Journal of Scientific Research
Banaras Hindu University, Varanasi

Vol. 63, 2019 : 89-103
ISSN : 0447-9483

RIBOSOMAL GENE BASED COMPARATIVE PHYLOGENIES FOR THE GENUS *MYCOBACTERIUM:* AN *IN-SILICO*APPROACH

Aayush Rastogi[1], Shruti Gautam[1], Manish Kumar[1*], and Rajesh Singh Tomar[1]
[1]Amity Institute of Biotechnology, Amity University, Gwalior, Madhya Pradesh
*mkumar@gwa.amity.edu

Abstract:

With the help of computational biology, scientists are now equipped to provide a proper classification of the organism and also to correct the previously wrongly placed organisms. Mycobacterium is a very significant genus in the world of prokaryotes as it has some major species that are tremendously pathogenic to humans, such as Mycobacterium tuberculosis. Hence the members under this genus need to be placed at appropriate positions in the phylogenetic tree and need to be assessed that which species is close to whom and which are widely dissimilar. In the current investigation, phylogenetic trees were constructed using ribosomal gene sequences. The tree constructed on the basis of 16S rRNA gene revealed two major groups and each group contain different 4 OTUs (operational taxonomic units) whereas tree constructed based on 23S contains 5 different OTUs and tree constructed based on 5S contain two different OTUs. The findings showed that the trees obtained on the basis of 23S, 16S, and 5S rRNA includes similarities along with some non-negotiable "discordance", which cannot be overlooked in the case of Mycobacterium. The results of phylogeny based on different ribosomal genes are not satisfying each other. This suggest that in the course of evolution there are changes directly in proportion to the length of nucleotide segment taken and on this basis it can be deduced that organisms show different relations regarding different sequence length taken under consideration, and for constructing a valid phylogeny, housekeeping genes should be taken along with conserved ribosomal gene sequences.

Keywords: OTUs, rRNA, Discordance, Phylogeny

Introduction

A biological system is usually a combination of cells, tissues, and organs that make it a unique and complex assembly. With the help of molecular biology which has revealed a massive amount of biological scenarios, including DNA, RNA, genome sequences, gene, and properties of proteins, it's much better to characterize biological systems as "symbiotic systems" rather than complex. A multitude of biological facts has been revealed by molecular biology, for example, genome sequences and properties of the protein, but this as lone to interpret biological systems is still not sufficient. Due to the inherent complexity of biological systems, a blend of computational and experimental approaches is projected to resolve this issue. Along with molecular disciplinary, computational biology, via practical modeling and conjectural investigation, provides a powerful base for addressing critical scientific queries head-on (Kitano, 2002).

Over several years, a momentous shift was observed in the field of microbiology that changed the determinative taxonomy to nucleotide based phylogeny. The field of medical biotechnology and microbiology is based on the microbial culture and their identification by biochemical and molecular characterization to diagnose different diseases. After DNA structure determination by Watson & Crick (1953), bacteria related research have seen a huge deflection from functional to molecular characterization based identification (Towner & Cockayne, 1993; Meena et al 2015; Kumar et al 2015; Kumar et al 2016; Kumar et al 2018).

To observe the evolutionary relationship of a large group of living organisms, a macromolecule based molecular study is required. None of the proteins can be used for this type of study. However, the conserved ribosomal rRNA does. The ribosomal gene sequences are evolved slowly in such a way that sequences are conserved and allow the phylogenetic relatedness and identity of close and distant species. The structure of ribosomal structural genes has been characterized well in a large group of organisms (Woese et al., 1977). Moreover, the phylogeny is mostly reliable based on these ribosomal gene sequences and their analysis. This now became the recent basis for new edition of Bergey's Manual of Systematic Bacteriology and Prokaryotes. Therefore the ribosomal gene based phylogenetic analysis provides a better understanding of taxonomy and systematics including species relatedness.

The ribosomal genes are very crucial and required for the survival which is more conserved in the kingdom bacteria and others. The comparison of rRNA sequences is a powerful tool for deducing the evolutionary tree for the relatedness of eubacteria, primitive archaebacteria, and eukaryotic organisms (Weisburg et al., 1991). As a result, the molecular characterization based on 16S rRNA gene is accepted now as a well defined and standard method for the characterization, identification and phylogeny of organisms of different domains (Amann et al., 1995; Woese, 1987). The application of ribosomal gene reflected on the phylogenetic relatedness, provided a better understanding of evolutionary relationship at species level (Olsen et al., 1991). But according to Dewhirst et al., 2005, a number of cases are there to be reinvestigated about closely related species or strains with the use of other molecules. In case of prokaryotic taxonomy and systematic, ribosomal gene based phylogeny appears to conflict with other molecular, morphological and phenotypic data.

Further to the medical microbiology, a genus Mycobacterium strikes down the mind as it has some deadliest organisms from the past (like *Mycobacterium tuberculosis*). The aerobic Mycobacteria are acid fast staining, non-mobile bacteria that have62–72 mol% GC in their DNA. They belong to bacterial family Mycobacteriaceae within the same genus Mycobacterium (Ade´Kambi et al., 2004). The Mycobacterium species is a very important species from pathogenicity point of view. Therefore, there actual identification (apart from preliminary identification processes) and their actual phylogenetic relatedness with several other Mycobacterial species are essential. Several

wet lab experiments indicated this pathogen at species level after their identification based on universal gene sequences.

Several years back, the ribosomal gene-based phylogeny of prokaryotes or eukaryotes has shown a variation and shifts in the determinative taxonomy along with discordant evolutionary relatedness among individuals. The current trends for the prokaryotic phylogenetic analysis rely on 16S rRNA gene sequences at the larger extent and have been accepted widely along with Bergey's Manual of Systematic Bacteriology. 16S rRNA gene-based identification and phylogeny was accepted by researchers but some discordant data conflict was appeared and observed with other ribosomal and housekeeping gene-based phylogeny. Many reports illustrated almost similar phylogenetic relatedness based on 16S rRNA sequences within-species differences.

The current research scenario facilitates the identification of microbes based on molecular characterization along with conventional and next-generation sequence analysis. A current investigation is an *in-silico* approach for placing various Mycobacterium species to a phylogenetic position based on three different types of universal conserved ribosomal gene sequences to observe their phylogenetic positions and evolutionary relatedness. Also further extending this approach to the comparison between graphical representations of the clustered fragments of these three universally conserved ribosomal gene sequences with the phylogenetic tree obtained respectively will provide an insight to the proper identification of several species of Mycobacterium done in different wet lab experiments. The purpose of this study was to compare the evolutionary position of Mycobacterium strains with the help of extensive 23S rRNA, 5S rRNA, and 16S rRNA genes sequences along with data based on *in-silico* RFLP.

Materials and Methods

Sequence Data

The publically available gene sequences and particularly relevant species with a significant number of ribosomal gene sequences were considered for the detailed phylogenetic analysis. These include isolates from different species of Mycobacterium genus such as *M. gilvum*, M. *intracellulare*, *M. africanum*, and others, documented in the text (Table 1).

All closely associated species and about more than 100 species were also downloaded from NCBI and representative accession numbers were used for the construction of phylogenetic tree, mentioned (Table 1).

Phylogenetic analysis

Phylogenetic analysis was carried out using the 16S rDNA, 5S rDNA and 23S rDNA sequences.

The ribosomal sequences of representative strains of Mycobacterium were utilized for the construction of the evolutionary tree. The BLAST homology revealed

the identified strains of Mycobacterium based on ribosomal gene sequence similarity ≤97%. The pair wise sequence alignment followed by multiple sequence alignment was done by the ClustalW program (Thompson et al 1994) with the default or inbuilt settings, and the sequence data is converted to MEGA format or PHYLIP format. Few changes have been done manually based on conserved sequences and the aligned columns with more than 50% gaps were removed. The software MEGA 5.0 was used that was based on a neighbor-Joining method for tree construction (Tamura et al 2007)

The small changes were done manually by removing 50% gaps and on the basis of conserved domains. The phylogenetic trees were constructed on the aligned datasets using MEGA 4.0.2 (Tamura et al 2007) using the neighbor-joining method (Saitou and Nei 1987). The Bootstrap value analysis was done by taking 1,000 random samples from multiple sequence alignments (Felsenstein 1985).

In-silico Restriction Enzyme Analysis

The ribosomal gene sequences were subjected to digestion by a restriction endonuclease. Virtual restriction enzyme digests of different ribosomal gene sequences were done through web-based Restriction digestion application 'Codon-code aligner' (Hazard et al 2013). Restriction enzymes which are tetra-cutters are generally frequent cutters and were selected for the digestion of ribosomal gene sequences to obtain a DNA fingerprinting pattern. The uniqueness of a restriction enzyme for a particular species was investigated. The ribosomal gene sequences were cut by EcoRI, AluI and Hae III restriction endonucleases and different fragments of different length are generated (Fig.1).

Table 1. The details of nucleotide sequences of ribosomal genes retrieved for the study of phylogeny among *Mycobacterium* strains.

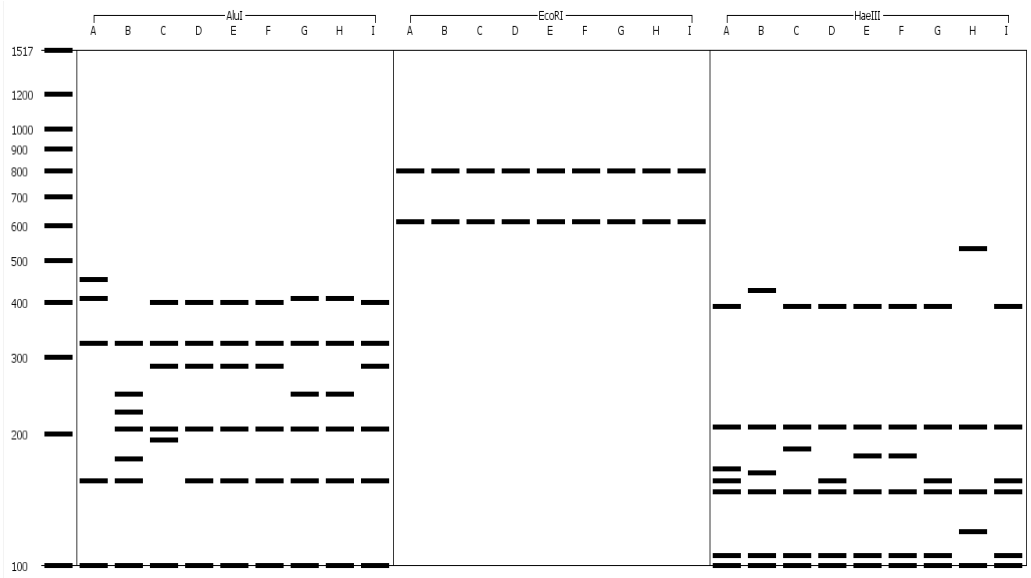
S. No.	Mycobacterium Species	rRNA Genes	Size of the rRNA gene	Representative sequences	NCBI Accession for representative sequences
1	*Mycobacterium gilvum* ATCC 43909, *M. intracellulare* strain DSM 43223, *M. africanum* strain ATCC 25420, *M.smegmatis* strain ATCC 19420, *M.marinum* strain ATCC 927, *M. abscessus* strain	16S rRNA	1500 bp	9	NR_025311, NR_025584, NR_042164, NR_029293, NR_042917, NR_113366, NR_025238, NR_042165, NR_118915,

	Hauduroy L948, *M. vanbaalenii* strain PYR-1, *M. kansasii* strain DSM 44162, *M. avium* strain DSM 44156				
2	*Mycobacterium smegmatis* strainMC2 155l, *M.avium* strain 104, *M.kansasii* strain ATCC 12478, *M. abscessus*, *Mycobacterium marinum* strain M, *M.africanum* strain GM041182, *M.intracellulare* strain MOTT-02, *Mycobacterium gilvum* strain PYR-GCK	23S rRNA	>3Kb	8	NR_076654.1,NR_103001.2, NR_121961.1,NR_077010.1, NR_076117.1,NR_076153.1, NR_076151.1, NR_076110.1
3	*Mycobacterium smegmatis* str. MC2 155, *M.avium* 104 strain 104, *M.kansasii* ATCC 12478, *M.vanbaalenii* PYR-1 strain, *M. abscessus*, *M. marinum* M strain M, *M. africanum* GM041182 strain GM041182, *M.intracellulare* MOTT-02, *M.gilvum* PYR-GCK strain	5S rRNA	< 500 bp	9	NR_075650.1,NR_103288.1, NR_121846.1,NR_075440.1, NR_076030.1,NR_075528.1, NR_076067.1,NR_075972.1, NR_075471.1

Results and Discussion

Phylogenetic tree based on 16S rRNA gene sequences

The Phylogenetic tree was constructed based on 16S rRNA gene sequences revealed two major groups i.e., group 1 and group 2. These two groups contain different operational taxonomic units (OTUs). In first major group, two distinct OTUs were obtained. First OTU consisted species such as *Mycobacterium marinum*, *Mycobacterium africanum* and *Mycobacterium kansasii*. The second OTU under this major group consisted species such as *Mycobacterium avium* and *Mycobacterium intracellulare*. These two OTUs are bifurcated with 100 bootstrap value. In the second major group 2, again two OTUs were obtained and designated as OTU 3 and 4. The species such as *Mycobacterium abscessus* was found under OTU number 3 while species such as *Mycobaterium vanbaalenii*, *Mycobacterium smegmatis* and *Mycobacterium gilvum* under OTU number 4. OTU 3 and 4 were bifurcated with bootstrap value about 50.



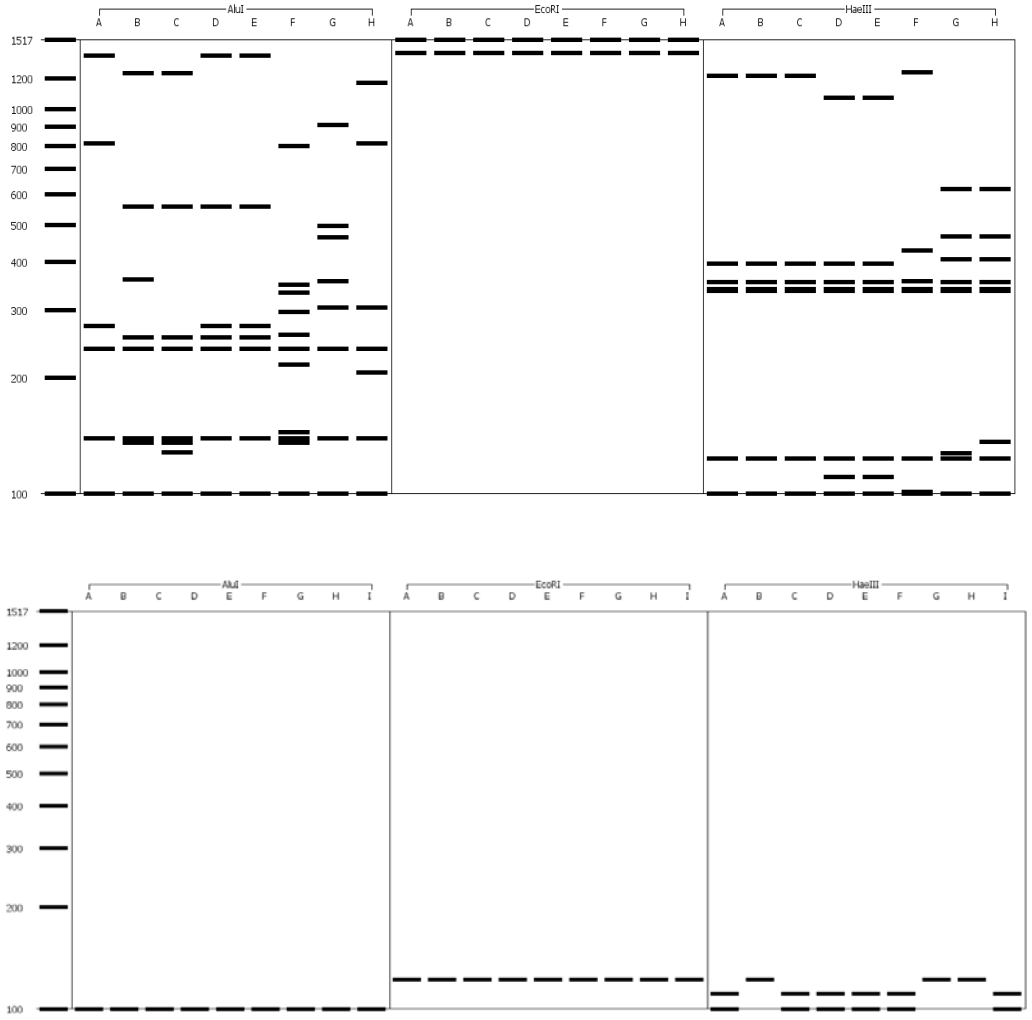


Fig.1. The virtual RFLP pattern was obtained after restriction digestion of all three ribosomal genes

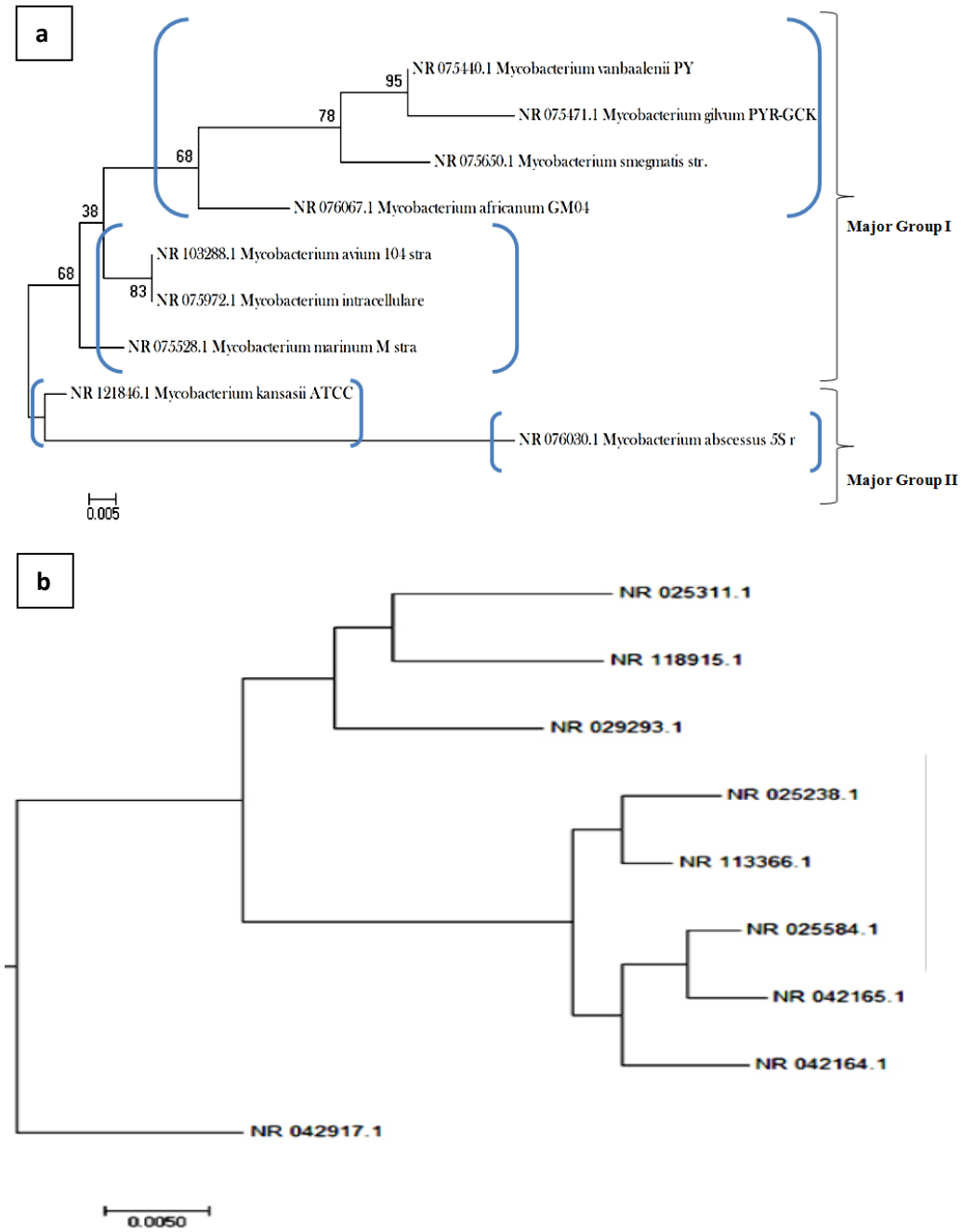


Fig 2. Phylogenetic tree construction based on 5S rRNA gene sequences (Blue parenthesis showing different OTUs) compared with the dendrogram based on RFLP pattern

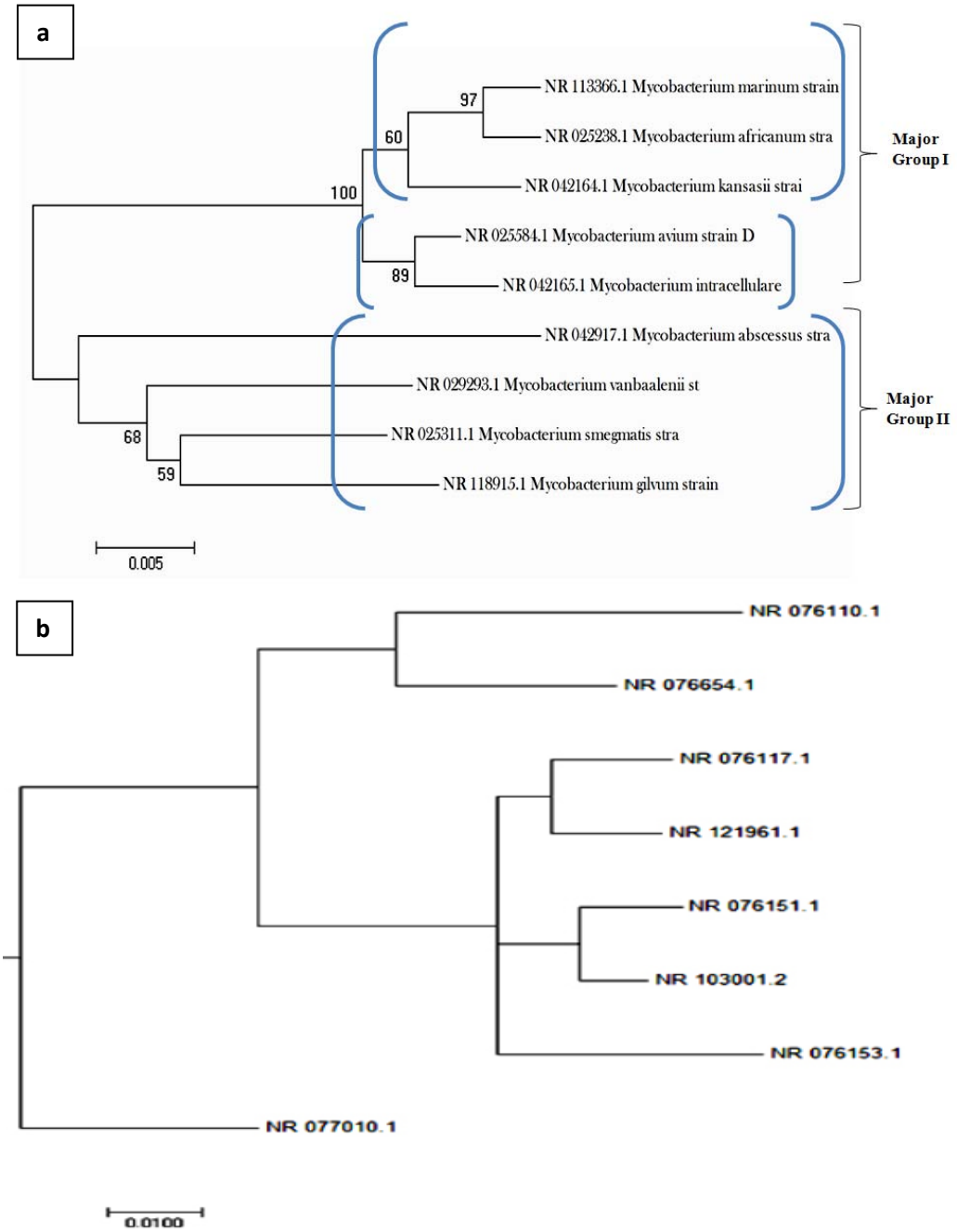


Fig 3. Phylogenetic tree construction based on 16S rRNA gene sequences (Blue parenthesis showing different OTUs) compared with the dendrogram based on RFLP pattern

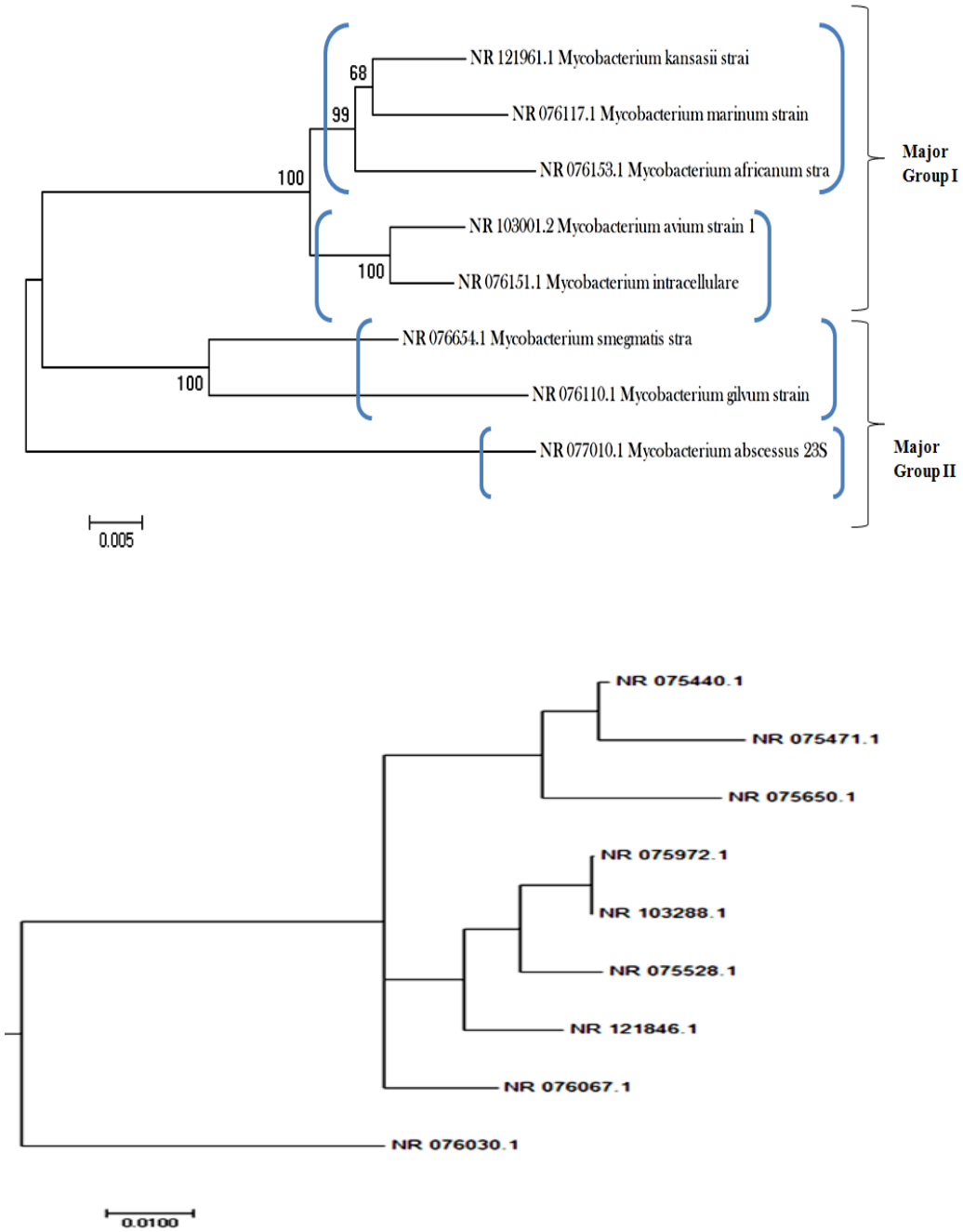


Fig 4. Phylogenetic tree construction based on 23S rRNA gene sequences (Blue parenthesis showing different OTUs) compared with the dendrogram based on RFLP pattern

Phylogenetic tree based on 23S rRNA gene sequences

In the phylogenetic tree based on 23S rRNA sequences, all species presenting major group 1 is matching with species of major group 1 in phylogenetic tree obtained based on 16S rRNA gene sequences with little bit differences in the bootstrap value of the branches. If this phylogenetic tree is compared with 23S rRNA gene sequence-based tree, the data for the second major group was obtained "discordant" while the first major group of first phylogenetic tree based on 16S rRNA was with an appropriate agreement with that of major group 1 of second phylogenetic tree based on 23S rRNA. For example species such as *M. avium* and *M. intracellulare* are bifurcated with a bootstrap value of 100 in the tree based on 23S rRNA, whereas same species in a tree based on 16S rRNA are bifurcated with the bootstrap value of 89. After comparison of these two phylogenetic trees, the 2nd major group was giving the discordant position of Mycobacterium species. In the tree obtained based on 23S rRNA, a separate single species M. abscessus was observed as separate OTU and it is appearing like an out-group. The total number of OTUs obtained in both 16S and 23S are similar i.e., 4 in number.

Phylogenetic tree based on 5S rRNA gene sequences

The final or 3rd tree based on 5S rRNA gene sequences again increasing the disagreement with the trees obtained based on 16S and 23S. The phylogenetic position of each and every species in this tree illustrate the discordant positions of Mycobacterium species if comparing with the trees based on 23S and 16S rRNA gene sequences. Here also two major groups were obtained and most of the species of Mycobacterium were placed in group 1. The major subgroup 1 was not subdivided into different OTUs as it was observed in earlier trees constructed. Only two OTUs were obtained with respect to major groups. Mycobacterium species such as M. kansasii and M. abscessus gives different positions in different trees constructed. The overall study elucidates that discordant phylogenetic tree was obtained for the same species, if constructed by using three different conserved universal ribosomal gene sequences. The M. abscessus appeared as an outgroup species in the tree constructed using 23S rRNA while it was phylogenetically placed in the same OTU containing M. kansasii in the tree based on 5S rRNA gene sequences. In another tree based on 16S rRNA gene sequences, M. abscessus was in major group 2 as a separate OTU.

Virtual RFLP and Dendrogram construction based on 0-1 matrix

The in-silico RFLP of ribosomal genes provided different banding pattern after restriction digestion with three restriction endonucleases. A polymorphic banding pattern was obtained in case of enzyme HaeIII and AluI while EcoRI has not given a significant banding pattern. The overall banding pattern was utilized for the construction of dendrogram for their grouping on the basis of similarity (0-1 matrix based). The RFLP pattern in case of 5S rRNA gene was found to be very poor with lesser bands separated and this may be due to little or absence of restriction sites for the restriction enzymes used in this study. Three dendrograms for three different ribosomal

genes were constructed using RFLP banding pattern. The dendrogram was compared with the phylogenetic tree constructed for the respective ribosomal genes. The bacterial groups were formed on the basis of variants formed due to variation in the banding pattern (Fig. 2,3,4). Overall, when phylogenetic tree constructed based on ribosomal gene sequences are compared, there is discordance in the phylogenetic positions of the Mycobacterium sp. whereas the tree constructed based on DNA fingerprinting pattern when compared with the sequence-based tree, there was very little or negligible discordance.

Discussion

Mycobacterium is a very important pathogenic bacterium and therefore it is very important and essential to know their proper identification and phylogenetic relationship among different Mycobacterium strains. The current investigation is an effort to analyze the phylogeny among different *Mycobacterium* sp. by considering different ribosomal gene sequences and fingerprinting technique such as RFLP. The phylogenetic tree constructed based on ribosomal gene sequences was compared with the tree (dendrogram) constructed based on RFLP banding pattern. The results suggested that the dendrogram either constructed based on ribosomal gene sequences or based on fingerprinting technique has provided overall similar relatedness among species with little or no discordant positions. The dendrogram constructed based on DNA fingerprinting banding pattern, giving an agreement with the tree constructed based on gene sequences. However, when three different trees were constructed based on 16S rRNA, 23S rRNA and 18S rRNA gene sequences, the phylogenetic tree based on 5S rRNA gives no agreement with other two trees. This may be due to the smaller size of this gene sequence that cannot give proper identification and evolutionary position.

The 23S rRNA gene-based tree has three distinct OTUs i.e. OUT 1 (accession 76654.1, 76110.1), OUT2 (077010.1) and OTU3 (Rest of the sequences) and this grouping was found to be approximately similar with the similarity based dendrogram (tree based on RFLP banding pattern). In a similar way, when 5S rRNA gene and 16S rRNA gene-based tree was compared with the RFLP based tree for the respective ribosomal gene, the overall phylogenetic position and relatedness of *Mycobacterium* sp. were found to be again approximately similar. This ‘approximate similarity’ refers to the discordant position of only one or two species among trees. The RFLP banding pattern obtained in the case of 5S rRNA gene was not found to be appropriate or significant as compared to 16S and 23S rRNA. Since the size of 5S rRNA gene is smaller (less than 500 bp), may be a reason for lesser or rare availability of restriction sites. There are several studies illustrating the in silico phylogeny of bacteria (Puri et al 2016; Pramanik et al 2017; Karakülah and Pavlopoulou 2018; Mirzaei et al 2014; Satpathy et al 2016) either by considering only conserved genes or different functional genes.

The congruence of 5S rRNA gene-based phylogenetic tree with another ribosomal gene-based tree may be due to comparatively very smaller size and little polymorphism obtained after virtual RFLP. In the phylogenetic study of Vibrio isolates described by Farhadi et al (2015), it was emphasized that the congruence in the phylogeny was probably due to horizontal gene transfer.

Conclusion

The current investigation that was based on computer simulation illustrates the phylogeny of *Mycobacterium* sp. based on different ribosomal gene sequences. The tools used in the current study are a reliable source of knowledge for conserved gene sequences. The reconstructed phylogenetic tree based on DNA sequences of ribosomal genes provided an insight for the identification and proper phylogeny of Mycobacterium sp. However, these tools were found to be highly successful in discriminating Mycobacterium sp. from its other relative species. The discordant data and results insisted us to utilize a combination of both conserved and other housekeeping genes such as DNA gyrase and RNA polymerase. The abundant and sufficient sequence data of ribosomal genes along with various housekeeping genes can be used for the supplementation of ribosomal gene sequence data to facilitate the identification, proper evolutionary relationship and exact phylogenetic position of widespread pathogen like Mycobacterium. This approach, therefore, may trigger the less time-consuming way and efforts for the characterization and identification along with exact OTU (Operational taxonomic unit) based phylogenetic position of new bacterial strains of this group.

Acknowledgement

Authors are highly obliged to Dr. Ashok K. Chauhan, Aseem Chauhan and Amity University Madhya Pradesh, Gwalior for their research encouragement and proper facilities available at Institute of Biotechnology, AUMP.

References

Ade´kambi, T., and Drancourt,M. (2004) Dissection of phylogenic relationships among nineteen rapidly growing Mycobacterium species by 16S rRNA, hsp 65,sodA, recA, and rpoB gene sequencing. *Int. J. Syst. Evol. Microbiol.* 54 (6), 2095-105.

Amann,R., Ludwig,W., and Schleifer, K.H. (1995) Phylogentic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143-169.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evol*. 39, 783-791.

Farhadi, T., Nezafat, N., and Ghasemi Y. (2015) Insilico phylogenetic analysis of Vibrio cholerae isolates based on three housekeeping genes. *International Journal of Computational Biology and Drug Design* 8 (1), 62-74.

Hazard, C., Gosling, P., Van der Gast, C.J., Mitchell, D.T., Doohan, F.M., and Bending, G.D. (2013) The role of local environment and geographical distance in determining community composition of arbuscularmycorrhizal fungi at the landscape scale. *The ISME Journal* 7, 498-508.

Karakülah, G. and Pavlopoulou, A. (2018) Insilico Phylogenetic Analysis of hAT Transposable Elements in Plants. *Genes* 9, 284.

Kumar, M., Srivastava, A.K., Pandey, A.K. (2015) Biocontrol activity ofsome potent Methylotrophs isolated from Bhitarkanika mangrove sediment. *Int J Curr Res Biosci Plant Biol* 2, 101-106.

Kumar, M., Tomar, R. S., Lade, H., and Paul, D. (2016) Methylotrophic bacteria in sustainable agriculture. *World J. Microbiol. Biotechnol.* 32, 1-9.

Kumar, M., Saxena, R., Parihar, S.S., Rai, P.K., and Tomar, R.S., (2018) Molecular Characterization and Phylogeny of Some Cyanobacterial Strains Isolated from Soil and Freshwater Ecosystem. *Journal of Pure and Applied Microbiology* 12(2), 897-904

Mirzaei, K., Bahramnejad, B., Shamsifard, H.S. and Zamani,W. (2014) Insilico identification, phylogenetic and bioinformatic analysis of argonaute genes in plants. *Int J Genomics* 967461.

Pramanik, K., Ghosh, P.K., Soumyajit, R., Sarkar, A., Mitra, S., Maiti, T.K.(2017) An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of Pseudomonas aeruginosa. *Journal of Genetic Engineering and Biotechnology* 15, 527-537

Meena, K. K., Kumar, M., Mishra, S., Ojha, S. K., Wakchaure, G.C., Sarkar, B. (2015) Phylogenetic study of methanol oxidizers from Chilika-Lake sediments using genomic and metagenomic approaches.*Indian J Microbiol.* 55, 151-162

Olsen, G. J., Larsen, N. & Woese, C. R. (1991) The ribosomal RNA database project. *Nucleic Acids Res.* 19, 2017-2021

Puri, A., Rai, A., Dhanaraj, P.S., Lal, R., Patel, D.D., Kaicker and Verma, M. (2016) An In Silico Approach for Identification of the Pathogenic Species, Helicobacter pylori and Its Relatives. *Indian J Microbiol.* 56(3), 277-286.

Satpathy, R., Konkimalla, V.B., and Ratha, J. (2016) In Silico Phylogenetic Analysis and Molecular Modelling Study of 2-Haloalkanoic Acid Dehalogenase Enzymes from Bacterial and Fungal Origin Advances in Bioinformatics. Volume 2016.

Saitou, N., Nei, M. (1987) The neighbor-joining method a new method for reconstructing phylogenetic trees. *Mol. Biol.Evol.*4, 406-425.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTAL W: Improving sensitivity of progressive multiple sequence alignments through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680

Tamura, K., Dudley, J., Nei, M., Kumar, S. (2007) MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.

Towner, K. J. & Cockayne, A. (1993) Molecular Methods for Microbial Identification and Typing. London: Chapman & Hall. In book: Molecular Methods for Microbial Identification and Typing. pp. 1-27.

Woese, C. R. (1987) Bacterial evolution. *Microbiol. Rev.* 51, 221-271.

Woese, C. R. (1977) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya, *Proc. Nati. Acad. Sci.*87, 4576-4579.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., (1991) 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173, 697–703