

# New Insights into Dissemination and Variation of the Health Care-Associated Pathogen *Acinetobacter baumannii* from Genomic Analysis

Meredith S. Wright,<sup>a</sup> Daniel H. Haft,<sup>b</sup> Derek M. Harkins,<sup>b</sup> Federico Perez,<sup>c</sup> Kristine M. Hujer,<sup>c,d</sup> Saralee Bajaksouzian,<sup>e</sup> Michael F. Benard,<sup>f</sup> Michael R. Jacobs,<sup>e,g</sup> Robert A. Bonomo,<sup>c,d,h</sup> Mark D. Adams<sup>a</sup>

J. Craig Venter Institute, San Diego, California, USA<sup>a</sup>; J. Craig Venter Institute, Rockville, Maryland, USA<sup>b</sup>; Louis Stokes Cleveland Veterans Affairs Medical Center, Cleveland, Ohio, USA<sup>c</sup>; Department of Medicine, Case Western Reserve University, Cleveland, Ohio, USA<sup>d</sup>; Department of Pathology, University Hospitals Case Medical Center, Cleveland, Ohio, USA<sup>e</sup>; Department of Biology, Case Western Reserve University, Cleveland, Ohio, USA<sup>f</sup>; Department of Pathology, Case Western Reserve University, Cleveland, Ohio, USA<sup>g</sup>; Departments of Pharmacology, Molecular Biology and Microbiology and Center for Proteomics, Case Western Reserve University, Cleveland, Ohio, USA<sup>h</sup>

**ABSTRACT** *Acinetobacter baumannii* is a globally important nosocomial pathogen characterized by an increasing incidence of multidrug resistance. Routes of dissemination and gene flow among health care facilities are poorly resolved and are important for understanding the epidemiology of *A. baumannii*, minimizing disease transmission, and improving patient outcomes. We used whole-genome sequencing to assess diversity and genome dynamics in 49 isolates from one United States hospital system during one year from 2007 to 2008. Core single-nucleotide-variant-based phylogenetic analysis revealed multiple founder strains and multiple independent strains recovered from the same patient yet was insufficient to fully resolve strain relationships, where gene content and insertion sequence patterns added additional discriminatory power. Gene content comparisons illustrated extensive and redundant antibiotic resistance gene carriage and direct evidence of gene transfer, recombination, gene loss, and mutation. Evidence of barriers to gene flow among hospital components was not found, suggesting complex mixing of strains and a large reservoir of *A. baumannii* strains capable of colonizing patients.

**IMPORTANCE** Genome sequencing was used to characterize multidrug-resistant *Acinetobacter baumannii* strains from one United States hospital system during a 1-year period to better understand how *A. baumannii* strains that cause infection are related to one another. Extensive variation in gene content was found, even among strains that were very closely related phylogenetically and epidemiologically. Several mechanisms contributed to this diversity, including transfer of mobile genetic elements, mobilization of insertion sequences, insertion sequence-mediated deletions, and genome-wide homologous recombination. Variation in gene content, however, lacked clear spatial or temporal patterns, suggesting a diverse pool of circulating strains with considerable interaction between strains and hospital locations. Widespread genetic variation among strains from the same hospital and even the same patient, particularly involving antibiotic resistance genes, reinforces the need for molecular diagnostic testing and genomic analysis to determine resistance profiles, rather than a reliance primarily on strain typing and antimicrobial resistance phenotypes for epidemiological studies.

Received 10 September 2013 Accepted 5 December 2013 Published 21 January 2014

**Citation** Wright MS, Haft DH, Harkins DM, Perez F, Hujer KM, Bajaksouzian S, Benard MF, Jacobs MR, Bonomo RA, Adams MD. 2014. New insights into dissemination and variation of the health care-associated pathogen *Acinetobacter baumannii* from genomic analysis. *mBio* 5(1):e00963-13. doi:10.1128/mBio.00963-13.

**Editor** Paul Keim, Northern Arizona University

**Copyright** © 2014 Wright et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Mark D. Adams, madams@jvci.org, or Robert A. Bonomo, robert.bonomo@va.gov.

Nosocomial infections are a significant public health concern and economic cost to health care systems (1, 2). Therefore, a critical need exists for a better understanding of nosocomial pathogen population dynamics and epidemiology to improve diagnostics and infection control efforts. Comparative whole-genome sequencing (WGS) offers opportunities to address these issues, with analyses expected to also lead to increased understanding of pathogen transmission routes and the movement of targeted genetic elements such as antimicrobial resistance or virulence-associated genes (3–7).

One nosocomial pathogen to emerge in recent years is *Acinetobacter baumannii*, which is of particular concern in light of the global occurrence of multidrug-resistant (MDR) and pan-drug-resistant strains (8–11). *A. baumannii* was an uncommon patho-

gen in health care settings until recent decades, but is now a leading cause of ventilator-associated pneumonia and surgical and urinary tract infections, among other illnesses (1). This increase in the prevalence of *A. baumannii* in health care-associated infections has occurred in conjunction with an increase in the prevalence of MDR strains, with 60% of the isolates in the United States reported as MDR according to recent surveillance data (1).

Drug resistance is a major factor contributing to the success of *A. baumannii* in hospital settings (12–15). Genetic characterizations of *A. baumannii* strains have revealed that they possess a diverse and extensive arsenal of chromosomal and plasmid-borne resistance genes (16, 17). For example, *A. baumannii* possesses the intrinsic chromosomal  $\beta$ -lactamase genes *bla*<sub>ADC</sub> and *bla*<sub>OXA-51-like</sub>, and overexpression of these genes driven by pro-

motors in upstream insertion sequence (IS) elements can confer resistance to extended-spectrum cephalosporins and carbapenems, respectively (18–21). Several allelic variants of each gene from clinical *A. baumannii* strains have been described (22–24), but less is known about the dynamics of alleles within *A. baumannii* populations.

Another hallmark of *A. baumannii* antibiotic resistance mechanisms is the prevalence of resistance islands (RIs) where transposon- and integron-derived modules have played a role in mobilizing genes conferring resistance to several classes of drugs, including the frontline carbapenem and aminoglycoside antibiotics. Many recent studies highlight the diversity in the genomic location, architecture, and content of these RIs, demonstrating the dynamic nature of *A. baumannii* antibiotic resistance mechanisms and the adaptive significance of these elements (25–27). Plasmid-borne resistance genes are also reported in *A. baumannii*, where the association of resistance genes with IS and site-specific recombination systems facilitates their dispersal (28–30). Though the analysis of the origin and movement of antibiotic resistance mechanisms remains a research priority, the diversity and distribution of resistance genes circulating among the *A. baumannii* strains within one hospital and their genetic context remain poorly resolved.

Despite increased research efforts into *A. baumannii* epidemiology and evolution (15, 31, 32), large gaps remain in our understanding of the evolutionary processes that contribute to strain diversification within hospital environments. Processes at this scale are important for understanding transmission routes and whether infections are primarily a result of patient-to-patient contact (including via mediators such as health care workers or environmental surfaces) or the result of multiple founder events. Previous examinations of *A. baumannii* epidemiology and diversity have focused on the examination of strain dynamics and gene content across time and space by using coarser measures of relatedness such as multilocus sequence typing (MLST), pulsed-field gel electrophoresis fingerprinting, or multilocus variable-number tandem-repeat analysis profiles (15, 33–38) and have demonstrated the successive clonal spread of three primary lineages, global clones I to III, in hospitals worldwide (34). While several sequence types (STs) of *A. baumannii* strains can coexist within hospitals (39, 40), the limited resolution of these typing schemes masks the extent of lateral gene transfer (LGT) and recombination, which are critical for driving strain differentiation. A more recent whole-genome comparative analysis of within-hospital strain dynamics suggested that multiple founder *A. baumannii* strains can be present simultaneously and noted the role of genome-wide recombination in within-hospital strain diversification (32).

To further expand what is known about *A. baumannii* evolutionary processes and population dynamics, we used comparative analysis of genome sequences from 49 *A. baumannii* strains to examine strain level genetic diversity and gene content variation over a 1-year period within one United States hospital system. By studying many strains from an interconnected health care environment, we aimed to gain a better understanding of routes of transmission between patients and within and among hospitals, the diversity of the population of founder strains, and the extent to which LGT, recombination, and mutation contribute to *A. baumannii* evolution.

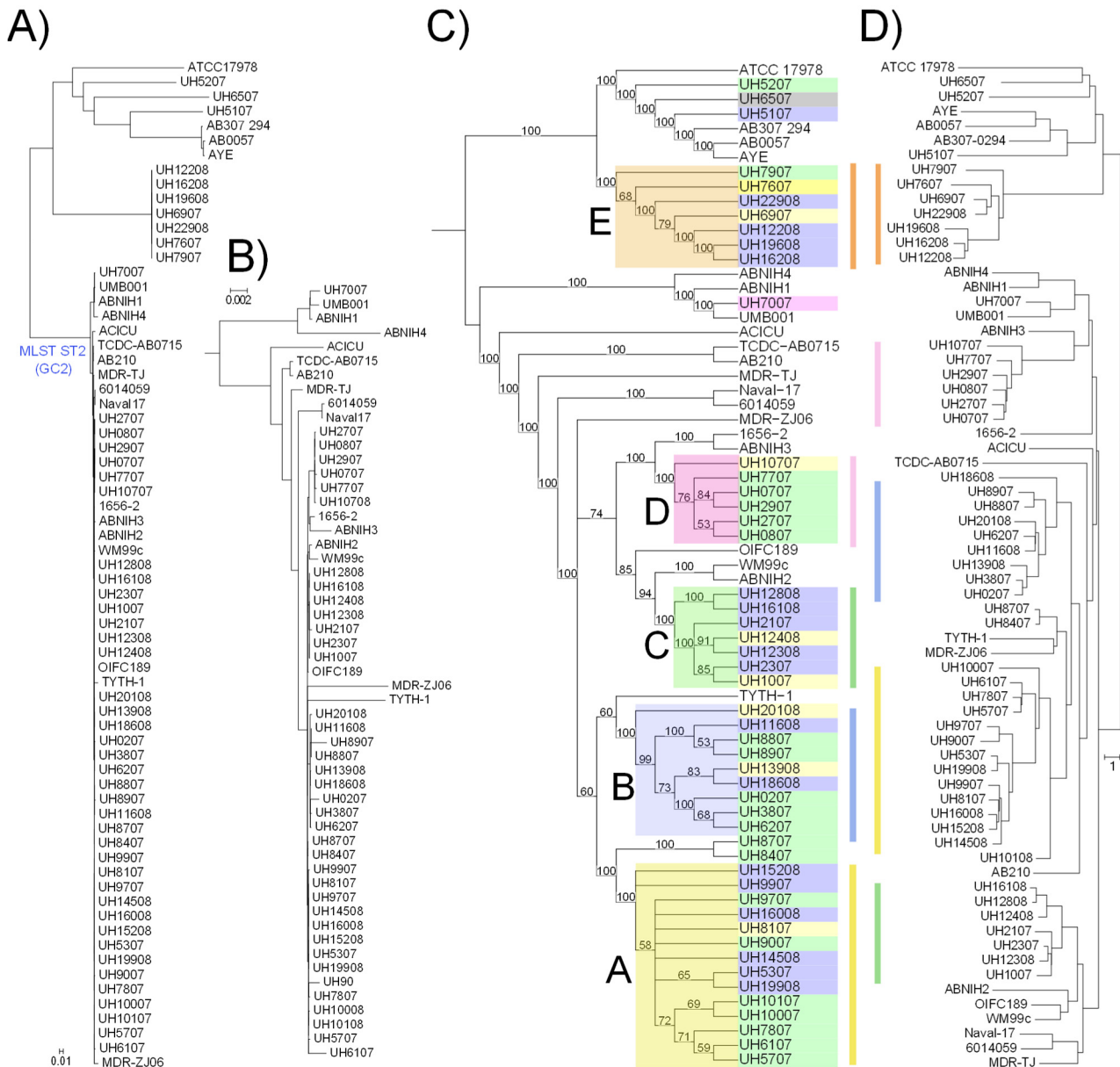
## RESULTS

We sequenced the genomes of 49 *A. baumannii* strains that were isolated from patients in one tertiary-care hospital, three regional hospitals, and an extended-care facility that are part of a single integrated hospital system. Earlier work showed that the majority of the isolates are from a single MLST group, implying that the outbreak was primarily clonal, with patient-to-patient transmission explaining most of the new infections (37). Genome sequencing confirmed that most of the strains belonged to global clone 2 (GC2) (ST2 in the nomenclature of Diancourt et al. [34]), with additional strains making up either new (three strains) or different MLST groups (seven strains were ST79) (see Table S1 in the supplemental material). To facilitate detailed analysis of strain relationships and correlation with clinical and phenotypic features, we first developed a robust phylogeny based on single-nucleotide variants (SNVs) that are present in core regions of the genome to represent ancestral relationships among strains. Because recombination in *A. baumannii* was previously reported (32), we also excluded regions of elevated SNV density from the analysis, where SNV patterns reflect recent recombination and not shared ancestry. Removal of these ~30,000 SNVs altered the topology of the phylogenetic tree by changing both the placement of a major lineage (i.e., clade D was initially on the same branch with ACICU) and the placement of interior branches (i.e., UH5307 and UH19908 in the original tree were in clade B).

The core SNV tree revealed five well-supported primary subclades, with four of them (A to D) representing highly similar GC2 isolates (Fig. 1A and B). Each clade was composed of strains from more than one hospital (Fig. 1C) and spanning the study period (see Table S1 in the supplemental material). The clades were interspersed with *A. baumannii* strains reported by others that were isolated from disparate geographic locations worldwide.

**Pangenome: core and accessory genes.** To explore additional genetic features that might shed light on strain relatedness, we compared the gene contents of the 49 University Hospitals (UH) and 20 reference *A. baumannii* strains by using the pangenome PanOCT analysis software. We determined that 2,651 open reading frames (ORFs) were common to all 69 strains and 2,906 ORFs were common to all 55 GC2 isolates. An additional 2,427 ORFs were present in subsets of at least two GC2 strains, illustrating the extent of gene content variability within this group. A gene content tree based on a pairwise distance matrix of shared gene content has a topology much different from that of the core SNV tree, yet the well-supported SNV-based clades are preserved (Fig. 1D). The difference in tree topology is driven primarily by gene gain of laterally acquired plasmid- and phage-associated genes and by IS-mediated gene loss, as discussed below. Significant variation in gene content was found, with only a few strains having identical genetic complements. In fact, in the 39 UH GC2 genomes analyzed here, there are 24 distinct gene sets. In the following sections, we describe the nature of these gene gain and loss events and their association with antibiotic resistance genes and genes that may be associated with other traits of clinical interest.

**Chromosomal gene gains: RIs.** Variations at three RI locations and structures were found in the UH strains (Fig. 2A). Two of the RIs were similar to previously identified GC2 RIs, with the third representing a previously undescribed RI location in *A. baumannii*. RI distribution among clades was generally consistent with the

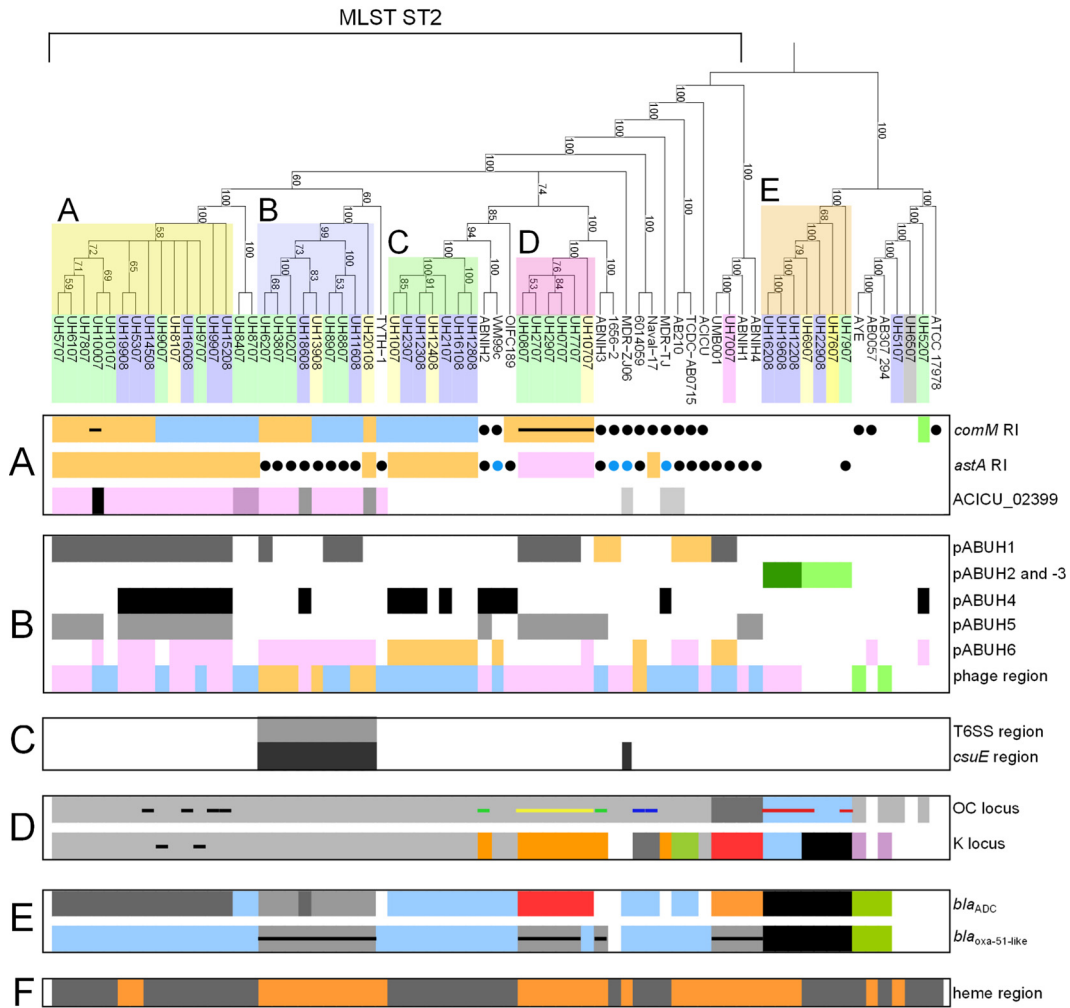


**FIG 1** Phylogenetic trees. (A) Core SNV maximum-likelihood phylogenetic tree constructed from a whole-genome alignment in Mauve and filtered to remove recombinant regions, resulting in 129,899 SNVs used for tree building. (B) Core SNV phylogenetic tree showing relationships among only the ST2 (or GC2) genomes. (C) Core SNV phylogenetic tree topology showing the relationships among all of the strains. Confidence values from 100 bootstrap iterations are shown at the nodes. Clade color boxes represent well-supported primary lineages containing UH strains. Strain name boxes are color coded by the hospital of origin, with reference genomes left uncolored, where green indicates a tertiary hospital, blue indicates an ECC, yellow indicates community hospital A, pink indicates community hospital B, and gray indicates community hospital C. (D) Gene content tree. Information about the percentage of shared genes from PanOCT pangenome clustering for each pair of strains was converted into a distance matrix and used to construct a neighbor-joining tree. Vertical color bars between panels C and D highlight the placement of primary clades in core SNV and gene content trees.

phylogenetic tree, but with a few important differences. Long reads from Pacific Biosciences RS sequencer (PacBio) single-molecule real-time (SMRT) sequencing aided in defining the genetic structure and chromosomal position of RIs in UH9907 and UH10707.

AbaR4-type RIs (also called AbGR1 [26]) were inserted into the *comM* gene of all GC2 UH strains, except for the UH7007 clade which had no RI at this position, and are primarily absent from non-GC2 strains, except for reference strains ATCC 17978, AYE,

and AB0057. Non-GC2 UH strain UH5207 contained genes associated with metal resistance in this RI. The primary difference within the GC2 strains centered on whether a second copy of a Tn6021-like transposon was present in the RI. This version of the RI, as determined from the UH9907 PacBio assembly, was identical to RI<sub>TYTH-1</sub> of TYTH-1, which was isolated in Taiwan in 2008 (41). The other variant, as determined from the UH6107 and UH10707 assemblies, was identical to AbaR4a (42), but the UH10707 version had an additional *ISAbal* copy downstream



**FIG 2** Genome features. The phylogenetic tree from Fig. 1C is shown at the top with the well-supported clades labeled. (A) RI gene content. Bar colors represent different variants; blank means that no RI is present. Horizontal bars in the *comM* RI row indicate the presence of *ISAbal1* at the same location. Black circles here indicate that an RI is present but is not similar to the UH strain RIs in content or organization. In the *astA* RI row, black circles indicate that an IS26-mediated deletion was present while blue circles indicate that an AbGRI2-like RI is present here. In the ACICU\_02399 row, bar colors show whether Tn1548 is present at the ACICU\_02399 location (pink) or at an undetermined location (gray) (see text for further details of variations). (B) Mobile genetic element distribution, where blank indicates that an element is not present. pABUH1 (gray) carries *bla*<sub>OXA-23</sub> and *aphA6*, while pACICU2 (orange) does not. pABUH2 and pABUH3 color differences reflect the presence of *bla*<sub>OXA-40</sub> on different plasmid backbones. pABUH4 and pABUH5 are large plasmids (~110 kb) carrying phage-related sequences. pABUH6 color differences reflect two variations, where pABUH6a (pink) is essentially identical to pAB0057 (CP001183) and pABUH6b (orange) has an additional 3 kbp, including a putative *mob* ORF (Fig. 1). Phage content refers to the phage-related region located at ACICU 1.11 Mbp, where blue indicates that both phages are present, pink indicates that the first phage is present, orange indicates that the second phage is present, and green indicates that a different phage is present (see the text for details). (C) Variation in the presence of a T6SS gene cluster and the *csuE* gene cluster. Shading, present; blank, absent. (D) Surface polysaccharide variants for the OC and capsular polysaccharide (K) loci. Blank, the structure is unique to that strain; horizontal bar color represents *ISAbal1* insertion locations within each locus. (E) Intrinsic chromosomal  $\beta$ -lactamase alleles. Blank indicates unique alleles, blue indicates that *bla*<sub>ADC</sub> is ADC30 variant, and horizontal bars indicates the presence of an upstream *ISAbal1* for *bla*<sub>OXA-51-like</sub>. (F) Variant region showing recombination around the heme utilization region. Orange indicates that the heme region is present, and gray indicates that the region is absent.

from the *sul* ORF. The distribution of the two variants among UH GC2 strains did not strictly conform to the phylogeny.

An AbGRI2-type island (sensu reference 26) was located adjacent to the arginine *N*-succinyltransferase gene (*astA*) (1.318-Mb A1S\_1093 in ATCC 17978) and associated with an ~40-kb chromosomal deletion (A1S\_1093 to A1S\_1126 in ATCC 17978), as found in other GC2 strains (26). The only non-GC2 strain to have a deletion in this region was UH7607. Within the GC2 strains, there were two variants whose distributions were primarily consistent with the phylogeny. The structure and content of one vari-

ant (clades A and C) were similar to those of the *AbaR*-like region of AbGRI2-1 from Australian reference strain WM99c containing *bla*<sub>TEM</sub>, which confers resistance to ampicillin, and *aphA1*, which confers resistance to kanamycin and neomycin, except that it lacked the class 1 integron portion of the element (26). Genomes in clade B had a larger deletion at this location (65 kb relative to 1.264 Mb [A1S\_1083] to 1.329 Mb [A1S\_1138] in ATCC 17978), with an IS26 copy at the deletion site but no apparent antibiotic resistance genes (e.g., UH8907). The second variant (clade D) did not have the 40-kb chromosomal deletion but had a smaller dele-

tion in the opposite direction (1.318 Mb [A1S\_1127] to 1.324 Mb [A1S\_1134] in ATCC 17978) and contained *aphA1* and *aac3IIa* (resistance to gentamicin) (e.g., UH10707).

The third and previously undescribed RI was located in a region corresponding to ~2.53 Mb in ACICU between genes encoding a major facilitator superfamily transporter and a putative GNAT family acetyltransferase (ACICU\_02398 and ACICU\_02399), where the content and organization are similar to those of Tn1548 (43). This RI has been reported on plasmid pZJ06 (CP001938) (44). This chromosomal location was confirmed by using PCR to verify RI flanking regions (data not shown). Several resistance genes are located within this element that confer resistance to aminoglycosides through target modification (*armA*) and enzymatic modification [*aadA1*, *aac(6')-Ib*, *aphA1*] and to chloramphenicol (*catB8*). The distribution of the Tn1548-like RI is largely consistent with the phylogeny and present in GC2 reference strain TYTH-1. Tn1548 is also present in the draft genomes of MDR-TJ and AB210, but its location could not be confirmed. The exception to this is UH10007, which does not carry the Tn1548-like element but has a deletion at the same location. However, the RI is present in closely related isolate UH10107, which was recovered from the same patient as UH10007. UH8407 and UH8707 have an IS26 element at this location and some Tn1548 genes, but the draft assembly state precludes precise definition of the structure. UH20108 and UH18608 have an IS26 copy at this location but no apparent RI.

**Mobile gene content differences.** Plasmid distribution among the UH and reference strains reveals a dynamic exchange occurring across hospital locations and within patients (Fig. 2B). One large pACICU2-like plasmid (designated pABUH1) present in the UH strains carries *bla<sub>OXA-23</sub>* and *aphA6*, both associated with IS elements (Fig. 2B; see Fig. S1 in the supplemental material). A fragment of *ISAbal* is located upstream of *aphA6*, whereas *bla<sub>OXA-23</sub>* is flanked by *ISAbal25* copies in the same orientation and inserted at the location where pACICU2 has a single *ISAbal25* copy. Other non-UH strains carry *bla<sub>OXA-23</sub>* in RIs, except *A. baumannii* UMB001, where it is also on a pACICU2-like plasmid. The plasmid occurs in two GC2 clades, A and D, and is variably present in a third, B. The dynamic nature of plasmid transfer is illustrated in three strains derived from a single patient over a 14-day period. Strains UH3807 (day 1) and UH6207 (day 14) have essentially identical genome sequences (clade B), except that UH6207 carries pABUH1. UH6107, isolated on day 7, has a clade A sequence. Our analysis suggests that pABUH1 was transferred from UH6107 to the UH3807-UH6207 background in the context of a mixed infection in this patient. The other clade B genomes carrying pABUH1 were isolated after UH6207.

Antibiotic resistance plasmids harboring *bla<sub>OXA-24/40</sub>* were restricted to clade E but found across hospital locations. Strains in clade E each carried two novel plasmids (pABUH2 and pABUH3) that have different *rep* and *mob* genes (Fig. 2B; see Fig. S1 in the supplemental material), where the *rep* gene in pABUH2 was similar to GR12 in the replicon typing scheme of Bertini et al. (45), but the *rep* gene in pABUH3 was novel and most similar to *rep* genes from plasmids in *A. lwoffii*. In one subclade, *bla<sub>OXA-24/40</sub>* occurred on pABUH2 (e.g., UH19608), and in the other subclade, it occurred on pABUH3 (e.g., UH7607). The *bla<sub>OXA-24/40</sub>* gene is flanked by XerC/XerD recombination sites on both plasmids, consistent with previous work demonstrating that this site-specific recombination system is mobilizing *bla<sub>OXA-24/40</sub>* in *A. bau-*

*mannii* (28, 29). The same alternative gene encoding a putative acetyltransferase is located at the corresponding location on the other non-*bla<sub>OXA-40</sub>* plasmid, also flanked by XerC/XerD sites. The XerC/XerD inverted repeat and intervening sequences from the two plasmids are nearly identical (see Fig. S1 in the supplemental material), with two nucleotide differences in the upstream XerC/XerD site that suggest a recombination event between the two plasmids at the upstream XerD location.

Two additional large plasmids, pABUH4 and pABUH5, each ~110 kbp and containing primarily phage-related sequences, were identified in the UH strains. pABUH-4 is similar to pABTJ2 from Chinese isolate MDR-TJ (contig accession number CP004359). The pABUH-5 plasmid in the assembly of UH10707 obtained by the hierarchical genome assembly process (HGAP) also carries the element IS26-*aphA1*-IS26-*mph2*-*msrE*-IS26-IS26, which encodes kanamycin and neomycin (*aphA1*) and macrolide (*mph2* and *msrE*) resistance. Some strains carry both of these plasmids (clade B), and the two plasmids may be joined across two phage integrase junctions (V426\_1793 and V426\_1913) in these strains, as indicated by the UH9907 HGAP PacBio assembly (AYOH00000000). The distribution of these plasmids is primarily consistent with the phylogeny, and they are present in reference genomes. Additional plasmids in the UH strains include two pAB0057-like (CP001183) plasmids of 8.7 and 10 kb showing highly variable distribution in the UH and reference strains (pABUH6, Fig. 2B).

Phage-related regions are another source of variability contributing to differentiation among *A. baumannii* strains. In GC2 genomes, three primary variants of a phage region are all located within the same chromosomal location corresponding to 1.11 to 1.16 Mbp in ACICU (ACICU\_00997 to ACICU\_01077) or 1.45 to 1.53 Mbp in TYTH-1 (M3Q\_1334 to M3Q\_1458), where variant distribution within the GC2 strains lacks a discernible geographic or phylogenetic pattern (Fig. 2B). One primary variant typified by the finished reference genome of TYTH-1 consists of two likely phage insertion events flanked by phage integrases. The other variant typified by ACICU has only one of the two phage elements (~45 kb). A third variant (e.g., UH0207-like) possesses only the second of the two phage elements (~36 kb). Strains in non-GC2 clade E vary in whether a phage region is present here. Interestingly, the four strains without a phage insertion at this location (i.e., UH6907, UH7607, UH7907, and UH22907) are also the only four UH strains to possess clustered regularly interspaced short palindromic repeat systems. This location is not the only phage-associated region in the genomes examined, as the UH9907 and UH10707 HGAP assemblies suggest that each has at least two other chromosomal prophage regions.

**Chromosomal gene losses: IS-mediated deletions.** There are multiple instances of chromosomal gene loss that are likely mediated by an IS adjacent to the deletion (Fig. 2C). One clade has two large deletions adjacent to *ISAbal* elements that have not been previously described, including a 40-kb region (1.377 to 1.426 Mbp in ACICU, corresponding to ORFs ACICU\_01272 to ACICU\_01320) encoding the entire type VI secretion system (T6SS) (Fig. 2C). A second *ISAbal*-associated deletion of ~20 kbp (2.538 to 2.558 Mbp in ACICU, corresponding to ORFs ACICU\_02401 to ACICU\_02418) occurred in a region of adhesion genes (*csuE*) and aspartate metabolism near the insertion of the Tn1548-like RI at ACICU\_02399. The *csuE* region is also absent from reference strain MDR-ZJ06.

**Surface polysaccharides.** The UH strains show substantial variation in the content and organization of loci involved in surface polysaccharide synthesis (Fig. 2D), yet there is no evidence of recombination occurring among the UH strains at either region involved in surface polysaccharide production as has been reported in other strain collections (32). Core lipooligosaccharide (LOS) loci (outer core [OC] locus sensu reference 46) consist of one type, OCL1, that is present in most strains, but UH strains have experienced a number of independent *ISAbal* insertions across the clades (horizontal bars in Fig. 2D). GC2 strains in clades A, B, and C have identical capsular (K locus) loci most closely related to ACICU (99.5% identity at the nucleotide level, KL2 in reference 46), except for two strains isolated from the same patient (UH9007 and UH9707) that have a unique *ISAbal* insertion at the same position (Fig. 2D). The K locus of clade D strains is most similar to that of MDR-TJ (98.6% nucleotide identity) or KL9, while all other clades have unique K loci.

**Intrinsic chromosomal resistance genes.** The phylogenetic distribution of different alleles of *bla*<sub>OXA-51-like</sub> and *bla*<sub>ADC</sub> show evidence of recombination and mutation. Each primary UH clade has a different *bla*<sub>ADC</sub> allele, yet the gene is always associated with upstream *ISAbal* oriented to allow overexpression of the gene (except for UH51007, UH5207, and UH6507, which do not have an upstream *ISAbal* copy, Fig. 2E) and present at the same chromosomal location. Sequences from strains A to D are either identical to the extended-spectrum ADC30 *bla*<sub>ADC</sub> variant (47) (clade C) or differ in the  $\Omega$  loop by either the substitution or the addition of another amino acid (see Fig. S2 in the supplemental material). However, non-GC2 clade E has a nucleotide insertion at bp 102 of the coding regions, likely resulting in a truncated ADC protein. Unlike the *bla*<sub>ADC</sub> allelic distribution, the allelic distribution of the *bla*<sub>OXA-51-like</sub> gene, and whether *ISAbal* is present upstream, is best explained by a recombination event. The coding variant in clades B and D (*bla*<sub>OXA-82</sub>), as in reference 22, is linked to the presence of an upstream insertion element, *ISAbal*, suggesting that the entire region was replaced by homologous recombination in one or both lineages. The most common *bla*<sub>OXA-51-like</sub> variant present in the other GC2 strains is *bla*<sub>OXA-66</sub>, with other variants differing by one allele or by more than one in the more distant clades (see Fig. S3 in the supplemental material).

**Recombinant heme utilization region.** Recombination is contributing to strain diversification at multiple scales in the UH strains, including the *bla*<sub>OXA-51-like</sub> region and plasmids described above. Additionally, a large region of 50 kb was identified as a zone of elevated SNV density between 0.965 and 1.015 Mbp in the ACICU genome coordinates. A 10-kb section of this region was previously identified as a recombination hot spot by Snitkin et al. (32), but the inclusion of a broader range of strains reveals that it extends much farther (ORFs corresponding to ACICU\_00866 to ACICU\_00912). There are two major variant types that differ in the presence of a 12-kb region containing several genes involved in heme utilization (ACICU\_00871 to ACICU\_00882, Fig. 2F). Other large regions where recombination is apparent are around the origin of replication in ACICU and the K locus, as described by Snitkin et al. (32).

**Coinfection dynamics.** There were six sets of isolates obtained from the same patients (a total of 13 strains; see Table S1 in the supplemental material). The UH8807-UH8907, UH8407-UH8707, and UH8107-UH9907 pairs were indistinguishable on the basis of SNV patterns and gene contents. One triplet of strains

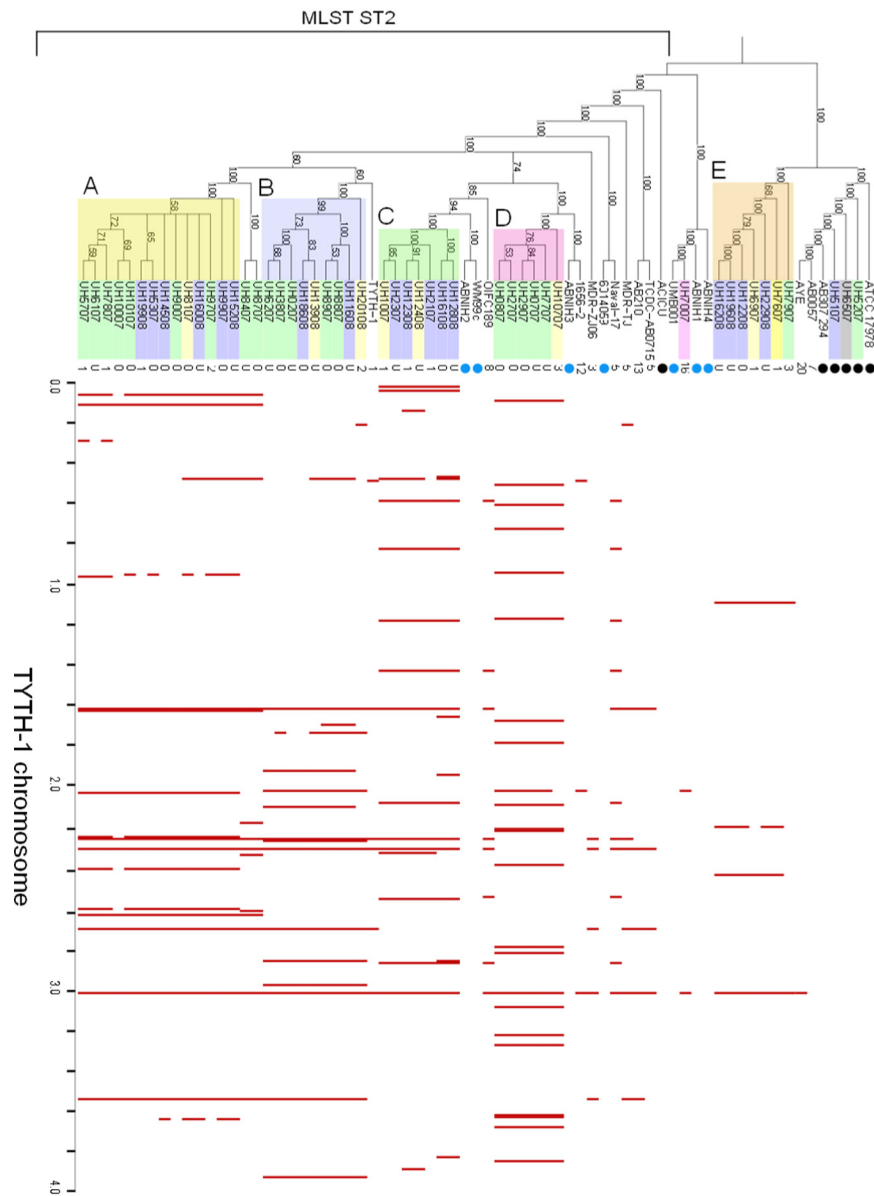
from the same patients was composed of two nearly identical strains of the same clade (UH3807 and UH6207) and a third of a different clade (UH6107) that enabled detection of the apparent plasmid transfer event described above. UH9007 and UH9707, isolated 15 days apart, are indistinguishable by core SNV analysis and have the same unique *ISAbal* insertion in the lipopolysaccharide (LPS) region that no other strains examined possess. Notably, they do differ in that UH9707 carries pABUH6 while UH9007 does not. UH10007 and UH10107, isolated on successive days, are also indistinguishable by core SNV analysis but differ markedly in gene content. For example, UH10007 carries pABUH5 and *aphA6*, which UH10107 lacks, but lacks the Tn1548-like RI that is present in UH10107. Furthermore, UH10007 and UH10107 differ at the *comM* RI in that UH10007 has an *ISAbal* copy inserted at the same location as the strains in clade D.

***ISAbal* insertion locations.** The distribution of *ISAbal* insertion sites among strains reinforced the major strain groups defined by SNV typing (Fig. 3). The pattern of shared sites was largely consistent with the phylogeny and in fact provides additional phylogenetic resolution in several cases. The relationships among several strains in clade A are poorly resolved by SNV analysis, but *ISAbal* insertion locations indicate that UH5707 and UH7807 should have the same terminal node, for example. Strains in clade D had *ISAbal* locations very distinct from those of the other GC2 strains examined, where the only insertion sites that are also present in other clades are adjacent to the *bla*<sub>ADC</sub> and *bla*<sub>OXA-51-like</sub> genes.

## DISCUSSION

A critical aspect of preventing nosocomial infections is understanding the transmission patterns of pathogens. Transmission patterns are particularly important for infections caused by MDR organisms that are difficult to treat. WGS analysis is now used to characterize nosocomial outbreaks (3, 4, 6) and obtain unprecedented insights. For example, Didelot et al. (3) showed that many *Clostridium difficile* infections could not be explained by transmission between symptomatic cases, which suggests that alternative transmission routes and the role of asymptomatic carriers need to be explored further. In contrast, smaller studies of *Klebsiella pneumoniae* and *A. baumannii* isolates have shown clear evidence of patient-to-patient transmission (7, 31).

Our approach was designed to evaluate the diversity of strains present during a short time period in a tertiary-care hospital along with its affiliated community hospitals and an extended-care center (ECC). We found extensive microdiversity among these strains, such that most of the strains examined represented distinct variants. Typing by conventional methods such as MLST suggests that the primary ST representing 39 of 49 strains arose from a single founder and spread clonally. The robust phylogeny constructed from core SNVs and gene content comparisons, however, revealed a much more complicated relationship among *A. baumannii* strains within the hospital system. The strains are composed of highly similar yet distinct lineages within ST2 (or GC2). The presence of reference strains from disparate geographic locations worldwide interspersed among UH strains in the core SNV phylogeny supports the idea that these branches diverged prior to entering the UH hospital system and long enough ago that the descendants are globally dispersed. The limited geographic clustering of strains within the UH hospital system suggests the



**FIG 3** *ISAbal* distribution. Locations of *ISAbal* insertion sites relative to the TYTH-1 chromosome for locations occurring in more than two genomes. The number of unique IS locations in each genome is listed in the first column to the right of the tree. Blue circles, not able to extract IS information from genome assemblies; black circles, no copies of *ISAbal* present.

existence of either considerably more independent founders or rapid mixing of strains among hospitals.

There is also evidence of clonal transmission within the hospital system, as observed in clade D, where five out of six strains originated from the same hospital in the same time frame and have indistinguishable gene contents. Compared to the other clade D strains, UH10707 is more divergent in gene content in lacking the pABUH1 plasmid carrying *bla*<sub>OXA-23</sub> and *aphA6*, carries a different *bla*<sub>OXA-51-like</sub> allele, and was isolated later and in a different hospital. With the available information, it is not possible to say whether UH10707 lost the plasmid from a recent ancestor it has in common with other clade D strains or whether the plasmid was acquired after UH10707 split from the other clade D

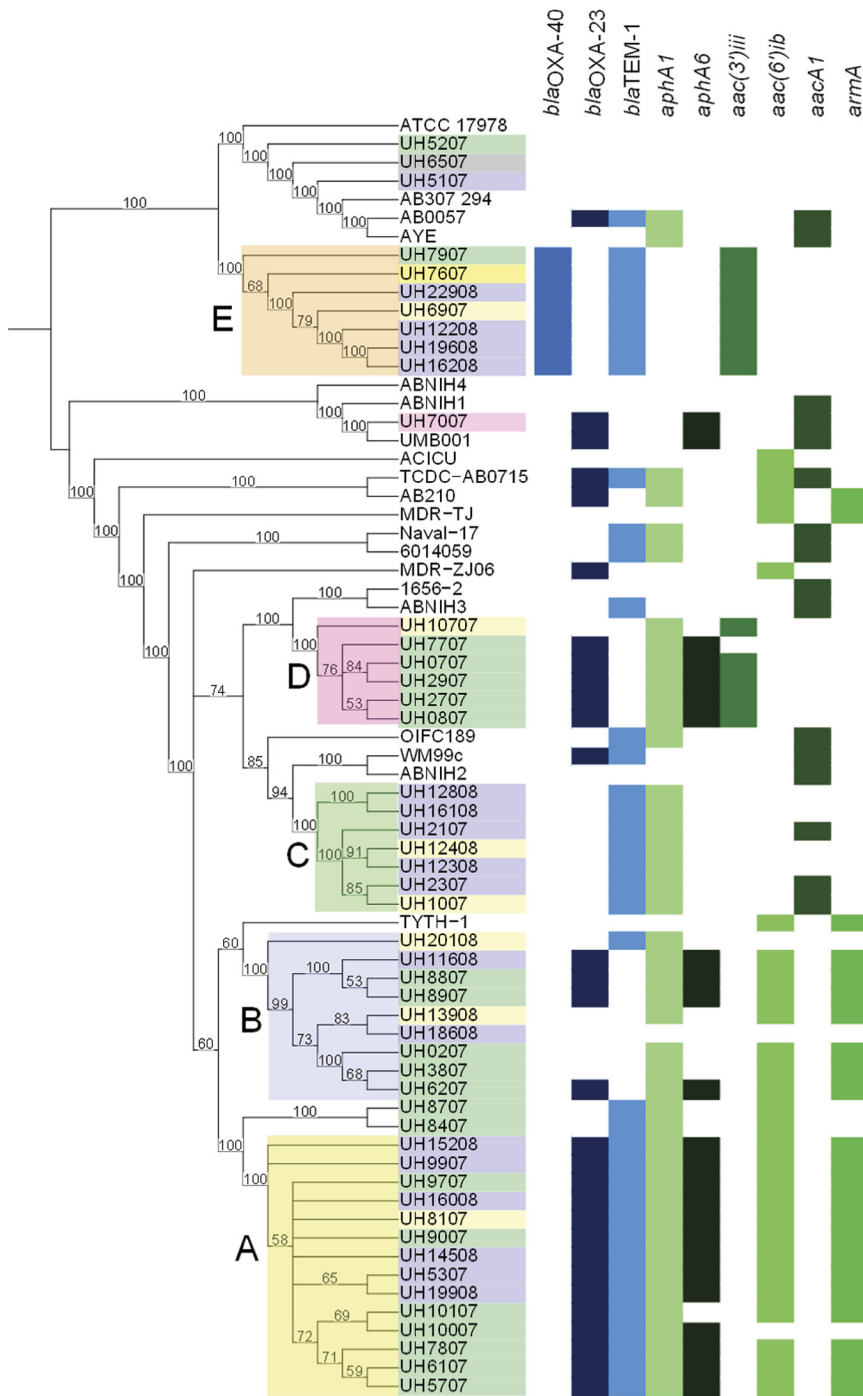
strains. This example highlights the complexity in determining clonality and divergence underlying the phylogenetic patterns and transmission routes observed in these strains.

#### Antibiotic resistance mechanisms.

The *A. baumannii* strains investigated here were all MDR strains, except for the three with novel STs (i.e., UH5107, UH5207, and UH6507), and genomic analyses identified an extensive and redundant repertoire of resistance genes (Fig. 4). The allelic distributions of two intrinsic chromosomal  $\beta$ -lactamase resistance genes were generally consistent within each primary clade but variable even across the closely related strains within the ST2 group. The *bla*<sub>ADC</sub> allelic distribution is suggestive of at least some mutation events leading to variation, but elevated SNV density (10 SNVs/kb) in clade D and other GC2 clades for 15 kb surrounding *bla*<sub>ADC</sub> reveals that recombination is also present (data not shown). Hamidian and Hall (48) have argued that *ISAbal* was independently acquired upstream of *bla*<sub>ADC</sub> variants, but it is also possible that the insertion of IS elements predated subsequent variation through mutation and recombination. Strains in clade E, which have a truncated and most likely nonfunctional ADC because of a frameshift, are generally susceptible to the cephalosporins (ceftazidime and cefepime) tested, and no other cephalosporinase genes are present. Other UH GC2 strains possess the extended-spectrum ADC variant ADC30, which confers broader substrate activity, which agrees with the observed cephalosporin resistance phenotypes (47). The allelic distribution of the *bla*<sub>OXA-51-like</sub>  $\beta$ -lactamase gene, and whether *ISAbal* is present upstream, is best explained by a recombination event, as the coding variant in clades B and D is linked to the presence of an upstream *ISAbal* element, sug-

gesting that the entire region was replaced by homologous recombination in one or both of the lineages. The UH strains carrying the *bla*<sub>OXA-51-like</sub> variant associated with an upstream *ISAbal* element (i.e., those in clades B and D) are all resistant to the carbapenems tested, consistent with the effect of a promoter provided by the IS.

RIs are a major source of variability contributing to microdiversity within the strains examined. To our knowledge, this is the first report of an RI at the ACICU\_02399 location. Tn1548 has been reported in a number of members of the family *Enterobacteriaceae*, and similar sequences have been observed in draft *A. baumannii* genomes and plasmid pZJ06. The chromosomal location of the closely related sequences in ACICU\_02399 reinforces the



**FIG 4** Distribution of the  $\beta$ -lactamase (blue) and aminoglycoside resistance (green) genes present in the UH strains in this study.

role that antimicrobial selection continues to play in driving strain differentiation. This is further evidenced by the presence of redundant genetic resistance mechanisms in RIs and elsewhere in the UH strains, suggesting continued pressure on the pathogen to enhance its ability to escape antimicrobial therapy. For example, several strains in clades B and D have both an acquired carbapenemase gene (*bla*<sub>OXA-23</sub>) and IS*Abal*-driving *bla*<sub>OXA-51</sub>-like. Aminoglycoside resistance mechanisms are particularly redun-

dant. Tn1548 contains the *armA* gene for an rRNA methyltransferase (target modification) that confers resistance to amikacin, gentamicin, and kanamycin (49) and genes that encode drug-specific aminoglycoside-modifying enzymes, *aac(6)-Ib* (gentamicin) and *aphA1* (kanamycin). There are also examples of genotypic redundancy, as is the case for those strains (e.g., clade A UH8107, UH9707, UH9907, UH15208, and UH16008) carrying copies of *aphA1* in both the *comM* and *astA* RIs.

Mobile elements show pronounced variability among strains such that the collective distribution of the plasmids and phage results in unique gene sets for many highly similar strains (Fig. 2B). This includes pABUH1, the pACICU-2-like plasmid that harbors *bla*<sub>OXA-23</sub> and *aphA6*. The pACICU2-like backbone is distributed worldwide, but among reference strains, only UMB001 has the pABUH1-like variant that carries the resistance genes (Fig. 2B). The association of pACICU2-like plasmids and *bla*<sub>OXA-23</sub> but not *aphA6*, has been noted before by Bertini et al. (45), who also demonstrated the ability of the plasmid to be transferred by conjugation. The two large (~110-kb) plasmids carrying phage-derived sequences, pABUH4 and pABUH5, have not been previously reported, although they are present in multiple reference strains. It is difficult to predict whether they confer any fitness advantage because of the overwhelming number of ORFs annotated as hypothetical and conserved hypothetical proteins, but the UH10707 assembly obtained by the HGAP does place a copy of *aphA1* and two macrolide resistance genes in an element containing four copies of IS26 on pABUH5. The detection of highly similar plasmids in reference strains from Australia and China indicates that this plasmid is circulating globally.

**Recombination.** Recombination has been reported to contribute to *A. baumannii* genome change (32), but the extent of recombination has not been studied in depth. Most of the genome regions in the strains sequenced here exhibited low levels of SNV density (<0.1 SNV/kbp). One exception is a recombinant region centered around a 50-kb span in which a putative heme utilization region, including a gene coding for a heme oxygenase (*hemO*) (50, 51), is variably present. The 12-kb heme utilization region is also highly variable among *A. baumannii* reference strains, making it difficult to assess whether this region was present in a common ancestor and lost by some strains or whether it was gained and subsequently



transferred via recombination to different lineages as initially observed by Antunes et al. (50). The region is present in some, but not all, non-*A. baumannii* Acinetobacter isolates, further complicating inference of the ancestral state and whether this region has contributed to the success of *A. baumannii* as a pathogen. Two strains that are grouped together by SNV analysis in clade A, UH5307 and UH19908, were isolated 5 months apart from different patients, but both were from the same extended-care facility and are an example of a recombination event in this region that is likely to have occurred within the UH hospital setting. There is also evidence of smaller-scale recombination events in the allelic distribution of *bla*<sub>ADC</sub> and *bla*<sub>OXA-51-like</sub> variants. Additionally, recombination events are not restricted to the chromosome, as observed by comparison of the gene contents of the two plasmids carrying *bla*<sub>OXA-40</sub>. However, it is not possible to determine the full extent of chromosomal homologous recombination by this approach, as exchanges that involve closely related chromosomal segments are difficult to detect by using an SNV density metric.

**Surface polysaccharide variation.** While there was no evidence of recombination occurring around the surface polysaccharide locus among the GC2 clades within the hospital system, there were multiple types encountered when the more diverse reference strains were examined, consistent with the previous characterization of these regions as a significant source of variability within *A. baumannii* (46). The large number of unique variants of these two loci when including non-UH strains for comparison emphasizes the potential selective pressure acting on these regions that play a role in interaction with the host immune system. However, the presence of reference strains with the same organization as that observed in the UH strains indicates that, on some level, these regions are stable. The OC locus is less variable, with one predominant type detected, but interestingly, there are multiple cases of independent *ISAbal* insertion events within the OC region. It is not clear whether these are simply random insertions that are selectively neutral or whether they affect the cell surface and potentially the host response.

One primary difference between GC2 strains and the other lineages examined here could have implications for surface polysaccharide structure and host interactions. Previous bioinformatic and biochemical analyses suggested that *A. baumannii* produces not LPS but LOS, with the distinction being that LOS lacks the O-antigen sugar repeat unit ligated to the core oligosaccharide because of the lack of an identifiable *waaL* O-antigen ligase gene (46). Our analysis showed that the GC2 strains, except the UH7007 clade, have two adjacent ORFs with WaaL conserved domains near the pilus locus (at 3.58 Mb in ACICU), where previous sequence analyses have detected only one such ORF at that location (46). The WaaL domain is associated both with genes encoding O-antigen ligases and with *pglL* genes involved in O-linked protein glycosylation (52), but assigning a definite function to a gene containing this domain on the basis of sequence data alone has previously not been possible. Recently, a hidden Markov model (HMM) protein family was developed to discriminate between WaaL proteins involved in protein glycosylation and those involved in O-antigen linkage (52). Of the two adjacent ORFs in the GC2 strains, one (ACICU YP\_001848035) is most certainly *pglL* on the basis of this HMM and previous characterization of protein glycosylation in *A. baumannii* (53). However, the second ORF with a WaaL domain (ACICU YP\_001848036) does not have the *pglL*-specific domain, as assessed by using the HMM, leading

us to speculate that it encodes the O-antigen ligase rather than a second protein involved in mediating protein glycosylation. Further work is necessary to establish whether this second ORF is expressed and whether strains harboring this second ORF are capable of producing LPS or, alternatively, glycosylating different surface proteins.

***ISAbal* distribution.** IS elements are important drivers of genome change in *A. baumannii* and other pathogens, as demonstrated by the roles they play in mobilizing antibiotic resistance genes, modulating gene expression, and mediating gene deletions, as observed in this study and others (54). The analysis undertaken here represents the first attempt to map the distribution of an IS within an *A. baumannii* genome and compare its locations among strains. The chromosomal distribution of *ISAbal* locations reveals not only that it is abundant in UH strains, with strains in clade D possessing >20 copies, but also that the whole genome is impacted. *ISAbal* may be a recent acquisition by GC2 strains, as it is absent from ACICU, and may have contributed to the spread of this lineage. The facts that the locations of *ISAbal* insertion sites can be used for phylogenetic analysis and are concordant with the core phylogeny suggest that, once inserted, these elements tend to stay in the chromosome, as no clear cases of *ISAbal* loss could be inferred from the genomes analyzed here.

**Other gene content variability with potential fitness implications.** Beyond variable antibiotic resistance gene and mobile element presence, there is substantial variation among strains in gene content that may have fitness implications. In addition to the putative heme utilization recombinant region and the region deleted with the insertion of AbGRI2 in many of the GC2 strains, there are multiple examples of IS-mediated gene deletions, including the deletion of the entire T6SS operon, and the *csuE*/aspartate metabolism regions from strains in clade C. These regions have been shown to be involved in interbacterial interactions (55, 56), and the initial adherence of cells to abiotic surfaces (57), respectively. One hypothesis is that the MDR phenotype conferred by the extensive antibiotic resistance gene repertoire renders the T6SS regions less important, as interbacterial competition is most likely reduced over the course of antibiotic therapy. The T6SS region is conserved among all of the other *A. baumannii* isolates examined, reinforcing the monophyletic structure of clade D. The persistence of cells with the *csuE* region deletion as indicated by the recovery of isolates with these deletions over the span of several months from multiple source types, including catheters, is surprising given the prediction of the region's importance for persistence on hospital surfaces (14, 57, 58). Furthermore, the *csuE* deletion is hinted at in *A. baumannii* strains from Latvia (59), and these genes are absent from the MDR-ZJ06 assembly, suggesting that this loss occurred before strains entered the UH hospital system and that these strains are persisting in hospital environments despite this loss. The deletion of the *csuE* gene also has diagnostic significance, as it has been proposed as a molecular marker for strain typing efforts (21).

**Coinfection as a facilitator of LGT.** For recombination and LGT to occur, two strains with differing gene contents must interact in a way that facilitates DNA exchange. Most of the studies done to date have characterized single isolates from a patient or assumed single strain infections when developing treatment strategies for patients. Data from the UH strains suggest that patients can be colonized or infected by multiple strains and that they are capable of interacting within the patient. This is best exemplified

by the likely transfer of pABUH1 among the UH3807-UH6107-UH6207 isolates. Other candidates for LGT in the context of coinfection are strains UH10007 and UH10107, which have differences in RIs and plasmid contents that cannot be explained by gene loss alone despite being isolated from the same patient only 1 day apart and being nearest neighbors in the SNV phylogeny. We hypothesize that an unsampled strain was present as a coinfecting strain and transferred genes to UH10007. UH9007 and UH9707 also varied in plasmid content, with UH9707 carrying pABUH6, which is missing from UH9007. In this case, it is difficult to determine whether these were initially two different strains that colonized the patient or, alternatively, whether this is an example of plasmid loss from one isolate occurring within the patient. Repeated infection with genetically distinct *A. baumannii* strains has been reported in the context of colistin treatment as well (60). Taken together, these findings suggest *in vivo* genetic exchange among different strains occurring within patients and provide a mechanism by which *A. baumannii* obtains new genetic information, including antibiotic resistance genes. Moreover, divergent strains with different STs were also encountered in the UH strains and represent potential opportunities for new genetic material to be introduced into GC2 strains through such a mechanism.

**Conclusion.** The presence of closely related strains in Asia and Europe and the presence of mobile elements recovered from geographically widespread areas highlight the global dissemination of these organisms and genes. By investigating dynamics within one hospital system, we were able to identify evolutionary mechanisms that contribute to this process at a local scale. There was limited spatial or temporal clustering of strain types and gene contents within different hospital components, indicating that an endemic and interacting *A. baumannii* population exists either within the UH hospital system or in patients colonized with the bacteria. The movement of patients and staff between the affiliated hospital locations may contribute to strain mixing and diversification. Previous work has hypothesized that ECCs represent reservoirs for health care-associated pathogens, including *A. baumannii* (61, 62). The observation that the same lineages and gene contents observed in the tertiary-care hospital and regional hospitals are also detected in the ECCs provides support for this hypothesis. Alternatively, asymptomatic carriers could facilitate a standing *A. baumannii* population circulating in the community. While the lack of high-resolution patient data in this study precludes a rigorous assessment of specific transmission routes, the data indicate that transmission and gene flow occurred among the hospitals.

**Summary.** Strain level whole-genome analyses of *A. baumannii* isolated from one integrated United States hospital system demonstrate that nearly every strain was unique despite being indistinguishable by conventional sequence typing methods and in some cases by core SNV typing. This study improves our understanding of the evolutionary processes that contribute to the emergence of MDR nosocomial pathogens, including genomic mixing during coinfection events. The analyses reported here lead to an improved conceptual model of *A. baumannii* population dynamics within hospitals that suggests that endemic strains exist and interact with one another, with the additional periodic influx of novel strains that potentially bring in new genetic material. These findings highlight the importance of identifying and screening high-risk patients, such as those who come from extended-care facilities or have previous antibiotic exposure, and the impor-

tance of developing rapid diagnostic tools to characterize antibiotic gene content in individual isolates. In addition to variability in antibiotic resistance determinants, other genomic regions demonstrate dynamic genomic change over short evolutionary time spans and may point to other aspects of *A. baumannii* physiology that contribute to its success as a nosocomial pathogen.

## MATERIALS AND METHODS

**Strain isolation and genotypic and phenotypic characterization.** Strains were isolated between 2007 and 2008 from an integrated health care system, the UH of Cleveland in Cleveland, OH. The UH are composed of a main tertiary-care facility affiliated regional hospitals and an ECC. The hospital system detected an increased prevalence of MDR *A. baumannii* strains beginning in 2007. At that point, isolates were subjected to additional analysis to characterize antibiotic resistance phenotypes and STs via MLST (37). From this collection of isolates, 49 isolates were selected for sequencing to capture a range of hospital locations, dates, genotypes, and antibiotic resistance phenotypes (see Table S1 in the supplemental material). Additionally, sets of strains isolated from the same patient were selected to investigate potential within-host strain interactions.

To put the sequenced strains (i.e., UH strains) in a phylogenetic context and assess shared and clade-specific gene contents, genomes were compared to reference draft and complete *A. baumannii* genomes available from NCBI as of January 2013 (see Table S2 in the supplemental material).

**DNA preparation, library construction, sequencing, and assembly.** DNA was isolated with the MasterPure Gram-positive DNA purification kit (Epicenter Biosciences). Illumina sequencing libraries were prepared by using either Nextera or TruSeq kits with indexed-encoded adapters from Illumina, according to the manufacturer's instructions. Libraries were pooled for sequencing on Illumina GAIIx or HiSeq, and paired-end sequence reads were obtained representing 50- to 200-fold genome coverage. Strains representing two distinct ST2 lineages (UH9907 and UH10707) were also subjected to SMRT sequencing on a PacBio. PacBio sequencing resulted in long reads (median, ~3.5 kbp) and ~10× to 20× coverage of error-corrected reads. Illumina sequence data were assembled by using Velvet (63). A range of k-mer values was evaluated, and the assembly with the largest N50 was selected for annotation and analysis. The PacBio sequence was assembled by using the HGAP (64). Several hybrid assemblies combining PacBio with Illumina sequence data were not sufficiently superior to the assembly obtained by the HGAP (data not shown). For information on assembly quality, see Table S3 in the supplemental material. Assemblies of Illumina data generally had contig N50 values of >100 kbp. The PacBio assemblies had contig N50 values of >1 Mbp.

**Genome annotation.** Genes were annotated in each genome assembly by using an automated annotation system (65). This annotation pipeline defines protein- and RNA-coding genes and assigns names (66) and functional classifications based on TIGRFAMs (67), Pfams (68), and CharProtDBs (69).

**Core phylogeny construction.** SNVs were identified on the basis of whole-genome alignment of 69 UH and reference assemblies with the SNV export functionality within Mauve (70). The full list of SNVs was then filtered by requiring that at least 67 genomes had a sequence at the variable position. These 162,180 candidate core SNVs were subsequently examined for evidence of recombination by plotting SNV density across 1-kb bins using the finished genome of the GC2 *A. baumannii* ACICU (NC\_010611.1) as a reference, where regions with elevated SNV density (pairwise, >10 SNVs/kb) were then excluded from core phylogeny construction. This filtering process yielded 129,899 presumed nonrecombinant SNVs, which were used to generate a core phylogeny (see Table S4 in the supplemental material). A maximum-likelihood tree was constructed by using RAxML (71) with 100 bootstrap replicates.

**Pangenome characterization: PanOCT and the gene content tree.**

Genes were clustered into ortholog sets with the pangenome analysis program PanOCT (72), which considers the gene neighborhood when identifying orthologs. A minimum identity of 70% was required for ORFs to be placed in the same cluster. A genome-wide measure of shared gene content was calculated by computing a distance matrix based on the number of genes shared by each pair of genomes. This distance matrix was used to generate a phylogenetic tree with FASTME (73).

**ISAbal1 analysis.** Insertion sites for the ISAbal1 element were inferred from draft genome assemblies by identifying fragments of the element at contig edges and mapping the adjacent flanking sequence to the finished reference genome most closely related to the ST2 UH strains, *A. baumannii* TYTH-1 (NC\_018706.1) (41).

**Nucleotide sequence accession numbers.** The sequences obtained in this study were deposited in the GenBank database under accession numbers AYGS00000000 to AYE00000000 for the Illumina Velvet assemblies and AYOI00000000 (UH9907) and AYOI00000000 (UH10707) for the PacBio HGAP assemblies.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00963-13/-/DCSupplemental>.

- Figure S1, EPS file, 0.2 MB.
- Figure S2, EPS file, 0.6 MB.
- Figure S3, EPS file, 0.3 MB.
- Table S1, XLSX file, 0.1 MB.
- Table S2, XLSX file, 0.1 MB.
- Table S3, XLSX file, 0.1 MB.
- Table S4, TXT file, 18.0 MB.

**ACKNOWLEDGMENTS**

This work was supported by National Institute of General Medical Sciences grant R01GM094403 to M.D.A. Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under grants R01AI072219 and R01AI063517 awarded to R.A.B. This study was supported in part by funds and/or facilities provided by the Cleveland Department of Veterans Affairs, the Veterans Affairs Merit Review Program, and the Geriatric Research Education and Clinical Center VISN 10 to R.A.B. F.P. is a Louis Stokes Scholar at Case Western Reserve University and supported by the Case Western Reserve University/Cleveland Clinic Clinical and Translational Science Collaborative U1TR000439.

We thank Erin Beck for assistance with PanOCT pangenome analysis, the J. Craig Venter Institute sequencing group and the Case Western Reserve University Genomics Core for producing Illumina sequence data, and the University of California San Diego BioGEM facility for PacBio sequencing.

**REFERENCES**

1. Sievert DM, Ricks P, Edwards JR, Schneider A, Patel J, Srinivasan A, Kallen A, Limbago B, Fridkin S, National Healthcare Safety Network (NHSN) Team and Participating NHSN Facilities. 2013. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009–2010. *Infect. Control Hosp. Epidemiol.* 34:1–14. <http://dx.doi.org/10.1086/668770>.
2. Zimlichman E, Henderson D, Tamir O, Franz C, Song P, Yamin CK, Keohane C, Denham CR, Bates DW. 2013. Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system. *JAMA Intern. Med.* 173:–2039–2046.
3. Didelot X, Eyre DW, Cule M, Ip CL, Ansari MA, Griffiths D, Vaughan A, O'Connor L, Golubchik T, Batty EM, Piazza P, Wilson DJ, Bowden R, Donnelly PJ, Dingle KE, Wilcox M, Walker AS, Crook DW, A Peto TE, Harding RM. 2012. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol.* 13:R118. <http://dx.doi.org/10.1186/gb-2012-13-12-r118>.
4. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CL, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TE, Walker AS, Crook DW. 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open* 2:e001124.
5. Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J, Peacock SJ. 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 8:e1002824.
6. Reuter S, Ellington MJ, Cartwright EJ, Koser CU, Torok ME, Gouliouris T, Harris SR, Brown NM, Holden MT, Quail M, Parkhill J, Smith GP, Bentley SD, Peacock SJ. 2013. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. *JAMA Intern. Med.* 173:1397–1404.
7. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program Group, Henderson DK, Palmore TN, Segre JA. 2012. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci. Transl. Med.* 4:148ra116.
8. Adams MD, Chan ER, Molyneaux ND, Bonomo RA. 2010. Genome-wide analysis of divergence of antibiotic resistance determinants in closely related isolates of *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54:3569–3577. <http://dx.doi.org/10.1128/AAC.00057-10>.
9. Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ, MacDonald IJ, Martin KM, Russo T, Campagnari AA, Hujer AM, Bonomo RA, Gill SR. 2008. Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. *J. Bacteriol.* 190: 8053–8064. <http://dx.doi.org/10.1128/JB.00834-08>.
10. Arroyo LA, Herrera CM, Fernandez L, Hankins JV, Trent MS, Hancock RE. 2011. The *pmrCAB* operon mediates polymyxin resistance in *Acinetobacter baumannii* ATCC 17978 and clinical isolates through phosphoethanolamine modification of lipid A. *Antimicrob. Agents Chemother.* 55: 3743–3751. <http://dx.doi.org/10.1128/AAC.00256-11>.
11. Fournier PE, Vallenet D, Barbe V, Audic S, Ogata H, Poirel L, Richet H, Robert C, Mangenot S, Abergel C, Nordmann P, Weissenbach J, Raoult D, Claverie JM. 2006. Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLoS Genet.* 2:e7. <http://dx.doi.org/10.1371/journal.pgen.0020007>.
12. Peleg AY, Seifert H, Paterson DL. 2008. *Acinetobacter baumannii*: emergence of a successful pathogen. *Clin. Microbiol. Rev.* 21:538–582. <http://dx.doi.org/10.1128/CMR.00058-07>.
13. Perez F, Hujer AM, Hujer KM, Decker BK, Rather PN, Bonomo RA. 2007. Global challenge of multidrug-resistant *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 51:3471–3484. <http://dx.doi.org/10.1128/AAC.01464-06>.
14. Sahl JW, Gillece JD, Schupp JM, Waddell VG, Driebe EM, Engelthaler DM, Keim P. 2013. Evolution of a pathogen: a comparative genomics analysis identifies a genetic pathway to pathogenesis in *Acinetobacter*. *PLoS One* 8:e54287. <http://dx.doi.org/10.1371/journal.pone.0054287>.
15. Zarrilli R, Pournaras S, Giannouli M, Tsakris A. 2013. Global evolution of multidrug-resistant *Acinetobacter baumannii* clonal lineages. *Int. J. Antimicrob. Agents* 41:11–19. <http://dx.doi.org/10.1016/j.ijantimicag.2012.09.008>.
16. Dijkshoorn L, Nemec A, Seifert H. 2007. An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. *Nat. Rev. Microbiol.* 5:939–951. <http://dx.doi.org/10.1038/nrmicro1789>.
17. Poirel L, Bonnin RA, Nordmann P. 2011. Genetic basis of antibiotic resistance in pathogenic *Acinetobacter* species. *IUBMB Life* 63: 1061–1067. <http://dx.doi.org/10.1002/iub.532>.
18. Héritier C, Poirel L, Fournier PE, Claverie JM, Raoult D, Nordmann P. 2005. Characterization of the naturally occurring oxacillinase of *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 49:4174–4179. <http://dx.doi.org/10.1128/AAC.49.10.4174-4179.2005>.
19. Héritier C, Poirel L, Nordmann P. 2006. Cephalosporinase overexpression resulting from insertion of ISAbal1 in *Acinetobacter baumannii*. *Clin. Microbiol. Infect.* 12:123–130. <http://dx.doi.org/10.1111/j.1469-0691.2005.01320.x>.
20. Hujer KM, Hamza NS, Hujer AM, Perez F, Helfand MS, Bethel CR, Thomson JM, Anderson VE, Barlow M, Rice LB, Tenover FC, Bonomo RA. 2005. Identification of a new allelic variant of the *Acinetobacter baumannii* cephalosporinase, ADC-7 beta-lactamase: defining a unique family of class C enzymes. *Antimicrob. Agents Chemother.* 49:2941–2948. <http://dx.doi.org/10.1128/AAC.49.7.2941-2948.2005>.

21. Turton JF, Ward ME, Woodford N, Kaufmann ME, Pike R, Livermore DM, Pitt TL. 2006. The role of ISAbal in expression of OXA carbapenemase genes in *Acinetobacter baumannii*. *FEMS Microbiol. Lett.* 258: 72–77. <http://dx.doi.org/10.1111/j.1574-6968.2006.00195.x>.
22. Evans BA, Brown S, Hamouda A, Findlay J, Amyes SG. 2007. Eleven novel OXA-51-like enzymes from clinical isolates of *Acinetobacter baumannii*. *Clin. Microbiol. Infect.* 13:1137–1138. <http://dx.doi.org/10.1111/j.1469-0691.2007.01828.x>.
23. Tian GB, Adams-Haduch JM, Taracila M, Bonomo RA, Wang HN, Doi Y. 2011. Extended-spectrum AmpC cephalosporinase in *Acinetobacter baumannii*: ADC-56 confers resistance to cefepime. *Antimicrob. Agents Chemother.* 55:4922–4925. <http://dx.doi.org/10.1128/AAC.00704-11>.
24. Villalón P, Valdezate S, Medina-Pascual MJ, Carrasco G, Vindel A, Saez-Nieto JA. 2013. Epidemiology of the *Acinetobacter*-derived cephalosporinase, carbapenem-hydrolysing oxacillinase and metallo- $\beta$ -lactamase genes, and of common insertion sequences, in epidemic clones of *Acinetobacter baumannii* from Spain. *J. Antimicrob. Chemother.* 68: 550–553. <http://dx.doi.org/10.1093/jac/dks448>.
25. Bonnin RA, Poirel L, Nordmann P. 2012. AbaR-type transposon structures in *Acinetobacter baumannii*. *J. Antimicrob. Chemother.* 67: 234–236. <http://dx.doi.org/10.1093/jac/dkr413>.
26. Nigro SJ, Farrugia DN, Paulsen IT, Hall RM. 2013. A novel family of genomic resistance islands, AbGR12, contributing to aminoglycoside resistance in *Acinetobacter baumannii* isolates belonging to global clone 2. *J. Antimicrob. Chemother.* 68:554–557. <http://dx.doi.org/10.1093/jac/dks459>.
27. Post V, White PA, Hall RM. 2010. Evolution of AbaR-type genomic resistance islands in multiply antibiotic-resistant *Acinetobacter baumannii*. *J. Antimicrob. Chemother.* 65:1162–1170. <http://dx.doi.org/10.1093/jac/dkq095>.
28. D'Andrea MM, Giani T, D'Arezzo S, Capone A, Petrosillo N, Visca P, Luzzaro F, Rossolini GM. 2009. Characterization of pABVA01, a plasmid encoding the OXA-24 carbapenemase from Italian isolates of *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 53:3528–3533. <http://dx.doi.org/10.1128/AAC.00178-09>.
29. Merino M, Acosta J, Poza M, Sanz F, Beceiro A, Chaves F, Bou G. 2010. OXA-24 carbapenemase gene flanked by XerC/XerD-like recombination sites in different plasmids from different *Acinetobacter* species isolated during a nosocomial outbreak. *Antimicrob. Agents Chemother.* 54: 2724–2727. <http://dx.doi.org/10.1128/AAC.01674-09>.
30. Mugnier PD, Poirel L, Naas T, Nordmann P. 2010. Worldwide dissemination of the blaOXA-23 carbapenemase gene of *Acinetobacter baumannii*. *Emerg. Infect. Dis.* 16:35–40.
31. Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D, Pallen MJ. 2010. High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. *J. Hosp. Infect.* 75:37–41. <http://dx.doi.org/10.1016/j.jhin.2010.01.012>.
32. Snitkin ES, Zelazny AM, Montero CI, Stock F, Mijares L, NISC Comparative Sequence Program, Murray PR, Segre JA. 2011. Genome-wide recombination drives diversification of epidemic strains of *Acinetobacter baumannii*. *Proc. Natl. Acad. Sci. U. S. A.* 108:13758–13763. <http://dx.doi.org/10.1073/pnas.1104401108>.
33. Decker BK, Perez F, Hujer AM, Hujer KM, Hall GS, Jacobs MR, Gebreyes WA, Zoll ST, Massire C, Eshoo MW, Ecker DJ, Rather PN, Bonomo RA. 2012. Longitudinal analysis of the temporal evolution of *Acinetobacter baumannii* strains in Ohio, USA, by using rapid automated typing methods. *PLoS One* 7:e33443. <http://dx.doi.org/10.1371/journal.pone.0033443>.
34. Diancourt L, Passet V, Nemeš A, Dijkshoorn L, Brisse S. 2010. The population structure of *Acinetobacter baumannii*: expanding multi-resistant clones from an ancestral susceptible genetic pool. *PLoS One* 5:e10034. <http://dx.doi.org/10.1371/journal.pone.0010034>.
35. Hamouda A, Evans BA, Towner KJ, Amyes SG. 2010. Characterization of epidemiologically unrelated *Acinetobacter baumannii* isolates from four continents by use of multilocus sequence typing, pulsed-field gel electrophoresis, and sequence-based typing of bla(OXA-51-like) genes. *J. Clin. Microbiol.* 48:2476–2483. <http://dx.doi.org/10.1128/JCM.02431-09>.
36. Hujer KM, Hujer AM, Hulten EA, Bajaksouzian S, Adams JM, Donosky CJ, Ecker DJ, Massire C, Eshoo MW, Sampath R, Thomson JM, Rather PN, Craft DW, Fishbain JT, Ewell AJ, Jacobs MR, Paterson DL, Bonomo RA. 2006. Analysis of antibiotic resistance genes in multidrug-resistant *Acinetobacter* sp. isolates from military and civilian patients treated at the Walter Reed Army Medical Center. *Antimicrob. Agents Chemother.* 50:4114–4123. <http://dx.doi.org/10.1128/AAC.00778-06>.
37. Perez F, Endimiani A, Ray AJ, Decker BK, Wallace CJ, Hujer KM, Ecker DJ, Adams MD, Toltzis P, Dul MJ, Windau A, Bajaksouzian S, Jacobs MR, Salata RA, Bonomo RA. 2010. Carbapenem-resistant *Acinetobacter baumannii* and *Klebsiella pneumoniae* across a hospital system: impact of post-acute care facilities on dissemination. *J. Antimicrob. Chemother.* 65:1807–1818. <http://dx.doi.org/10.1093/jac/dkq191>.
38. Wisplinghoff H, Hippler C, Bartual SG, Haefs C, Stefanik D, Higgins PG, Seifert H. 2008. Molecular epidemiology of clinical *Acinetobacter baumannii* and *Acinetobacter genomic species 13TU* isolates using a multilocus sequencing typing scheme. *Clin. Microbiol. Infect.* 14:708–715. <http://dx.doi.org/10.1111/j.1469-0691.2008.02010.x>.
39. Marchaim D, Navon-Venezia S, Leavitt A, Chmelnitsky I, Schwaber MJ, Carmeli Y. 2007. Molecular and epidemiologic study of polyclonal outbreaks of multidrug-resistant *Acinetobacter baumannii* infection in an Israeli hospital. *Infect. Control Hosp. Epidemiol.* 28:945–950. <http://dx.doi.org/10.1086/518970>.
40. Valenzuela JK, Thomas L, Partridge SR, van der Reijden T, Dijkshoorn L, Iredell J. 2007. Horizontal gene transfer in a polyclonal outbreak of carbapenem-resistant *Acinetobacter baumannii*. *J. Clin. Microbiol.* 45: 453–460. <http://dx.doi.org/10.1128/JCM.01971-06>.
41. Liu CC, Tang CY, Kuo HY, Lu CW, Chang KC, Liou ML. 2013. The origin of *Acinetobacter baumannii* TYTH-1: a comparative genomics study. *Int. J. Antimicrob. Agents* 41:318–324. <http://dx.doi.org/10.1016/j.ijantimicag.2012.12.010>.
42. Seputiene V, Povilonis J, Suziedeliene E. 2012. Novel variants of AbaR resistance islands with a common backbone in *Acinetobacter baumannii* isolates of European clone II. *Antimicrob. Agents Chemother.* 56: 1969–1973. <http://dx.doi.org/10.1128/AAC.05678-11>.
43. Galimand M, Sabtcheva S, Courvalin P, Lambert T. 2005. Worldwide disseminated *armA* aminoglycoside resistance methylase gene is borne by composite transposon Tn1548. *Antimicrob. Agents Chemother.* 49: 2949–2953. <http://dx.doi.org/10.1128/AAC.49.7.2949-2953.2005>.
44. Zhou H, Zhang T, Yu D, Pi B, Yang Q, Zhou J, Hu S, Yu Y. 2011. Genomic analysis of the multidrug-resistant *Acinetobacter baumannii* strain MDR-ZJ06 widely spread in China. *Antimicrob. Agents Chemother.* 55:4506–4512. <http://dx.doi.org/10.1128/AAC.01134-10>.
45. Bertini A, Poirel L, Mugnier PD, Villa L, Nordmann P, Carattoli A. 2010. Characterization and PCR-based replicon typing of resistance plasmids in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54: 4168–4177. <http://dx.doi.org/10.1128/AAC.00542-10>.
46. Kenyon JJ, Hall RM. 2013. Variation in the complex carbohydrate biosynthesis loci of *Acinetobacter baumannii* genomes. *PLoS One* 8:e62160. <http://dx.doi.org/10.1371/journal.pone.0062160>.
47. Rodríguez-Martínez JM, Poirel L, Nordmann P. 2010. Genetic and functional variability of AmpC-type  $\beta$ -lactamases from *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54:4930–4933. <http://dx.doi.org/10.1128/AAC.00427-10>.
48. Hamidian M, Hall RM. 2013. ISAbal targets a specific position upstream of the intrinsic ampC gene of *Acinetobacter baumannii* leading to cephalosporin resistance. *J. Antimicrob. Chemother.* 68:2682–2683.
49. Galimand M, Courvalin P, Lambert T. 2003. Plasmid-mediated high-level resistance to aminoglycosides in *Enterobacteriaceae* due to 16S rRNA methylation. *Antimicrob. Agents Chemother.* 47:2565–2571. <http://dx.doi.org/10.1128/AAC.47.8.2565-2571.2003>.
50. Antunes LC, Imperi F, Towner KJ, Visca P. 2011. Genome-assisted identification of putative iron-utilization genes in *Acinetobacter baumannii* and their distribution among a genotypically diverse collection of clinical isolates. *Res. Microbiol.* 162:279–284. <http://dx.doi.org/10.1016/j.resmic.2010.10.010>.
51. Vallenet D, Nordmann P, Barbe V, Poirel L, Mangenot S, Bataille E, Dossat C, Gas S, Kreimeyer A, Lenoble P, Oztaş S, Poulain J, Segurens B, Robert C, Abergel C, Claverie JM, Raoult D, Médigue C, Weissenbach J, Cruveiller S. 2008. Comparative analysis of acinetobacters: three genomes for three lifestyles. *PLoS One* 3:e1805. <http://dx.doi.org/10.1371/journal.pone.0001805>.
52. Schulz BL, Jen FE, Power PM, Jones CE, Fox KL, Ku SC, Blanchfield JT, Jennings MP. 2013. Identification of bacterial protein O-oligosaccharyltransferases and their glycoprotein substrates. *PLoS One* 8:e62768. <http://dx.doi.org/10.1371/journal.pone.0062768>.
53. Lees-Miller RG, Iwashkiw JA, Scott NE, Seper A, Vinogradov E, Schild S, Feldman MF. 2013. A common pathway for O-linked protein-

- glycosylation and synthesis of capsule in *Acinetobacter baumannii*. *Mol. Microbiol.* 89:816–830. <http://dx.doi.org/10.1111/mmi.12300>.
54. Mugnier PD, Poirel L, Nordmann P. 2009. Functional analysis of insertion sequence IS*Aba1*, responsible for genomic plasticity of *Acinetobacter baumannii*. *J. Bacteriol.* 191:2414–2418. <http://dx.doi.org/10.1128/JB.01258-08>.
  55. Carruthers MD, Nicholson PA, Tracy EN, Munson RS, Jr. 2013. *Acinetobacter baumannii* utilizes a type VI secretion system for bacterial competition. *PLoS One* 8:e59388. <http://dx.doi.org/10.1371/journal.pone.0059388>.
  56. Weber BS, Miyata ST, Iwashkiw JA, Mortensen BL, Skaar EP, Pukatzki S, Feldman MF. 2013. Genomic and functional analysis of the type VI secretion system in *Acinetobacter*. *PLoS One* 8:e55142. <http://dx.doi.org/10.1371/journal.pone.0055142>.
  57. Tomaras AP, Dorsey CW, Edelmann RE, Actis LA. 2003. Attachment to and biofilm formation on abiotic surfaces by *Acinetobacter baumannii*: involvement of a novel chaperone-usher pili assembly system. *Microbiology* 149:3473–3484. <http://dx.doi.org/10.1099/mic.0.26541-0>.
  58. Peleg AY, de Breij A, Adams MD, Cerqueira GM, Mocali S, Galardini M, Nibbering PH, Earl AM, Ward DV, Paterson DL, Seifert H, Dijkshoorn L. 2012. The success of *Acinetobacter* species; genetic, metabolic and virulence attributes. *PLoS One* 7:e46984. <http://dx.doi.org/10.1371/journal.pone.0046984>.
  59. Saule M, Samuelsen Ø, Dumpis U, Sundsfjord A, Karlson A, Balode A, Miklasevics E, Karah N. 2013. Dissemination of a carbapenem-resistant *Acinetobacter baumannii* strain belonging to international clone II/sequence type 2 and harboring a novel *AbaR4*-like resistance island in Latvia. *Antimicrob. Agents Chemother.* 57:1069–1072. <http://dx.doi.org/10.1128/AAC.01783-12>.
  60. Snitkin ES, Zelazny AM, Gupta J, NISC Comparative Sequencing Program, Palmore TN, Murray PR, Segre JA. 2013. Genomic insights into the fate of colistin resistance and *Acinetobacter baumannii* during patient treatment. *Genome Res.* 23:1155–1162. <http://dx.doi.org/10.1101/gr.154328.112>.
  61. Munoz-Price LS. 2009. Long-term acute care hospitals. *Clin. Infect. Dis.* 49:438–443. <http://dx.doi.org/10.1086/600391>.
  62. Sengstock DM, Thyagarajan R, Apalara J, Mira A, Chopra T, Kaye KS. 2010. Multidrug-resistant *Acinetobacter baumannii*: an emerging pathogen among older adults in community hospitals and nursing homes. *Clin. Infect. Dis.* 50:1611–1616. <http://dx.doi.org/10.1086/652759>.
  63. Zerbino DR. 2010. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinformatics* Chapter 11:Unit 11.5.
  64. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
  65. Davidsen T, Beck E, Ganapathy A, Montgomery R, Zafar N, Yang Q, Madupu R, Goetz P, Galinsky K, White O, Sutton G. 2010. The comprehensive microbial resource. *Nucleic Acids Res.* 38:D340–D345. <http://dx.doi.org/10.1093/nar/gkp912>.
  66. Goll J, Montgomery R, Brinkac LM, Schobel S, Harkins DM, Sebastian Y, Shrivastava S, Durkin S, Sutton G. 2010. The Protein Naming Utility: a rules database for protein nomenclature. *Nucleic Acids Res.* 38:D336–D339. <http://dx.doi.org/10.1093/nar/gkp958>.
  67. Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White O. 2007. TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res.* 35:D260–D264. <http://dx.doi.org/10.1093/nar/gkl1043>.
  68. Sonnhammer EL, Eddy SR, Durbin R. 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28:405–420. [http://dx.doi.org/10.1002/\(SICI\)1097-0134\(199707\)28:3](http://dx.doi.org/10.1002/(SICI)1097-0134(199707)28:3).
  69. Madupu R, Richter A, Dodson RJ, Brinkac L, Harkins D, Durkin S, Shrivastava S, Sutton G, Haft D. 2012. CharProtDB: a database of experimentally characterized protein annotations. *Nucleic Acids Res.* 40:D237–D241. <http://dx.doi.org/10.1093/nar/gkr1133>.
  70. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403. <http://dx.doi.org/10.1101/gr.2289704>.
  71. Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463. <http://dx.doi.org/10.1093/bioinformatics/bti191>.
  72. Fouts DE, Brinkac L, Beck E, Inman J, Sutton G. 2012. Panoct: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Res.* 40:e172. <http://dx.doi.org/10.1093/nar/gks757>.
  73. Desper R, Gascuel O. 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J. Comput. Biol.* 9:687–705. <http://dx.doi.org/10.1089/106652702761034136>.