P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2024,(18), (01):13 – 25

Comparative study on 16S rRNA gene sequences for clinical and environmental isolates of *legionella pneumophila*

Mohammed Ibrahim Khalid , Ibrahim A. Abdulrahman

Department of Biology ,College of Science, University of Anbar , Anbar,Iraq;
University of Fallujah / Headquarter; Fallujah , Iraq;

ARTICLE INFO

Received: 28 / 08 /2023
Accepted: 25 / 09/ 2023
Available online: 20/ 06 / 2024

DOI: 10.37652/juaps.2023.142996.112

Keywords:
Legionella pneumophila, 16S rRNA gene sequence, phylogenetic tree.

ABSTRACT

The objective of this study was to determine and describe the identity and phylogenetic position of two isolates of *Legionella pneumophila*, one clinical (assigned M1) was isolated from a patient with pneumonia symptoms, and one environmental (assigned H1) was isolated from the surface of potable water pipe close-ended, both isolates were obtained from Falluja city, Iraq. One genetic locus of the 16S rRNA gene sequences was included in the present study to evaluate the variation in genetic patterns for each of these ribosomal sequences. Direct sequencing was performed on the amplified segments in order to determine the pattern of genetic variability present in the bacterial samples that were obtained. Subsequently, a detailed and accurate tree was constructed in order to investigate the validity of the observed variations' phylogenetic placements and the discrimination that was made between them. Our results revealed the presence of two nucleic acid variants (141A>G and 247T>C) that were not found in this reference sequence. The inference was made from the analysis of the tree that the investigated M1 sample occupied a distinct phylogenetic position and they suited in the vicinity of several strains that were deposited from European sources. It was also inferred that the M1 sample was derived from the other incorporated strains – including H1 samples – within the same tree. Accordingly, the nucleic acid variants that were observed in the M1 sample were discovered to exert a remarkable effect on changing the phylogenetic positioning of the investigated M1 sample within the incorporated clades of *L. pneumophila*. The positioning of H1 samples in the vicinity of many non-clinical strains indicated a high ability for the utilized ribosomal RNA to identify the actual isolation source from which they were collected. The current study found the concerned capacity of 16S rRNA sequences to be able to identify the currently investigated sequences of *L. pneumophila* in the investigated samples. The utilization of this genetic locus can be extended to detect the phylogenetic distributions of the other types of bacterial sources. These potentially valuable amplicons are also able to be investigated to discover additional details through the identified isolates in other bacterial instances obtained from the researched sources.

INTRODUCTION

Legionella pneumophila is a gram-negative, aerobic bacteria in the Gamma proteobacteria order. They survive and thrive in protozoan hosts, especially amoebae, and produce biofilms in naturally aquatic (rivers and lakes) and manmade (cooling towers, spas, hot springs, and bath settings) waters[1]. Legionnaires' disease (LD) was first documented in 1976, and *L. pneumophila* is still responsible for over 95% of the cases [2,3]. LD occurs mainly by inhaling aerosols or aspirating water containing *Legionella* [4].

Legionellosis, a result of *Legionella* bacterium infection, involves both LD and the flu-like disease known as Pontiac Fever[5]. LD is an increasingly prevalent atypical pneumonia that is possibly acquired in the community, through travel, or nosocomial infection. Besides being older, having a compromised immune system, or having a chronic lung disorder, both current and former smokers are at higher risk for LD [6].

Among the most broadly utilized current microbial profiling approaches is the sequencing of a very significant and useful bacterial gene, known as the 16S rRNA gene [7]. The 16S rRNA sequence possesses a significant value in the identification of microbes due

*Corresponding author at: Department of Biology ,College of Science, University of Anbar , Anbar,Iraq
ORCID:https:// https://orcid.org/0000-0002-2026-8114 ,
Tel: +9647713435440
Email: mohammedalani23@gmail.com

to its well-characterized and strongly conserved organization along with the amount that exists in cells that are metabolically active [8]. Microorganism genotypic identification using the sequencing of 16S rRNA genes has evolved as an increasingly accurate, objective, and dependable approach to microbiological identification, with the additional function of determining taxonomic relationships, especially among bacteria[9]. The 16S rRNA gene sequence is extensively employed as a molecular marker in bacterial phylogenetic investigations[10]

Genomes encode rRNA molecules, which play a critical role in the translation of mRNAs, or messenger RNAs, through the ribosome. The 16S rDNA coding is responsible for the 16S rRNA, which, together with many proteins, makes up the 30S small subunit of ribosomes in bacteria[11]. The 16S rDNA gene is approximately 1500 nucleotides in length and possesses conserved regions, indicating that the sequence of nucleotides in the conserved regions is similar across the majority of bacteria, even those with drastically varying genome compositions and phenotypic features. It also has nine regions, which vary significantly with nucleotide conformation among the bacterial species[12]

The 16S rDNA gene is thought to have evolved principally in response to its ability to function inside the ribosomal translating machinery. When a nucleotide sequence's alignment shows a difference at a certain place (such as a transition or transversion), it is thought that such nucleotide differences produce alternative secondary or tertiary structures, which influence translation[13].

Investigating the genetic polymorphism of *Legionella pneumophila* can provide insights into its epidemiology, transmission patterns, and potential sources of infection [14] This knowledge is crucial for developing effective strategies to control and prevent outbreaks of *Legionella*-associated diseases. Identifying the genetic polymorphism of *L. pneumophila* can aid in tracing the sources of infection and identifying specific strains responsible for outbreaks[15]. By analyzing the genetic fingerprints of different isolates, researchers can determine whether certain strains are more prevalent in certain regions or environments, helping in the identification of potential reservoirs of the bacterium. As well, investigating genetic polymorphism can contribute to the development of diagnostic methods[16]. Despite the importance of genetic variation and its impact on the clinical and environmental aspects, there is a lack of studies that reported to describe such sort of association. Due to the lacking of studies that investigated the genetic polymorphisms of the clinical and environmental isolates of *L. pneumophila* in Iraq, this study aimed to identify and characterize both identity as well as phylogenetic position (a clinical sample) and *L. pneumophila* isolated from a plumping water system (an environmental sample) in Falluja city, Iraq. to provide insights into its epidemiology, transmission dynamics, diagnostic methods, and potential interventions, ultimately contributing to the improved management and control of *Legionella*-related diseases in the region.

METHODS

2.1 PCR assay: A conventional PCR technique was performed for the detection of 16s rRNA genes. The specific *L. pneumophila* 16s rRNA gene primers sequences were as follows; F-16srRNA: 5'AGGGTTGATAGGTTAAGAGC-3', R-16srRNA: 5'-CCAACAGCTAGTTGACATCG-3'. DNA amplification was performed in a thermocycler (Thermo Fisher Scientific, USA) under conditions of the initial denaturation at 95°C for 5 minutes, followed by 30 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, extension at 72°C for 1 minute and a final step of extension at 72°C for 7 minutes.

2.2. The sequencing of nucleic acid PCR amplicons

The resolved PCR amplicons were subjected to forward commercial sequencing, following the protocols provided by Macrogen Inc. Geumchen, Seoul, South Korea. The next analysis specifically concentrated on clear chromatographs obtained from ABI (Applied Biosystem) sequenced documents, in order to minimize the potential impact of PCR or sequencing artifacts on the annotation and variations. The process of identifying virtual locations and obtaining information about polymerase chain reaction (PCR) segments and their characteristics involved comparing the nucleic acid sequences of the bacterial samples under study with the retrieved nucleic acid sequences.

2.3The interpretation of the sequencing data

The targeted PCR products were subjected to sequencing, and the resulting data was edited, aligned, and analyzed alongside the corresponding sequences in the databases. The analysis was performed utilizing BioEdit Sequence Alignment Editor Software Version 7.1 (DNASTAR, Madison, WI, USA). The discovered discrepancies in every sequenced sample were assigned numerical values in both PCR amplicons and their respective positions within the reference genome. The nucleic acids that were observed were assigned numerical values in both the PCR amplicons and their respective places within the referenced genome. Snap Gene Viewer version 4.0.4 was utilized to annotate each identified variant within the bacterial sequences (https://www.snapgene.com).

2.4. Deposition of sequences to GenBank

All the newly investigated and evaluated sequences were submitted to the NCBI Bank portal, and every one of the requirements outlined by the portal was adhered to in accordance with the server's guidelines [17]. The submitted sequence was provided as nucleic acid sequences in the NCBI to get a unique GenBank accession number for the investigated sequences.

2.5. The construction of a comprehensive phylogenetic tree

The present study involved the construction of a phylogenetic tree utilizing the neighbor-joining approach. The identified variations were compared to their adjacent homologous reference sequences using the NCBI-BLASTn server [18]. The resulting tree was then visualized as a circular cladogram utilizing the iTOL suite [19]The sequences of all the species included in the comprehensive tree were assigned distinct colors to facilitate differentiation from the other species.

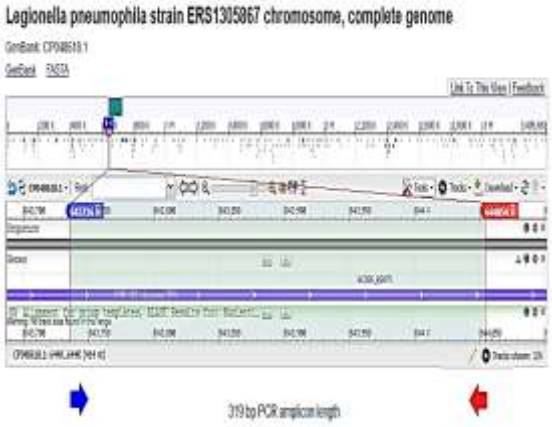
3. RESULTS AND DISCUSSION

The present investigation contained a total of two samples, which were designated as H1 and M1. The provided samples underwent screening in order to partially amplify the 16S rRNA sequences. Thus, the variation of these ribosomal sequences can be used for the description of these bacterial species due to the possible ability of rRNA sequences to adapt to variable genetic diversity. With regard to the amplicons of H1 amplicons. The PCR amplicons were subjected to NCBI blastn analysis, which revealed the exact matching of the sequences.[18].

The NCBI BLASTn engine showed an entire homology between the sequenced sample of H1 and *L. pneumophila* reference target sequences (GenBank acc. CP048618.1), The precise positions and the cumulative size of the specific loci were determined using the NCBI server, and the precise coordinates of the loci's beginning and end were verified based on the closest bacterial target with the highest degree of similarity. (Fig. 1a).

The NCBI BLASTn algorithm demonstrated a homology of up to 99% between the M1 sample and the reference target DNA sequences of *L. pneumophila* (GenBank acc. KF537568.1). The specific positions, supplementary details, and dimensions of the particular spots of interest were also determined on the NCBI server, and the precise locations of the targeted region were confirmed within the most closely related bacterial reference. (Fig. 1b).

A) H1



B) M1

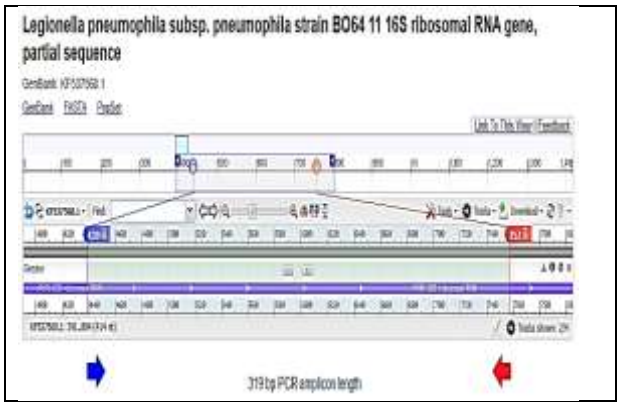


Fig. 1. The exact positions of the investigated samples covered a portion of the 16S rRNA sequences within *Legionella pneumophila*. Branch A) and B) refer to the most homologous reference sequences to H1 and M1 samples, respectively.

While the alignment of the H1 sample with its reference sequences showed an entire homology (GenBank acc. no. CP048618.1) (Fig. 2a), the alignment results of the M1 sample revealed the presence of two nucleic acid variations in comparison with the most similar referring reference nucleic acid sequences (GenBank acc. KF537568.1) (Fig. 2b).

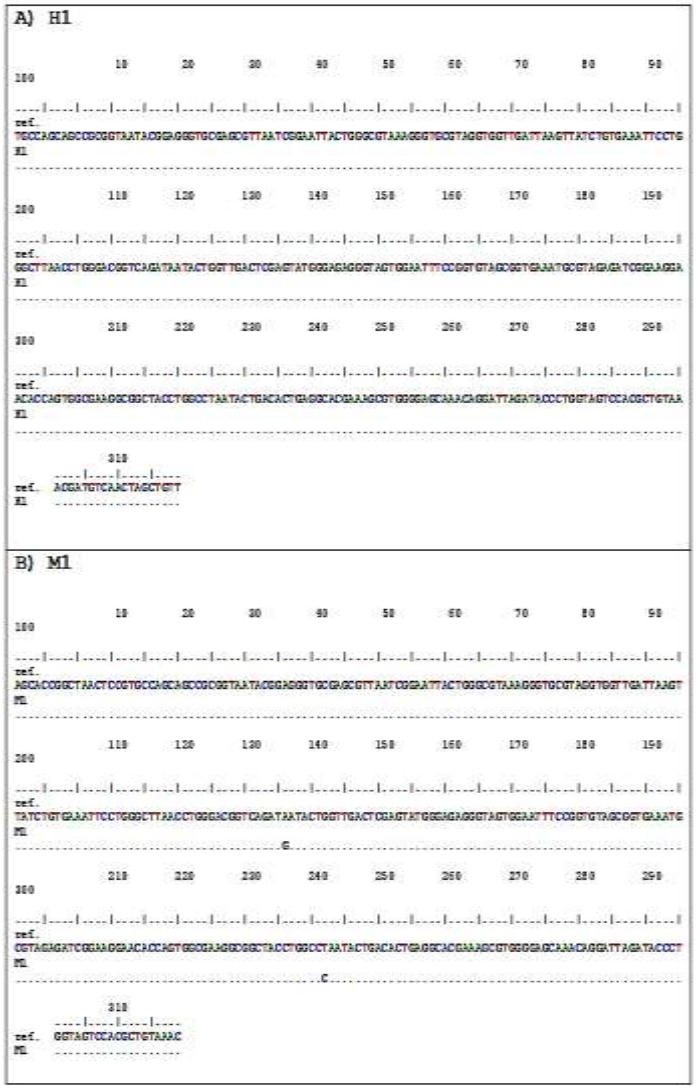


Fig. 2. Nucleic acid sequences alignment of two samples of *Legionella pneumophila* with their corresponding reference sequences (GenBank acc. no. CP048618.1 for H1 sample in branch A, and GenBank acc. no. KF537568.1 for M1 sample in branch B). The symbol "ref" refers to the NCBI referring sequence.

To confirm the identified variations of this study, the DNA sequence chromatograms of the analyzed samples were carefully examined and recorded, together with comprehensive annotations. The chromatograms depicting the sequences were presented in accordance with their respective positions within the PCR amplicons for both *L. pneumophila* isolates. The presence of each of these variants was confirmed in its original chromatogram and the absence of any possible technical error was also confirmed. Whereas all other non-confirmed nucleic acid variations were eliminated from any further consideration. The sequencing reaction indicated that the H1 samples did not exhibit any

variation with its reference sequences (Fig. 3a). Whereas the sequencing results confirmed the presence of two nucleic acid variants (141A>G and 247T>C) identified in the M1 sample compared with the reference sequences of *L. pneumophila* (GenBank acc. no. KF537568.1) (Fig. 3b).

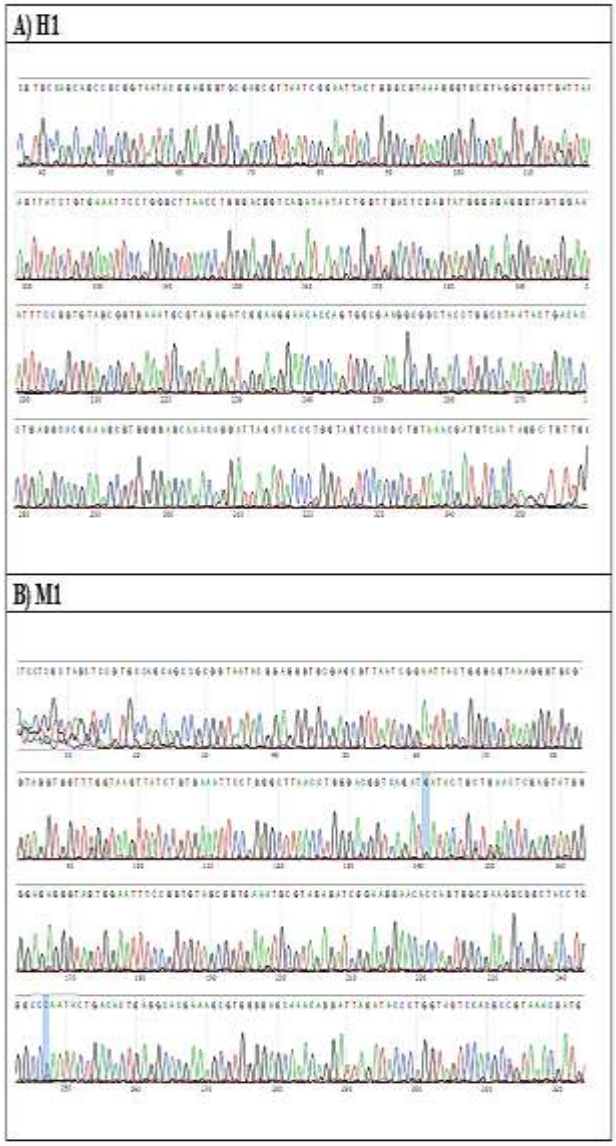


Fig. 3. The chromatograms of the investigated bacterial samples of *Legionella pneumophila*. Branch A) and B) refer to the chromatogram sequences of H1 and M1 samples, respectively. The blue highlighted peaks indicated the positions in which nucleic acid polymorphism was detected.

These differences observed in the currently analyzed M1 sample were not found in the corresponding reference sequences with which these sequences were compared (GenBank acc. no. KF537568.1). This sort of distribution was related to the different sources from which both bacterial samples were isolated. Both ribosomal sequences were deposited in the NCBI web server, and unique accession numbers were obtained for the analyzed sequences. GenBank OQ832595.1 and OQ609611.1 were deposited in NBCI to represent the H1 and M1 samples, respectively.

16S rRNA sequencing is a widely used molecular technique for identifying *Legionella* at the species level, based on the sequence variations in the 16S rRNA gene, which is present in all bacterial genomes. Many studies have used 16S rRNA sequencing to identify the presence of *L.* pneumophila in clinical and non-clinical samples, as well as to identify genetic variations within *L. pneumophila* populations. There are several examples available to demonstrate its importance for the identification and characterization purposes of *L. pneumophila*. It has been found that partial 16S rRNA gene sequencing can accurately identify all *L. pneumophila* isolates and classify non-*pneumophila Legionella* isolates to the species level [20]. Moreover, a PCR-based method that targets single-nucleotide polymorphisms (SNPs) within the 16S rRNA gene proves effective in detecting and differentiating *L. pneumophila* and non-*L. pneumophila* strains. This method relies on the presence of specific genetic variations to accurately identify and classify different strains of *L. pneumophila*. The same study reported that *Legionella spp*. can be isolated and identified through in vitro culture and the utilization of multilocus sequence analysis targeting 16S rRNA sequences. It was deduced that direct sequencing of the amplicon produced by the 16S rRNA gene primers can be performed to identify non-*L. pneumophila* species of *legionella* [21]. Another study used fluorescent resonance energy transfer probes targeting the 16S rRNA gene to construct a sensitive and specific real-time PCR for the identification and differentiation of *L. pneumophila* from other *Legionella spp.* Accordingly, it can be stated that the utilization of 16S rRNA sequencing offers several advantages in the identification of *L. pneumophila*. Firstly, it enables the accurate identification of all *L. pneumophila* isolates [22]. By targeting specific regions within the 16S rRNA gene, this sequencing method provides a reliable means to differentiate and confirm the presence of *L. pneumophila* strains[23]. Furthermore, 16S rRNA

sequencing facilitates the rapid classification of non-*pneumophila Legionella* isolates to the species level[24]. While 16S rRNA sequencing is a valuable tool for identifying bacterial species and studying genetic diversity, it has certain limitations. It may not provide sufficient resolution for distinguishing closely related strains or differentiating strains within a single species[25]. In such cases, additional genomic techniques, such as whole-genome sequencing, may be necessary for a more comprehensive analysis of *L. pneumophila* populations [26,27].

In order to provide a comprehensive understanding of the phylogenetic distances between the bacterial samples obtained from two distinct sources, this study generated a phylogenetic tree based on the sequences of nucleic acids observed through the amplified PCR products of the ribosomal amplicons of *L. pneumophila*. Phylogenetic trees can be used to analyze the genetic diversity of *L. pneumophila* taken from different sources[28,29]. Phylogenetic trees are graphical representations of the evolutionary relationships among organisms based on their genetic information. In the case of *L. pneumophila*, genetic diversity can be analyzed by sequencing specific regions of the bacterial genome, such as ribosomal sequences. These sequences can then be used to construct phylogenetic trees that depict the relatedness and evolutionary history of different *L. pneumophila* strains with high efficiency [30]. The phylogenetic analysis allows for the classification of *L. pneumophila* isolates into distinct clades, which can provide information about the spread and transmission of specific strains. Furthermore, they can also assess the evolutionary importance of the identified genetic variations among different *L. pneumophila* isolates. Additionally, the trees can help identify potential outbreak strains or closely related strains that may have emerged from a common ancestor. The analysis of genetic diversity using phylogenetic trees can provide valuable information for understanding the epidemiology, transmission dynamics, and evolution of *L. pneumophila* [31]. It can aid in tracking the sources and routes of infection, identifying potential environmental reservoirs, and informing public health interventions and preventive measures.

This study constructed a phylogenetic tree based on nucleic acid variations found in the isolates that formed distinct clusters based on their source. The clinical isolates were more closely related to each other than to isolates from non-clinical sources, suggesting that they shared a common ancestry and may have evolved in response to selective pressures unique to the environment.

Due to the ability of *Legionella pneumophila* isolates to exhibit significant genetic diversity as reflected in their placement within phylogenetic trees, the generation of a tree could provide several replies for such biological distribution. This genetic diversity may be driven by different selective pressures in different environments, including host immune defenses and exposure to antibiotics. Understanding the genetic diversity of *L. pneumophila* isolates obtained from clinical and non-clinical sources can also be important for guiding the development of effective treatments and infection control strategies. Thus, a phylogenetic tree was generated to incorporate these two investigated samples alongside other relative nucleic acid sequences of the referring sequences of *L. pneumophila* to assess the possible role of the isolation sources in their phylogenetic distribution among the other related strains.

The current constructed tree was represented in two cladograms, which were made to generate a rectangular cladogram and curved to generate a circular cladogram (Fig. 4a and b). In each form, a particular phylogenetic distribution of the incorporated sequences was notified. The samples that were the subject of the current investigation were aligned with the other sequences that are closely related, resulting in the inclusion of these sequences in the current cladogram. The complete tree had a total of thirty-six aligned nucleic acid sequences. The samples included in the produced cladogram were grouped into two primary phylogenetic clades within the species *L. pneumophila*. Significantly, the substantial phylogenetic positioning seen between clade-1 and clade-2 indicates a considerable phylogenetic divergence between these two clades. Clade 1 was situated at a distance from clade-2. The divergence of the two clades depicted in the phylogenetic tree can be attributed to dissimilarities in the respective ancestral origins from which the samples

were derived. The first clade was made of our investigated H1 sample positioned with the other incorporated samples of *L. pneumophila* sequences to constitute the major clade with thirty-two organisms (clade-1). The organisms incorporated within this clade were made of variable strains of *L. pneumophila* sequences that were deposited from variable sources worldwide, such as hot spring water in Tunisia (GenBank JQ478489.1), water from hotel pumping system in Greece (GenBank JX827099.1), fresh water in USA (GenBank NR_074231.1, and CP015950.1), faucet water biofilm in Palestine (GenBank KF537568.1, KX778106.1), the river of Taiwan (GenBank KY451032.1), urban air samples in Spain (GenBank JX827099.1, and MH412932.1), environmental water samples in Japan (GenBank LC491286.1, and LC491286.1), and water from the shower head in the UK (GenBank LR134380.1). The positioning of the investigated H1 sample beside these strains has added another layer of confirmation for the environmental source of this sample. As well, the currently investigated sequences of the M1 sample had shown no phylogenetic separation from the other strains in the clade-1 due to the absence of any identified polymorphism. Accordingly, it can be stated that this positioning gives an accurate description of the exact environmental source from which this strain is obtained.

Apart from this clade, our investigated M1 sample was positioned beside three strains within clade-2. This sample occupied a phylogenetic position with a further distance. The reason for this faraway position was attributed to the existence of two nucleic acid variations (141A>G and 247T>C) in the M1 sample, which is the main reason behind this distribution. Due to these variations, the M1 sample occupied unique phylogenetic positions within the clade-2. Due to the isolation of the M1 sample from sputum, the observed differences in the distribution of nucleic acid variations may also be associated with the clinical source from which each sample was isolated. Noteworthy, two strains that occupied the nearest phylogenetic position to our samples were deposited from Italy (GenBank FR799701.1 and FR799702.1). However, the source from which these strains were isolated was not shown. In a relative vicinity to these strains, one French strain was also positioned (GenBank FQ958211.1). As in the case of the other two Italian strains, no source of isolation was shown for the French sample. In both cases, the European source of this sample is indicated.

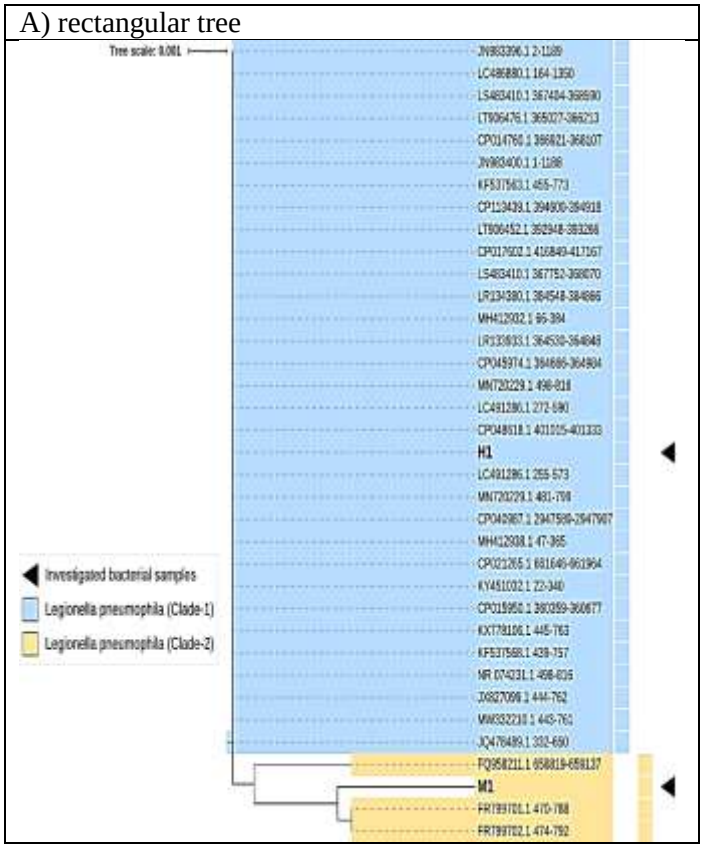
Based on the identified ribosomal variations, it might be possible that the clinical M1 sample has developed a new strategy to avoid the host immune defense. One important aspect of *Legionella pneumophila* infections is the bacterium's capacity to withstand host defenses and antimicrobial agents. Accordingly, these bacterial organisms can develop special mechanisms to resist host immunity that differ according to the isolation source from which it was isolated. Many studies have investigated the genetic basis of this resistance and have found that the isolation source of *L. pneumophila* may be associated with the genetic variation this organism takes to resist the host. For example, a study conducted revealed that *L. pneumophila* possesses genetic characteristics that are associated with disease and offer a biological foundation for its ability to resist complement-mediated killing [32]. A further investigation revealed that *L. pneumophila* possesses the ability to impede the proliferation of other *Legionella* species through the use of a secreted inhibitory agent known as HGA (homogentisic acid). The investigation additionally discovered lpg1681 as a genetic element that provides resistance to HGA.[33]. A research team published on the genetic variation of *L. pneumophila* strains prevalent throughout the Comunidad Valenciana in Spain [34]A study showed significant genetic variation among *L. pneumophila* outbreak isolates as a result of environmental mutation, recombination, and gene flow. [35]. A study identified antibiotic resistance and virulence genes in *L. pneumophila* serogroup 1 and other *Legionella* species using next-generation sequencing[36]. Therefore, it is rational to state that the differences in the isolation source have led these samples to take two different nucleic acid variations with respect to each other within the same clade. The reason for these identified differences was attributed to the differences in the conditions of the clinical source from which these strains were collected. The differences in these sources have forced this sample to take a particular ribosomal variation to resist the host immunity against it. Owing to

the variable conditions of the infected sputum, M1 samples have given certain nucleic acid variations that differ from other reference strains to withstand host resistance. The currently observed data was largely sustained in literature since many studies found that the differences in the isolation source of a bacterial organism may be associated with the genetic variation this organism takes to resist the host. Several studies have identified variations in the genetic compositions of *L. pneumophila* strains obtained from diverse origins, encompassing humans, animals, and the surrounding milieu. The present study conducted an analysis on the genetic variability of strains of *L. pneumophila* that are distributed in various locations. The findings highlight the significance of comprehending the scope and dispersion patterns of strain variations of this bacterium within the environment. Such understanding is crucial for the development and implementation of effective control measures, as well as for investigating the origins of outbreaks [34]. A separate investigation has identified genetic variations among *L. pneumophila* strains through the utilization of PCR and Southern analysis, perhaps associated with their pathogenicity [37]. A study that investigated the frequency of genes associated with virulence and sequence-based types within *L. pneumophila* isolates from the water systems of a tertiary care hospital in India discovered an elevated percentage of virulence genes as well as a genetic diversity in the isolates from the environment of a hospital [38]. These findings suggest that the isolation source of bacterial organisms may be an important factor in the development of antibiotic resistance. Bacterial strains that are frequently exposed to antibiotics in a clinical setting, such as those isolated from human clinical samples, may be more likely to develop resistance to antibiotics than strains isolated from other sources. In addition, bacterial strains that are adapted to a specific host may have different genetic profiles that affect their ability to resist the host.

The distinct positioning of the M1 sample referred to the noticeable effects of the detected variants in this sample to cause a noticeable phylogenetic alteration within the currently generated tree. This data suggested that the detected variants have had a significant impact on the tree's structure, causing a noticeable phylogenetic alteration. This alteration may mean that the detected variants have led to the creation of a possible new branch or the reorganization of existing ones within the tree, indicating changes in the evolutionary relationships between the bacterial samples. This observation indicated a possible presence of a significant role of the identified variants in inducing these noticeable tilts in comparison with the referring sequences of *L. pneumophila*. Hence, a notable divergence is observed in relation to the original position of these bacterial genomes into clade-1. Moreover, the categorization of all examined bacterial specimens under clade-2 of *L. pneumophila* signifies the existence of extensive phylogenetic dispersion patterns among these microorganisms, potentially associated with their respective isolation sources.

Due to the positioning of the clade-1 near the root of the tree, it is suggested that this clade can be considered the ancestral clade from which other sequences were descended. This data suggests that the H1 sample may represent ancestral strains for other strains that are developed from it over time. Due to its position away from the roots, the M1 clinical sample represents one of these developed strains.

A) rectangular tree



B) circular tree

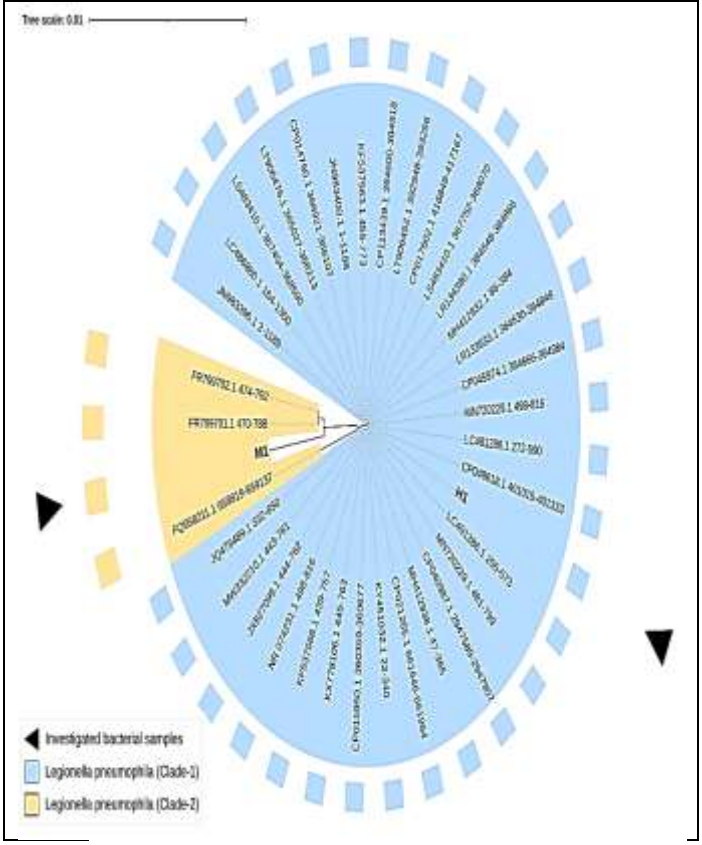


Fig. 4. The phylogenetic tree of the 16S ribosomal fragments for two samples of Legionella pneumophila is presented in a complete manner, with branch A representing a rectangular cladogram and branch B representing a circular cladogram. The triangular shape that is colored black is representative of the bacterial sequences that have been subjected to analysis. The numbers provided correspond to the GenBank accession numbers assigned to each respective species described. The numerical values located at the upper section of the tree correspond to the extent of variation in scale magnitude among the species classified within the all-encompassing tree structure.

The present study of this tree verified the reactions of the sequence as it revealed the precise location based on the neighbor-joining method in the analyzed sequences. The incorporation of ribosomal sequences in the present study has provided additional evidence for the accurate identification and characterization of these bacterial species. The 16S rRNA gene is a widely used marker gene for bacterial phylogenetic studies due to its high level of conservation across bacteria and its ability to provide insights into the evolutionary history of bacterial species. Our study demonstrated the utility of 16S rRNA gene sequencing for investigating the genetic diversity and population structure of *L. pneumophila* strains and providing insights into the evolutionary history of these important opportunistic pathogens. The current construction of a comprehensive tree based on 16S rRNA has yielded a valuable tool for effectively distinguishing among the bacterial isolates currently undergoing investigation, due to the better discriminatory capacity of these pieces. This finding provides additional evidence about the efficacy of the currently employed ribosomal sequence in distinguishing the examined *L. pneumophila* strains and accurately determining their phylogenetic relationships. The data presented in this study demonstrate a significant potential for the effective identification of bacterial sequences using genetic fragments based on the 16S rRNA gene. Overall, highlighting the importance of understanding the genetic variations present in bacterial samples and how these variations can affect the evolutionary relationships between different groups of bacteria is highly important to assess the actual biological diversity of each isolate of *L. pneumophila*. Therefore, the use of 16S rRNA sequencing offers significant advantages in the identification of *L. pneumophila*. It allows for the accurate identification of *L. pneumophila* isolates, precise classification of non-*pneumophila Legionella* species, and the detection of specific genetic markers for distinguishing between different strains of *L. pneumophila*. These advantages contribute to improved understanding and management of *Legionella*-related infections and environmental surveillance.

4. CONCLUSIONS

1. In this investigation, two different isolates of *Legionella pneumophila* were compared in terms of the range of phylogenetic diversity. Each fragment has given a particular piece of information regarding the isolation source from which each strain was isolated.
2. 16S rRNA sequencing is a powerful tool for identifying and characterizing clinical samples of *Legionella pneumophila* and has several advantages, including its distinctive sensitivity and specificity, low costs, and expeditious processing duration.
3. The observed ability of the employed 16S rRNA sequences to classify the samples that were studied

(H1 and M1) into distinct clades indicates that the 16S rRNA possesses an efficient capability to differentiate bacterial samples into multiple phylogenetic groups. Due to the close positioning of our investigated H1 sample in the vicinity of many environmental strains, the non-clinical source of this sample is confirmed.

4. Due to the close positioning of the M1 sample in the vicinity of European strains, the European source of this sample is expected.
5. Owing to the presence of two nucleic acid variations compared with the nearest strain, the distinct phylogenetic position for the M1 sample was obtained. This suggests a large role for the utilized ribosomal sequences to identify a wide spectrum of nucleic acid variations within the investigated samples.

RECOMMENDATIONS

1. This study recommends using the 16S rRNA fragment to give much more details in pathogens identification, pathogens discrimination from other closely related organisms, and a better description of pathogens hosts and their isolation sources.
2. Due to its high genetic variability, further researches are highly recommended to be conducted on the 16S rRNA fragment using large-scale investigations to unravel more details that need more samples to be unraveled.
3. advising the later researchers to use Rt-PCR for identifying the effect of variation on gene expression.

5. Acknowledgement:

This study was supported by the Department of Biology- College of Science/ / Anbar University and with the assistance of the Anbar Governorate Water Directorate

References:

[1] Okunaga M, Kushiro K, Horie R, Kondo A, Abe T. Identification of microbes coexisting with Legionella spp. in bathwaters. NPJ Clean Water. 2022 Dec 1;5(1).

[2] Salinas MB, Fenoy S, Magnet A, Vaccaro L, Gomes TD, Hurtado C, et al. Are pathogenic Legionella non-pneumophila a common bacteria in Water Distribution Networks? Water Res. 2021 May 15;196.

[3] Walker JT, McDermott PJ. Confirming the presence of Legionella pneumophila in your water system: a review of current Legionella testing methods. Journal of AOAC International. 2021 Jul 1;104(4):1135-47.

[4] Tan LT, Tee WY, Khan TM, Ming LC, Letchumanan V. Legionella pneumophila—The causative agent of Legionnaires' disease. Progress In Microbes & Molecular Biology. 2021 Apr 8;4(1).

[5] Moffa MA, Rock C, Galiatsatos P, Gamage SD, Schwab KJ, Exum NG. Legionellosis on the rise: A scoping review of sporadic, community-acquired incidence in the United States. Epidemiology & Infection. 2023 Jul 28:1-5.

[6] Mentula S, Kääriäinen S, Jaakola S, Niittynen M, Airaksinen P, Koivula I, et al. Tap water as the source of a Legionnaires' disease outbreak spread to several residential buildings and one hospital, Finland, 2020 to 2021. Euro Surveill. 2023 Mar 1;28(11).

[7] Hiergeist A, Reischl U, Gessner A, Garzetti D, Stecher B, Gálvez EJC, et al. Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. International Journal of Medical Microbiology. 2016 Aug 1;306(5):334–42.

[8] Melaine F, Saad M, Faucher S, Tabrizian M. Selective and High Dynamic Range Assay Format for Multiplex Detection of Pathogenic Pseudomonas aeruginosa, Salmonella typhimurium, and Legionella pneumophila RNAs Using Surface Plasmon Resonance Imaging. Anal Chem. 2017 Jul 18;89(14):7802–7.

[9] Petti CA, Polage CR, Schreckenberger P. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. J Clin Microbiol.2005Dec;43(12):6123–5.

[10] Gomez-Valero L, Rusniok C, Buchrieser C. Legionella pneumophila: Population genetics, phylogeny and genomics. Vol. 9, Infection, Genetics and Evolution. 2009. p. 727–39.

[11] Jay ZJ, Inskeep WP. The distribution, diversity, and importance of 16S rRNA gene introns in the order Thermoproteales. Biol Direct. 2015 Jul 9;10(1).

[12] Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019 Dec 1;10(1).

[13] Garcia-Mazcorro JF, Barcenas-Walls JR. Thinking beside the box: Should we care about the non-coding strand of the 16S rRNA gene? Vol. 363, FEMS Microbiology Letters. Oxford University Press; 2016.

[14] Hilbi H, Jarraud S, Hartland E, Buchrieser C. Update on Legionnaires' disease: Pathogenesis, epidemiology, detection and control: MicroMeeting. In: Molecular Microbiology. Blackwell Publishing Ltd; 2010. p. 1–11.

[15] Lapierre P, Nazarian E, Zhu Y, Wroblewski D, Saylors A, Passaretti T, et al. Legionnaires' disease outbreak caused by endemic strain of Legionella pneumophila, New York, New York, USA, 2015. Emerg Infect Dis. 2017 Nov 1;23(11):1784–91.

[16] Pascale MR, Mazzotta M, Salaris S, Girolamini L, Grottola A, Simone ML, et al. Evaluation of MALDI–TOF Mass Spectrometry in Diagnostic and Environmental Surveillance of Legionella Species: A Comparison With Culture and Mip-Gene Sequencing Technique. Front Microbiol. 2020 Dec 15;11.

[17] Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2017 Jan 1;45(D1):D37–42.

[18] Zhang Z, Schwartz S, Wagner L, Miller W. A Greedy Algorithm for Aligning DNA Sequences. Vol. 7, JOURNAL OF COMPUTATIONAL BIOLOGY. Mary Ann Liebert, Inc. Pp; 2000.

[19] Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019 Jul 1;47(W1).

[20] Wilson DA, Reischl U, Hall GS, Procop GW. Use of partial 16S rRNA gene sequencing for identification of Legionella pneumophila and non-pneumophila Legionella spp. J Clin Microbiol. 2007 Jan;45(1):257–8.

[21] Stølhaug A, Bergh K. Identification and differentiation of Legionella pneumophila and Legionella spp. with real-time PCR targeting the 16S rRNA gene and species identification by mip sequencing. Appl Environ Microbiol. 2006 Sep;72(9):6394–8.

[22] Foudeh AM, Daoud JT, Faucher SP, Veres T, Tabrizian M. Sub-femtomole detection of 16s rRNA from Legionella pneumophila using surface plasmon resonance imaging. Biosens Bioelectron. 2014 Feb 15;52:129–35.

[23] Baldan R, Cliff PR, Burns S, Medina A, Smith GC, Batra R, et al. Development and evaluation of a nanopore 16S rRNA gene sequencing service for same day targeted treatment of bacterial respiratory infection in the intensive care unit. Journal of Infection. 2021 Aug 1;83(2):167–74.

[24] Janczarek M, Palusińska-Szysz M. PCR method for the rapid detection and discrimination of Legionella spp. based on the amplification of pcs, pmtA, and 16S rRNA genes. J Appl Genet. 2016 May 1;57(2):251–61.

[25] Borthong J, Omori R, Sugimoto C, Suthienkul O, Nakao R, Ito K. Comparison of database search methods for the detection of Legionella pneumophila in water samples using metagenomic analysis. Front Microbiol. 2018 Jun 19;9(JUN).

[26] Reuter S, Harrison TG, Köser CU, Ellington MJ, Smith GP, Parkhill J, Peacock SJ, Bentley SD, Török ME. A pilot study of rapid whole-genome sequencing for the investigation of a Legionella outbreak. BMJ open. 2013 Jan 1;3(1):e002175.

[27] Church DL, Cerutti L, Gürtler A, Griener T, Zelazny A, Emler S. Performance and application of 16S rRNA gene cycle sequencing for routine identification of bacteria in the clinical microbiology laboratory. Vol. 33, Clinical Microbiology Reviews. American Society for Microbiology; 2020. p. 1–74.

[28] Wullings BA, Van Der Kooij D. Occurrence and genetic diversity of uncultured Legionella spp. in drinking water treated at temperatures below 15 C. Applied and Environmental Microbiology. 2006 Jan;72(1):157-66.

[29] Coscollá M, Gosalbes MJ, Catalán V, González-Candelas F. Genetic variability in environmental isolates of Legionella pneumophila from Comunidad Valenciana (Spain). Environ Microbiol. 2006 Jun;8(6):1056–63.

[30] Gomez-Valero L, Rusniok C, Buchrieser C. Legionella pneumophila: Population genetics, phylogeny and genomics. Vol. 9, Infection, Genetics and Evolution. 2009. p. 727–39.

[31] David S, Sánchez-Busó L, Harris SR, Marttinen P, Rusniok C, Buchrieser C, et al. Dynamics and impact of homologous recombination on the evolution of Legionella pneumophila. PLoS Genet. 2017;13(6).

[32] Wee BA, Alves J, Lindsay DSJ, Klatt AB, Sargison FA, Cameron RL, et al. Population analysis of Legionella pneumophila reveals a basis for resistance to complement-mediated killing. Nat Commun. 2021 Dec 1;12(1).

[33] Levin TC, Goldspiel BP, Malik HS. Density-dependent resistance protects Legionella pneumophila from its own antimicrobial metabolite, HGA. Elife. 2019 May 28;8:e46086.

[34] Sánchez-Busó L, Coscollá M, Pinto-Carbó M, Catalán V, González-Candelas F. Genetic Characterization of Legionella pneumophila Isolated from a Common Watershed in Comunidad Valenciana, Spain. PLoS One. 2013 Apr 25;8(4).

[35] McAdam PR, Vander Broek CW, Lindsay DSJ, Ward MJ, Hanson MF, Gillies M, et al. Gene flow in environmental Legionella pneumophila leads to genetic and pathogenic heterogeneity within a Legionnaires' disease outbreak. Genome Biol. 2014;15(11):504.

[36] Bakour S, Sankar SA, Rathored J, Biagini P, Raoult D, Fournier PE. Identification of virulence factors and antibiotic resistance markers using bacterial genomics. Vol. 11, Future Microbiology. Future Medicine Ltd.; 2016. p. 455–66.

[37] Samrakandi MM, Cirillo SLG, Ridenour DA, Bermudez LE, Cirillo JD. Genetic and phenotypic differences between Legionella pneumophila strains. J Clin Microbiol. 2002;40(4):1352–62.

[38] Sreenath K, Chaudhry R, Vinayaraj E V., Dey AB, Kabra SK, Thakur B, et al. Distribution of Virulence Genes and Sequence-Based Types Among Legionella pneumophila Isolated From the Water Systems of a Tertiary Care Hospital in India. Front Public Health. 2020 Nov 23;8.

P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2024,(18), (01):13 – 25

دراسة مقارنة على تسلسلات جين 16s rRNA للعزلات السريرية والبيئية لبكتيريا الفيلقية المستروحة

محمد ابراهيم خالد[1]* ، إبراهيم عبد الكريم عبد الرحمن العاني[2]

[1]جامعة الانبار، كلية العلوم، قسم علوم الحياة

[2]جامعة الفلوجة / رئاسة الجامعة

mohammedalani23@gmail.com , dr_ibrahim65@uofallujah.edu.iq

الخلاصة:

الهدف من هذه الدراسة هو تحديد ووصف للتماثل والتطابق والموقع التطوري من الشجرة الوراثية لعزلتين من البكتيريا الفيلقية المستروحة، لقد تم عزل العزلة السريرية (يرمز لها M1) من بصاق مريض يعاني من أعراض الالتهاب الرئوي، وتم عزل عزلة اخرى بيئية (يرمز لها H1) قد عزلت من بطانة السطح الداخلي لاحد أنابيب مياه الشرب ذات النهاية المغلقة، تم الحصول على العزلتين من مدينة الفلوجة، العراق. بواسطة نفس الباحثين الذين أجروا هذه الدراسة الحالية. تم تضمين موضع جيني واحد لتسلسل جين 16s rRNA في هذه الدراسة لتقييم التباين في الأنماط الجينية لكل من هذه التسلسلات الريبوسومية. وتم إجراء تجارب باستخدام التسلسل المباشر على القطع المضخمة (Amplicons) لكي يتم تحديد نمط التباين الوراثي الموجود في العينات البكتيرية التي تم الحصول عليها. وبعد ذلك، تم إنشاء شجرة مفصلة ودقيقة من أجل التحقق من صحة المواضع التطورية للاختلافات المرصودة والتمييز الذي تم بينها. من خلال محاذاة (alignment) تسلسل الأحماض النووية لعينة M1 مع التسلسلات المرجعية المودعة لبنك الجينات لبكتيريا الفيلقية الرئوية ، تم تحديد السلالة ذات رمز الانضمام (GenBank acc. no. KF537568.1)، والتي أظهرت انها على أعلى درجة من التماثل مع عينة M1 .وقد تم الكشف عن وجود نوعين مختلفين من تسلسل النيوكليوتيدات للحمض النووي، والتي كانت في المواقع التالية (A>G141 وT>C247) لم يكونا موجودين في التسلسل المرجعي للسلالة التي أظهرت اعلى نسبة تطابق بينهما. وقد تم الاستنتاج من تحليل الشجرة الوراثية أن عينة M1 التي تم فحصها احتلت موقعًا نشوئيًا متميزًا وكانت ذات قرب مناسب من العديد من السلالات التي تم توديعها في بنك الجينات من مصادر أوروبية. كما تم استنتاج أن عينة M1 مشتقة من السلالات المدمجة الأخرى في بنك الجينات – بما في ذلك عينة H1 – داخل نفس الشجرة. وفقًا لذلك، تم اكتشاف أن متغيرات الحمض النووي التي تمت ملاحظتها في عينة M1 لها تأثير ملحوظ على تغيير الوضع التطوري لعينة M1 التي تم فحصها داخل مجاميع (clades) مشتركة لبكتيريا الفيلقية المستروحة. وكذلك أشار تحديد موضع عينة H1 بالقرب من العديد من السلالات غير السريرية إلى قدرة عالية لتسلسل لحمض النووي 16s rRNA المستخدم على تحديد مصدر العزل الفعلي الذي تم جمعها منه. وقد توصلت الدراسة الحالية إلى أن قدرة تسلسلات جين 16s rRNA كانت كفؤة في تحديد الهوية لبكتيريا *Legionella pneumophila* في العينات التي تم فحصها. يمكن توسيع استخدام هذا الموقع الجيني للكشف عن التوزيعات التطورية للأنواع الأخرى من المصادر البكتيرية. ويمكن أيضًا فحص تسلسلات النواتج المضخمة لهذا الجين (amplicon) يحتمل أن تكون ذات قيمة لاكتشاف تفاصيل إضافية من خلال تحديد العزلات في أمثلة بكتيرية أخرى.

الكلمات المفتاحية: *بكتيريا الفيلقية المستروحة Legionella pneumophila، تسلسلات جين 16s rRNA، الشجرة الوراثية*