



A Curated, Comprehensive Database of Plasmid Sequences

Lauren Brooks,^b Mo Kaze,^a Mark Sistrom^a

^aUniversity of California Merced, Merced, California, USA

^bUtah Valley University, Orem, Utah, USA

ABSTRACT Plasmid sequences are central to a myriad of microbial functions and processes. Here, we have compiled a database of complete plasmid sequences and associated metadata curated from both NCBI's recent genome database update, which includes plasmids as organisms, and all available annotated bacterial genomes. The resultant database contains 10,892 complete plasmid sequences and associated metadata.

Plasmids are one of the key vectors of horizontal gene transfer in bacteria and archaea (1). Plasmids play a major role in bacterial genetic diversity (2), evolution (3), and adaptation (4). Conjugative exchange (i.e., the transfer of plasmids from one bacterium to another) can lead to the spread of a variety of functions, including degradation of heavy metals and anthropogenic toxic waste (5), bacteriocin and toxin production to ward off predators (6), and, alarmingly, antibiotic resistance and virulence plasmids that inhibit antibiotics and lead to novel and untreatable diseases (7). Plasmids are also extensively used as tools in genetic engineering (8).

To generate a comprehensive plasmid database, we started with the recent NCBI genome database update, which has a separate collection of plasmids as organisms. FASTA format files containing plasmid "genome" sequences were downloaded on 5 March 2018 from <ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plasmid/>, resulting in 11,677 plasmid sequences. Using the R package Rentrez (<https://cran.r-project.org/web/packages/rentrez/index.html>), we downloaded the metadata available from the nucleotide database for each entry based on the locus number contained in the header file for each plasmid. Metadata from the BioProject, BioSample, and Assembly databases were also pulled for each plasmid sequence when present. An initial review of the metadata demonstrated that not all sequences contained in the downloaded files were complete plasmid sequences. After downloading all sequences labeled as plasmids ($n = 11,677$), we filtered the database using the nucleotide metadata to remove partial plasmid sequences from the databases ($n = 9,763$) and again using the assembly metadata to remove incomplete assemblies ($n = 7,434$). Additionally, 8 sequences labeled as phages were found and removed from the database. This resulted in 7,426 complete and assembled plasmid sequences following this initial screening.

In addition to curating the predefined NCBI plasmid database, we extracted plasmid sequences from bacterial genomes with complete assemblies in NCBI's prokaryotic genome database (<https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/>). Genomic assemblies labeled as partially complete or in contigs were not included to ensure that only complete plasmid sequences were included in our final database. Sequences that were already included as part of the original plasmid downloads, as identified by their accession or locus numbers, were removed as duplicates. This allowed us to include an additional 3,466 complete, annotated plasmid sequences, resulting in our database of 10,892 complete and annotated plasmid sequences for subsequent analyses.

The two data sets described above were combined to result in a comprehensive,

Citation Brooks L, Kaze M, Sistrom M. 2019. A curated, comprehensive database of plasmid sequences. *Microbiol Resour Announc* 8:e01325-18. <https://doi.org/10.1128/MRA.01325-18>.

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine

Copyright © 2019 Brooks 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/).

Address correspondence to Mark Sistrom, msistrom@ucmerced.edu.

Received 2 October 2018

Accepted 16 November 2018

Published 3 January 2019

complete, and annotated plasmid database. Metadata for this final list were compiled using the accession version number provided in the header for each plasmid sequence as described above.

Data availability. The plasmid database is available in fasta format and associated metadata are available in csv format at <https://doi.org/10.15146/R33X2J>.

ACKNOWLEDGMENTS

This work received no specific grant from any funding agency. Lauren Brooks was responsible for the conceptualization, methodology, formal analysis, data curation, writing (original draft preparation), and editing. Mo Kaze was responsible for both writing (original draft preparation) and editing. Mark Sistrom provided resources, editing, and supervision.

We declare no conflicts of interest.

REFERENCES

1. Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 3:711–721. <https://doi.org/10.1038/nrmicro1234>.
2. Halary S, Leigh JW, Cheaib B, Lopez P, Baptiste E. 2010. Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci U S A* 107:127–132. <https://doi.org/10.1073/pnas.0908978107>.
3. Eberhard WG. 1990. Evolution in bacterial plasmids and levels of selection. *Q Rev Biol* 65:3–22. <https://doi.org/10.1086/416582>.
4. Heuer H, Smalla K. 2012. Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol Rev* 36:1083–1104. <https://doi.org/10.1111/j.1574-6976.2012.00337.x>.
5. Shahi A, Ince B, Aydin S, Ince O. 2017. Assessment of the horizontal transfer of functional genes as a suitable approach for evaluation of the bioremediation potential of petroleum-contaminated sites: a mini-review. *Appl Microbiol Biotechnol* 101:4341–4348. <https://doi.org/10.1007/s00253-017-8306-5>.
6. Riley MA, Wertz JE. 2002. Bacteriocins: evolution, ecology, and application. *Annu Rev Microbiol* 56:117–137. <https://doi.org/10.1146/annurev.micro.56.012302.161024>.
7. van Wintersdorff HCJ, Penders J, van Niekerk MJ, Mills ND, Majumder S, van Alphen BL, Savelkoul PHM, Wolffs PFG. 2016. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front Microbiol* 7:173. <https://doi.org/10.3389/fmicb.2016.00173>.
8. Simon R, Priefer U, Pühler A. 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram negative bacteria. *Nat Biotechnol* 1:784. <https://doi.org/10.1038/nbt1183-784>.