

Horizontal transfer in bacterial Methionyl tRNA synthetase is very common shown by Genus and phyla level phylogenetic analysis.

Ghorpade Prabhakar B., Pange Avinash D., Bhaskar Sharma

College of Veterinary And Animal Sciences, Parbhani

Abstract:-

Methionyl tRNA synthetase is single copy informational gene in *Salmonella typhimurium*. Informational genes are more conserved than operational genes. In this study we had analyzed HGT events within MetG sequences of different bacterial genera. A species tree based on 16srRNA sequences of the same genus was drawn evaluated against the generally accepted species tree of the bacteria. MetG phylogenetic tree was evaluated against the 16srRNAs tree and HGT event identified. Similarly phyla trees were made and HGT event identified. 24 HGT events were identified between genus and 11 within phyla. MetG is a considered as conserved gene finding so many HGT event in this gene indicate that horizontal gene transfer is very common in this gene. Manual tree making for phyla could help to understand phylogenetic relationships between very large trees.

Key Terms:-Methionyl tRNA synthetase, Topological difference, Robinson Fould Distance, MEGA.

Background:-

Phylogenetic trees constructed from different genes for same set of organism often gives different phylogeny. The difference in phylogeny could be due to insertion, deletion, sampling errors, differences in rate of substitution, lateral gene transfer etc within these genes; making it difficult to choose a gene to reconstruct phylogeny[1]. In higher animals mitochondrial genes are often chosen because these genes are inherited together and no recombination events are known between them [2]. In case of microorganism ribosomal genes are the key standard for phylogenetic analysis because 16srRNA genes undergo least HGT between species and are of sufficient length (1500 bases) contains both conserved and variable regions to enable species comparison[3-5]. Authentic species trees are used for tracing phylogeny[6] and also are for identifying horizontal gene transfer (HGT) between species [4]and phyla.

Horizontal gene transfer can be identified based on different codon usage pattern or GC content for a particular gene in comparison to other genes in organism [7]. Other approach involves comparison of gene specific tree with species or real time tree. In comparing the 16SrRNA is considered standard for that organism and treated as species tree then one can compare gene of interest phylogeny as gene tree[8, 9]. HGT is lateral gene transfer between different species[10, 11]. The HGT between different species can be

explained at phylum level by treating any bacterial species showing up in displaced location as separate phylum.

In bacteria horizontal transfer is a common phenomenon affecting true phylogeny[12], despite these problems species trees for bacteria have been drawn using multiple conserved gene. This tree is generally accepted as representing true phylogeny within bacterial species [4, 6,13].

Tree topologies generated by methods using different statistical computation are different. Generally trees are character based or distance based. Character types include Maximum Parsimony, Maximum Likelihood and Bayesian inference and distance methods include Minimum Evolution and Least squares. The clustering algorithm with distance as data type, have UPGMA, Neighbourhood Joining and Fitch- Margoliash as methods to draw phylogeny[14-17]. Studies comparing the trees generated by different methods with real time tree shows different fit with the true species tree.

Methods to compare different phylogenetic trees are based on dissimilarity or common information measurements. The dissimilarity methods include Robinson-Foulds (RF)[18], Tree-Bisection-Reconnect (TBR), Subtree-Prune-Regraft (SPR), Geodesic Tree Distance[19], Tanglegrams[20] and Rotation Distance [21].The Common information model for tree comparison includes Maximum Common Refinement Subtree and Maximum agreement subtree[22].

In bacteria generally accepted species tree is that of 16srRNA which is largely based on ribosomal RNA sequences [5]. Ribosomal RNA sequences are conserved but in recent studies it has been shown that they too are subjected to HGT. Another conserved gene is methionyl tRNA synthetase, only single copy of this gene is present in *Salmonella enetica* serovar typhimurium. Aminoacyl tRNA synthetases (AARS) are important for supplying amino acid charged tRNAs to facilitate translation. AARS are classified into Class I and ClassII based on differences in their catalytic sites [23]. MetG is classified into Class I owing to the presence of HIGH and KMSKS signature as amino acid sequence[24, 25]. Among AARS Methionyl tRNA synthetase (MetG) is required for initiator tRNA and supplying Methionine charged tRNA.

In this study we have taken 57 Bacterial Genus representing all the phyla of bacteria mentioned in Hugenholtz phylogeny [26, 27]and downloaded the 16srRNA [28]and MetG sequences (NCBI). Species tree was drawn based on 16srRNA sequences and this species tree was compared with accepted species tree[26]. The tree which gave minimum RF distance was taken as species tree and was used as species tree for finding HGT within species in MetG generated tree. The results are discussed specifically in the context of MetG and in general.

Results:-

Sequences:-

Sequences and accession no of sequences of MetG and 16srRNA used in this study are given in supporting file1 and file 2.

Trees created:-

The 16srRNA and MetG trees created by ML,MP, ME, NJ, UPGMA are available in supporting files(File 3).

Species tree:

The topological distance between tree generated by different methods and Hugenholtz phylogenetic tree (Fig1) as realtime tree are shown in Table1. Minimum Evolution (ME) and NJ tree gave lowest value of RF distance. These trees were identical trees since RF distance between them was zero. 16srRNA ME tree was taken as species tree for further analysis.

MetG tree:

Topological distance between various trees generated with MetG sequences and the species tree 16srRNA ME selected above are shown in table 2. The minimum topological distance with of all the methods was found between Maximum parsimony tree of metG and the species tree of 16srRNA.

Phyla tree:-

Phyla trees were generated by choosing the option collapse in Figtree and drawing phyla in tree builder. Phylum tree created in tree builder and viewed with tree viewer of T-REX for 16SrRNA is shown in Fig.4; for metG phylum tree is shown in Fig.5.

HGT

The HGT was calculated between phyla and species trees. As was expected the HGT within phyla tree was less 11 HGT (Fig.7) than species tree 24 HGT (Fig.6).

Discussion:-

With the increase in the number of bacterial sequences, our understanding of evolutionary relationship between bacterial genera tends to change. The new sequences when included for generating a phylogenetic tree may cause addition of clades in established or new location this may change relevance or irrelevance of phylogeny within established trees; and thus give new understanding to relatedness within genus, phyla or kingdom or even universal tree.

Generally accepted bacterial phylogeny is that of Hugenholtz [26, 27]. We had used that as species tree to evaluate 16srRNA tree generated in this study assuming that it would represent true species tree and by comparison with that we can find HGT in metG sequences. 16srRNA sequences are

generally accepted as conserved with no HGT but recently it has been shown that these sequences are not as conserved as assumed and HGT within 16srRNA sequences is present[35]. We used distance as measurement of dissimilarity between various trees. ME and NJ 16srRNA trees gave minimum dissimilarity with the species tree. ME [36]and NJ [37]generally give identical tree if the number of sequences are less [38-40]. A comparison of real time sequence tree with different phylogenetic trees for HIV reveal that best fit tree by phylogenetic methods were NJ, ML [40]. Based on these results we took ME 16srRNA tree as species tree for HGT estimation in MetG sequences. Comparison of distance measurement with the trees generated by different methods for MetG sequences was used to select the best fit tree. This time tree generated by Maximum parsimony gave lowest RF distance. Maximum parsimony tree is generally used for protein sequences and since protein sequences was used for generating MetG tree that could be the reason for getting lowest RF distance between MetG and 16srRNA species tree generated by us. It is difficult to predict true species tree from the different tree generating methods[40]. Alignment with a known species tree is the true measure for making such prediction.

We found 24 HGT events within MetG sequences, which is on the higher side considering that this protein is a product of informational gene and has high modular structure. The reasons for higher HGT event we observed are difficult to explain. One reason which may partly explain this anomaly could be the species selection. The number of clades/branches in the tree generated by us with MetG sequences had shown different branches particularly proteobacteria (Fig. 3, 5) whereas these were together in Hugenholtz tree (Fig1.) and in 16srRNA tree (Fig.2, Fig4). The high incidence of HGT in MetG sequences found in this study suggests that a relook on prevalence of HGT in other conserved gene sequences is needed.

When the HGT was compared between phyla 11 HGT (Fig.7) incidences were found, this suggest that HGT within phyla is less common than genus. It has been suggested that HGT is dependent on evolutionary distance [12, 41-44].

Our results are in accordance with this fact. This is the first description of HGT within phyla and genus tree. Phyla level HGT by virtue of its relative simplistic network in comparison to genus network may help to understand jumbled network created by intensive gene transfer events which some author predict could replace classical rRNA based phylogeny[10, 11, 45, 46].

Materials and Methods:-

Retrieval of MetG Sequences:-

Salmonella enterica subsp. Entericaserovar Typhimurium genome sequence and other 57 amino acid sequences in fasta file format were retrieved from NCBI.

The 16SrRNA for creating species phylogeny for corresponding genus mentioned in metG sequences were downloaded from Ribosomal Database-Silva from www.arb-silva.de [28].

Alignment of MetG and 16SrRNA sequences:-

The alignments of sequences were performed in MEGA [29] using Muscle. Unaligned region of alignment were removed.

The Evolutionary analysis

Evolutionary analysis was done in MEGA [30-32] and the methods used were Maximum Likelihood tree, Neighbour Joining tree, Minimum Evolution tree, Maximum Parsimony tree and UPGMA.

Test of reliability was Bootstrap method[33]with 500 replications. In order to keep things simple we have kept captions given in MEGA with minor changes for each figures to get information regarding methodology of tree analysis.

Species tree:-

The framework of Hugenholtz species tree as given in [26, 27], was used to construct corresponding species tree in Tree builder canvas of T-REX with addition of archaea as root. This tree was used to select 16srRNA species tree for identifying HGT events.

Analysis of Horizontal Gene transfer:-

The T-REX [34] was used for comparing tree generated by different methods. In the first step the tree corresponding to Hugenholz tree generated above was treated as standard species tree against 16srRNA gene tree generated by different methods. Then the 16srRNA gene tree showing lowest Robinson Fould distance was selected as species tree for identifying HGT events in MetG tree. MetG tree was drawn by different methods and each tree was compared with 16srRNA species tree. The tree which had shown smallest RF, least square and bipartite distance was used for identifying HGT within MetG sequences.

Collapsing species into Phylum:-

The bacterial species belonging to same phylum and falling on same cluster were collapsed into phylum in Figtree(<http://tree.bio.ed.ac.uk/software/figtree>). This action was performed on ME tree of 16SrRNA and MP tree of metG sequences.

Phylogenetic tree at Phylum level:-

Figtree was used to get collapsed phylum tree with individual species on branch represented as phylum by annotating. Then T-REX Newik Builder 0.3 was used to build the phylogenetic tree for phylum level representing corresponding species phylogeny.

Viewing Phylogenetic tree:-

Manually drawn phylogenetic tree was viewed with Tree Viewer from T-REX server.

Analysis of Horizontal Gene transfer at phylum level:-

HGT was calculated between species tree (16SrRNA ME tree) and gene tree (Met MP tree) in Trex server.

Acknowledgement:- We thank Indian Veterinary Research Institute for supporting this work. We also thank ICAR (GP, PA, BS) and CSIR (GP) for financial support for this research.

References:-

1. Löytynoja A, Goldman N: Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 2008, 320(5883):1632-1635.
2. Ingman M, Kaessmann H, PaÈaÈbo S, Gyllenstein U: Mitochondrial genome variation and the origin of modern humans. *Nature* 2000, 408(6813):708-713.
3. Olsen GJ, Woese CR: Ribosomal RNA: a key to phylogeny. *The FASEB journal* 1993, 7(1):113-123.
4. Maddison WP: Gene trees in species trees. *Systematic biology* 1997, 46(3):523-536.
5. Woese CR: Bacterial evolution. *Microbiological reviews* 1987, 51(2):221.
6. Pamilo P, Nei M: Relationships between gene trees and species trees. *Molecular biology and evolution* 1988, 5(5):568-583.
7. Lawrence JG, Ochman H: Amelioration of bacterial genomes: rates of change and exchange. *Journal of molecular evolution* 1997, 44(4):383-397.
8. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF: Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *Journal of Bacteriology* 2004, 186(9):2629-2635.
9. Boc A, Philippe H, Makarenkov V: Inferring and validating horizontal gene transfer events using bipartition dissimilarity. *Systematic biology* 2010, 59(2):195-211.
10. Doolittle W, Boucher Y, Nesbø C, Douady C, Andersson J, Roger A: How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2003, 358(1429):39-58.
11. Kurland C, Canback B, Berg OG: Horizontal gene transfer: a critical view. *Proceedings of the National Academy of Sciences* 2003, 100(17):9658-9662.
12. Ochman H, Lawrence JG, Groisman EA: Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000, 405(6784):299-304.
13. Chaudhary R, Burleigh JG, Fernández-Baca D: Inferring Species Trees from Incongruent Multi-Copy Gene Trees Using the Robinson-Foulds Distance. *arXiv preprint arXiv:12102665* 2012.
14. Saitou N, Imanishi T: Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol Biol Evol* 1989, 6(5):514-525.
15. Yang Z, Rannala B: Molecular phylogenetics: principles and practice. *Nature Reviews Genetics* 2012, 13(5):303-314.

16. Nei M: Phylogenetic analysis in molecular evolutionary genetics. *Annual review of genetics* 1996, 30(1):371-403.
17. Mount DW: Choosing a method for phylogenetic prediction. *Cold Spring Harbor Protocols* 2008, 2008(4):pdb. ip49.
18. Robinson D, Foulds LR: Comparison of phylogenetic trees. *Mathematical Biosciences* 1981, 53(1):131-147.
19. Amenta N, Godwin M, Postarnakevich N, St John K: Approximating geodesic tree distance. *Information processing letters* 2007, 103(2):61-65.
20. Buchin K, Buchin M, Byrka J, Nöllenburg M, Okamoto Y, Silveira RI, Wolff A: Drawing (complete) binary tanglegrams. In: *Graph Drawing: 2009*. Springer: 324-335.
21. John KS: Comparing phylogenetic trees. In: *EMBO Workshop on Current Challenges and Problems in Phylogenetics, Isaac Newton Institute, Cambridge, UK: 2007*.
22. Finden C, Gordon A: Obtaining common pruned trees. *Journal of Classification* 1985, 2(1):255-276.
23. Burbaum J, Schimmel P: Structural relationships and the classification of aminoacyl-tRNA synthetases. *J Biol Chem* 1991, 266(26):16965-16968.
24. Woese CR, Olsen GJ, Ibba M, Söll D: Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiology and Molecular Biology Reviews* 2000, 64(1):202-236.
25. Ibba M, Söll D: Aminoacyl-tRNA synthesis. *Annual review of biochemistry* 2000, 69(1):617-650.
26. Hugenholtz P, Goebel BM, Pace NR: Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology* 1998, 180(18):4765-4774.
27. Ecker DJ, Sampath R, Willett P, Wyatt JR, Samant V, Massire C, Hall TA, Hari K, McNeil JA, Büchen-Osmond C: The Microbial Rosetta Stone Database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents. *BMC microbiology* 2005, 5(1):19.
28. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO: The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* 2013, 41(D1):D590-D596.
29. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular biology and evolution* 2013, 30(12):2725-2729.
30. Kumar S, Nei M, Dudley J, Tamura K: MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics* 2008, 9(4):299-306.
31. Tamura K, Dudley J, Nei M, Kumar S: MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution* 2007, 24(8):1596-1599.
32. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 2011, 28(10):2731-2739.
33. Felsenstein J: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985:783-791.
34. Alix B, Vladimir M: T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic acids research* 2012, 40(W1):W573-W579.
35. Kitahara K, Yasutake Y, Miyazaki K: Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 2012, 109(47):19220-19225.
36. Nei M, Kumar S, Takahashi K: The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proceedings of the National Academy of Sciences* 1998, 95(21):12390-12397.

37. Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* 1987, 4(4):406-425.
38. Takahashi K, Nei M: Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular biology and evolution* 2000, 17(8):1251-1258.
39. Rzhetsky A, Kumar S, Nei M: Four-cluster analysis: a simple method to test phylogenetic hypotheses. *Molecular biology and evolution* 1995, 12(1):163-167.
40. Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J: Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proceedings of the National Academy of Sciences* 1996, 93(20):10864-10869.
41. Brügger K, Redder P, She Q, Confalonieri F, Zivanovic Y, Garrett RA: Mobile elements in archaeal genomes. *FEMS microbiology letters* 2002, 206(2):131-141.
42. Nakamura Y, Itoh T, Matsuda H, Gojobori T: Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature genetics* 2004, 36(7):760-766.
43. Ge F, Wang L-S, Kim J: The cobweb of life revealed by genome-scale estimates of horizontal gene transfer. *PLoS biology* 2005, 3(10):e316.
44. Boto L: Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society B: Biological Sciences* 2010, 277(1683):819-827.
45. de la Cruz F, Davies J: Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends in microbiology* 2000, 8(3):128-133.
46. Doolittle WF: Phylogenetic classification and the universal tree. *Science* 1999, 284(5423):2124-2128.

Figures Legends:-

Fig.1. Hugenholtz tree representation by Tree builder

Fig. 2.The 16srRNA Minimum Evolution tree

Fig.3 The metG tree made by Maximum Parsimony method

Fig.4. 16srrna Phylum tree made inTrex tree builder

Fig.5. metG(MP) Phylum tree made in Trex tree builder

Fig.6. Horizontal transfer in genus of 16SrRNA tree made by Minimum Evolution (species tree) with metGtree made by Maximum Parsimonymethod (gene tree)(HGT=24)

Fig.7.Horizontal transfer in genus of 16SrRNA Phyla tree (ME) (as species tree) with metGPhyla tree (MP) (gene tree)(HGT=11)

Table 1. Topological distance between Hugenholz tree and 16srRNA Trees

Trees	Robinson Fould distance	Least-squares coefficient(LS)	Bipartition dissimilarity
ML16srRNA	59	287613602416.595	93.5
MP16srRNA	58	1900590657932.631	109.5
ME16srRNA	55	287613563786.531	88.0
NJ16srRNA	55	287613563786.531	88.0
UPGMA16srRNA	59	287613567458.186	93.0

Table2a. Topological distance between 16srRNA tree and metGtre

Trees	Robinson Fould distance	Least-squares coefficient(LS)	Bipartition dissimilarity
ML-met	72	289277102268782	104
MP-met	66	5456770617520	81
ME-met	70	289277485646107	102
NJ-met	70	289277485646107	102
UPGM A-met	70	289277506971835	105

Fig.3 The metG tree made by Maximum Parsimony method

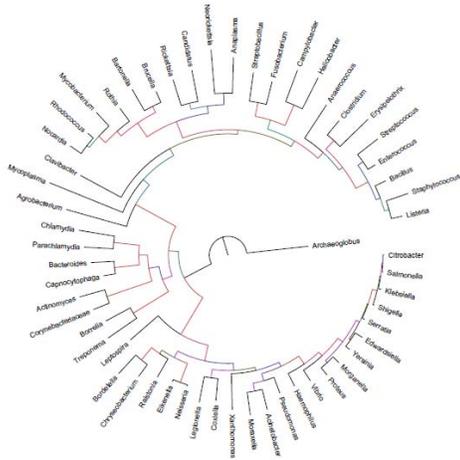


Fig.4. 16srrna Phylum tree made in Trex tree builder

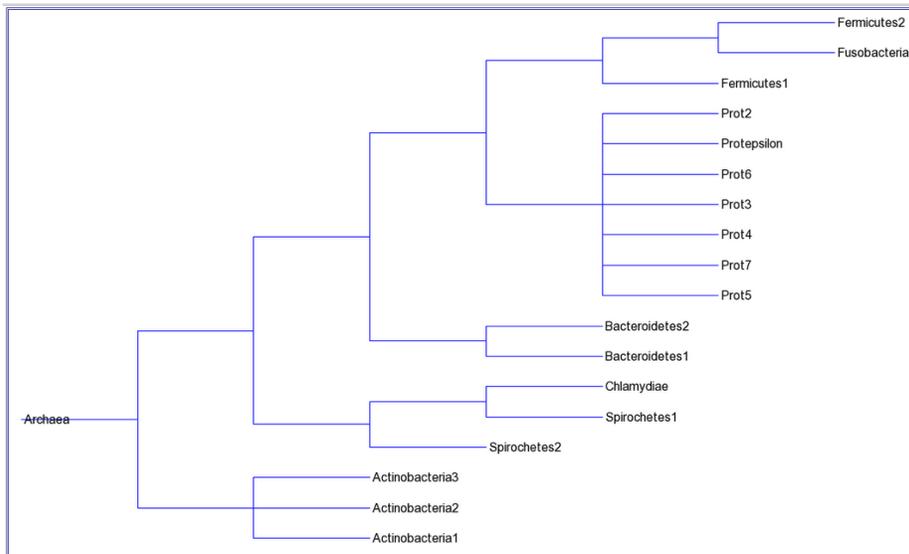


Fig.5. metG(MP) Phylum tree made in Trex tree builder

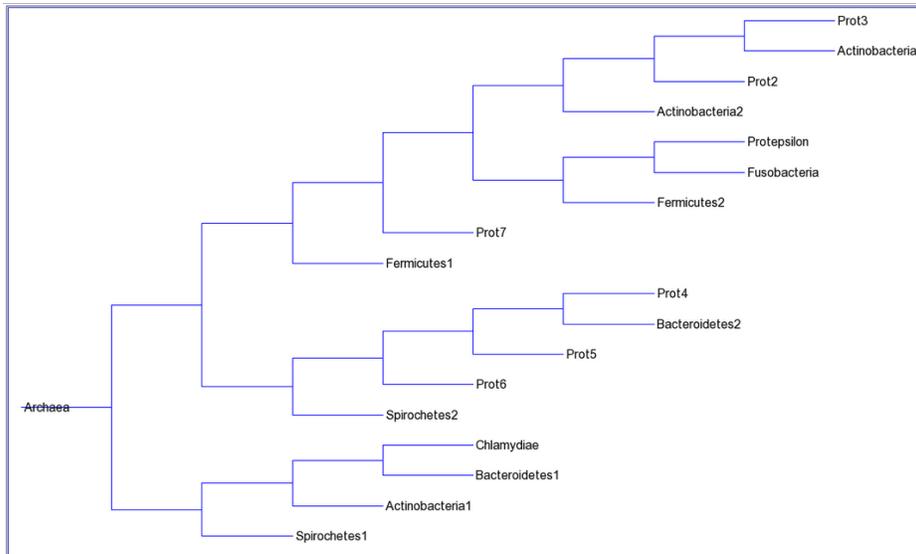


Fig.6. Horizontal transfer in genus of 16SrRNA tree made by Minimum Evolution (species tree) with metG tree made by Maximum Parsimony method (gene tree)(HGT=24)

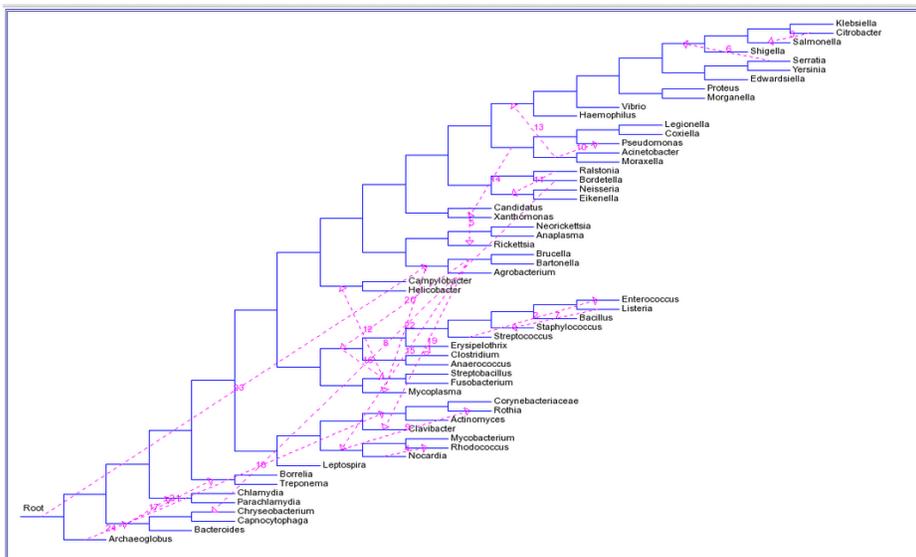


Fig.7. Horizontal transfer in genus of 16SrRNA Phyla tree (ME) (as species tree) with metGPhyla tree (MP) (gene tree)(HGT=11)

