University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Wick, RR;Heinz, E;Holt, KE;Wyres, KL

Title:
Kaptive web: User-Friendly capsule and lipopolysaccharide serotype prediction for Klebsiella genomes

Date:
2018-06-01

Citation:
Wick, R. R., Heinz, E., Holt, K. E. & Wyres, K. L. (2018). Kaptive web: User-Friendly capsule and lipopolysaccharide serotype prediction for Klebsiella genomes. Journal of Clinical Microbiology, 56 (6), https://doi.org/10.1128/JCM.00197-18.

Persistent Link:
https://hdl.handle.net/11343/254806

License:
CC BY

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

EPIDEMIOLOGY

Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes

Ryan R. Wick,[a] Eva Heinz,[b] Kathryn E. Holt,[a] Kelly L. Wyres[a]

[a]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Victoria, Australia

[b]Wellcome Trust Sanger Institute, Hinxton, South Cambridgeshire, United Kingdom

ABSTRACT As whole-genome sequencing becomes an established component of the microbiologist's toolbox, it is imperative that researchers, clinical microbiologists, and public health professionals have access to genomic analysis tools for the rapid extraction of epidemiologically and clinically relevant information. For the Gram-negative hospital pathogens such as *Klebsiella pneumoniae*, initial efforts have focused on the detection and surveillance of antimicrobial resistance genes and clones. However, with the resurgence of interest in alternative infection control strategies targeting *Klebsiella* surface polysaccharides, the ability to extract information about these antigens is increasingly important. Here we present Kaptive Web, an online tool for the rapid typing of *Klebsiella* K and O loci, which encode the polysaccharide capsule and lipopolysaccharide O antigen, respectively. Kaptive Web enables users to upload and analyze genome assemblies in a web browser. The results can be downloaded in tabular format or explored in detail via the graphical interface, making it accessible for users at all levels of computational expertise. We demonstrate Kaptive Web's utility by analyzing >500 *K. pneumoniae* genomes. We identify extensive K and O locus diversity among 201 genomes belonging to the carbapenemase-associated clonal group 258 (25 K and 6 O loci). The characterization of a further 309 genomes indicated that such diversity is common among the multidrug-resistant clones and that these loci represent useful epidemiological markers for strain subtyping. These findings reinforce the need for rapid, reliable, and accessible typing methods such as Kaptive Web. Kaptive Web is available for use at http://kaptive.holtlab.net/, and the source code is available at https://github.com/kelwyres/Kaptive-Web.

KEYWORDS *Klebsiella*, capsular polysaccharide, capsule, genome analysis, genomics, lipopolysaccharide, molecular epidemiology

Whole-genome sequencing (WGS) represents a powerful tool for the characterization and public health surveillance of bacterial pathogens. This technology is now routinely used by a number of public health laboratories (1–3), and there is increasing interest in its use in clinical labs (4–6). While there are several well-developed protocols which use WGS data for the determination of multilocus sequence types (STs), resistance gene profiling, and phylogenetic investigations (7–9), there remain gaps in the repertoire; e.g., the characterization of species-specific antigens is currently restricted to a small number of species (10–12). Furthermore, many WGS analyses rely on software via a command line interface that requires bioinformatics skills to install, execute, and interpret, thereby limiting their accessibility. Instead, we need tools that can extract information and present it in an easily interpretable manner to bioinformaticians, public health professionals, and clinicians alike (5, 7, 8).

Klebsiella pneumoniae is a major cause of health care-associated infections with high

Received 8 February 2018 **Returned for modification** 5 March 2018 **Accepted** 28 March 2018

Accepted manuscript posted online 4 April 2018

Citation Wick RR, Heinz E, Holt KE, Wyres KL. 2018. Kaptive Web: user-friendly capsule and lipopolysaccharide serotype prediction for *Klebsiella* genomes. J Clin Microbiol 56:e00197-18. https://doi.org/10.1128/JCM.00197-18.

Editor Daniel J. Diekema, University of Iowa College of Medicine

Copyright © 2018 Wick et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Ryan R. Wick, rrwick@gmail.com.

rates of multidrug resistance (MDR) (13). In particular, the emergence and global dissemination of extended-spectrum beta-lactamase (ESBL) and carbapenemase-producing (CP) clones are major concerns and have led to the recognition of *K. pneumoniae* as an urgent public health threat (14, 15). With the lack of new antimicrobial therapies, there has been a resurgence of interest in alternative strategies, such as phage therapy (16–19), monoclonal antibody therapy (20–23), and vaccination (24–26). Several therapeutic targets have been suggested, and the polysaccharide capsule (K antigen) and lipopolysaccharide (O antigen) are among the most frequent. Both are also considered key virulence determinants that are necessary to establish infection, primarily owing to their serum resistance and antiphagocytic properties (27–31). Of note, capsular serotypes vary substantially in the degrees of serum resistance they provide. For example, K1, K2, and K5 are highly serum resistant and are associated with hypervirulent strains that differ from classical *K. pneumoniae* in that they commonly cause community-acquired disease (32–34). Despite the importance of these loci, *K. pneumoniae* serotyping is not widely available, even in large central public health laboratories, and the most practical option for most laboratories is genotyping the loci involved in antigen biosynthesis via multiplex PCR (35, 36) or WGS (37, 38).

Lipopolysaccharide comprises three subunits: lipid A, the core oligosaccharide, and the O antigenic polysaccharide (39). With only one exception, the key determinants of the O antigenic polysaccharide are colocated at the O locus (previously known as the *rfb* locus) (36, 40–42). While 10 serologically distinct O antigens have been recognized, many isolates are nontypeable (23, 43), and investigations have identified 12 distinct O loci (25, 36). Interestingly, both the O1 and O2 antigens, which are by far the most common (23, 25, 43), are each associated with the same two loci, O1/O2v1 and O1/O2v2 (25). The expression of the v1 locus results in the production of D-galactan I, characteristic to a subset of O2 antigens (41, 44). The expression of the v2 locus results in the production of D-galactan III, associated with the remaining O2 subtypes (21, 45). Regardless of the subtype, any O2 antigen can be converted to O1 by the addition of D-galactan II, which requires the products of *wbbY* and *wbbZ* that are located outside the O locus, i.e., elsewhere in the genome (44).

The *Klebsiella* polysaccharide capsule is produced through a Wzy-dependent process (46), for which the synthesis and export machinery are encoded in a single 10 to 30-kbp region of the genome known as the K locus (47, 48). Seventy-seven distinct capsule phenotypes have been recognized by serological typing (49), but many isolates are serologically nontypeable. We recently explored the K loci among a large diverse *K. pneumoniae* WGS collection and were able to define 134 distinct loci on the basis of protein coding gene content, suggesting there are at least this many distinct capsule types circulating in the population (38).

Given the interest in targeting these diverse surface polysaccharides and the lack of accessible serotyping assays for *K. pneumoniae*, tools for WGS-based K and O locus typing will be essential for researchers, clinicians, and public health microbiologists. We previously developed Kaptive for K locus typing from WGS assemblies (38), which has become a key component of the *Klebsiella* bioinformatics tool kit (50–52), but it requires command-line skills and a degree of bioinformatics expertise to use. Here we present Kaptive Web, an easy-to-use web-based implementation of the Kaptive algorithm which has been extended to type both K and O loci. We demonstrate its utility (i) for the identification of K and O loci for serotype prediction and as epidemiological markers and (ii) to inform the design and implementation of control strategies targeting the capsules or lipopolysaccharides of *K. pneumoniae*.

MATERIALS AND METHODS

O locus definitions. Unlike K loci which were defined on the basis of gene content, O loci have been defined by sequence identity in the conserved *wzm* and *wzt* genes (25). For example, two O loci can have the same genes, but modestly divergent sequences (>5% divergence). Kaptive is compatible with these definitions because it first chooses a best locus on the basis of a nucleotide search. Only then does it tally the gene content of the locus.

A complication comes from the fact that O antigens O1 and O2 are encoded by the same two O loci. It is the presence or absence of two other genes elsewhere in the genome, *wbbY* and *wbbZ*, which determines the specific antigen. When both genes are present, D-galactan II is produced, leading to the O1 antigen. When they are absent, the O2 antigen is the result. We have added the appropriate logic to Kaptive (both command-line Kaptive and Kaptive Web), so it will report the locus as O1 or O2 on the basis of the presence/absence of these genes as determined by a tBLASTn search with coverage and identity thresholds of 90% and 80%, respectively. If Kaptive finds only one of the two genes, it will report the locus as O1/O2.

A Kaptive-compatible O locus reference database comprising the O loci reported in reference 25 plus the O8 locus (GenBank accession AB819963.1) is available for download via the command-line Kaptive GitHub page (https://github.com/katholt/Kaptive).

Kaptive Web. Kaptive Web is available for use at http://kaptive.holtlab.net/. It was developed using the web2py framework (53). The source code for the web implementation is available on GitHub, so users can host their own copy of the software (https://github.com/kelwyres/Kaptive-Web). Kaptive Web automatically populates the "reference database" selection with the contents of command-line Kaptive's database directory, enabling automatic compatibility with any new locus databases for *Klebsiella* or other bacteria. Kaptive Web utilizes a 16-core 64-GB RAM server hosted by the Australian NeCTAR cloud. A single genome analysis requires approximately 4 min and 20 s to complete for the K and O locus databases, respectively. The server can run up to 15 analyses simultaneously, enabling large data sets to be processed relatively quickly.

Genome data for K and O locus characterization. Sequence read data for 309 *K. pneumoniae* organisms were obtained as part of the global diversity study (54), and 13 O3 antigen-producing isolates (20) were assembled *de novo* using Unicycler v0.4.1 (55). Genome assemblies were uploaded to Kaptive Web in a single compressed data directory and analyzed with the *Klebsiella* primary K locus and the *Klebsiella* O locus databases. The total Kaptive Web analysis times for the global data set were 52 min (K locus) and 12 min (O locus). The results were inspected via the Kaptive Web graphical interface and downloaded in tabular format (see Data Set S1 in the supplemental material).

The same protocol was used for characterization of 201 publicly available CG258 genome assemblies (see Data Set S3). These genomes were identified among the complete set of *Klebsiella* genomes (downloaded from GenBank on 12 October 2017) on the basis of ST information generated using Kleborate (https://github.com/katholt/Kleborate). STs 11, 258, 340, 395, 437, 512, 855, and 895 were included in the analyses.

RESULTS AND DISCUSSION

Introducing Kaptive Web. Kaptive Web is a browser-based method for running Kaptive and visualizing the results. Users upload one or more assemblies and select their preferred typing database (Fig. 1A). There is no limit to the number of assemblies that can be uploaded for a single run, though multiple assemblies must be uploaded as a tar.gz or zip file. After the upload is complete, command-line Kaptive is automatically run on the remote server (Fig. 1B). Results appear in a table with one row per genome assembly showing key details: the best-matching locus from the reference database, the match confidence, nucleotide identity, and coverage compared to the reference (Fig. 1C). The rows are colored on the basis of the match confidence, with six possible levels.

(i) Perfect. The locus was found in a single piece (one alignment within a single contig) with 100% coverage and 100% nucleotide identity to the reference.

(ii) Very high. The locus was found in a single piece with ≥99% coverage and ≥95% nucleotide identity to the reference, with no truncated/missing genes and no extra genes compared to the reference.

(iii) High. The locus was found in a single piece with ≥99% coverage, with ≤3 truncated/missing genes and no extra genes compared to the reference.

(iv) Good. The locus was found in a single piece or with ≥95% coverage, with ≤3 truncated/missing genes and ≤1 extra gene compared to the reference.

(v) Low. The locus was found in a single piece or with ≥90% coverage, with ≤3 truncated/missing genes and ≤2 extra genes compared to the reference.

(vi) None. Did not qualify for any of the above.

The top two confidence levels, "very high" and "perfect," require the locus to be found uninterrupted in a single contig with the expected gene content. If the locus has truncated/missing genes or was found in multiple discontiguous pieces, a lower confidence level will result. Insertion sequence (IS) integrations are one possible cause of truncated/missing genes. When an IS interrupts a locus gene, that gene's function is likely lost (38), and for short-read Illumina sequencing, IS integration also typically



FIG 1 Kaptive Web screenshots. (A) Users choose a file or zipped directory on their local computer and select the desired reference database. (B) Job information screen detailing the number of assemblies uploaded and their status. (C) Kaptive Web results for four *Klebsiella* isolates, with the first result expanded.

causes assembly fragmentation (56). However, truncated/missing genes and assembly fragmentation can also result from poor read coverage or indel sequencing errors, in which case, the gene is likely still functional. This is why Kaptive is somewhat tolerant of truncated/missing genes and discontiguity—a locus with these issues can still achieve a confidence of "high." Extra genes in a locus (genes that are not in the best-matching locus reference but are in another locus reference) are strongly indicative of a biological change and thus reduce confidence more so than missing genes—a locus found with an extra gene can achieve a confidence of no greater than "good." When a locus has genuinely different gene content relative to the reference, the resulting serotype may be affected, but the Kaptive match can still be useful for phylogenetic typing purposes.

Clicking on an assembly row expands the view to show more detail, including a diagram of the best-matching locus with genes colored by tBLASTn coverage and identity (Fig. 1C). Beneath the locus diagram are two expandable lists of additional K or O locus genes identified inside and outside the locus region of the query genome (genes that are not usually present in the reference locus). It is common to see matches to a small number of additional genes outside the locus region of the query due to sequence homology (see reference 38 for further details). An additional gene within the locus region of the query genome may indicate that the genome has a novel locus type and will likely correspond to a large length discrepancy from the reference (shown on the right side of the display). In such a case, users may wish to perform further analyses outside Kaptive Web. To facilitate this, Kaptive Web lists the position of the locus in the query genome (shown on the left side of the display), along with a link that allows these assembly regions to be downloaded in FASTA format. For K locus typing, Kaptive Web will also report the alleles for the conserved *wzc* and *wzi* K locus genes, for compatibility with earlier schemes that focused on these genes (37, 57, 58).

Figure 1C shows Kaptive Web K locus results for four isolates from a global *Klebsiella* diversity study data set (54), with various degrees of data quality. The assembly for strain Pus_15987 has a "perfect" match for KL1. Strain D-026-I-b-1 has a best matching locus of KL107, though poor assembly quality resulted in very low identity and coverage and, consequently, a confidence level of "none." Strain QMP_M1-200 has a "high" match for KL11. It contains the entire KL11 sequence, but with a moderate amount of divergence (92% nucleotide sequence identity). In most cases, minor nucleotide divergence likely does not affect the capsule phenotype. However, it should be noted that even a single nonsense or frameshift mutation can have important implications; e.g., the key distinction between the K22 and K37 capsules is not gene content but rather a nonsense mutation in the acetyltransferase gene (48). Kaptive Web clearly identifies potential nonsense or frameshift mutations by marking such loci with "missing genes."

The results for strain AJ170 are expanded in Fig. 1C, showing the full Kaptive Web visualization. It has a very good coverage and identity match to KL38, yet the locus was not found in a single contiguous piece of the assembly. Kaptive was also unable to find a translated protein sequence for one of the KL38 genes, *wckR*, illustrated by the gray coloring in the locus diagram. In this instance, both issues (discontiguous locus sequence and missing gene) were caused by a break in the assembly, splitting the locus over two contigs. This may have resulted from poor read coverage, in which case *wckR* may have been intact and functional in the original isolate. Alternatively, the assembly break may have resulted from an insertion sequence interrupting *wckR*, in which case, gene function is likely lost; such interruptions have been characterized in a number of *Klebsiella* K loci (38). This uncertainty is why AJ170 only achieved a "good" confidence score for KL38.

O locus database and typing. The Kaptive algorithm was originally developed and validated for typing the K locus of *K. pneumoniae* (38), but it can in principle be used to type any variable locus that occurs no more than once per genome. In Kaptive Web, we apply it to O locus typing, which follows mostly the same procedure as K locus typing but with two unique aspects. First, serological types O1 and O2 are distinguished not by the O locus but by two genes, *wbbY* and *wbbZ*, elsewhere in the chromosome. Second, the O locus shared by the O1 and O2 serotypes comes in two varieties (v1 and v2), which are distinguishable using genomic data (they differ in terms of gene content) but are serologically cross-reactive. These aspects are incorporated into Kaptive as follows: (i) the relevant O locus variant is reported as v1 or v2, and (ii) an additional search is conducted for *wbbY* and *wbbZ* to decide whether the locus should be reported as O1 or O2. If only one of *wbbY* or *wbbZ* is found, Kaptive will give a label of O1/O2 (i.e., possibly either).

A recent study of O3 antigens identified several subtypes that can be distinguished serologically and genotypically (O3, O3a, and O3b [20]). The O3 and O3b loci corre-

spond to the O3l and O3s loci previously described in Follador et al. (25) and are distinguished by divergent sequences of the *wbdA* and *wbdD* genes (25). The O3 and O3a loci are distinguished by a single point mutation in *wbdA* (C80R) (20). While Kaptive does not aim to distinguish antigen subtypes, the O3b subtype has sufficient nucleotide divergence to necessitate a separate reference sequence. Kaptive therefore designates O3 loci as either O3/O3a (covered by the same reference sequence) or O3b. As more phenotypic data for antigen subtypes become available, we will consider adding broader antigen subtyping capabilities to Kaptive.

We assessed the accuracy of Kaptive O locus typing by applying the *Klebsiella* global diversity study genomes (54) for which O locus types were previously inferred on the basis of nucleotide variation in the universally conserved *wzm* and *wzt* genes (25). Of the 309 WGS assemblies, the numbers which matched each of the confidence levels were as follows: 3, perfect; 212, very high; 28, high; 56, good; 3, low; and 7, none (see Data Set S1 in the supplemental material). The assemblies with a confidence of "low" or "none" were possibly due to low assembly quality; seven had the O locus split over multiple contigs and three had very low coverage. There was very good agreement between the O locus types defined previously (25) and the Kaptive results—only 8/309 assemblies had discrepancies. Of those, four were cases where the previous type was O1 and Kaptive assigned O1/O2 (i.e., it only found one of *wbbY* and *wbbZ*) and one was where the previous type was O1 and Kaptive assigned O2. The remaining three discrepancies were all between O3 and OL104, which are distinguished by their *wbdD* genes (25). Isolates AJ031 and D-026-I-b-1 were mistyped due to poor assembly—not all of the O locus was represented. Isolate U_13792_2 was typed as O3b by Kaptive but previously assigned OL104, and manual inspection of the predicted WbdD amino acid sequence suggested that this strain produces a hybrid WbdD protein. Given the more subtle distinctions between the O3/O3a, O3b, and OL104 loci, we further tested the accuracy of Kaptive's O locus typing using 13 additional O3 *K. pneumoniae* for which genome data and O antigen phenotypes were previously determined (20) and found Kaptive correctly typed all 13 genomes (see Data Set S2).

Application of Kaptive Web to track K and O locus diversity in MDR clones. The increasing rates of antimicrobial resistance, particularly against last-line drugs such as carbapenems, has led to a resurgence of interest in phage and monoclonal antibody therapies and vaccinations targeting *K. pneumoniae* (16–25). There is particular interest in targeting the globally distributed MDR clones, including CG258 (21–23, 26). ST258, the most well-known member of CG258, has rapidly become the most common cause of CP *Klebsiella* infections in the United States (59). Recent studies suggest low lipopolysaccharide diversity in this clone, with the majority of ST258 isolates expressing the O2 antigen (21, 23). Similarly, early studies reported that ST258 harbored just two distinct capsule types (60). However, subsequent work has shown greater K locus variation (61), particularly among other members of the clonal group, e.g., ST11, ST340, and ST437, which are frequent causes of CP infections outside the United States (62).

The large number of publicly available CG258 assemblies make this group an ideal case for exploring the broader diversity of K and O loci within a single clonal group. To this end, we downloaded all CG258 genome assemblies available in GenBank ($n = 201$ as of 12 October 2017) (see Data Set S3) and analyzed them using Kaptive Web. "Good" or better K and O locus calls were obtained for 173 (86%) and 186 (93%), respectively (Data Set S3). Importantly, while there were dominant types (KL107 and O2v2, the combination of which accounted for 65 [32%] of the isolate genomes), there was also much diversity, with 25 K loci and 6 O loci in total (not distinguishing O1 and O2) (Fig. 2).

There is now emerging evidence that like CG258, other globally distributed MDR clones also harbor diverse K and O loci (50, 63). For example, Fig. 3 shows the core chromosomal phylogeny of the 309 global genomes (54), highlighting Kaptive Web's K and O locus calls for three additional globally distributed MDR lineages (54, 62). Locus exchange tends to result from recombination and demarcates diverging sublineages

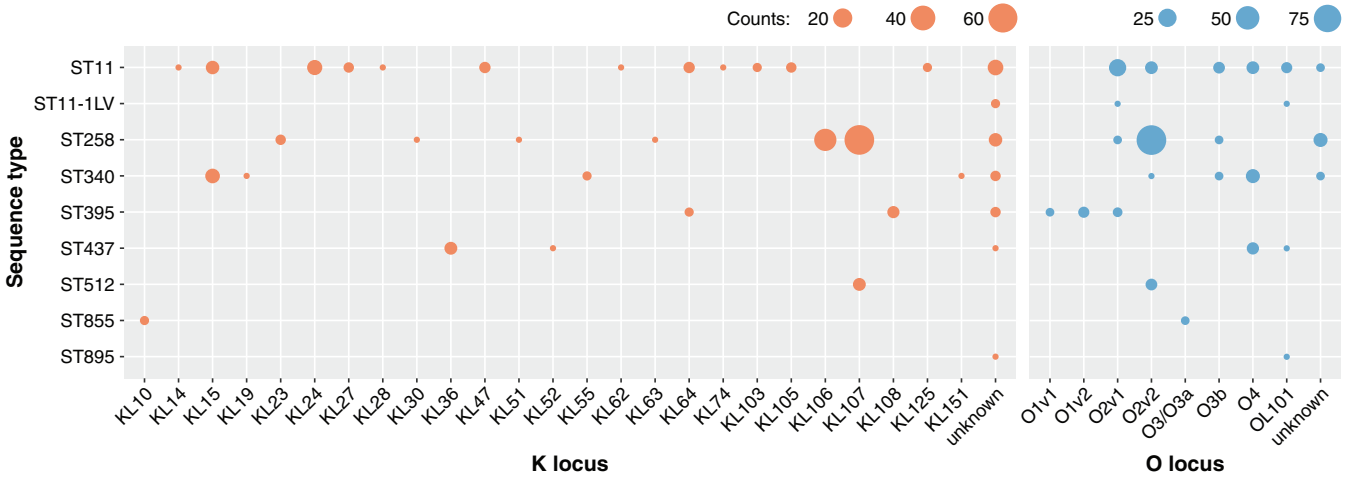


FIG 2 K and O locus diversity among 201 CG258 genomes. Counts of genomes representing each K locus (orange) and O locus (blue) are shown by multilocus sequence type (ST). ST11-1LV indicates an unassigned single locus variant of ST11. Genomes with "low" or "none" confidence Kaptive calls are shown as "unknown."

within the expanding MDR clones (60, 61); hence, the locus calls can also serve as epidemiological markers for the subtyping of MDR strains (58, 60). Therapies specific to the dominant K or O antigens may apply further selective pressure that shifts the population towards different types, as is well documented following the introduction

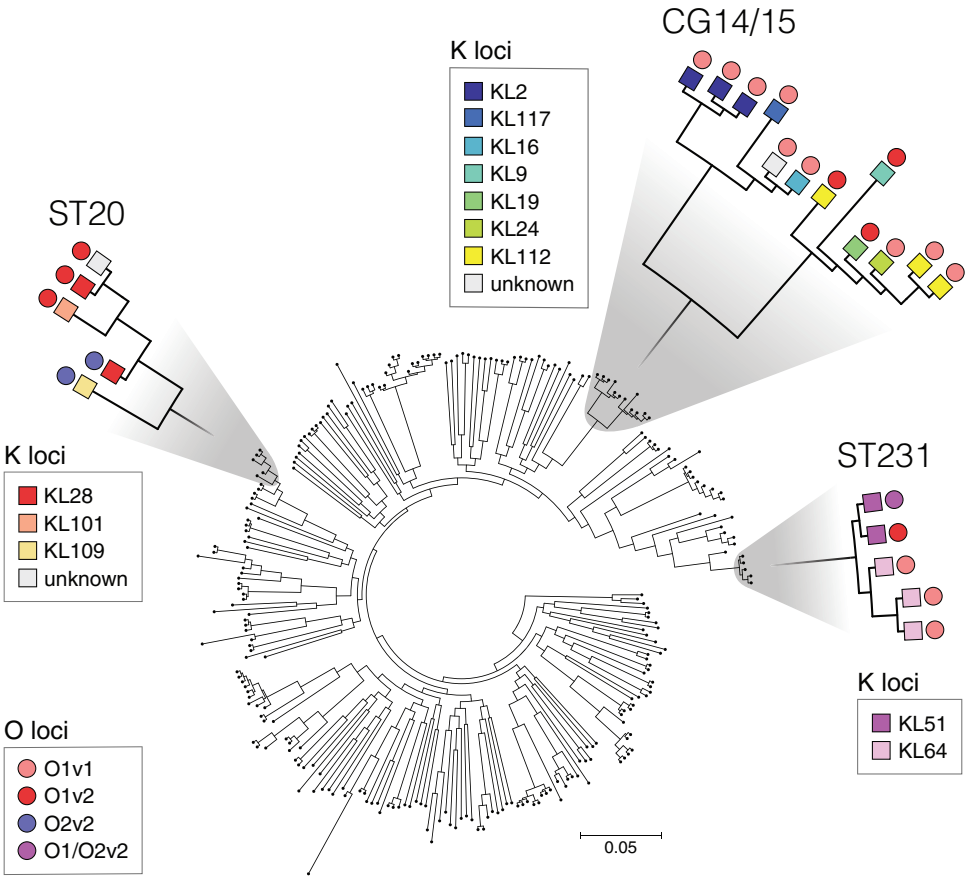


FIG 3 Phylogenetic tree of global *K. pneumoniae* isolates with Kaptive results for K and O loci noted on selected clonal groups. Only Kaptive locus calls with a confidence of "good" or better were included in this figure. Lower confidence matches were labeled "unknown."

of protein-conjugate vaccines targeting the *Streptococcus pneumoniae* polysaccharide capsule (64, 65). The success of new control measures directed at K and O loci will therefore depend on reliable tracking of K and O loci in the *K. pneumoniae* population. Kaptive and Kaptive Web provide simple WGS-based solutions to monitor these trends and ensure that therapies are well targeted and keep up with *K. pneumoniae* evolution. With its graphical interface and remote computation, Kaptive Web also makes these analyses accessible to the wider public health community.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JCM.00197-18.

SUPPLEMENTAL FILE 1, XLSX file, 0.1 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.
SUPPLEMENTAL FILE 3, XLSX file, 0.1 MB.

ACKNOWLEDGMENT

Web development services were provided by the eResearch group at the University of Melbourne.

REFERENCES

1. U.S. Food and Drug Administration. 2017. GenomeTrakr program. U.S. Food and Drug Administration, Silver Spring, MD. https://www.accessdata.fda.gov/scripts/fdatrack/view/track_project.cfm?program=cfsan&id=CFSAN-ORS-Genometrakr.
2. Jackson BR, Tarr C, Strain E, Jackson KA, Conrad A, Carleton H, Katz LS, Stroika S, Gould LH, Mody RK, Silk BJ, Beal J, Chen Y, Timme R, Doyle M, Fields A, Wise M, Tillman G, Defibaugh-Chavez S, Kucerova Z, Sabol A, Roache K, Trees E, Simmons M, Wasilenko J, Kubota K, Pouseele H, Klimke W, Besser J, Brown E, Allard M, Gerner-Smidt P. 2016. Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. Clin Infect Dis 63:380–386. https://doi.org/10.1093/cid/ciw242.
3. Kwong JC, Mercoulia K, Tomita T, Easton M, Li HY, Bulach DM, Stinear TP, Seemann T, Howden BP. 2016. Prospective whole-genome sequencing enhances national surveillance of *Listeria monocytogenes*. J Clin Microbiol 54:333–342. https://doi.org/10.1128/JCM.02344-15.
4. Török M, Peacock S. 2016. Microbial whole-genome sequencing: applications in clinical microbiology and public health, p 32–48. *In* Persing D, Tenover F, Hayden R, Ieven M, Miller M, Nolte F, Tang Y, van Belkum A (ed), Molecular microbiology: diagnostic principles and practice, 3rd ed. ASM Press, Washington, DC.
5. Witney AA, Cosgrove CA, Arnold A, Hinds J, Stoker NG, Butcher PD. 2016. Clinical use of whole genome sequencing for *Mycobacterium tuberculosis*. BMC Med 14:46. https://doi.org/10.1186/s12916-016-0598-2.
6. Robilotti E, Kamboj M. 2015. Integration of whole-genome sequencing into infection control practices: the potential and the hurdles. J Clin Microbiol 53:1054–1055. https://doi.org/10.1128/JCM.00349-15.
7. Kwong JC, McCallum N, Sintchenko V, Howden BP. 2015. Whole genome sequencing in clinical and public health microbiology. Pathology 47:199–210. https://doi.org/10.1097/PAT.0000000000000235.
8. Wyres KL, Conway TC, Garg S, Queiroz C, Reumann M, Holt K, Rusu LI. 2014. WGS analysis and interpretation in clinical and public health microbiology laboratories: what are the requirements and how do existing tools compare? Pathogens 3:437–458. https://doi.org/10.3390/pathogens3020437.
9. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
10. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J Clin Microbiol 53:2410–2426. https://doi.org/10.1128/JCM.00008-15.
11. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry NK. 2016. Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. PeerJ 4:e2477. https://doi.org/10.7717/peerj.2477.
12. Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M, Stinear T, Levine MM, Robins-Browne RM, Holt KE. 2016. *In silico* serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. Microb Genom 2:e000064. https://doi.org/10.1099/mgen.0.000064.
13. Pendleton JN, Gorman SP, Gilmore BF. 2013. Clinical relevance of the ESKAPE pathogens. Exp Rev Anti Infect Ther 11:297–308. https://doi.org/10.1586/eri.13.12.
14. Centers for Disease Control and Prevention. 2013. Antibiotic resistance threats in the United States, 2013. Centers for Disease Control and Prevention, Atlanta, GA.
15. World Health Organization. 2014. Antimicrobial resistance: global report on surveillance. World Health Organization, Geneva, Switzerland.
16. Pan YJ, Lin TL, Lin YT, Su PA, Chen CT, Hsieh PF, Hsu CR, Chen CC, Hsieh YC, Wang JT. 2015. Identification of capsular types in carbapenem-resistant *Klebsiella pneumoniae* strains by *wzc* sequencing and implications for capsule depolymerase treatment. Antimicrob Agents Chemother 59:1038–1047. https://doi.org/10.1128/AAC.03560-14.
17. Aleshkin AV, Ershova ON, Volozhantsev NV, Svetoch EA, Popova AV, Rubalskii EO, Borzilov AI, Aleshkin VA, Afanas'ev SS, Karaulov AV, Galimzyanov KM, Rubalsky OV, Bochkareva SS. 2016. Phagebiotics in treatment and prophylaxis of healthcare-associated infections. Bacteriophage 6:e1251379. https://doi.org/10.1080/21597081.2016.1251379.
18. D'Andrea MM, Marmo P, Henrici De Angelis L, Palmieri M, Ciacci N, Di Lallo G, Demattè E, Vannuccini E, Lupetti P, Rossolini GM, Thaller MC. 2017. φBO1E, a newly discovered lytic bacteriophage targeting carbapenemase-producing *Klebsiella pneumoniae* of the pandemic clonal group 258 clade II lineage. Sci Rep 7:2614. https://doi.org/10.1038/s41598-017-02788-9.
19. Pallavali RR, Degati VL, Lomada D, Reddy MC, Durbaka VRP. 2017. Isolation and *in vitro* evaluation of bacteriophages against MDR-bacterial isolates from septic wound infections. PLoS One 12:e0179245. https://doi.org/10.1371/journal.pone.0179245.
20. Guachalla LM, Stojkovic K, Hartl K, Kaszowska M, Kumar Y, Wahl B, Paprotka T, Nagy E, Lukasiewicz J, Nagy G, Szijártó V. 2017. Discovery of monoclonal antibodies cross-reactive to novel subserotypes of *K. pneumoniae* O3. Sci Rep 7:6635. https://doi.org/10.1038/s41598-017-06682-2.
21. Szijártó V, Guachalla LM, Hartl K, Varga C, Banerjee P, Stojkovic K, Kaszowska M, Nagy E, Lukasiewicz J, Nagy G. 2016. Both clades of the epidemic KPC-producing *Klebsiella pneumoniae* clone ST258 share a modified galactan O-antigen type. Int J Med Microbiol 306:89–98. https://doi.org/10.1016/j.ijmm.2015.12.002.
22. Motley MP, Fries BC. 2017. A new take on an old remedy: generating antibodies against bacteria in a postantibiotic world. mSphere 2:e00397-17. https://doi.org/10.1128/mSphere.00397-17.

23. Pennini ME, De Marco A, Pelletier M, Bonnell J, Cvitkovic R, Beltramello M, Cameroni E, Bianchi S, Zatta F, Zhao W, Xiao X, Camara MM, DiGiandomenico A, Semenova E, Lanzavecchia A, Warrener P, Suzich J, Wang Q, Corti D, Stover CK. 2017. Immune stealth-driven O2 serotype prevalence and potential for therapeutic antibodies against multidrug resistant *Klebsiella pneumoniae*. Nat Commun 8:1991. https://doi.org/10.1038/s41467-017-02223-7.
24. Ahmad TA, El-Sayed LH, Haroun M, Hussein AA, El Ashry el SH. 2012. Development of immunization trials against *Klebsiella pneumoniae*. Vaccine 30:2411–2420. https://doi.org/10.1016/j.vaccine.2011.11.027.
25. Follador R, Heinz E, Wyres KL, Ellington MJ, Kowarik M, Holt KE, Thomson NR. 2016. The diversity of *Klebsiella pneumoniae* surface polysaccharides. Microb Genom 2:e000073. https://doi.org/10.1099/mgen.0.000073.
26. Seeberger PH, Pereira CL, Khan N, Xiao G, Diago-Navarro E, Reppe K, Opitz B, Fries BC, Witzenrath M. 2017. A semi-synthetic glycoconjugate vaccine candidate for carbapenem-resistant *Klebsiella pneumoniae*. Angew Chem Int Ed Engl 56:13973–13978. https://doi.org/10.1002/anie.201700964.
27. Merino S, Camprubí S, Albertí S, Benedí VJ, Tomas JM. 1992. Mechanisms of *Klebsiella pneumoniae* resistance to complement-mediated killing. Infect Immun 60:2529–2535.
28. Evrard B, Balestrino D, Dosgilbert A, Bouya-Gachancard JL, Charbonnel N, Forestier C, Tridon A. 2010. Roles of capsule and lipopolysaccharide O antigen in interactions of human monocyte-derived dendritic cells and *Klebsiella pneumoniae*. Infect Immun 78:210–219. https://doi.org/10.1128/IAI.00864-09.
29. March C, Cano V, Moranta D, Llobet E, Pérez-Gutiérrez C, Tomás JM, Suárez T, Garmendia J, Bengoechea JA. 2013. Role of bacterial surface structures on the interaction of *Klebsiella pneumoniae* with phagocytes. PLoS One 8:e56847. https://doi.org/10.1371/journal.pone.0056847.
30. Hsieh P, Lin T, Yang F, Wu M, Pan Y, Wu S, Wang J. 2012. Lipopolysaccharide O1 antigen contributes to the virulence in *Klebsiella pneumoniae* causing pyogenic liver abscess. PLoS One 7:e33155. https://doi.org/10.1371/journal.pone.0033155.
31. Cortés G, Borrell N, Astorza D, Gómez C, Sauleda J, Albertí S. 2002. Molecular analysis of the contribution of the capsular polysaccharide and the lipopolysaccharide O side chain to the virulence of *Klebsiella pneumoniae* in a murine model of pneumonia. Infect Immun 70:2583–2590. https://doi.org/10.1128/IAI.70.5.2583-2590.2002.
32. Brisse S, Fevre C, Passet V, Issenhuth-Jeanjean S, Tournebize R, Diancourt L, Grimont P. 2009. Virulent clones of *Klebsiella pneumoniae*: identification and evolutionary scenario based on genomic and phenotypic characterization. PLoS One 4:e4982. https://doi.org/10.1371/journal.pone.0004982.
33. Luo Y, Wang Y, Ye L, Yang J. 2014. Molecular epidemiology and virulence factors of pyogenic liver abscess causing *Klebsiella pneumoniae* in China. Clin Microbiol Infect 20:O818–O824. https://doi.org/10.1111/1469-0691.12664.
34. Lee IR, Molton JS, Wyres KL, Gorrie C, Wong J, Hoh CH, Teo J, Kalimuddin S, Lye DC, Archuleta S, Holt KE, Gan Y-H. 2016. Differential host susceptibility and bacterial virulence factors driving *Klebsiella* liver abscess in an ethnically diverse population. Sci Rep 6:29316. https://doi.org/10.1038/srep29316.
35. Turton JF, Perry C, Elgohari S, Hampton CV. 2010. PCR characterization and typing of *Klebsiella pneumoniae* using capsular type-specific, variable number tandem repeat and virulence gene targets. J Med Microbiol 59:541–547. https://doi.org/10.1099/jmm.0.015198-0.
36. Fang C, Shih Y, Cheong C, Yi W. 2015. Rapid and accurate determination of lipopolysaccharide O-antigen types in *Klebsiella pneumoniae* with a novel PCR-based O-genotyping method. J Clin Microbiol 54:666–675. https://doi.org/10.1128/JCM.02494-15.
37. Brisse S, Passet V, Haugaard AB, Babosan A, Kassis-Chikhani N, Struve C, Decré D. 2013. *wzi* gene sequencing, a rapid method for determination of capsular type for *Klebsiella* strains. J Clin Microbiol 51:4073–4078. https://doi.org/10.1128/JCM.01924-13.
38. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, Holt KE. 2016. Identification of *Klebsiella* capsule synthesis loci from whole genome data. Microb Genom 2:e000102. https://doi.org/10.1099/mgen.0.000102.
39. Whitfield C, Trent MS. 2014. Biosynthesis and export of bacterial lipopolysaccharides. Annu Rev Biochem 83:99–128. https://doi.org/10.1146/annurev-biochem-060713-035600.
40. Whitfield C, Richards JC, Perry MB, Clarke BR, Maclean LL. 1991. Expression of two structurally distinct D-galactan O antigens in the lipopolysaccharide of *Klebsiella pneumoniae* serotype O1. J Bacteriol 173:1420–1431. https://doi.org/10.1128/jb.173.4.1420-1431.1991.
41. Clarke BR, Whitfield C. 1992. Molecular cloning of the *rfb* region of *Klebsiella pneumoniae* serotype O1:K20: the *rfb* gene cluster is responsible for synthesis of the D-galactan I O polysaccharide. J Bacteriol 174:4614–4621. https://doi.org/10.1128/jb.174.14.4614-4621.1992.
42. Sugiyama T, Kido N, Kato Y, Koide N, Yoshida T, Yokochi T. 1997. Evolutionary relationship among *rfb* gene clusters synthesizing mannose homopolymer as O-specific polysaccharides in *Escherichia coli* and *Klebsiella pneumoniae*. Gene 198:111–113. https://doi.org/10.1016/S0378-1119(97)00300-4.
43. Trautmann M, Ruhnke M, Rukavina T, Held TK, Cross AS, Marre R, Whitfield C. 1997. O-antigen seroepidemiology of *Klebsiella* clinical isolates and implications for immunoprophylaxis of *Klebsiella* infections. Clin Diagn Lab Immun 4:550–555.
44. Hsieh PF, Wu MC, Yang FL, Chen CT, Lou TC, Chen YY, Wu SH, Sheu JC, Wang JT. 2014. D-Galactan II is an immunodominant antigen in O1 lipopolysaccharide and affects virulence in *Klebsiella pneumoniae*: implication in vaccine design. Front Microbiol 5:608. https://doi.org/10.3389/fmicb.2014.00608.
45. Kelly RF, Perry MB, MacLean LL, Whitfield C. 1995. Structures of the O-antigens of *Klebsiella* serotypes 02 (2a,2e), 02 (2a,2e,2h), and 02 (2a,2f,2g), members of a family of related D-galactan O-antigens in *Klebsiella* spp. Innate Immun 2:131–140. https://doi.org/10.1177/096805199500200208.
46. Rahn A, Drummelsmith J, Whitfield C. 1999. Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. J Bacteriol 181:2307–2713.
47. Arakawa Y, Wacharotayankun R, Nagatsuka T, Ito H, Kato N, Ohta M. 1995. Genomic organization of the *Klebsiella pneumoniae cps* region responsible for serotype K2 capsular polysaccharide synthesis in the virulent strain Chedid. J Bacteriol 177:1788–1796. https://doi.org/10.1128/jb.177.7.1788-1796.1995.
48. Pan Y-J, Lin T-L, Chen C-T, Chen Y-Y, Hsieh P-F, Hsu C-R, Wu M-C, Wang J-T. 2015. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. Sci Rep 5:15573. https://doi.org/10.1038/srep15573.
49. Ørskov IDA, Fife-Asbury MA. 1977. New *Klebsiella* capsular antigen, K82, and the deletion of five of those previously assigned. Int J Syst Evol Microbiol 27:386–387. https://doi.org/10.1099/00207713-27-4-386.
50. Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis J, Brettin T, Xia F, Musser J. 2017. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates, Houston, Texas: unexpected abundance of clonal group 307. mBio 8:e00489-17. https://doi.org/10.1128/mBio.00489-17.
51. Gorrie CL, Mirceta M, Wick RR, Judd LM, Wyres KL, Thomson NR, Strugnell RA, Pratt NF, Garlick JF, Watson KM, Hunter PC, McGloughlin SA, Spelman DW, Jenney AWJ, Holt KE. 11 January 2018. Antimicrobial resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital. Clin Infect Dis https://doi.org/10.1093/cid/ciy027.
52. Anes J, Hurley D, Martins M, Fanning S. 2017. Exploring the genome and phenotype of multi-drug resistant *Klebsiella pneumoniae* of clinical origin. Front Microbiol 8:1913. https://doi.org/10.3389/fmicb.2017.01913.
53. Di Pierro M. 2011. web2py for scientific applications. Comput Sci Eng 13:64–69. https://doi.org/10.1109/MCSE.2010.97.
54. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen Van K, Nguyen TV, Dao TT, Mensink M, Minh Le V, Nhu NTK, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR. 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci U S A 112:E3574–E3581. https://doi.org/10.1073/pnas.1501049112.
55. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
56. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132. https://doi.org/10.1099/mgen.0.000132.
57. Pan YJ, Lin TL, Chen YH, Hsu CR, Hsieh PF, Wu MC, Wang JT. 2013. Capsular types of *Klebsiella pneumoniae* revisited by *wzc* sequencing. PLoS One 8:e80670. https://doi.org/10.1371/journal.pone.0080670.
58. Satlin MJ, Chen L, Patel G, Gomez-simmonds A, Weston G, Kim AC, Seo

SK, Rosenthal ME, Sperber SJ, Jenkins SG, Hamula CL, Uhlemann A, Levi MH, Fries BC, Tang Y, Juretschko S, Rojtman D, Hong T, Mathema B, Jacobs M, Walsh T, Bonomo R, Kreiswirth B. 2017. Multicenter clinical and molecular epidemiological analysis of bacteremia epicenter of the United States. Antimicrob Agents Chemother 61:e02349-16. https://doi.org/10.1128/AAC.02349-16.

59. Kitchel B, Rasheed JK, Patel JB, Srinivasan A, Navon-Venezia S, Carmeli Y, Brolund A, Giske CG. 2009. Molecular epidemiology of KPC-producing *Klebsiella pneumoniae* isolates in the United States: clonal expansion of multilocus sequence type 258. Antimicrob Agents Chemother 53:3365–3370. https://doi.org/10.1128/AAC.00126-09.
60. Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, Chavda KD, Jacobs MR, Mathema B, Olsen RJ, Bonomo RA, Musser JM, Kreiswirth BN. 2014. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. Proc Natl Acad Sci U S A 111:4988–4993. https://doi.org/10.1073/pnas.1321364111.
61. Wyres KL, Gorrie C, Edwards DJ, Wertheim HFL, Hsu LY, Van Kinh N, Zadoks R, Baker S, Holt KE. 2015. Extensive capsule locus variation and large-scale genomic recombination within the *Klebsiella pneumoniae* clonal group 258. Genome Biol Evol 7:1267–1279. https://doi.org/10.1093/gbe/evv062.
62. Wyres KL, Holt KE. 2016. *Klebsiella pneumoniae* population genomics and antimicrobial-resistant clones. Trends Microbiol 24:944–956. https://doi.org/10.1016/j.tim.2016.09.007.
63. Zhou K, Lokate M, Deurenberg RH, Tepper M, Arends JP, Raangs EGC, Lo-Ten-Foe J, Grundmann H, Rossen JWA, Friedrich AW. 2016. Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum β-lactamase producing ST15 *Klebsiella pneumoniae*. Sci Rep 6:20840. https://doi.org/10.1038/srep20840.
64. Hausdorff WP, Brueggemann AB, Hackell JG, Scott AG. 2008. Pneumococcal serotype epidemiology, p 139–160. *In* Siber G, Klugman KP, Mäkeľa PH (ed), Pneumococcal vaccines: the impact of conjugate vaccine. ASM Press, Washington, DC.
65. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet 45:656–663. https://doi.org/10.1038/ng.2625.