Whole Genome Sequencing Elucidates the Epidemiology of Nosocomial Clusters of *Acinetobacter baumannii*

Stefanie Willems et al.

Supplemental Methods

Species and antibiotic resistance determination

Species identification of presumable isolates was performed using MALDI-TOF MS on the Bruker Biotyper system (Bruker Daltonic GmbH, Bremen, Germany). In addition, all species were confirmed using 16S rRNA sequences derived from the whole genome sequencing (WGS) data. Here, 16S rRNA sequences were extracted from the WGS data using the Ridom SeqSphere^{+} software version 2 beta (Ridom GmbH, Münster, Germany) and compared to published *A. baumannii* 16S rRNA sequences at NCBI GenBank (http://www.ncbi.nlm.nih.gov/nuccore).

Antibiotic resistance patterns were determined using VITEK II (BioMérieux, Marcy-l'Étoile, France); testing was performed in accordance to the EUCAST versions 3.1 to 5.0. Rating of isolates as multidrug-resistant (MDR) or non-MDR was conducted in accordance to the German national guideline (1). Here, *A. baumannii* isolates resistant against at least three of the following four substances / substance groups were rated MDR: piperacillin, cefotaxime/ceftazidime, ciprofloxacin, imipenem/meropenem. All other *A. baumannii* isolates were rated as non-MDR (1).

Whole genome sequencing and data analysis

For WGS on a MiSeq instrument (Illumina, San Diego, CA, USA) a single colony was inoculated in nutrient broth (Heipha, Eppelheim, Germany) and incubated overnight (37 °C).

Genomic DNA was purified using the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. Subsequently, one ng of genomic DNA was introduced into library preparation with the Nextera XT DNA Sample preparation Kit (Illumina) and paired-end sequenced with the 2x250 bp MiSeq Reagent Kit v2 (Illumina) with an average insert size of 300 bp. Libraries were scaled to reach 100-fold sequencing coverage. Only sequencing runs that fulfilled the manufacturers' specifications (Illumina) with respect to cluster density and Q30 were further analyzed. The resulting sequence files (fastq-file format) were quality trimmed at their 5'- and 3'-ends until an average base quality of 30 was reached in a window of 20 bases and *de novo* assembled using the Velvet assembler version 1.1.04 that is integrated in the SeqSphere^{+} software (Ridom GmbH) using the parameters "optimized k-mer size and coverage cutoff values based on the average length of contigs with > 1000 bp" as previously described (2). From the resulting sequence assembly files (ACE-file format), the coding sequences were extracted in a pairwise comparison with each target (see Table S1) of the reference genome if they showed a sequence similarity of 90% and an overlap of 100%, respectively. If the sequenced genes were not found using these parameters, they were rated as "missing data" irrespective whether they were really absent in the sequenced genome or below the sequence similarity and overlap thresholds. For allele designation, the detected gene sequence was compared to all previously detected allelic sequences of the same gene; identical alleles, i. e. 100% sequence similarity and 100% overlap, were given the same allele number, otherwise a new allele designation was assigned. The resulting allelic profiles of all isolates were then used to build the minimum-spanning tree within the SeqSphere^{+} software. Here, the option "pairwise ignoring missing values" was used.

For SNP typing, we extracted all SNPs that were present within any of the 3,319 cgMLST target genes (Table S1) using SeqSphere^{+} version 3.0 (Ridom GmbH). We then filtered for insertions/deletions and adjacent SNPs (in a window of 10 bp) that were not included for

further analysis. SNP genotypes were compared using the same algorithm for minimum-spanning tree construction as mentioned before for cgMLST.

For construction of a cgMLST-based minimum-spanning tree of the data of Fitzpatrick et al. (3), we downloaded the sequence data from NCBI, analyzed them using the same set of cgMLST targets and parameters for tree construction as for this study. Finally, we manually compared the clustering of both methods.

Supplemental References

1. **Kommission fur Krankenhaushygiene und Infektionspravention (KRINKO) beim Robert Koch-Institut (RKI).** 2012. [Hygiene measures for infection or colonization with multidrug-resistant gram-negative bacilli. Commission recommendation for hospital hygiene and infection prevention (KRINKO) at the Robert Koch Institute (RKI)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz **55:**1311-1354.
2. **Bletz S, Mellmann A, Rothgänger J, Harmsen D.** 2015. Ensuring backwards compatibility: traditional genotyping efforts in the era of whole genome sequencing. Clin Microbiol Infect **21:**347.e341-344.
3. **Fitzpatrick MA, Ozer EA, Hauser AR.** 2016. Utility of whole-genome sequencing in characterizing *Acinetobacter* epidemiology and analyzing hospital outbreaks. J Clin Microbiol **54:**593-612.

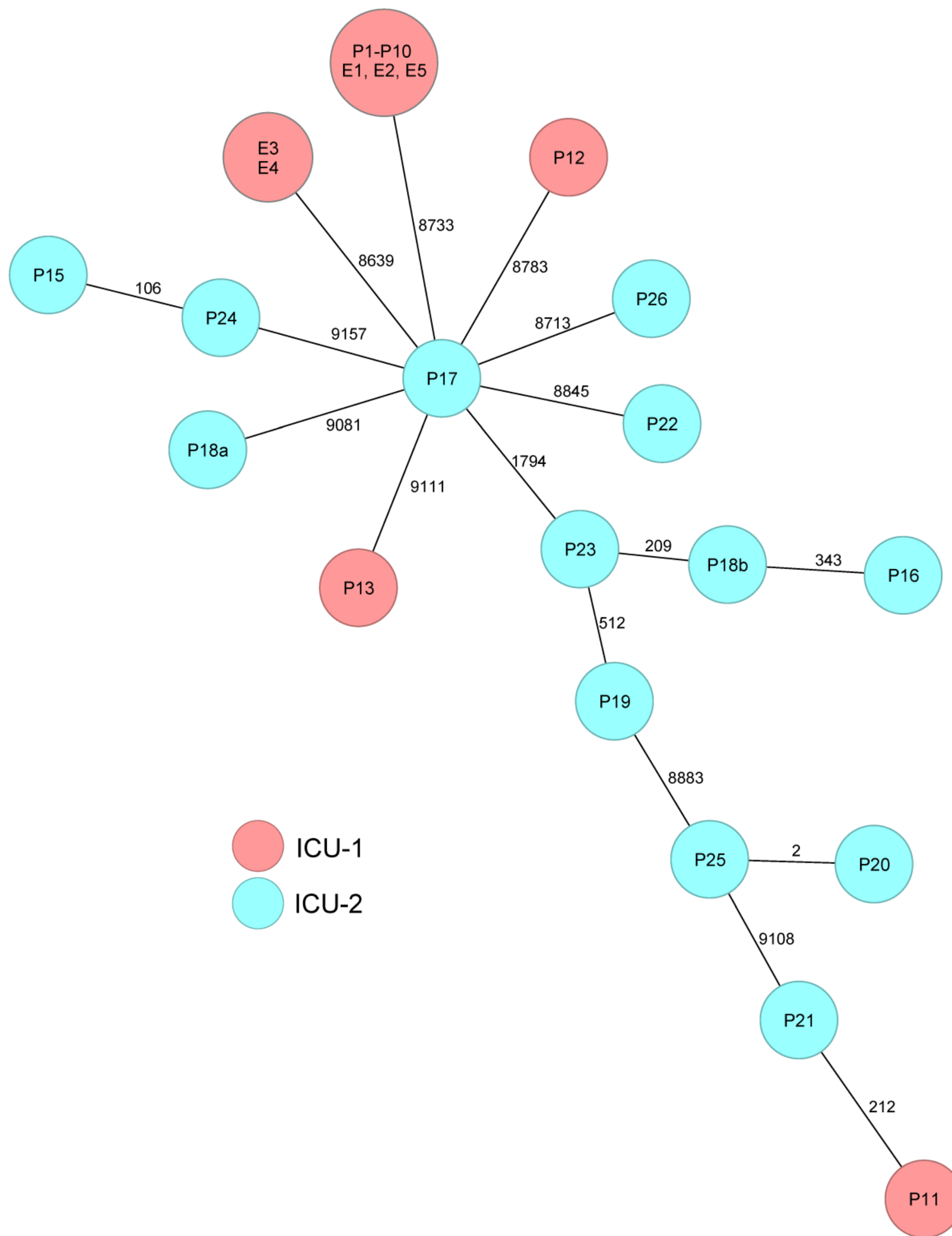

Figure S1. Minimum-spanning tree of single nucleotide polymorphism (SNP)-based typing of 32 *A. baumannii* isolates. SNPs were extracted from all genomic targets used also for cgMLST analysis (see Table S1) using the SeqSphere$^+$ software version 3.0 (Ridom GmbH, Münster, Germany). Insertions/deletions and adjacent SNPs (in a window of 10 bp) were filtered and not included for further analysis. Each circle represents a SNP genotype based on in total 33,766 SNPs and is named by the isolate(s). The number on connecting lines gives the number of differing SNPs. The size of the circle is proportional to the number of isolates with

an identical SNP genotype and the color of the circles represents the different wards. SNP genotypes were listed in Table S2.

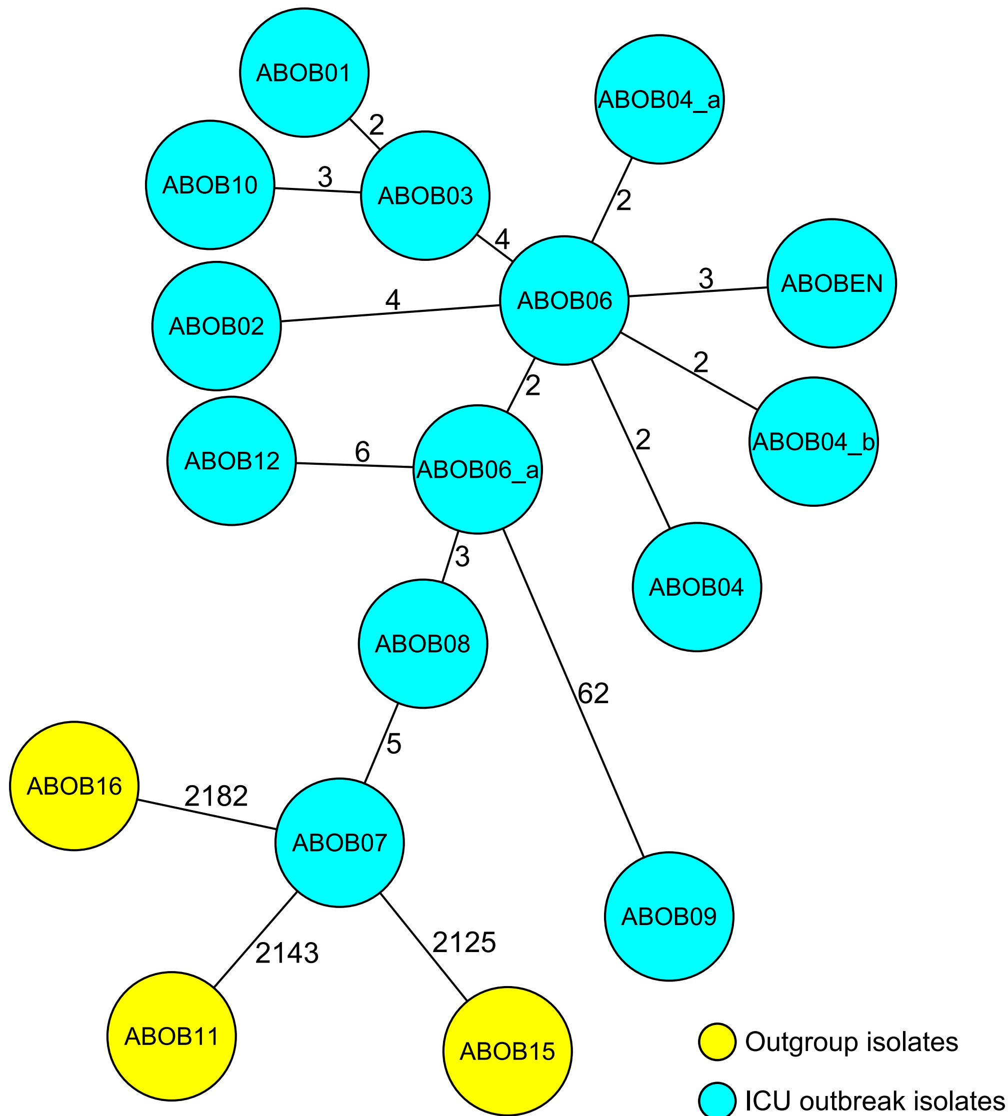

Figure S2. Minimum-spanning tree of the 14 outbreak isolates and 3 outgroup isolates of the study from Fitzpatrick et al. based on cgMLST analysis. Each circle represents an allelic profile, i. e. the genotype, based on allelic profiles of up to 3,319 target genes (Table S3) and is named by the isolate(s). The minimum-spanning tree is constructed with the “pairwise ignoring missing values” option turned on in the SeqSphere^{+} software during comparison. The number on connecting lines gives the number of differing alleles. The size of the circle is proportional to the number of isolates with an identical genotype and the color of the circles differentiates between outbreak and outgroup isolates.