Lawrence Berkeley National Laboratory
Recent Work

Title
The evolution of microbial species - a view through the genomic lens

Permalink
https://escholarship.org/uc/item/9567h4q7

Authors
Varghese, Neha
Mukherjee, Supratim
ivanova, Natalia
et al.

Publication Date
2014-03-19
The evolution of microbial species – a view through the genomic lens

Neha Varghese¹*, Supratim Mukherjee¹, Natalia Ivanova¹, Konstantinos Mavromatis³, Konstantinos Konstantinidis², Nikos Kyrpides¹ and Amrita Pati¹

¹ LBNL - Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, USA
² Georgia Institute of Technology, Atlanta, GA USA
³Celgene Corporation, San Francisco, CA USA

*To whom correspondence should be addressed: Email: Neha Varghese: njvarghese@lbl.gov, Supratim Mukherjee: supratimmukherjee@lbl.gov, or Amrita Pati: apati@lbl.gov

March 21, 2014

ACKNOWLEDGMENTS:

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was conducted under the LDRD grant YLD012: Computational, Data Management and Analysis Methods for the Study of a Rapidly Expanding Genome and Metagenome Sequence Data Space (YLD012). Development was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 and Los Alamos National Laboratory under contract DE-AC02-06NA25396.

DISCLAIMER:

LBNL: This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service
The evolution of microbial species – a view through the genomic lens

by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.
The evolution of microbial species – a view through the genomic lens

Neha J. Varghese1, Supratim Mukherjee1, Natalia N. Ivanova1, Konstantinos Mavromatis1, Konstantinos T. Konstantinidis2, Nikos C. Kyrpides3* and Amrita Pati4*

(1) DOE Joint Genome Institute, Walnut Creek, CA (2) Georgia Institute of Technology, Atlanta, GA (3) Celgene Corp, San Francisco, CA

ABSTRACT

For a long time prokaryotic species definition has been under debate and a constant source of turmoil in microbiology. This has recently prompted the ASM to call for a scalable and reproducible technique, which uses meaningful commonalities to cluster microorganisms into groups corresponding to prokaryotic species. Whole-genome Average Nucleotide Identity (gANI) was previously suggested as a measure of genetic distance that generally agrees with prokaryotic species assignments based on the accepted best practices (DNA-DNA hybridization and 16S rDNA similarity). In this work, we prove that gANI is indeed the meaningful commonality based on which microorganisms can be grouped into the aforementioned clusters. By analyzing 1.76 million pairs of genomes we find that identification of the closest relatives of microorganisms cluster in groups with meaningful commonalities or to determine what commonalities may be best used to cluster microorganisms into meaningful groups.

METHODS

For a long time prokaryotic species definition has been under debate and a constant source of turmoil in microbiology. This has recently prompted the ASM to call for a scalable and reproducible technique, which uses meaningful commonalities to cluster microorganisms into groups corresponding to prokaryotic species. Whole-genome Average Nucleotide Identity (gANI) was previously suggested as a measure of genetic distance that generally agrees with prokaryotic species assignments based on the accepted best practices (DNA-DNA hybridization and 16S rDNA similarity). In this work, we prove that gANI is indeed the meaningful commonality based on which microorganisms can be grouped into the aforementioned clusters. By analyzing 1.76 million pairs of genomes we find that identification of the closest relatives of microorganisms cluster in groups with meaningful commonalities or to determine what commonalities may be best used to cluster microorganisms into meaningful groups.

RESULTS

We used the thresholds of gANI=96.5% and AF=0.5 to group the genomes using Maximal Clique Enumeration (MCE). For each pairwise comparison, gANI was applied across all genomes of microorganisms cluster in groups with meaningful commonalities or to determine what commonalities may be best used to cluster microorganisms into meaningful groups. For the first time, gANI was applied across all available sequenced prokaryotic genomes and its potential to cluster microorganisms into such “natural groups” was explored. We demonstrate that gANI, which maximally utilizes the commonalities between microbial genomes, is a robust measure of genetic relatedness for establishing accurate evolutionary relationships. The gANI-based cliques were validated by comparisons with “named” species, similarity of genomes to named species, and its potential for moving forward with taxonomy. Further, the groupings identified are presented in detail at http://ani.jgi-psf.org to facilitate comprehensive downstream analysis for researchers across different disciplines.

CONCLUSIONS

Several reports have already illustrated that microbial taxonomic assignments are inconsistent with emerging genetic, systematic, and phenotypic information for a large number of species. According to a recent ASM report, “in moving forward with microbial taxonomy, it is critical to determine whether microorganisms cluster in groups with meaningful commonalities or to determine what commonalities may be best used to cluster microorganisms into meaningful groups.” For the first time, gANI was applied across all available sequenced prokaryotic genomes and its potential to cluster microorganisms into such “natural groups” was explored. We demonstrate that gANI, which maximally utilizes the commonalities between microbial genomes, is a robust measure of genetic relatedness for establishing accurate evolutionary relationships. The gANI-based cliques were validated by comparisons with “named” species, similarity of genomes to named species, and its potential for moving forward with taxonomy. Further, the groupings identified are presented in detail at http://ani.jgi-psf.org to facilitate comprehensive downstream analysis for researchers across different disciplines. The integrated microbial genomes (IMG) system, Nucleic Acids Res. 2006;34:D344-D348.

REFERENCES


ACKNOWLEDGMENT

This work was conducted under the LR08 grant YLD12-1. Computational, Data Management and Analysis Methods for the Study of a Rapidly Expanding Genome and Metagenome Sequence Data Space (YLD12012), Development was performed under the auspices of the US Department of Energy’s Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract DE-AC52- 07NA27344 and Los Alamos National Laboratory under contract DE-AC02-08NA25396. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department under contract DE-AC02-05CH11231.

CONTACT

For questions or comments or more detailed information, please contact:
Neha Varghese <njvarghese@lbl.gov>
Supratim Mukherjee <supratim.mukherjee@lbl.gov>
Amrita Pati <apati@lbl.gov>