



# MSA( Multiple sequence alignment) and phylogenetic relationship of 16 s rRNA in bacteria causing infective endocarditis

P. Keshav Krishna\*

First year BDS student, Saveetha Dental College, Saveetha University, Chennai-600007

V.Vishnu Priya

Associate Professor, Department of Biochemistry, Saveetha Dental College, Saveetha University, Chennai-600007

## Abstract:

**Topic:** MSA and phylogenetic relationship of 16s rRNA sequence in bacteria causing infective endocarditis

**Aim:** To estimate the phylogenetic between the organisms in question.

**Objective:** To analyze the 16s rRNA nucleotide sequence and establish a relationship among these species which cause infective endocarditis

**Background:** Endocarditis is an inflammation of the inner layer of the heart, the endocardium. It usually involves the heart valves. There are several ways to classify endocarditis. The simplest classification is based on cause: either infective or non-infective, depending on whether a microorganism is the source of the inflammation or not. To name a few bacteria which cause endocarditis are *Haemophilus segnis*, *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens* and *Kingella kingae*, these organisms are together grouped as HACEK organisms which are being used as part of the study

**Reason:** Infective endocarditis is a growing risk factor and the breakthrough with a broad spectrum cure would be helpful for which a phylogenetic analysis would assist in the process

**Key words:** MSA ; Phylogenetic analysis ; Infective endocarditis ; HACEK

## INTRODUCTION:

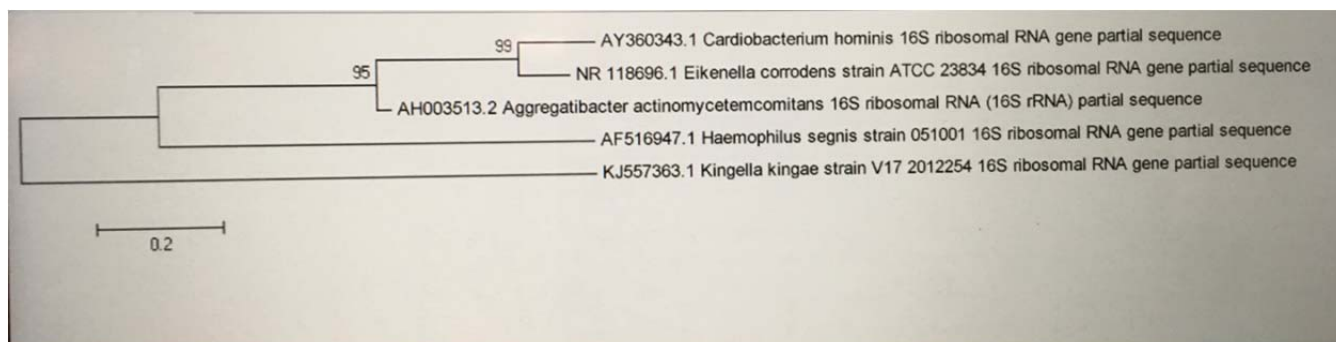
Endocarditis is an inflammation of the inner layer of the heart, the endocardium<sup>[1]</sup>. It usually involves the heart valves. There are several ways to classify endocarditis. The simplest classification is based on cause: either infective or non-infective, depending on whether a microorganism is the source of the inflammation or not. Endocarditis is a growing risk among people of greater age. Bacteria are a major cause for infective endocarditis (IE) as given by the AHA (American Heart Association)<sup>[2][3][4]</sup>. But it is to be noted that no specific bacteria has been found to cause IE. To name a few bacteria which cause endocarditis are *Haemophilus segnis*, *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens* and *Kingella kingae*. Although the exact treatment plan which works for all these is yet to be observed. Even with the latest advancements in the medical field, a few relations still remain a mystery. The relation has not yet been proved statistically but bacterial endocarditis is common with people who have had their tooth extracted between the ages of 20 and 40, As published by Brian E D Cooke in Proceedings of the royal society of medicine in 1970. Recent literature suggests that these HACEK organisms are only responsible for about 5-10% of all cases of infective endocarditis<sup>[5]</sup>. They have been a frequent cause of culture-negative endocarditis.<sup>[6]</sup> Culture negative refers to an inability to produce a colony on regular agar plates; this is because these bacteria are fastidious (requiring a specific nutrient). However the many organisms of HACEK (*Haemophilus segnis*,

*Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens* and *Kingella kingae* which live in the oral cavity) can be treated equally if a phylogenetic relationship can be established<sup>[7]</sup>. MSA (Multiple Sequence Alignment) helps in obtaining a similarity or relationship among the organisms and hence this article could provide of use while drug designing for the same.

## MATERIALS AND METHODS:

### Sequence Analysis:

The initial step in obtaining the MSA and obtaining the phylogenetic analysis of any species or specimen is obtaining the nucleotide sequence for the organism in question. The species which were included in the study were *Haemophilus segnis*, *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens* and *Kingella kingae*, all of which have been proven to be potential causes of acute infective endocarditis. These organisms are collectively called as HACEK organisms which are a normal part of the human microflora, living in the oral-pharyngeal region. The nucleotide sequences were retrieved from the NCBI data base using the access codes AF516947.1 for *Haemophilus segnis*, AH003513.2 for *Aggregatibacter actinomycetemcomitans*, AY360343.1 for *Cardiobacterium hominis*, NR118696.1 for *Eikenella corrodens* and KJ557363.1 for *Kingella kingae*. For all the above species the nucleotide partial sequence of 16s rRNA was obtained in the FASTA format.



### Multiple Sequence Alignment:

Then the multiple sequence alignment was done for the same using the online software called, Clustal Omega. In Clustal Omega the conserved symbol(\*) denoted that the particular gene was same for all the nucleotide sequences which were used as a part of the study. In the same Clustal Omega the phylogenetic relationship was observed which is distance based. Each branch in the tree is called a Clade and each unit is called a taxon. Once the MSA is done, the phylogenetic analysis has to be done.

### Phylogenetic Analysis:

Prior to 1990, phylogenetic inferences were generally presented as narrative scenarios. Such methods are often ambiguous and lack explicit criteria for evaluating alternative hypotheses.<sup>[8][9][10]</sup> The phylogenetic analysis as a whole can be classified into distance based and character based analysis. However, MSA is the raw material for both.<sup>[11]</sup> Once entering MEGA and a new alignment has been chosen, the FASTA format of the same genes which are to be used are retrieved. The retrieved format is not completely aligned. Hence the alignment is done using the same MEGA software. The alignment is then saved in the MEGA format. The final file is completely aligned and shows the similar nucleotide sequences that has been taken. The tree is generated by combining ClustalW<sup>[12]</sup> and TreeDyn<sup>[13]</sup> using either all sites of the BLAST-reconstructed multiple-alignment or gap-free sites only. Then the phylogenetic tree is constructed for the same set of given data and accordingly the phylogenetic relationship can be analysed and used for further usage, the length of each clade is in correlation with the time units which is given in the phylogenetic tree.

### RESULTS AND DISCUSSION:

The multiple sequence alignment obtained by using the FASTA format of the nucleotide sequence was analysed and so was the phylogenetic tree which was obtained through the MEGA5 software. In this study the method used is the maximum likelihood method which is part of the character based phylogenetic analysis. The phylogenetic analysis was done using a software known as MEGA. Lesser the distance in the phylogenetic tree closer the homologue. Genes which are similar are known as Homologues. Any gene which occurs in different forms due to gene duplication are known as paralogs. Since only the 16s rRNA is used no paralogs are observed. Through analysis of the phylogenetic tree one can obtain how

evolved one organism is from another. Through phylogenetic analysis we can also establish an evolutionary relationship among the HACEK organisms. It is observed that through the phylogenetic tree that the organism *Kingella kingae* split into a separate lane through the road of evolution. It is also noted that *Cardiobacterium hominis* and *Eikenella corrodens* are the most closely related organisms among all the five organisms. It is also a significant distance between *Aggregatibacter actinomycetemcomitans* and the rest of the organisms which implies that the organism is not as evolved as the rest of the organisms which are a part of the HACEK (*Haemophilus segnis*, *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens* and *Kingella kingae*). Thus as a conclusive analysis, since *Cardiobacterium hominis* and *Eikenella corrodens* are so closely related, a similar treatment would work most of the time once the causative organism has been identified.

### CONCLUSION:

As a conclusive statement due to the close relationship between *Cardiobacterium* and *Eikenella*, these causative agents can be treated in a similar manner.<sup>[14]</sup> Thus, this would in turn help to cure the disease

### REFERENCES:

1. Heiro M, Helenius H, Mäkilä S, Hohenthal U, Savunen T, Engblom E, Nikoskelainen J, Kotilainen P (2006). "Infective endocarditis in a Finnish teaching hospital: a study on 326 episodes treated during 1980–2004". *Heart*. 92 (10): 1457–62. doi:10.1136/hrt.2005.084715. PMC 1861063 PMID 16644858
2. Bayer AS, Bolger AF, Taubert KA. et al. Diagnosis and management of IE and its complications. AHA Scientific Statement. *Circulation*. 1998;98:2936–48.[PubMed]
3. Dajani AS, Taubert KA, Wilson W. et al. Prevention of Bacterial Endocarditis: Recommendations by the American Heart Association. *Circulation*. 1997;96:358–66. [PubMed]
4. Golden RL. William Osler at 150. *JAMA*. 1999;282:2252–8. [PubMed]
5. Epaulard O, Roch N, Potton L, Pavese P, Brion JP, Stahl JP. Infective endocarditis-related stroke: diagnostic delay and prognostic factors. *Scand J Infect Dis*. 2009. 41(8):558-62. [Medline].
6. para, childcount:2Durack DT, Lukes AS, Bright DK. New criteria for diagnosis of infective endocarditis: utilization of specific echocardiographic findings. Duke Endocarditis Service. *Am J Med*. 1994 Mar. 96(3):200-9. [Medline].
7. para, childcount:1Brusch JL. Infective endocarditis and its mimics in the critical care unit. Cunha BA, ed. *Infectious Diseases in Critical Care*. 2nd ed. New York, NY: Informa Healthcare; 2007. 261-2.
8. Gotoh O; Consistency of optimal sequence alignment. *Bull Math Biol* 1990, 52:509-525

9. Richard C. Brusca & Gary J. Brusca (2003). *Invertebrates* (2nd ed.). Sunderland, Massachusetts: Sinauer Associates. ISBN 978-0-87893-097-5.
10. Bock, W.J. (2004). Explanations in systematics. Pp. 49-56. In Williams, D.M. and Forey, P.L. (eds) *Milestones in Systematics*. London: Systematics Association Special Volume Series 67. CRC Press, Boca Raton, Florida.
11. Auyang, Sunny Y. (1998). Narratives and Theories in Natural History. In: *Foundations of complex-system theories: in economics, evolutionary biology, and statistical physics*. Cambridge, U.K.; New York: Cambridge University Press.
12. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994, 22 (22): 4673-4680. 10.1093/nar/22.22.4673. Pubmed
13. Chevenet F, Brun C, Banuls AL, Jacq B, Christen R: TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics.* 2006, 7: 439-10.1186/1471-2105-7-439.
14. Ashton JC; *Phylogenetic Methods in Drug delivery*. *Curr Drug Discov Technol.* 2013 Dec;10(4):255-62. PMID: 23895575