

Article**Phylogenetic tree of *Proteus spp.* Based on partial *rpoB* gene sequence analysis**Hussein Ali Mutlag^{1,*}, and Intesar N. Khelkal^{2,*}¹ Biology Department, College of Science, Al-Mustansiriyah University, Iraq.*Correspondence: Intesarnkshaibani2021@Uomustansiriyah.edu.iq, Husseinalsaleem84@gmail.comAvailable from: <http://dx.doi.org/10.21931/RB/CSS/2023.08.01.73>

Abstract: Due to the importance of the proposed *rpoB* gene as an alternative biomarker for microbial community studies, this study has focused on phylogenetic relationships among local *Proteus* clinical isolates. Fifty bacterial isolates were collected and identified phenotypically according to the culture, microscopic examination and biochemical tests. VITEK 2 compact system was used to confirm identification. Genotypic identification was performed after DNA extraction for 10 selected isolates and amplification with *rpoB* gene-specific primer and gel electrophoresis. The products were detected with a (1090 bp) molecular size band, which was sent for Sanger sequencing using an ABI3730XL automated DNA sequencer, and data were analyzed and compared with standard sequences in GenBank. The isolates have been registered in the National Center for Biotechnology Information (NCBI) with accession numbers and named (HIMUS1, HIMUS2, HIMUS3, HIMUS4, HIMUS5, HIMUS6, HIMUS7, HIMUS8, HIMUS9 and HIMUS10). The phylogenetic tree was constructed using partial (895 bp) *rpoB* gene sequences for those ten strains. Evolutionary distances were calculated using the method of Maximum Composite Likelihood with 1000 bootstrap replicates using GENEIOUS software. The sequences presented a similarity percentage ranging between (98.76% and 100%) when compared with the sequences of standard strains in NCBI.

Keywords: *rpoB* gene, *Proteus spp.* Sequencing, Phylogenetic analysis.

Introduction

Sequencing of the 16S ribosomal RNA gene has been begun for investigating and identifying bacterial communities since the 1970s, which targets the 16S rRNA gene; the 16S rRNA gene is around ~1,550 base pairs (bp) long. It has highly conserved and incredibly variable regions among bacterial species. The highly conserved regions permit the design of "universal" primers in PCR to target and amplify the 16S rRNA sequence, while the variable regions allow for differentiating among diverse bacterial clades. These characters permit 16S rRNA sequencing investigations to capture nearly all the bacteria in a microbial community, which can then be compared to large 16SrRNA databases to determine their characteristics ¹.

Together with the 16S rRNA gene, the *rpoB* gene has helped to describe new bacterial species and upgrade bacterial community analysis. Furthermore, these housekeeping genes commonly exist as single copies in bacterial genomes, regulating the overestimation of operational taxonomical units (OTUs) in microbial groups ². The *rpoB* gene encodes the beta subunit of bacterial RNA polymerase. Its sequence is more distinguished than the 16S rRNA gene to

discriminate various species of bacteria owing to the higher divergence levels of the *rpoB* gene sequence than the 16S rRNA gene. Furthermore, the partial *rpoB* gene sequence exhibits the exact reading frame, leading to the easy confirmation of sequence accuracy ³.

The *rpoB* gene has been recognized as one of the little possible candidates suitable for bacterial phylogeny; thus, sequence analysis of the RNA polymerase β subunit has been suggested as a powerful tool for universal bacterial identification, and the usage of the *rpoB* gene presents many potential benefits above standard 16SrRNA gene-based methods ⁴.

The genus *Proteus*, belonging to the family Enterobacteriaceae, is a motile Gram-negative bacterium that survives in soil, water, garbage, rotten organic matter and the intestinal tracts of mammals. Most members of the genus *Proteus* are opportunistic pathogens that cause various infections in humans, including urinary tract infections, wounds, respiratory tract, skin, eye, ear, nose, and throat infections ⁵.

Materials

Bacterial isolates: Fifty bacterial isolates suspected to belong to *Proteus* spp. and have been collected from different clinical sources: stool, urine, ear, and wound.

Markers

Promega/USA provided DNA ladder (100-1500 bp).

Oligonucleotide Primers

The RPO B-specific primer F: AACCAGTTCGCGTTGGCCTGG R: CCTGAACAACACGCTCGGA was provided by Macrogen /Korea (Giammanco et al., 2011).

Standard Strain

Standard strain *Proteus mirabilis* ATCC 29245 was obtained from Al-Kindy Hospital/Baghdad.

Phylogenetic tree program analysis

GENEIOUS software

Methods

Identification of bacterial isolates

Isolation and identification of bacterial isolates were performed according to standard bacteriological techniques ⁶. All colonies suspected as *Proteus* in primary cultures were purified by subculture on different media. Species were identified phenotypically according to the morphological features on culture media, microscopic examination, and biochemical tests ^{7, 8}. VITEK 2 compact system was used to confirm identification.

Detection of *rpoB* gene

DNA extraction and *rpoB* amplification

According to the protocol of ABIO pure extraction, Genomic DNA was isolated from the bacterial growth of all bacterial isolates. DNA was quantified, and primer stock was prepared. PCR master was mixed, and primer solutions were stirred by vortex at room temperature. The components of the tubes were mixed by vortex and then placed in the PCR apparatus.

Gel electrophoresis

The integrity of extracted DNA was examined by 1% agarose gel electrophoresis and viewed using a UV trans illuminator. Gel electrophoresis was adopted to confirm the presence of amplification. The horizontal agarose gel was cast, and DNA was loaded.

Sequencing of rpoB gene

The PCR product from amplifying the rpoB gene for both the forward and reverse strands of (10) selected species was sent for Sanger Sequencing using an ABI3730XL automated DNA sequencer by Macrogen Corporation – Korea. The results were analyzed and edited by using GENEIOUS software. The edited sequence was compared with the database using BLAST (Basic Local Alignment Search Tool) to detect the close relation with submitted sequences.

Phylogenetic analyses based on rpoB gene sequences

Using the branch lengths contained in the indirect tree, the analysis involved (10) nucleotide sequences. The analysis was performed using GENEIOUS prime⁹ Following rpoB sequencing using the Sanger method.

Results

Identification of bacterial isolates

Several conventional biochemical tests were done to characterize the suspected Proteus isolates, and the results indicated that these isolates were primarily identified as Proteus spp. To confirm the identification results, the VITEK 2 Compact automated system was used; results showed that 49 isolates were identified as P. mirabilis (98%) that showed a confidence value of 99-96% (excellent identification) and only one isolate was identified as P. Hauser (2 %).

Detection of rpoB gene

The rpoB gene encodes the β subunit of bacterial RNA polymerase, also known as housekeeping genes. It has emerged as a core gene candidate for phylogenetic analyses and identification of bacteria, especially when studying closely related isolates. To detect the rpoB gene in Proteus isolates, DNA was extracted from all samples, and conventional PCR was carried out in order to amplify this gene by using a specific rpoB primer; when the two aligned together to form a visible band by gel electrophoresis, all of these isolates (100%) indicated by the presence of band with molecular size (1090 bp) when compared with ladder figure 1

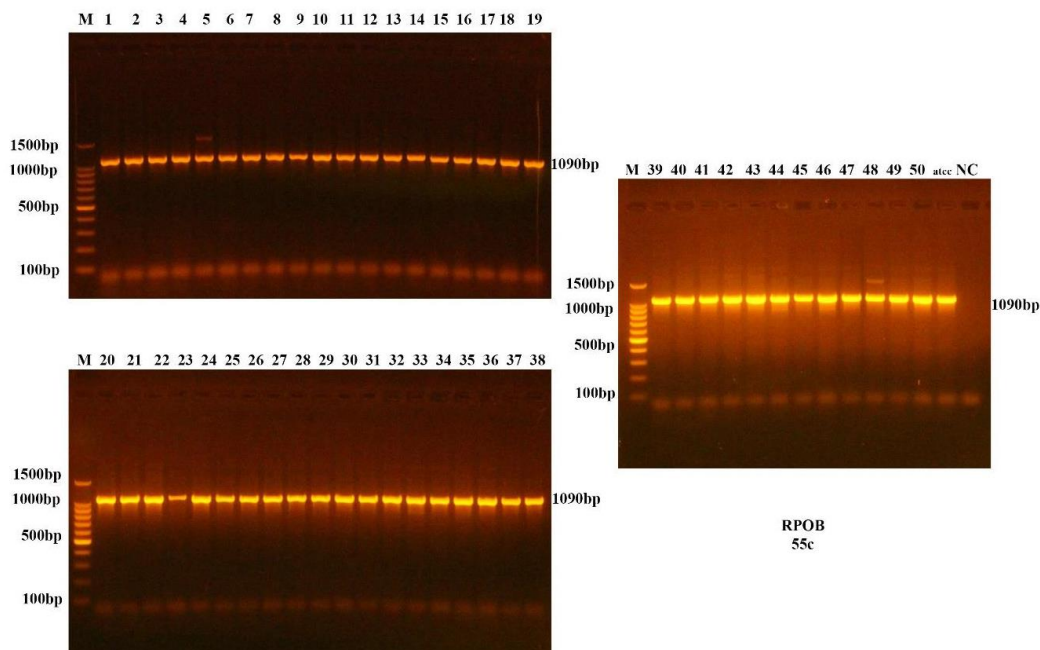


Figure 1. The amplification results of the *rpoB* gene of *Proteus spp.* Samples (1-51) were fractionated on 1.5% agarose gel electrophoresis and stained with Eth. Br. M: 100bp ladder marker.

Proteus mirabilis ATCC 29245

Sequence analysis of rpoB gene

Sequencing of *rpoB* gene using Sanger method has been performed for ten *Proteus* isolates and the sequences that obtained were deposited in the GenBank with the following accession no: LC699880, LC699881, LC699882, LC699883, LC699884, LC699885, LC699886, LC699887, LC699888, LC699889. The obtained DNA sequences were compared with standard GenBank sequences and the numerical chain maximum identification ratio. The sequences have been gave a similarity range between (98.76 % - 100%). According to *rpoB* gene sequences in GenBank, The ten strains have been recognized into two species, *P.mirabilis* and *P.terrae* Figure 2.

```

tgcattaatg ccacaagata tgatcaatgc aaaaccgatt tctgctgcag ttaaagagtt
ctttggctct agccagttat cgcagtttat ggatcagaat aaccgctgt cagaaattac
ccataaacgt cgtatttctg cattaggccc tgggtgctctg actcgtgaac gtgcaggctt
cgaagttcga gacgtacacc caactcacta tggctgtgtg tgtccaatcg aaacacctga
aggtccaaat atcggcttga ttaactcatt atctgtttat gcacagacta acgagtacgg
attcctagaa acgccttacc gtgttgtaga aaacaacgct gtaacagatg aaatccacta
tctgtctgca attgaagaag gtaacttcatt cattgctcag gctaactccg tattagatga
tgatggccac ttcgtagaag aattagtaac ttgccgtcat aaaggtgagt caagtttatt
tagtctgat caagttcagt acatggacgt ttaactcaa caggtgtttt ctgtcgggtgc
ttactgatc ccattccttg aacacgatga cgccaaccgt gcattgatgg gtgcgaacat
gcaacgtcag gctgttccta cacttcgtgg tgataaacca ctggtaggta caggtatgga
acgtgctgta gcggttgact ccggtgttac tgcggttagca aaacgtggtg gtactgttca
gtatgttgat gcttctcgtg tcgttatcaa agtaaacgaa gatgagactt atgctgtgga
agcaggcatt gatatttaca gcttaactaa atataccgc tctaaccaaa acacatgtat

```

taaccaaaca ccttggttt ctttaggtga acctgttgaa cgtggtgatg tact
 A
 tgcattgatg ccacaagata tgatcaacgc aaaaccgatt tctgctgctg tcaaagaatt
 ctttggtct agtcagttat cacagttat ggatcagaat aaccgctat cagaaattac
 gcacaaacgt cgtatttctg cattaggccc aggtggttg acccgtgaac gtgcaggctt
 tgaagttcga gacgtacacc caactcacta cggtcgtgtg tgcccaatcg aaacgcctga
 aggtccaaac atcggcttta tcaactcatt atctgtttat gcacagacta acgagtacgg
 tttcctagaa acgccttacc gtgtgttgaa aaacaacgct gtaacagatg aaattcacta
 tctgtctgca atcgaagaag gtaacttcatt cattgtcag gctaactccg tattagacga
 tgacaatcac tttgtgaag aattagttac ttgccgtcat aaaggtagt caagcttatt
 tagtcgtgac caagttcagt acatggacgt ttcaacccaa caggtgttt ctgctggtgc
 ttcactgatc ccattccttg aacacgatga cgccaaccgt gcattgatgg gtgcgaacat
 gcaacgtcag gctgttccta cacttcgtgg tgataaacca ctagtaggta caggtatgga
 acgtgcagta gcggttgact ccggtgttac tgcggttgct aaacgtggtg gtgtgttca
 gtatgttgat gcttctcgta tcgttatcaa agtaaacgaa gaagagactt acgctggtga
 agcaggcatt gatatttaca gcttgactaa atacaccgc ttaacaaa acacatgtat
 taaccaaact ccttggttt ctttaggtga acctgttgaa cgtggtgatg tactt
 B

Figure 2. Sequences of rpoB gene: A- Proteus mirabilis HIMUS1 B- Proteus terrae HIMUS7.

3.4. Phylogenetic analysis based on rpoB gene sequences

Using the branch lengths contained in the indirect tree, the analysis involved

(10) nucleotide sequences. Following rpoB sequencing using the Sanger method, The obtained DNA sequences were compared with other sequences in GenBank, also with numerical chain and maximum identification ratio. The sequences have been gave

a similarity range between (98.76 % - 100%).

According to rpoB gene sequences in GenBank. The ten strains have been recognized into two species, P.mirabilis and P.terrae.

In the current study, the phylogenetic tree of Proteus spp. was generated to show the relationships and comparison between our sequence data and previously published data from NCBI (GenBank). The rpoB gene for these isolates has been deposited and registered in the National Center for Biotechnology Information (NCBI) with the accession numbers. These strains named as (HIMUS1, HIMUS2, HIMUS3, HIMUS, HIUS5, HIMUS6, HIMUS7, HIMUS8, HIMUS9 and HIMUS10).

The sequences obtained in this study were deposited in the GenBank with the following accession no: LC699880, LC699881, LC699882, LC699883, LC699884, LC699885, LC699886, LC699887, LC699888, LC699889.

The PCR products were approximately 1090 kb in size, corresponding to this group of bacteria, after entering and aligning these sequences of rpoB. Furthermore, by GENEIOUS software, phylogenetic trees were generated in Figure 3.

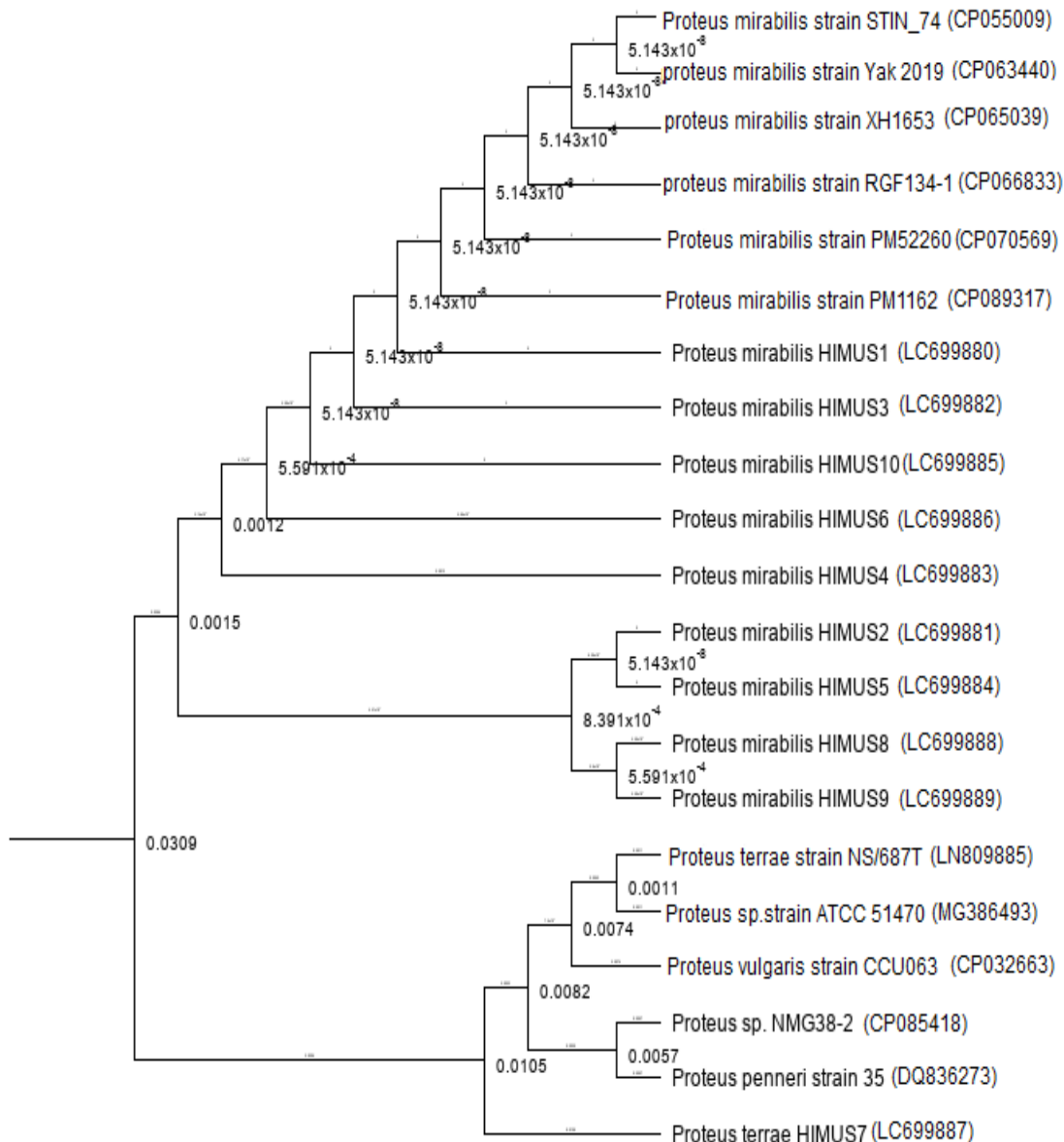


Figure 3. Phylogenetic analysis of partial nucleotide sequences of the *rpoB* gene of reference and clinical strains of *Proteus* spp. Evolutionary distances were calculated using the method of Maximum Composite Likelihood with 1000 bootstrap replicates using GENEIOUS software.

The extent of difference between homologous DNA sequences in different strains is used to measure how much these bacterial genera or species have diverged from one another evolutionarily. The sequence should be long enough to provide information. Every cell has one copy of the *rpoB* gene in bacteria.

The results of phylogenetic analysis of *Proteus* spp. Strains have been shown to identify percentages ranging between 93.6% - 100% Table 1. The highest percentage of similarity was between *rpoB* sequences of *P.mirabilis* HIMUS1, HIMUS2, HIMUS5 strains and standard strains *P.mirabilis*; STIN_74, Yak2019, XH1653, RGF143-1, PM52260, PM1162 while the lowest identity of sequences was between *P.mirabilis* strains HIMUS2, HIMUS5, HIMUS6, HIMUS8 and standard strain *Proteus* sp.NMG38-2 .All strains of our study were almost grouped in two branches more closely related to *P. mirabilis* reference strains installed from the GenBank.

One branch (group 1) included 9 new strains: *P.mirabilis* HIMUS1,2,3,4,5,6,8,9,10 Figure 3.

The other branch (group 2) differs from the first, including *P.terrae* HIMUS7, and there are differences between *rpoB* gene sequences between the two branches and between the same branch.

	CP032663	CP055009	CP063440	CP065039	CP066833	CP070569	CP085481	CP089317	DQ836273	LC699880	LC699881	LC699882	LC699883	LC699884	LC699885	LC699886	LC699887	LC699888	LC699889	LN809885	MG386493
CP032663		94.1%	94.1%	94.1%	94.1%	94.1%	96.5%	94.1%	97.9%	94.1%	93.9%	94.1%	94.1%	93.9%	94.1%	93.9%	97.8%	93.9%	94.1%	98.4%	98.6%
CP055009	94.1%		100%	100%	100%	100%	93.7%	100%	93.9%	100%	99.7%	100%	99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
CP063440	94.1%	100%		100%	100%	100%	93.7%	100%	93.9%	100%	99.7%	100%	99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
CP065039	94.1%	100%	100%		100%	100%	93.7%	100%	93.9%	100%	99.7%	100%	99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
CP066833	94.1%	100%	100%	100%		100%	93.7%	100%	93.9%	100%	99.7%	100%	99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
CP070569	94.1%	100%	100%	100%	100%		93.7%	100%	93.9%	100%	99.7%	100%	99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
CP085481	98.5%	93.7%	93.7%	93.7%	93.7%	93.7%		93.7%	98.9%	93.7%	93.6%	93.7%	93.7%	93.6%	93.7%	93.6%	98.0%	93.6%	93.7%	98.6%	98.9%
CP089317	94.1%	100%	100%	100%	100%	100%	93.7%		93.9%	100%	99.7%	100%	99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
DQ836273	97.9%	93.9%	93.9%	93.9%	93.9%	93.9%	98.9%	93.9%		93.9%	93.8%	93.9%	93.9%	93.8%	93.9%	93.8%	97.8%	93.8%	93.9%	98.1%	98.3%
LC699880	94.1%	100%	100%	100%	100%	100%	93.7%	100%	93.9%		99.7%	100%	99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
LC699881	93.9%	99.7%	99.7%	99.7%	99.7%	99.7%	93.6%	99.7%	93.8%	99.7%		99.7%	99.7%	100%	99.7%	99.8%	94.7%	99.8%	99.9%	93.9%	94.1%
LC699882	94.1%	100%	100%	100%	100%	100%	93.7%	100%	93.9%	100%	99.7%		99.8%	99.7%	100%	99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
LC699883	94.1%	99.8%	99.8%	99.8%	99.8%	99.8%	93.7%	99.8%	93.9%	99.8%	99.7%	99.8%		99.7%	99.8%	99.7%	94.9%	99.7%	99.8%	94.0%	94.3%
LC699884	93.9%	99.7%	99.7%	99.7%	99.7%	99.7%	93.6%	99.7%	93.8%	99.7%	100%	99.7%	99.7%		99.7%	99.8%	94.7%	99.8%	99.9%	93.9%	94.1%
LC699885	94.1%	100%	100%	100%	100%	100%	93.7%	100%	93.9%	100%	99.7%	100%	99.8%	99.7%		99.9%	94.9%	99.7%	99.8%	94.0%	94.3%
LC699886	93.9%	99.9%	99.9%	99.9%	99.9%	99.9%	93.6%	99.9%	93.8%	99.9%	99.8%	99.9%	99.7%	99.8%	99.9%		94.7%	99.8%	99.9%	93.9%	94.1%
LC699887	97.8%	94.9%	94.9%	94.9%	94.9%	94.9%	98.0%	94.9%	97.8%	94.9%	94.7%	94.9%	94.9%	94.7%	94.9%	94.7%		94.7%	94.9%	98.1%	98.1%
LC699888	93.9%	99.7%	99.7%	99.7%	99.7%	99.7%	93.6%	99.7%	93.8%	99.7%	99.8%	99.7%	99.7%	99.8%	99.7%	99.8%	94.7%		99.9%	93.9%	94.1%
LC699889	94.1%	99.8%	99.8%	99.8%	99.8%	99.8%	93.7%	99.8%	93.9%	99.8%	99.9%	99.8%	99.8%	99.9%	99.8%	99.9%	94.9%	99.9%		94.0%	94.3%
LN809885	98.4%	94.0%	94.0%	94.0%	94.0%	94.0%	98.6%	94.0%	98.1%	94.0%	93.9%	94.0%	94.0%	93.9%	94.0%	93.9%	98.1%	93.9%	94.0%		99.8%
MG386493	98.6%	94.3%	94.3%	94.3%	94.3%	94.3%	98.9%	94.3%	98.3%	94.3%	94.1%	94.3%	94.3%	94.1%	94.3%	94.1%	98.1%	94.1%	94.3%	99.8%	

Table 1. Similarity ratio in sequences of *rpoB* gene between *Proteus* strains. In the current study and reference strain. Strains in this table were prepared by Sanger sequencing using ABI3730XL, automated DNA sequences, by GENEIOUS software.

In this table, it was observed that the difference between the *rpoB* gene sequence of *P.mirabilis* strains ranged between (0.01 – 0.03 %) and the difference between the *rpoB* gene sequence of *P. mirabilis* and *P. terrae* (5.1 – 5.3 %). This corresponds with ¹², who established similarities and differences between *rpoB* gene sequences of *Proteus* spp.

Discussion

Confirmation of the bacterial identification was achieved by molecular genotypic identification. Sequence analysis of the RNA polymerase β subunit encoding gene *rpoB* has been proposed as a novel tool for bacterial identification. Also, the partial *rpoB* gene sequence displays the exact reading frame, leading to the easy verification of sequence accuracy ¹⁰.

Strain no. 36, which was identified as *P. Hauser* by VITEK 2; it appeared that this isolate belonged to the species *P. terrae* after conducting a gene sequencing for bacterial isolates under study because the VITEK 2 compact system could not detect the species *P.terrae* according to the instructions for the device that is identifying only four species; *P.mirabilis*, *P. vulgaris*, *P. penner* and *P Hauser* ¹¹. The analysis was performed using GENEIOUS prime ⁹.

Conclusion

Partial sequencing of the *rpoB* gene assists in detecting genetic similarities and differences among *Proteus* spp. A phylogenetic tree is essential in understanding

and determining the evolutionary relationships between bacterial isolates and developing new treatments.

Acknowledgments

Thanks to the bacteriology labs. Of Al-Kindi Teaching Hospital and some other hospitals in Baghdad and Dyala for their assistance in providing the samples. Also, we thank the Biology Department / College of Science / Mustansiriyah University (<http://uomustansiriyah.edu.iq>) / Baghdad for advice and support.

References

- 1 Weinroth, M. D.; Belk, A. D.; Dean, C.; Noyes, N.; Dittoe, D. K.; Rothrock Jr, M. J., ... & Wells, J. E. Considerations and best practices in animal science 16S. ribosomal RNA gene sequencing microbiome studies. *Journal of Animal Science*, **2022**; 100(2), skab346.
- 2 Ogier, J. C.; Pages, S.; Galan, M.; Barret, M., & Gaudriault, S. rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC microbiology*, **2019**; 19(1), 1-16
- 3 Rungsirivanich, P.; Inta, A.; Tragoolpua, Y., & Thongwai, N. Partial rpoB gene sequencing identification and probiotic potential of *Floricoccus penangensis*. ML061-4 isolated from Assam tea (*Camellia sinensis* var. *assamica*). *Scientific Reports*, **2019**; 9(1), 1-12.
- 4 Jamel, A. N.; Allami, R. H., & Hamza, S. J. Molecular Study Of RpoB Gene In *Proteus mirabilis* Isolated From Urinary Tract Infection From Different Hospitals In Baghdad. *Plant Archives*, **2020**; 20(1), 2379.2383
- 5 Zheng, X.; Ma, J.; Lu, Y.; Sun, D.; Yang, H.; Xia, F., & Tang, B. Detection of tet (X6) variant-producing *Proteus terrae* subsp. *cibarius* from animal cecum in Zhejiang, China. *Journal of Global Antimicrobial Resistance*, **2022**; 29, 124-130.
- 6 Pathirana, H. N. K. S.; De Silva, B. C. J.; Wimalasena, S. H. M. P.; Hossain, S., & Heo, G. J. Comparison of virulence genes in *Proteus* species isolated from human and pet turtle. *Iranian journal of veterinary research*, **2018**; 19(1), 48.
- 7 Forbes, B. A.; Daniel, F. S., and Alice, S. W. B. Scott's. *Diagnostic Microbiology*. 12th ed., Mosby Elsevier Company. **2007**; USA pp62-465.
- 8 Brooks, G. F.; Carroll, K. C.; Butel, J. S.; Morse, S. A., and Mietzner, T. A. *Medical Microbiology*. **2013**; 26th edit. pp46-48
- 9 Abou-Shaara, H. F. How can the Sequence Alignment Method Affect the Phylogenetic Relationship Results? *Scientific Papers: Animal Science & Biotechnologies/Lucrari Stiintifice: Zootehnie si Biotehnologie*, **2020**; 53(1).
- 10 Lin, J. N.; Lai, C. H.; Yang, C. H., & Huang, Y. H. *Elizabethkingia* infections in humans: from genomics to clinics. *Microorganisms*, **2019**; 7(9), 295.
- 11 BioMérieux. VITEK 2 Identification Cards | bioMérieux United Kingdom & Ireland[Internet]. Available from: <https://www.biomerieux.co.uk/product/vitek-2-identification-cards>. **2021**.
- 12 El-Sokkary, M. M. A. Genotyping of New and Old *Proteus mirabilis* Isolates. from Mansoura Hospitals in Egypt by rpoB Sequence Analysis. *Advances in Microbiology*, **2015**, 5(07), 549..

Received: May 15, 2023/ Accepted: June 10, 2023 / Published: June 15, 2023

Citation: Mutlag , H.A.; Khelkal, I. N. Phylogenetic tree of *Proteus* spp. Based on partial rpoB gene sequence analysis. *Revis Bionatura* 2023;8 (1) 73. <http://dx.doi.org/10.21931/RB/CSS/2023.08.01.73>