

Metagenome-assembled genomes reveal novel diversity and atypical sources of a superbug

Rafael López-Sánchez,¹ Alejandro Aguilar-Vera,¹ Santiago Castillo-Ramírez¹

AUTHOR AFFILIATION See affiliation list on p. 4.

ABSTRACT The genomic epidemiology of *Acinetobacter baumannii* has been based on culture-dependent whole-genome sequencing. However, this approach neglects sources that are not amenable to microbial culture. Here, we show that metagenome-assembled genomes (MAGs) are useful in appraising the neglected diversity of *A. baumannii* from unconventional sources. We analyzed 22 MAGs in the context of genomes from the main human international clones and some animal and plant isolates. Given the known ecology of the species, some of these MAGs came from unexpected sources, such as the New York City subway or a kitchen counter. Our global phylogeny, with 240 genomes, demonstrated that although some MAGs were associated with some human international clones, most of them were not and were distantly related to the human clones. Furthermore, we found that these MAGs have antibiotic-resistance genes. Our results show that many MAGs represent novel lineages with some important resistance genes coming from unexpected sources. Our work illustrates a powerful approach to studying the neglected diversity of this superbug from uncommon sources. We anticipate that in the coming years, MAGs will be employed to analyze unconventional sources for other important superbugs.

IMPORTANCE The genomic epidemiology of *Acinetobacter baumannii* has been based on cultivated bacterial isolates. This disregards *A. baumannii* diversity from sources not amenable to microbial culture. In this study, we demonstrate that metagenome-assembled genomes (MAGs) are very helpful in assessing the understudied diversity of *A. baumannii* from atypical sources. Our results show that many MAGs represent novel lineages with important resistance genes coming from unexpected sources. We anticipate that in the coming years, approaches similar to ours will be employed to delve into different sources for other important superbugs.

KEYWORDS superbug, genomic epidemiology, metagenomics, *Acinetobacter baumannii*, metagenome-assembled genomes

Antimicrobial drug resistance is a major public health crisis. In this respect, the ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) bacteria are a major cause of human deaths due to antibiotic-resistant infections. Whole-genome sequencing (WGS) has been essential in characterizing the genetic structure and transmission patterns of these bacteria. However, given the culture dependence of standard WGS, many potential sources of these superbugs have not been explored.

In this respect, metagenome-assembled genomes (MAGs) inferred from metagenomic culture-independent data have been shown to be useful in understanding bacterial diversity from different environments that are not very amenable to culture-dependent approaches (1). As a proof of concept, we gathered high-quality MAGs of *A. baumannii* and demonstrated that they can be a helpful tool for studying

Editor Paolo Visca, Università Roma Tre, Roma, Italy

Address correspondence to Santiago Castillo-Ramírez, iago@ccg.unam.mx.

The authors declare no conflict of interest.

Received 10 January 2025

Accepted 14 February 2025

Published 18 March 2025

Copyright © 2025 López-Sánchez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

atypical sources of superbugs and uncovering untapped diversity within these bacteria. According to some suggestions from the Genomic Standard Consortium (2), we included only MAGs with completeness >95% and contamination <5%, as well as those with defined geographic location. We chose *A. baumannii* as it is one of the most relevant bacteria in terms of human nosocomial infections (3), and it is even considered a One Health problem for which we need to increase our research efforts on its non-human populations (4, 5). Our data show that many MAGs represent novel lineages coming from unexpected sources.

To make reliable inferences, we selected only MAGs that were already available on the National Center for Biotechnology Information (NCBI) database and had sound completeness (>95%) and hardly any contamination (<5%) (see Table S1). We found 22 MAGs coming from four different countries, namely, Australia, China, India, and the USA. These MAGs were recovered from 10 distinct sources (see Table S1; Fig. 1): human feces, oral plaque biofilm, a kitchen counter, metal, wood, dental calculus, the New York City (NYC) subway, sputum, and activated sludge. Most of the MAGs were inferred from the samples isolated between 2008 and 2022, but one was from 1950, and 3 MAGs did not have information. Importantly, many of these MAGs were recovered from very atypical sources, that is, environments not previously described or hardly ever described as sources for this species. For instance, they were found in activated sludge, a kitchen counter, or the NYC subway. To put these MAGs in the context of the global *A. baumannii* population and better understand their diversity, we constructed a phylogeny following previous studies (6, 7), including genomes from the main human international clones (ICs) and some isolates recently described for animals (pigs and cattle) and grass (8, 9). Table S1 lists all the animal, ICs, and grass isolates selected (228 genomes), which were downloaded from the NCBI. A global phylogeny consisting of 250 genomes was constructed (see Methods). The phylogeny (Fig. 1) shows that most MAGs (13/22, 59%) form a diverse clade that is not closely related to any of the major human ICs or animal

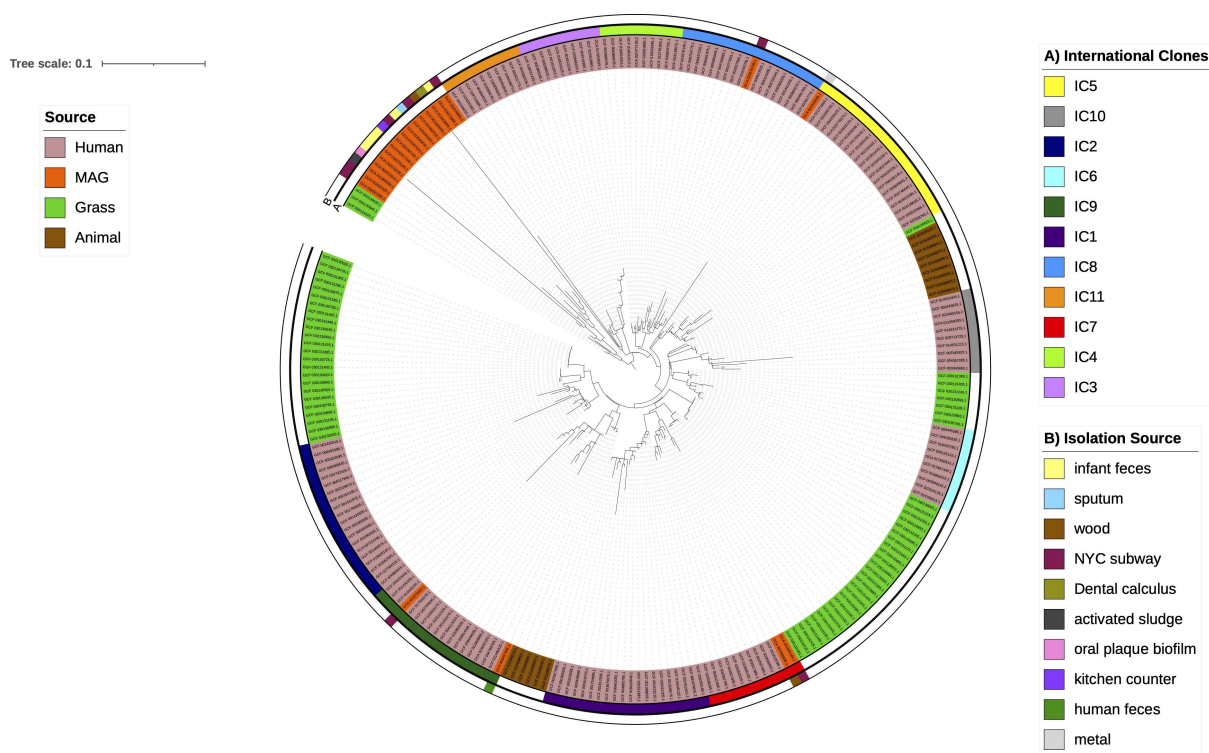


FIG 1 Phylogeny showing the MAGs in the context of the known diversity of *A. baumannii*. Maximum likelihood phylogeny based on the single gene families without recombination. Isolates are colored-coded, showing human, animal, and grass lineages and MAGs. The outer ring (B) provides the source of the MAGs, whereas the inner ring (A) highlights the human international clones. The scale shows the phylogenetic distance, measured in substitutions per site.

or plant isolates, implying that this clade has some overlooked diversity of this species. However, we also observed that few MAGs clustered within or next to major human ICs. For instance, we noted two MAGs were associated with IC7, two with IC9, one with IC8, and another with IC5 (see Fig. 1). Thus, while a few MAGs are closely related to some important human ICs, many MAGs are very distantly related to the human ICs, and we consider they represent novel diversity within species. We also determined the resistome for the MAGs conducting an *in silico* prediction by means of CARD (10) (see Methods). Figure 2 shows a heat map visualizing the resistome of the MAGs. The MAGs have important intrinsic resistance genes, such as efflux pumps and intrinsic oxacillinases (OXA-51-like family). We noted two types of efflux pumps and several allelic variants of the OXA-51-like family. These MAGs also have several ADC β -lactamase variants. Of note, compared to *A. baumannii* isolates of human clinical origin, which have a vast amount of acquired resistance genes (11), the MAGs analyzed here had hardly any acquired resistance genes. Nonetheless, these MAGs have a few acquired resistance genes of clinical relevance, such as *sul2* or *mphE* (see Fig. 2). We acknowledge that the paucity of acquired antibiotic resistance genes might also be due to the very nature of the MAGs, as in some instances, not all the actual genes for an isolate might be recovered in a MAG. Collectively, these data show that although MAGs do not have as many antibiotic resistance genes as human clinical isolates, they do have important intrinsic antibiotic resistance genes and few acquired antibiotic resistance genes.

In conclusion, these results show that many MAGs represent novel lineages with relevant resistance genes coming from unexpected sources, given the current known ecology of the pathogen. If we want to understand the global, multi-host epidemiology of *A. baumannii* (12), we should go beyond cultivated isolate genomes and also consider MAGs. However, we should ensure compliance with some minimum standards of quality

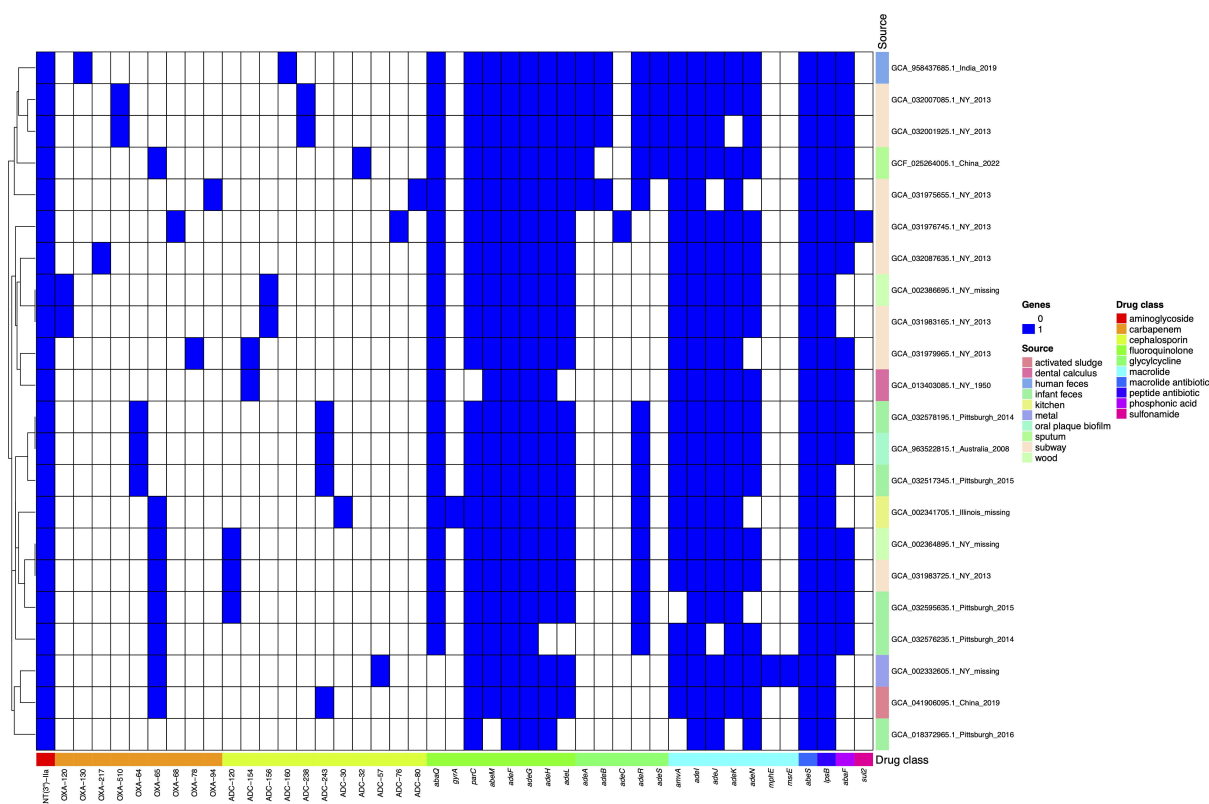


FIG 2 Prediction of the antibiotic resistance genes. The different antibiotic resistance genes are shown in the heat map. The *in silico* prediction was conducted using CARD. In the case of *gyrA*, blue means that the isolate has the SNP associated with the resistant phenotype. MAGs are to the right of the heat map, and the antibiotic resistance genes are at the bottom. The vertical vector to the right shows the source of the MAGs, whereas the bottom horizontal vector denotes the drug class.

and metadata (2). Of note, MAGs can be useful beyond this species and help uncover genomic diversity for other species within the genus. For instance, in the recently conducted largest genomic epidemiology study of *Acinetobacter junii* (13), we noted that several genomes included were reconstructed from metagenomes (i.e., MAGs; see Table S1 in Aguilar-Vera et al.[13]). In a broader context, our study illustrates that MAGs can be a powerful approach to studying the neglected diversity of superbugs from uncommon sources. We anticipate that in the coming years, MAGs will be employed to analyze unconventional sources for other ESKAPE pathogens.

All the genomes were downloaded from the NCBI database and are listed in Table S1. We used CheckM v.1.1.3 (14) to evaluate the completeness and contamination of all the genomes analyzed. Values of completeness and contamination, as well as metadata for all the genomes, are reported in Table S1. To ensure high-quality MAGs, only those showing values of completeness >95%, contamination <5%, and data on geographic location were selected. For consistency, all the genomes were re-annotated using PROKKA v1.12 (15). Single gene families were identified by running a pan-genome analysis via Roary (16). Individual gene alignments were constructed using FSA (17) and concatenated with an in-house PERL script. The global phylogeny was constructed on the concatenated alignment of all the single gene families without recombination, and the maximum likelihood was run on the concatenated alignment employing RAXML v8 (18). We used the GTR + I model to run the phylogeny. We also determined the resistome for the MAGs using CARD (10), considering a coverage of $\geq 70\%$ and an identity percentage of $\geq 95\%$.

ACKNOWLEDGMENTS

We warmly thank Alfredo J. Hernández-Alvarez for installing some of the libraries and programs used in this study.

AUTHOR AFFILIATION

¹Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

AUTHOR ORCIDs

Rafael López-Sánchez  <http://orcid.org/0000-0003-3115-1250>

Alejandro Aguilar-Vera  <http://orcid.org/0000-0003-0665-3355>

Santiago Castillo-Ramírez  <http://orcid.org/0000-0003-3922-7088>

AUTHOR CONTRIBUTIONS

Rafael López-Sánchez, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – review and editing | Alejandro Aguilar-Vera, Data curation, Resources, Software, Validation, Visualization | Santiago Castillo-Ramírez, Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All the genomes used in this study are listed in Table S1. This table provides the GenBank or RefSeq assembly accession numbers.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Table S1 (Spectrum00106-25-s0001.xls). Genomes and MAGs used in this study.

REFERENCES

1. Setubal JC. 2021. Metagenome-assembled genomes: concepts, analogies, and challenges. *Biophys Rev* 13:905–909. <https://doi.org/10.1007/s12551-021-00865-y>
2. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloie-Fadrosch EA, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>
3. Organization WH. 2024. WHO bacterial priority pathogens list, 2024. World Health Organization.
4. Castillo-Ramírez S. 2023. The importance of *Acinetobacter baumannii* from non-human sources. *Lancet Microbe* 4:e761–e762. [https://doi.org/10.1016/S2666-5247\(23\)00246-X](https://doi.org/10.1016/S2666-5247(23)00246-X)
5. Hernández-González IL, Castillo-Ramírez S. 2020. Antibiotic-resistant *Acinetobacter baumannii* is a One Health problem. *Lancet Microbe* 1:e279. [https://doi.org/10.1016/S2666-5247\(20\)30167-1](https://doi.org/10.1016/S2666-5247(20)30167-1)
6. Graña-Miraglia L, Evans BA, López-Jácome LE, Hernández-Durán M, Colín-Castro CA, Volkow-Fernández P, Cevallos MA, Franco-Cendejas R, Castillo-Ramírez S. 2020. Origin of OXA-23 variant OXA-239 from a recently emerged lineage of *Acinetobacter baumannii* international clone V. *mSphere* 5:e00801-19. <https://doi.org/10.1128/mSphere.00801-19>
7. Mateo-Estrada V, Fernández-Vázquez JL, Moreno-Manjón J, Hernández-González IL, Rodríguez-Noriega E, Morfin-Otero R, Alcántar-Curiel MD, Castillo-Ramírez S. 2021. Accessory genomic epidemiology of cocirculating *Acinetobacter baumannii* clones. *mSystems* 6:e0062621. <https://doi.org/10.1128/mSystems.00626-21>
8. Mateo-Estrada V, Tyrrell C, Evans BA, Aguilar-Vera A, Drissner D, Castillo-Ramírez S, Walsh F. 2023. *Acinetobacter baumannii* from grass: novel but non-resistant clones. *Microb Genom* 9:001054. <https://doi.org/10.1099/mgen.0.001054>
9. Mateo-Estrada V, Vali L, Hamouda A, Evans BA, Castillo-Ramírez S. 2022. *Acinetobacter baumannii* from grass: novel but non-resistant clones. *Microbiol Spectr* 10:e0128922. <https://doi.org/10.1128/spectrum.01289-22>
10. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, et al. 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>
11. Hernández-González IL, Mateo-Estrada V, Castillo-Ramírez S. 2022. The promiscuous and highly mobile resistome of *Acinetobacter baumannii*. *Microb Genom* 8:000762. <https://doi.org/10.1099/mgen.0.000762>
12. Castillo-Ramírez S. 2023. Genomic epidemiology of *Acinetobacter baumannii* goes global. *MBio* 14:e0252023. <https://doi.org/10.1128/mbio.02520-23>
13. Aguilar-Vera A, Bello-López E, Pantoja-Núñez GI, Rodríguez-López GM, Morales-Erasto V, Castillo-Ramírez S. 2024. *Acinetobacter junii*: an emerging One Health pathogen. *mSphere* 9:e0016224. <https://doi.org/10.1128/msphere.00162-24>
14. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
15. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
16. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>
17. Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, Holmes I, Pachter L. 2009. Fast statistical alignment. *PLoS Comput Biol* 5:e1000392. <https://doi.org/10.1371/journal.pcbi.1000392>
18. Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>