Sherry *et al*., 2021). Mapping data can be found in Table S2, while pairwise SNVs can be found in Table S3, ranging from 0 to 715.

There were three occurrences of putative intra-colony transmission at the Camellia gardens colony, including *K. pneumoniae* ST1017 (*n* = 2 isolates, three pairwise SNVs), *K. africana* ST4938 (*n* = 3 isolates, zero pairwise SNVs) and *K. pneumoniae* ST5033 (*n* = 4 isolates, ≤2 pairwise SNVs), and one occurrence at the captive Mylor colony of *K. pneumoniae* ST5035 (*n* = 15 of 16 isolates recovered from this colony, ≤3 pairwise SNVs) (Fig. 2). A single putative *K. africana* ST4938 inter-colony transmission event was observed between the Camellia Gardens and Blackalls Park colonies (*n* = 2 isolates, 21 pairwise SNVs). Notably, this inter-colony cluster was separated from the Camellia Gardens ST4938 intra-colony cluster and two unclustered isolates by just 36–58 pairwise SNVs, which may be indicative of longer-term circulation within the colony and/or acquisition from a common long-term reservoir (also see ST4938 phylogeny below).

Two *K. pneumoniae* ST4919 isolates, one from the Adelaide colony and the other from the Camellia Gardens colony (approximately 1150 km apart), were separated by 715 SNVs.

Plasmid diversity and transmission between species

A total of 11 plasmids were identified across the 13 completed genomes. Two isolates (*K. africana* ST4938, strain FF1003 and *K. pneumoniae* ST5033, strain FF979) harboured two plasmids each, seven isolates harboured a single plasmid each, while four isolates contained none
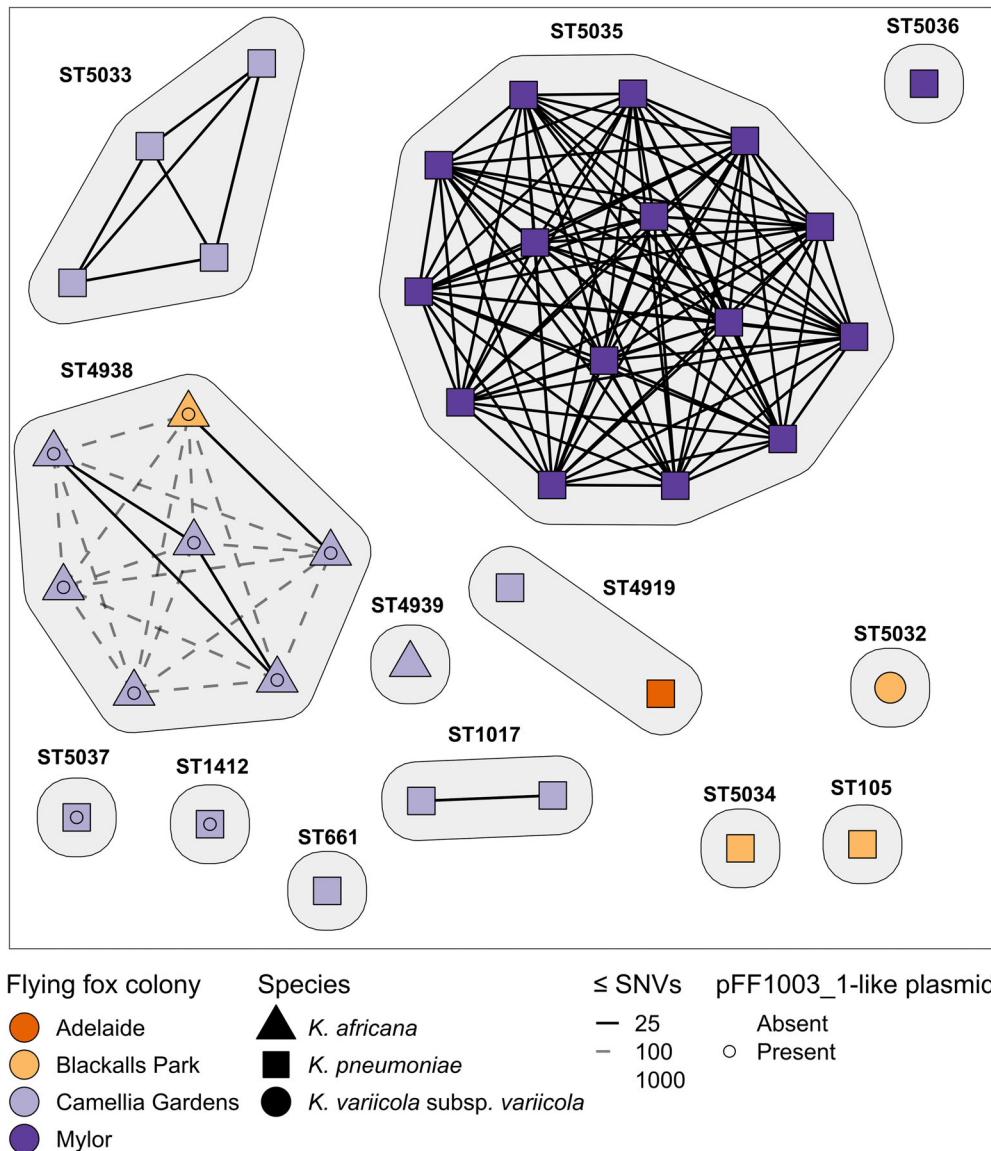
Fig. 2. Network graph showing pairwise SNV distances between isolates of the same ST. Isolates are shown as nodes, grouped by ST with the shape representing the species. The SNV distances are shown as lines between nodes. Isolates are considered part of a transmission cluster if they have ≤25 pairwise SNVs, represented by a solid line. Dashed lines show 26–100 pairwise SNVs. Isolates are coloured based on their grey-headed flying fox colony isolation source as indicated in the figure. The presence of the pFF1003-like plasmid is also indicated.

(Table S2). The plasmids ranged from approximately 5 to 175 kb in length (Fig. 3 and Table S4). At least seven of the 11 plasmids were predicted to be either self-conjugative or mobilizable and only one contained any antimicrobial resistance genes (*K. pneumoniae* ST1017, strain FF1009_1, carrying *qnrS1* and a variant of *dfrA*14) and none contained any of the well-characterized *Kp*SC virulence genes (Wyres *et al*., 2020). Nonetheless, all plasmids demonstrated high identity (≥95%) with generally high coverage (37%–98%; mean 73%) to at least one plasmid from human-derived *Kp*SC isolates (Fig. 3).

A reference-based mapping approach was used to determine the presence or absence of each of our completed plasmids among host isolates of the same ST, which revealed a high level of conservation (≥78% mapping coverage in all but three instances, comprised of three out of seven *K. africana* ST4938 isolates missing the pFF1003_2 plasmid) (Table S2). At most, plasmids differed among themselves by eight pairwise SNVs (Table S3). Notably, these included the pFF1009_1 AMR plasmid identified in both *K. pneumoniae* ST1017 isolates, and the ~5.5 kb pFF1043_1 plasmid identified
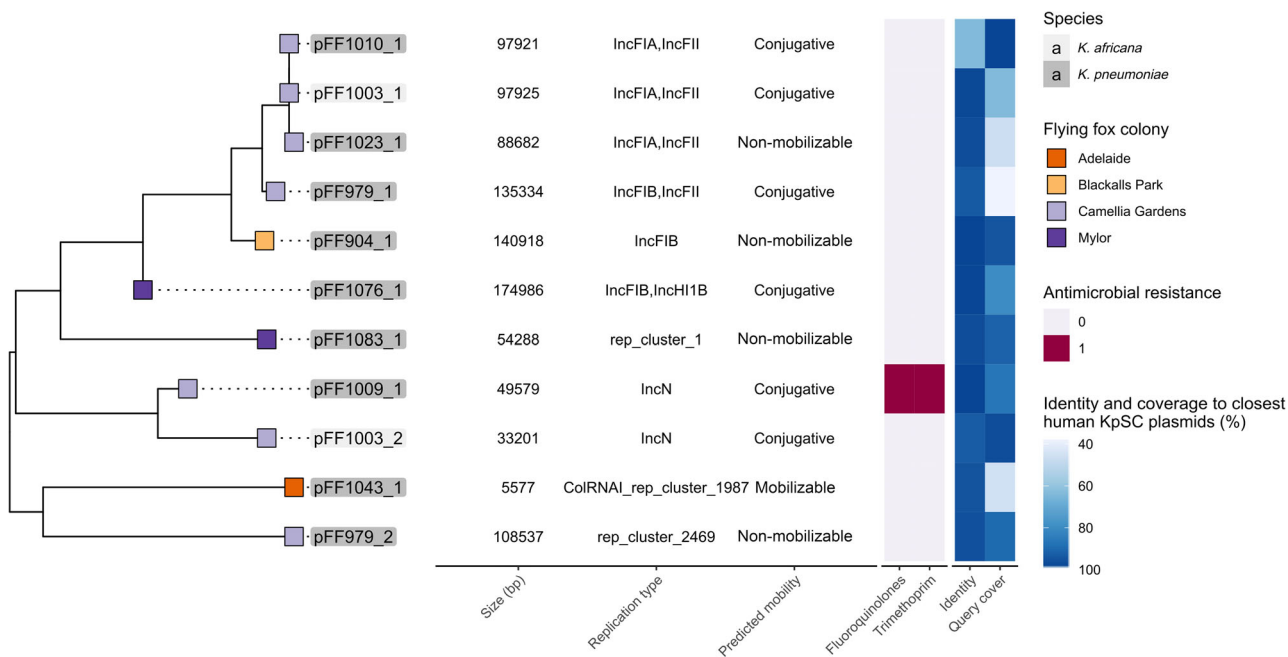
Fig. 3. MASH distance tree of plasmids found in the long read flying fox *Kp*SC genomes. Tree was inferred from MASH distances using the FastME algorithm. Tip colour indicates grey-headed flying fox colony, while colour of plasmid label denotes the species as per legend in figure. The corresponding heatmap shows the presence of antimicrobial resistance genes in pink, along with identity and coverage to the closest *Kp*SC plasmid isolated from a human source, identified by GenBank's BLASTn. Accession numbers can be found in Table S4.

among the *K. pneumoniae* ST4919 isolates collected from two independent colonies (Adelaide and Camellia Gardens). These contained zero and one pairwise SNVs to the completed pFF1009_1 and pFF1043_1 respectively.

Read mapping analyses indicated that a pFF1003_1-like conjugative plasmid was present in nine of the total 39 isolates (23%) (including those from both the Camellia Gardens and Blackalls Park colonies, maximum eight pairwise SNVs). These included representatives of *K. africana* ST4938, as well as *K. pneumoniae* ST5037 and ST1412 (Figs 3 and 4). Comparison of the completed sequences for pFF1003_1 (mapping reference from FF1003, a *K. africana* ST4938 isolate) and pFF1010_1 (from FF1010, a *K. pneumoniae* ST1412) showed 99% BLASTn coverage and 99% identity, indicative of recent plasmid transmission, though SNV distances were not calculated. However, a comparison between the pFF1003_1 reference and the completed sequence for pFF1023_1 (from FF1023, a *K. pneumoniae* ST5037) showed only 77% BLASTn coverage and 99% identity. Approximately 9 kb of pFF1023_1 was deleted compared to pFF1003_1 (at least six independent deletion events), including part of the conjugation locus. As such, this plasmid was predicted to be non-conjugative. The pFF1023_1 *nan* locus [consisting of *nanA*, *nanT*, *nanE2*, *nanK*, *yhcH*, a putative N-acetylneuraminic acid porin (GenBank accession:

HAT1683918.1), *nanC* and *nanM*, flanked by IS3 family transposases] was also inverted compared to the pFF1003_1 reference plasmid (Fig. 4C).

pFF1003_1 (97 kb) showed 64% BLASTn coverage and 99% identity to the publicly available 141 kb human-derived *Kp*SC plasmid (pINF116_1). This plasmid was identified from *K. pneumoniae* ST36 strain INF116, isolated from an Australian human urine sample in 2013 (GenBank accession: CP031793.1). Compared with pINF116, pFF1003_1 lacked many of the amino acid transport genes such as *hisQ*, *hisJ* and *glnQ*, but contained the additional tyrosine recombinase *xerC*, tRNA(fMet)-specific endonuclease *vapC* and the putatively intact sialic acid degradation *nan* locus (Fig. 4C).

Kp*SC from flying foxes were not distinguished by metabolic capabilities

*Kp*SC have specific nutrient and metabolic requirements for growth, which likely influence their host range. To predict whether or not flying fox isolates could potentially colonize humans or other sources and vice versa, we constructed genome-scale metabolic models for the grey-headed flying fox isolates along with *Kp*SC from other sources to identify metabolic commonalities and differences. We then simulated growth phenotypes *in silico* and overlaid these on the core genome phylogeny of
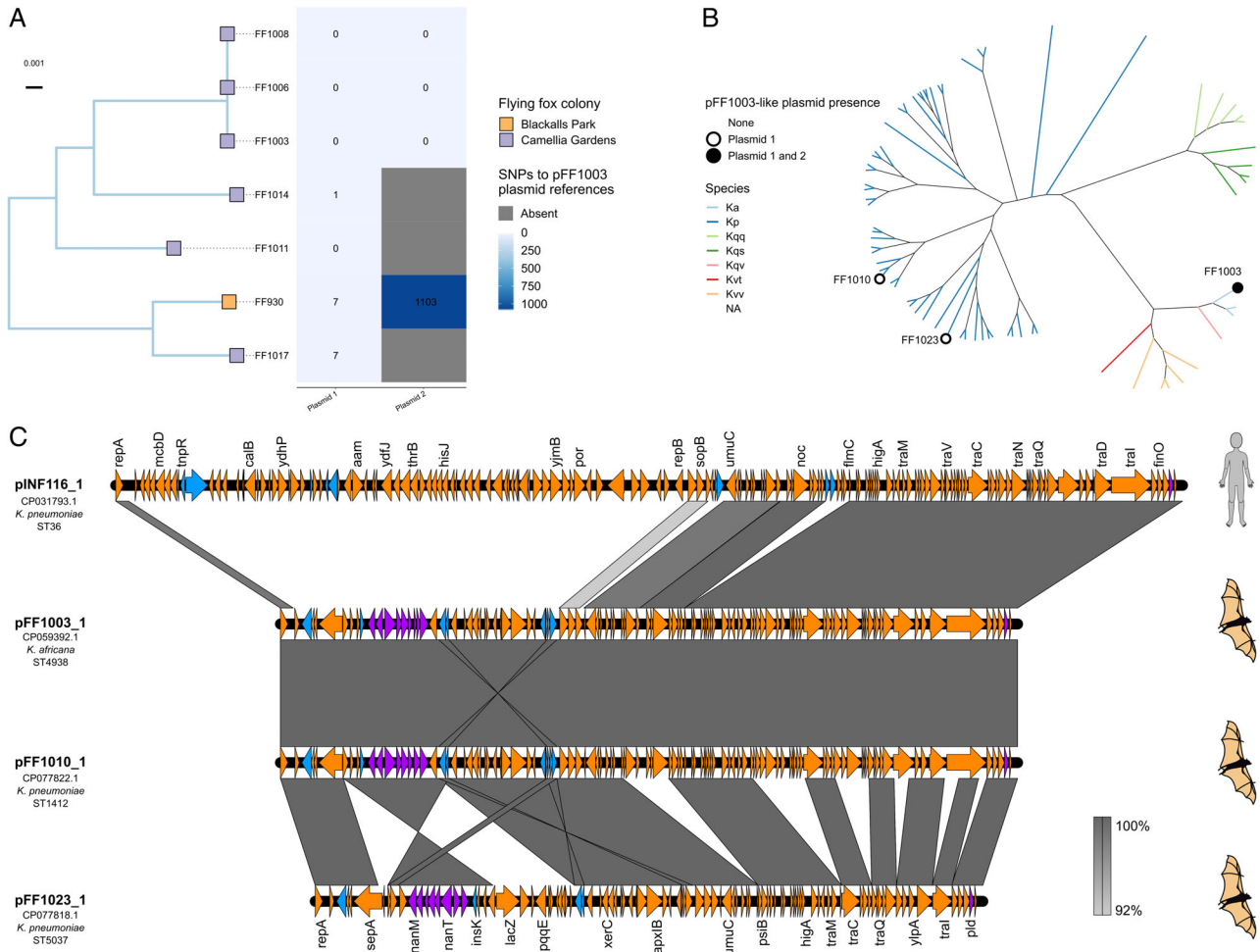
Fig. 4. Comparison and distribution of ST4938 strains and associated plasmid FF1003_1.
A. Mid-point rooted core SNV chromosomal maximum-likelihood phylogeny of seven *K. africana* ST4938 isolates. Colony is indicted by tip colour. The heatmap shows plasmid SNV counts compared to the pFF1003_1 and pFF1003_2 completed reference plasmids.
B. Unrooted core chromosomal maximum likelihood daylight phylogeny of all flying-fox derived isolates included in the study with dots showing the presence of pFF1003-like plasmids. Branch lengths were left unscaled to fit figure proportions.
C. Plasmid comparisons between pINF116_1 carried by *K. pneumoniae* INF116, isolated from a human source and flying fox plasmids pFF1003_1, pFF1010_1 and pFF1023_1. The orange arrows indicate genes, blue arrows indicate transposons and purple arrows indicate metabolic genes such as the sialic acid catabolism *nan* locus. The grey bars between maps indicate nucleotide identity as per the legend in the figure.

72 *Kp*SC genomes. All *Kp*SC isolates were predicted to utilize a core set of substrates accounting for 49.6% of carbon, 40% nitrogen, 72.9% of phosphorus and 40% of sulfur substrates tested (Tables S5 and S6). Percentage of total nutrient sources tested which displayed variable usage (non-core) across the *Kp*SC isolates was much lower, consisting of 10.7% carbon, 3.2% nitrogen, 1.7% phosphorus and 0% sulfur sources.

We generated a core gene phylogeny and overlaid the variable substrate usage profiles to examine the relationship between lineage and metabolic substrate usage (Fig. 5 and Fig. S1). Isolates from flying foxes were distributed throughout the phylogeny and predicted growth phenotypes appeared to cluster by species, then more

granularly by ST (Fig. 5), rather than isolate source. Hierarchical clustering of isolates based on predicted growth phenotypes was also consistent with this observation (Fig. S2), with isolates of the same species generally clustering together. One isolate, *K. pneumoniae* DHQP1002001 appeared as an outgroup to the other *K. pneumoniae* isolates in the phylogeny. This appeared to be a genetic species hybrid of mostly *K. pneumoniae* with roughly 1 Mb from a *K. quasipneumoniae* subsp. *quasipneumoniae* strain (Fig. S3). Notably, unlike the rest of the *K. pneumoniae* isolates, this species hybrid was predicted to be unable to utilize ethanolamine and D-glucarate as sole sources of carbon. While *K. quasipneumoniae* subsp. *quasipneumoniae* isolates
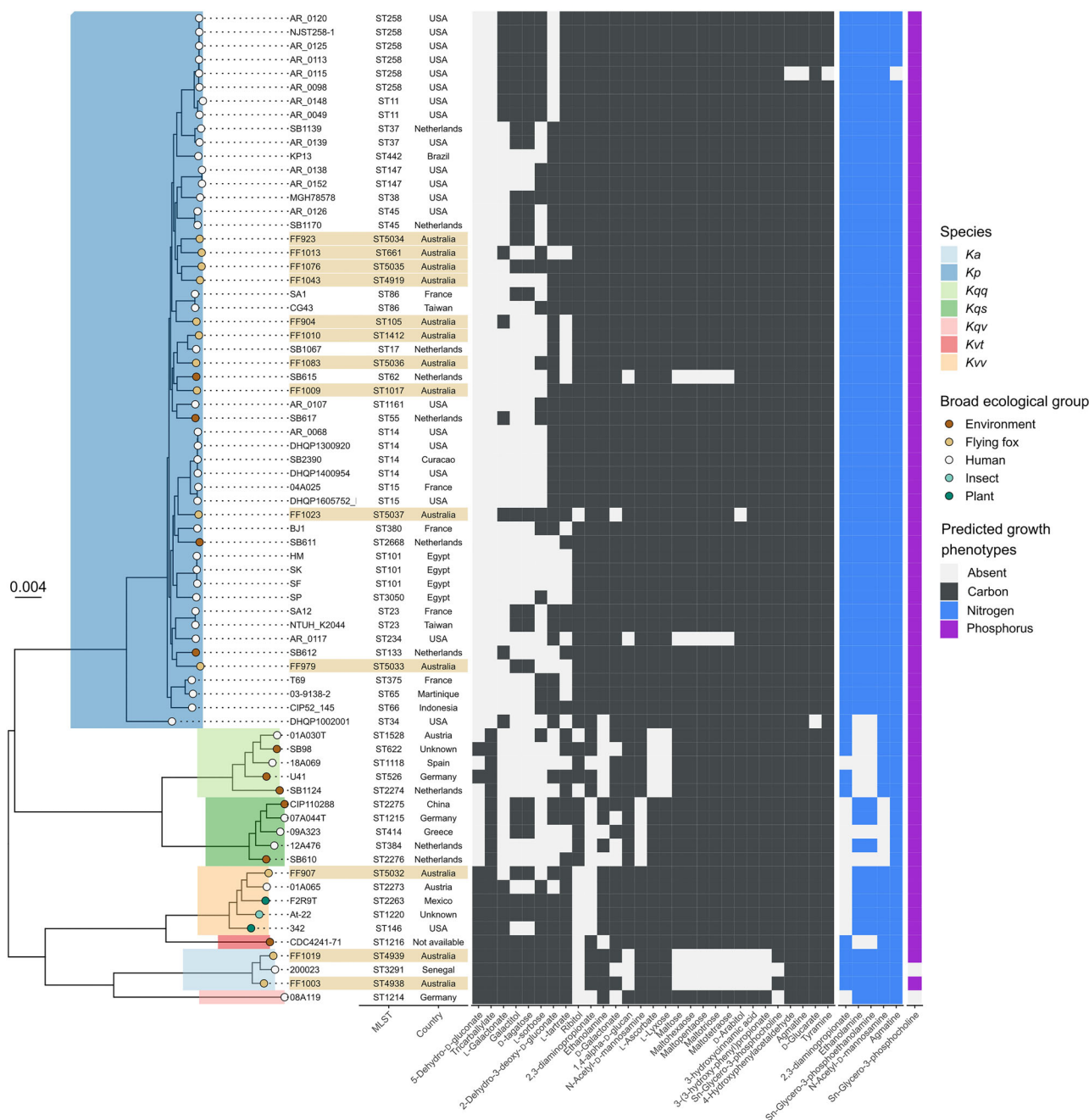
Fig. 5. Mid-point rooted core genome maximum likelihood phylogeny of 72 KpSC isolates and variable metabolic substrate use. Species are shaded on the phylogeny branches as per legend in the figure; *Ka* refers to *K. africana*, *Kp* refers to *K. pneumoniae*, *Kqq* refers to *K. quasipneumoniae* subsp. *quasipneumoniae*, *Kqs* refers to *K. quasipneumoniae* subsp. *similipneumoniae*, *Kqv* refers to *K. quasivariicola*, *Kvt* refers to *K. variicola* subsp. *tropica*, *Kvv* refers to *K. variicola* subsp. *variicola*. Broad ecological group/isolation source is shown by tip colour. Flying fox isolates are shown as light brown strips. Bootstrap values are shown in Fig. S1. Variable predicted growth substrate utilization is shown in the heatmap, coloured by substrate type as indicated. No variable sulfur usage was predicted.

were universally predicted to be unable to utilize the ethanolamine, they were all predicted to utilize tricarballylate and D-glucarate. It is unclear whether this pattern of substrate usage loss exists for *Klebsiella* hybrid species in general or was limited due to the data used.

Discussion

Our high-resolution genomic comparison of *Kp*SC from grey-headed flying fox colonies across Australia indicates strong evidence for strain transmission both within and between colonies (Fig. 2) because we identified near-

identical strains (very low number of chromosomal SNVs) from spatially separated samples that were likely derived from distinct animals (although it was not possible to associate any given sample with a given individual). While individual flying fox turnover within colonies was not accounted for, the small number of total transmission events was likely due to the rarity of *Kp*SC isolates from flying foxes in general, despite extensive sampling (McDougall *et al*., 2021b). Five putative strain transmission clusters were identified, four of which involved the wild Camellia Gardens colony in the state of New South Wales. This included three clusters contained within this colony and one transmission cluster between this colony and the wild colony at Blackalls Park. The latter was associated with *K. africana* ST4938 for which isolates FF1017 (Camellia Gardens) and FF930 (Blackalls Park), were separated by just 21 chromosomal SNVs (Fig. 2) and shared a pFF1003_1-like (~98 kb) plasmid with 0 pairwise SNVs, indicating they descended from a very recent common ancestor (although a second plasmid present in FF1017 was absent from FF930). Notably, these colonies are situated 123 km apart (Fig. 1) and were both sampled within a 23-day period. As grey-headed flying foxes have been shown to travel large distances of ~50 km (Tidemann and Nelson, 2004) and up to >1000 km in some cases (Roberts *et al*., 2012), it is possible for individual flying foxes to move between the Blackalls Park and Camellia Gardens colonies contributing bacterial strains to samples derived from both locations and/or causing direct inter-colony transmission. Alternatively, transmission may have occurred indirectly via a common source (food/water etc.). Additionally, our analyses indicated that the two ST4938 clusters, as well as two additional unclustered ST4938 isolates from Camellia Gardens, differed by no more than 58 pairwise SNVs (Fig. 2 and Table S3). These values are much lower than would be expected for randomly selected *Kp*SC of the same ST (Gorrie *et al*., 2018; David *et al*., 2019) and similar to those observed for isolates representing longer-term transmission within hospital networks (Gorrie *et al*., 2018, David *et al*., 2019). Hence, these findings are consistent with longer-term transmission in the flying fox population or acquisition from the same long-term reservoir rather than replicate sampling from a single animal, although additional longitudinal sampling of these colonies is required for confirmation.

The fifth putative strain transmission cluster was identified at the captive Mylor colony where animals were individually sampled. The colony was located within a rehabilitation centre, which at the time of sampling was a closed population of 30–40 adult grey-headed flying foxes from the local wild Adelaide Botanic Park colony that had been affected by a heat stress event. These unhealthy individuals may have suffered increased susceptibility to opportunistic colonization and dissemination of *Kp*SC isolates, e.g. from other wildlife or domestic animals that were housed in close proximity, which we speculate facilitated the broad-scale dissemination of *K. pneumoniae* ST5035 in the population [isolated from 15 of 20 sampled individuals (McDougall *et al*., 2021b)].

There is currently very limited data on the dynamics of bacterial transmission among flying foxes and other bat species; however, our results align with a previous study of viral dynamics (Epstein *et al*., 2020), which found transmission of Nipah virus in Indian flying foxes within Bangladesh occurred within and between colonies. This study also noted the cycles of Nipah virus prevalence could be influenced by colony demographics, such as individual bat turnover which reduced herd immunity (Epstein *et al*., 2020). Grey-headed flying foxes in colonies across eastern Australia have an average daily individual turnover of 17.5% as they travel between roosts via a node-based network (Welbergen *et al*., 2020), indicating high influx and efflux of individuals. This behavioural activity may make grey-headed flying fox colonies susceptible to between-colony transmission of microbes such as *Kp*SC; however, further sampling is required to quantify these transmission events and determine if they represent rare, transient or ongoing phenomena.

In addition to strain transmissions, we identified putative transmission of a 98 kb plasmid, passing the species barrier between *K. africana* ST4938 and *K. pneumoniae* ST1412 and ST5037 (Figs 3 and 4). The plasmid variants harboured by FF1003 (*K. africana* ST4938) and FF1010 (*K. pneumoniae* ST1412) were highly similar (four pairwise SNVs, 100% coverage and 99% nucleotide identity), supporting a recent transmission event (independent parallel evolution of these replicons is possible, but unlikely). In contrast, the plasmid variant harboured by FF1023 (*K. pneumoniae* ST5037) showed at least six deletions as well as a structural rearrangement, indicating a plasmid transmission likely occurred at some point in the more distant evolutionary history of these strains (either directly or indirectly via intermediary hosts). In that regard, it is possible that the pFF1003-like plasmid may have been circulating within the Camellia Gardens colony for some time and/or has been introduced from a common reservoir multiple times. Aside from the presence of several selfish genetic elements including a *higAB* toxin–antitoxin system, intact conjugative *tra* locus (except in the case of pFF1023_1) (Fig. 4C), this pFF1103-like plasmid contained a putatively intact *nan* sialic acid locus that was conserved in all variants. The *nan* locus encodes degradation of N-acetylneuraminic acid, a sialic acid (Olson *et al*., 2013). Utilization and acquisition of N-acetylneuraminic acid improve colonization and provide competitive growth advantages to a variety of pathogens (Chang *et al*., 2004; Manco *et al*., 2006; Almagro-Moreno

and Boyd, 2009). It is possible that the presence of this locus provided some sort of selective advantage to promote the transmission and longer-term maintenance of the pFF1103-like plasmid in the Camellia Gardens population, although further investigations would be required to confirm this hypothesis.

Plasmid pFF1003 also showed high similarity to a plasmid identified from INF116, a *K. pneumoniae* clinical isolate causing a urinary tract infection in a patient in a Melbourne hospital in 2013 (Wyres *et al.*, 2019). In fact, our search of publicly available nucleotide sequence data indicated that all completed plasmids from flying fox-derived *Kp*SC shared high identity (≥95%) with generally high coverage (37%–98%; mean 73%) (Table S4) to at least one other plasmid identified from human clinical isolates. Hence, our data provide evidence for the flow of genetic material either directly or indirectly between *Kp*SC isolated from flying foxes and human specimens at some point in their evolutionary history. In order for direct transfer to occur these isolate populations would need to coexist, at least transiently, within the same spatial–temporal and ecological niche. Alternatively, indirect gene flow through intermediate plasmid hosts would also explain the presence of these highly similar plasmids. However, we cannot speculate on the directionality, frequency or time frames over which such transfers may occur without larger, contemporaneous sample collections from both flying fox and human populations.

While much more work is required to fully understand the distribution and flow of *Kp*SC strains and genetic material between niches (including detailed genomic analysis of large contemporaneous isolate collections), the results of the preliminary metabolic analyses presented here are consistent with niche overlap (i.e. *Kp*SC from flying foxes did not appear to be metabolically distinct). We hypothesised that there may be metabolic differences between *Kp*SC from flying foxes and those from other sources because occupation of different niches/hosts is likely associated with different nutrient availability, which are impacted by elements such as host diet. We note that our analysis was constrained to a small sample size and that there are likely many metabolic traits which are uncaptured within the metabolic models used in this study, as well as limitations of the current input database. Over time, this approach should become more robust as databases become more complete (https://www.genome.jp/kegg/docs/upd_map.html) (Kanehisa and Goto, 2000). There may also be other genetic differences between flying fox and human strains of *Kp*SC such as K/O loci, previously discussed within the context of flying fox isolates by McDougall *et al.* (2021b); however, due to a paucity of grey-headed flying fox isolates, we lack the statistical power to discern this.

Data Availability

These sequence data have been submitted to the GenBank databases under BioProject accession number PRJNA646592. All other data used in this study can be found in supplemental tables and at Figshare (https://doi.org/10.6084/m9.figshare.16609054) (Vezina *et al.*, 2021).

References

Ahmed, H.A., Awad, N.F.S., Abd El-Hamid, M.I., Shaker, A., Mohamed, R.E., and Elsohaby, I. (2021) Pet birds as potential reservoirs of virulent and antibiotic resistant zoonotic bacteria. *Comp Immunol Microbiol Infect Dis* **75**: 101606.

Almagro-Moreno, S., and Boyd, E.F. (2009) Sialic acid catabolism confers a competitive advantage to pathogenic *Vibrio cholerae* in the mouse intestine. *Infect Immun* **77**: 3807–3816.

Anaconda-Software-Distribution. (2016) Anaconda.

Auguie, B. and Antonov, A. (2017) Package 'gridExtra'.

Awasthi, S., Srivastava, P., Singh, P., Tiwary, D., and Mishra, P.K. (2017) Biodegradation of thermally treated high-density polyethylene (HDPE) by *Klebsiella pneumoniae* CH001. *3 Biotech* **7**: 332.

Bushnell, B. (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner.

Campitelli, E. (2021) ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'.

Chang, D.-E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., *et al.* (2004) Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc Natl Acad Sci U S A* **101**: 7427–7432.

Csardi, G., and Nepusz, T. (2005) The Igraph software package for complex network research. *InterJournal Complex Systems* **1695**: 1–9.

David, S., Reuter, S., Harris, S.R., Glasner, C., Feltwell, T., Argimon, S., *et al.* (2019) Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nat Microbiol* **4**: 1919–1929.

Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013) COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst Biol* **7**: 74.

Edwards, D., Pope, B., and Holt, K. (2016) RedDog.

Epstein, J.H., Anthony, S.J., Islam, A., Kilpatrick, A.M., Ali Khan, S., Balkey, M.D., *et al*. (2020) Nipah virus dynamics in bats and implications for spillover to humans. *Proc Natl Acad Sci U S A* **117**: 29190–29201.

Fraser, G.C., Hooper, P.T., Lunt, R.A., Gould, A.R., Gleeson, L.J., Hyatt, A.D., *et al*. (1996) Encephalitis caused by a lyssavirus in fruit bats in Australia. *Emerg Infect Dis J* **2**: 327–331.

Gharout-Sait, A., Touati, A., Ahmim, M., Brasme, L., Guillard, T., Agsous, A., and de Champs, C. (2019) Occurrence of Carbapenemase-producing *Klebsiella pneumoniae* in Bat Guano. *Microb Drug Resist* **25**: 1057–1062.

Gorrie, C.L., Mirčeta, M., Wick, R.R., Edwards, D.J., Thomson, N.R., Strugnell, R.A., *et al*. (2017) Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clin Infect Dis* **65**: 208–215.

Gorrie, C.L., Mirceta, M., Wick, R.R., Judd, L.M., Wyres, K.L., Thomson, N.R., *et al*. (2018) Antimicrobial-resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital. *Clin Infect Dis* **67**: 161–170.

Håkonsholm, F., Hetland, M.A.K., Svanevik, C.S., Sundsfjord, A., Lunestad, B.T., and Marathe, N.P. (2020) Antibiotic sensitivity screening of *Klebsiella* spp. and *Raoultella* spp. isolated from marine bivalve molluscs reveal presence of CTX-M-producing *K. pneumoniae*. *Microorganisms* **8**: 1909.

Halpin, K., Young, P.L., Field, H.E., and Mackenzie, J.S. (2000) Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. *J Gen Virol* **81**: 1927–1932.

Hamza, E., Dorgham, S.M., and Hamza, D.A. (2016) Carbapenemase-producing *Klebsiella pneumoniae* in broiler poultry farming in Egypt. *J Glob Antimicrob Resist* **7**: 8–10.

Hawkey, J., Vezina, B., Monk, J.M., Judd, L.M., Harshegyi, T., López-Fernández, S., *et al*. (2022) A curated collection of *Klebsiella* metabolic models reveals variable substrate usage and gene essentiality. *Genome Research*. https://doi.org/10.1101/gr.276289.121

Hladicz, A., Kittinger, C., and Zarfel, G. (2017) Tigecycline resistant *Klebsiella pneumoniae* isolated from Austrian River Water. *Int J Environ Res Public Health* **14**: 1169.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**: W5–W9.

Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27–30.

Knittel, M.D., Seidler, R.J., Eby, C., and Cabe, L.M. (1977) Colonization of the botanical environment by *Klebsiella* isolates of pathogenic origin. *Appl Environ Microbiol* **34**: 557–563.

Kolde, R. (2019) pheatmap: pretty Heatmaps.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**: 540–546.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Lam, M.M.C., Wick, R.R., Watts, S.C., Cerdeira, L.T., Wyres, K.L., and Holt, K.E. (2021) A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat Commun* **12**: 4188.

Lefort, V., Desper, R., and Gascuel, O. (2015) FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* **32**: 2798–2800.

Liao, X., Yang, R.-S., Xia, J., Chen, L., Zhang, R., Fang, L.-X., *et al*. (2019) High colonization rate of a novel carbapenem-resistant *Klebsiella* lineage among migratory birds at Qinghai Lake, China. *J Antimicrob Chemother* **74**: 2895–2903.

Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., *et al*. (2020) MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol* **38**: 272–276.

Manco, S., Hernon, F., Yesilkaya, H., Paton, J.C., Andrew, P.W., and Kadioglu, A. (2006) Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infect Immun* **74**: 4014–4020.

Marques, C., Menezes, J., Belas, A., Aboim, C., Cavaco-Silva, P., Trigueiro, G., *et al*. (2019) *Klebsiella pneumoniae* causing urinary tract infections in companion animals and humans: population structure, antimicrobial resistance and virulence genes. *J Antimicrob Chemother* **74**: 594–602.

Massé, J., Dufour, S., and Archambault, M. (2020) Characterization of *Klebsiella* isolates obtained from clinical mastitis cases in dairy cattle. *J Dairy Sci* **103**: 3392–3400.

McDougall, F.K., Boardman, W.S.J., and Power, M.L. (2021a) Characterization of beta-lactam-resistant *Escherichia coli* from Australian fruit bats indicates anthropogenic origins. *Microb Genomics* **7**: 000571.

McDougall, F.K., Wyres, K.L., Judd, L.M., Boardman, W.S.J., Holt, K.E., and Power, M.L. (2021b) Novel strains of *Klebsiella africana* and *Klebsiella pneumoniae* in Australian Fruit Bats (*Pteropus poliocephalus*). *Res Microbiol* **172**: 103879.

Melo-Nascimento, A.O.D.S., Treumann, C., Neves, C., Andrade, E., Andrade, A.C., Edwards, R., *et al*. (2018) Functional characterization of ligninolytic *Klebsiella* spp. strains associated with soil and freshwater. *Arch Microbiol* **200**: 1267–1278.

Monk, J.M., Charusanti, P., Aziz, R.K., Lerman, J.A., Premyodhin, N., Orth, J.D., *et al*. (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A* **110**: 20338–20343.

Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., *et al*. (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol* **35**: 904–908.

nanoporetech. (2020) *"Medaka." 1.0.3*. URL https://github.com/nanoporetech/medaka.

Navon-Venezia, S., Kondratyeva, K., and Carattoli, A. (2017) *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol Rev* **41**: 252–275.

Neuwirth, E. (2014) RColorBrewer: ColorBrewer Palettes.

Norsigian, C.J., Attia, H., Szubin, R., Yassin, A.S., Palsson, B.Ø., Aziz, R.K., and Monk, J.M. (2019) Comparative genome-scale metabolic modeling of metallo-beta-lactamase–producing multidrug-resistant *Klebsiella pneumoniae* clinical isolates. *Front Cell Infect Microbiol* **9**: 161.

Norsigian, C.J., Fang, X., Seif, Y., Monk, J.M., and Palsson, B.O. (2020) A workflow for generating multi-strain genome-scale metabolic models of prokaryotes. *Nat Protoc* **15**: 1–14.

Olson, M.E., King, J.M., Yahr, T.L., and Horswill, A.R. (2013) Sialic acid catabolism in *Staphylococcus aureus*. *J Bacteriol* **195**: 1779–1788.

Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* **17**: 132.

Paradis, E., Claude, J., and Strimmer, K. (2004) APE: analyses of Phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

Pedersen, T. L. (2021a) Package 'ggforce'.

Pedersen, T. L. (2021b) Package 'ggraph'.

Pendleton, J.N., Gorman, S.P., and Gilmore, B.F. (2013) Clinical relevance of the ESKAPE pathogens. *Expert Rev Anti Infect Ther* **11**: 297–308.

Philbey, A.W., Kirkland, P.D., Ross, A.D., Field, H.E., Srivastava, M., Davis, R.J., and Love, R.J. (2008) Infection with Menangle virus in flying foxes (*Pteropus* spp.) in Australia. *Aust Vet J* **86**: 449–454.

R Core Team. (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roberts, B.J., Catterall, C.P., Eby, P., and Kanowski, J. (2012) Long-distance and frequent movements of the flying-fox *Pteropus poliocephalus*: implications for management. *PLoS One* **7**: e42532.

Robertson, J., and Nash, J.H.E. (2018) MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genomics* **4**: e000206.

Rodrigues, C., Passet, V., Rakotondrasoa, A., Diallo, T.A., Criscuolo, A., and Brisse, S. (2019) Description of *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and *Klebsiella variicola* subsp. *variicola* subsp. nov. *Res Microbiol* **170**: 165–170.

Rosenblueth, M., Martínez, L., Silva, J., and Martínez-Romero, E. (2004) *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. *Syst Appl Microbiol* **27**: 27–35.

Rossum, G.v., and Python-Software-Foundation. (1995) *Python Tutorial, Technical Report CS-R9526*. Amsterdam: Centrum voor Wiskunde en Informatica (CWI).

RStudio-Team. (2020) *RStudio: Integrated Development for R*. Boston: MA RStudio.

Schliep, K.P. (2011) Phangorn: phylogenetic analysis in R. *Bioinformatics* **27**: 592–593.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Seif, Y., Kavvas, E., Lachance, J.-C., Yurkovich, J.T., Nuccio, S.-P., Fang, X., et al. (2018) Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nat Commun* **9**: 3771.

Sherry, N.L., Lee, R.S., Gorrie, C.L., Kwong, J.C., Stuart, R. L., Korman, T.M., et al. (2021) Pilot study of a combined genomic and epidemiologic surveillance program for hospital-acquired multidrug-resistant pathogens across multiple hospital networks in Australia. *Infect Control Hosp Epidemiol* **42**: 573–581.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009–1010.

Thiele, I., and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* **5**: 93–121.

Tidemann, C.R., and Nelson, J.E. (2004) Long-distance movements of the grey-headed flying fox (*Pteropus poliocephalus*). *J Zool* **263**: 141–146.

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., et al. (2020) Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* **21**: 180. https://doi.org/10.1186/s13059-020-02090-4.

Vezina, B., Judd, L. M., McDougall, F. K., Boardman, W. S. J., Power, M. L., Brisse, S., Monk, J. M., Holt, K. E., and Wyres, K. L. (2021) *Klebsiella* genome-scale metabolic models isolated from grey-headed flying foxes. Figshare.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.

Watts, S. (2018) GenomePainter.

Welbergen, J.A., Meade, J., Field, H.E., Edson, D., McMichael, L., Shoo, L.P., et al. (2020) Extreme mobility of the world's largest flying mammals creates key challenges for management and conservation. *BMC Biol* **18**: 101.

Wickham, H. (2007) Reshaping data with the reshape package. *J Stat Softw* **21**: 1–20.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wyres, K.L., Lam, M.M.C., and Holt, K.E. (2020) Population genomics of *Klebsiella pneumoniae*. *Nat Rev Microbiol* **18**: 344–359.

Wyres, K.L., Wick, R.R., Judd, L.M., Froumine, R., Tokolyi, A., Gorrie, C.L., et al. (2019) Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. *PLoS Genet* **15**: e1008114.

Yu, G. (2020) aplot: Decorate a 'ggplot' with Associated Information.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**: 28–36.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Supporting Information.