

Phylogeny of different *Helicobacter pylori* strains could not be explained by 16S rRNA gene due to high similarity in gene sequence across strains

Wenfa Ng (✉ ngwenfa771@hotmail.com)

National University of Singapore

Short Report

Keywords: *Helicobacter pylori*, 16S rRNA gene, phylogenetic tree, multiple sequence alignment, evolutionary pressure

Posted Date: January 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1223822/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Understanding the evolutionary relatedness of different strains of a species helped identify strain-specific differences that may be useful for disease diagnosis and treatment. Typically, such strain level typing would be augmented by molecular assays such as DNA sequencing, and phylogenetic tree analysis. This work utilizes public data on the 16S rRNA gene sequence of different strains of *Helicobacter pylori* to help plot the phylogenetic tree that describes the evolutionary trajectories of the different strains. Results from multiple sequence alignment reveals high level of conservation in 16S rRNA gene sequence across strains. This then translates into a phylogenetic tree structure that suggests very close evolutionary relationships of the different strains except for one outlier strain. Even in the case of the outlier strain, its evolutionary distance from other brethren was also not large. Overall, the results obtained in this study indicates that 16S rRNA gene may not capture strain-level phylogeny between different strains of the same species, and point to efforts in elucidating this phylogenetic effect in other genes of the species. Such genes may be involved in virulence during pathogenesis in humans, and may thus be subjected to higher evolutionary pressure and natural selection.

Full Text

Evolution constantly exerts selection pressure on organisms due to physiological and cellular effects of mutations in the genome [1]. This then creates a driving force for evolution of features or characteristics in the organism along certain trajectories. These trajectories could be encoded in one or more genes. Hence, determining the gene sequence of the organism or strain of interest in relation to others could offer a glimpse into the evolutionary relationships between the organisms or strains under study.

One useful approach to determine such evolutionary relationships is through the combined use of multiple sequence alignment and phylogenetic tree analysis [2] [3] [4] [5]. Briefly, multiple sequence alignment sought to detect, at the single nucleotide level, differences and similarities in the gene sequence of the same gene under investigation from different species or strains. Upon determining this ensemble set of similarities and differences, phylogenetic tree analysis help visualize the evolutionary distance between the strains and species. Depending on the structure of the phylogenetic tree obtained, further information on the number of phylogenetic gene cluster could also be obtained.

16S rRNA gene has been well-established as a good phylogenetic marker for prokaryotes at the species level [6] [7] [8]. This meant that 16S rRNA gene encodes sufficient sequence diversity to help chronicle the individual evolutionary trajectories of different species [9] [10] [11]. However, little is known about the ability of this molecule in encoding strain level evolutionary relationships. To understand this task better, it is important to comprehend the molecular underpinnings that demarcates different strains of the same species at the genome level.

Specifically, strain level differences in the genome could arise in both phylogenetic markers such as 16S rRNA gene or other genes [12] [13] [14]. Depending on the prevailing evolutionary selection marker of the

strains, this could be metabolic genes or virulence genes, as these hold the most significance for the fitness of the strain. Take, for example, the human gastric pathogen, *Helicobacter pylori*. As a pathogen, virulence genes play an important role in determining the success of the organism in colonizing human stomach epithelial lining. Hence, evolutionary pressure would exert its effect on the virulence genes to help select for fitter strains that have higher level of success in colonizing human stomachs and causing other complications.

This work sought to examine the utility of 16S rRNA gene of *H. pylori* in chronicling the evolutionary relationships between different strains of the species through the combined approach of multiple sequence alignment and phylogenetic tree analysis. 16S rRNA gene sequence information of 7 strains of *Helicobacter pylori* was obtained from de Silva database and European Nucleotide Archive. These were subjected to multiple sequence alignment and phylogenetic tree analysis using an in-house MATLAB software using built-in MATLAB functions.

Results from multiple sequence alignment were plotted in Figure 1. The data shows high level of sequence conservation in the 16S rRNA gene of different strains of *H. pylori*, thereby, suggesting that the molecule could not encode sufficient sequence diversity to account for strain level genetic differences for the different strains investigated. Hence, 16S rRNA gene of *H. pylori* is not a good biomarker to demarcate different strains of the species.

Phylogenetic tree analysis as depicted in Figure 2 reveals high level of similarity in the 16S rRNA gene of different strains of *H. pylori*. This comes about from the close evolutionary relationships between different strains of the species except for one outlier strain. Combining the results from multiple sequence alignment and phylogenetic tree analysis reveals that 16S rRNA gene is not a good candidate for describing the phylogeny of different strains of *H. pylori*.

Overall, evolutionary relationships between strains could be encoded in 16S rRNA gene or other metabolic and virulence genes. In the case of *H. pylori*, combined multiple sequence alignment and phylogenetic tree analysis shows that 16S rRNA gene could not encode sufficient sequence diversity to chronicle the evolutionary divergence of different strains of the species. This could be due to the relatively weak evolutionary pressure on the 16S rRNA gene which is related to the selection pressure on ribosome function. If this is true, then most of the evolutionary pressure on *H. pylori* is not on metabolism, protein production or growth rate, rather the fitness of a strain is determined by the activity of virulence proteins encoded by virulence genes. Hence, if the latter hypothesis is true, strain level divergence may be chronicled by differential genetic sequences of virulence genes of the species.

Declarations

Conflicts of interest

The author declares no conflicts of interest.

Funding

No funding was used in this work.

References

1. R. Milo, J. H. Hou, M. Springer, M. P. Brenner, and M. W. Kirschner, "The relationship between evolutionary and physiological variation in hemoglobin," *Proc. Natl. Acad. Sci.*, vol. 104, no. 43, pp. 16998–17003, Oct. 2007, doi: 10.1073/pnas.0707673104.
2. B. Chowdhury and G. Garai, "A review on multiple sequence alignment from the perspective of genetic algorithm," *Genomics*, vol. 109, no. 5, pp. 419–431, Oct. 2017, doi: 10.1016/j.ygeno.2017.06.007.
3. J. Daugelaite, A. O' Driscoll, and R. D. Sleator, "An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics," *ISRN Biomath.*, vol. 2013, p. e615630, Aug. 2013, doi: 10.1155/2013/615630.
4. S. S. Roy, R. Dasgupta, and A. Bagchi, "A Review on Phylogenetic Analysis: A Journey through Modern Era," *Comput. Mol. Biosci.*, vol. 4, no. 3, Art. no. 3, Sep. 2014, doi: 10.4236/cmb.2014.43005.
5. G. Munjal, M. Hanmandlu, and S. Srivastava, "Phylogenetics Algorithms and Applications," in *Ambient Communications and Computer Systems*, Singapore, 2019, pp. 187–194. doi: 10.1007/978-981-13-5934-7_17.
6. B. Yang, Y. Wang, and P.-Y. Qian, "Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis," *BMC Bioinformatics*, vol. 17, no. 1, p. 135, Mar. 2016, doi: 10.1186/s12859-016-0992-y.
7. R. J. Case, Y. Boucher, I. Dahllöf, C. Holmström, W. F. Doolittle, and S. Kjelleberg, "Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies," *Appl. Environ. Microbiol.*, Jan. 2007, doi: 10.1128/AEM.01177-06.
8. J. S. Johnson *et al.*, "Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis," *Nat. Commun.*, vol. 10, no. 1, p. 5029, Nov. 2019, doi: 10.1038/s41467-019-13036-1.
9. J. P. Earl *et al.*, "Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes," *Microbiome*, vol. 6, no. 1, p. 190, Oct. 2018, doi: 10.1186/s40168-018-0569-2.
10. X. Gao, H. Lin, K. Revanna, and Q. Dong, "A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy," *BMC Bioinformatics*, vol. 18, no. 1, p. 247, May 2017, doi: 10.1186/s12859-017-1670-4.
11. J. Jeong *et al.*, "The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology," *Sci. Rep.*, vol. 11, no. 1, p. 1727, Jan. 2021, doi: 10.1038/s41598-020-80826-9.

12. F. Asnicar *et al.*, "Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0," *Nat. Commun.*, vol. 11, no. 1, p. 2500, May 2020, doi: 10.1038/s41467-020-16366-7.
13. D. T. Truong, A. Tett, E. Pasolli, C. Huttenhower, and N. Segata, "Microbial strain-level population structure and genetic diversity from metagenomes," *Genome Res.*, vol. 27, no. 4, pp. 626–638, Apr. 2017, doi: 10.1101/gr.216242.116.
14. F. Beghini *et al.*, "Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3," *eLife*, vol. 10, p. e65088, May 2021, doi: 10.7554/eLife.65088.

Figures

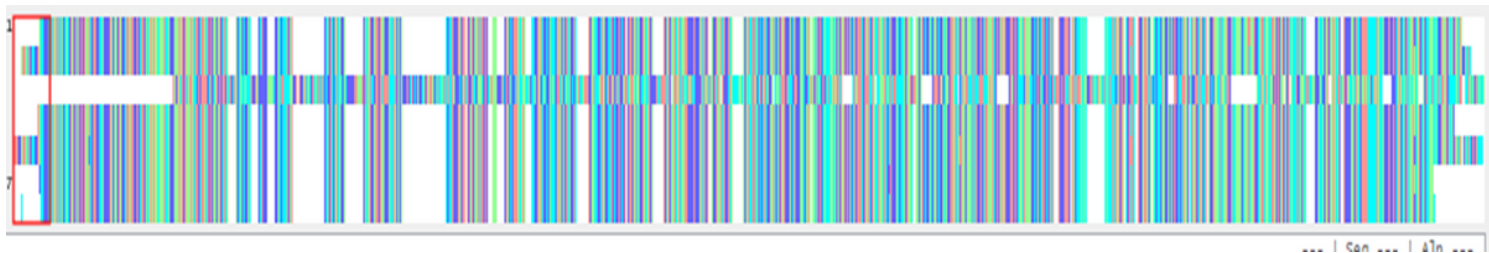


Figure 1

Multiple sequence alignment of 16S rRNA gene of different strains of *Helicobacter pylori* reveals high level of sequence conservation across the entire length of the gene except for one outlier strain.

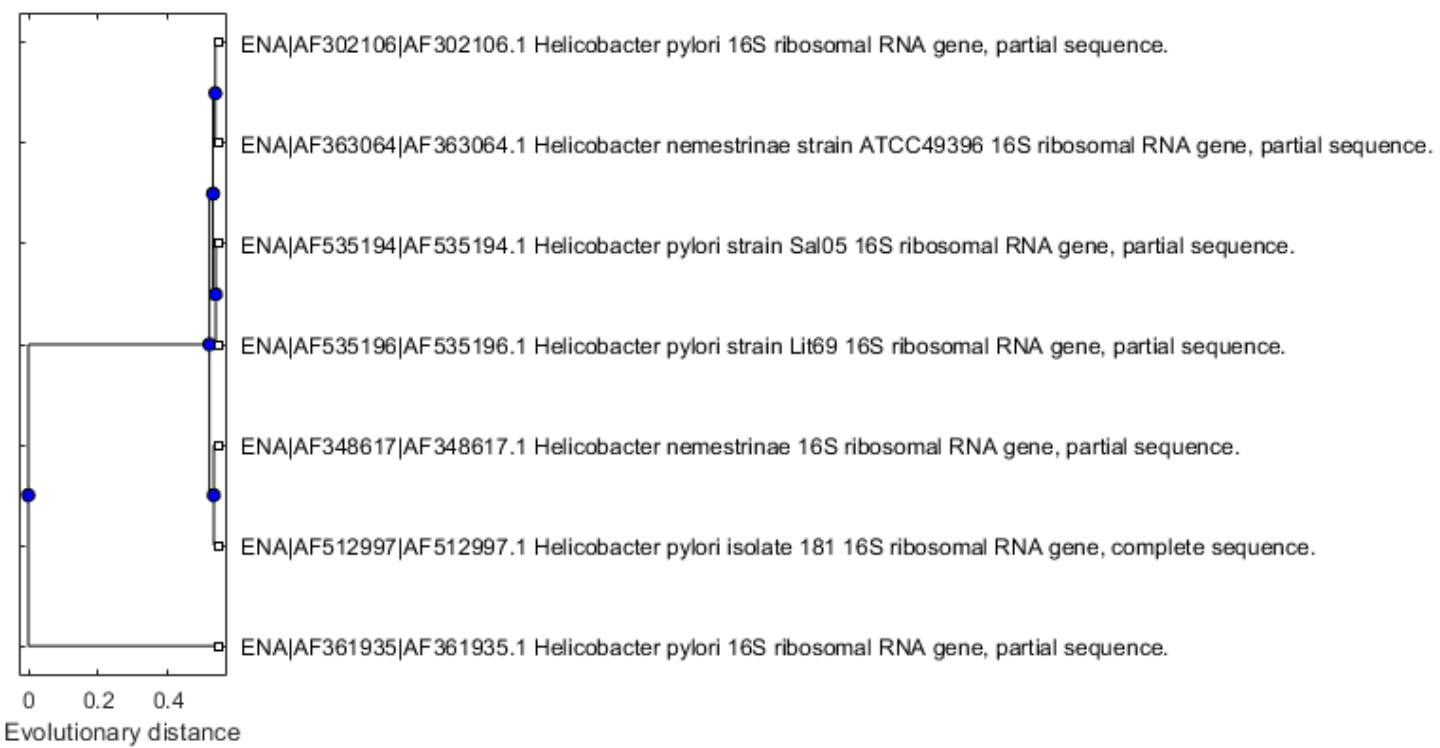


Figure 2

Phylogenetic tree of 16S rRNA gene of different strains of *H. pylori*. The data shows that there is little sequence divergence between different strains of *H. pylori* except for one outlier strain, which indicates that 16S rRNA gene could not be used to describe the phylogeny of different strains of *H. pylori*.