Phylogenetic tree analysis based on the 16S sequence alignment for Klebsiella spp. isolated from different sources

Israa A.J. Ibrahim1, Tuqa A. Kareem2, Yaseen M. Azeez3, Hawraa K. Falhi3
1Department of Microbiology, College of Science, Al-Karkh University of Science, Baghdad, Iraq
2Al-Mustafa University College, Baghdad, Iraq
3Al-Kindi Teaching Hospital, Baghdad, Iraq

Received: 18/6/2019 Accepted: 27/7/2019

Abstract
16S ribosomal RNA (16S rRNA) gene sequences used to study bacterial phylogeny and taxonomy have been by far the most common housekeeping genetic marker utilized for identification and ancestor determination. This study aimed to investigate, for the first time, the relationship between Klebsiella spp. isolated from clinical and environmental samples in Iraq.

Fifty Klebsiella spp. isolates were isolated from clinical and environmental sources. Twenty-five isolates were collected from a fresh vegetable (Apium graveolens) and 25 from clinical samples (sputum, wound swab, urine). Enteric bacteria were isolated on selective and differential media and identified by an automatic identification system, vitek-2. The total DNA was extracted and PCR amplified for selected isolates. The 16S rRNA gene was amplified by using the universal primer 27F (5' - AGAGTTTGATCCTGGCTCAG- 3') and 1492R (5' - GGTTACCTTGTTACGACTT- 3'). The 16SrRNA gene sequence was analysed among some local isolates, and the results were compared with the standard data of similar registered strains in NCBI. The most common species of Klebsiella was Klebsiella pneumoniae pneumoniae (Kpp), followed by Klebsiella pneumoniae ozaenae (Kpo) and Klebsiella oxytoca (Ko). The results of the identification of species and sub species by using the biochemical test (vitek-2) were more precise than those obtained by the use of the universal primer. Phylogenetic tree strategies have clearly indicated a relatively close similarity amongst all analysed Klebsiella isolates and revealed the intra-species genetic distance between the individual isolates of the Klebsiella spp. In conclusion, our results revealed the main advantage of using universal primers for the identification of Klebsiella spp. and their root from nature.

Keywords: Klebsiella spp., Apium graveolens, clinical sample, 16S sequence.

*Email: israa_ibrahim66@yahoo.com

*Email: israa_ibrahim66@yahoo.com

 DOI: 10.24996/ijs.2019.60.12

ISSN: 0067-2904
**Introduction**

16S rRNA gene sequences used to study bacterial phylogeny and taxonomy have been, for several reasons, by far the most commonly used housekeeping genetic marker. These reasons include the following; the 16S rRNA sequence is present in almost all bacteria, often existing as a multigene family, or operons; the function of the 16S rRNA gene has not changed over time, suggesting that random sequence changes are a more accurate measure of time (evolution); and the 16S rRNA gene (1,500 bp) is large enough for informatics purposes [1]. Phylogenetic markers include the presence of specific protein coding or structural genes, the combinations of such genes and their variants, and insertion and repeat elements. Among these molecular markers is the 16S rRNA, an ~1500 bp gene coding for a catalytic RNA that is part of the 30S ribosomal subunit, which has desirable properties that allowed it to become the most commonly used molecular marker. Of these desirable properties, the functional constancy of this gene assures its validity as a molecular chronometer, which is essential for a precise assessment of the phylogenetic relatedness of organisms [2]. 16S rRNA sequencing is routinely used in the clinical laboratory but rarely used for microbial limits test for nonparenteral products.

"Bacterial 16S rRNA genes contain nine “hypervariable regions” (V1-V9) that demonstrates considerable sequence diversity among different bacteria. Species-specific sequences within a given hypervariable region constitute useful targets for diagnostic assays and other scientific investigations" [3]. Molecular testing allows for highly specific and sensitive identification of a large number of pathogens from clinical isolates and clinical specimens. The ability of molecular techniques to identify pathogens directly from clinical samples makes the rapid identification without recourse to culture possible[4]. The species of *Klebsiella* are all gram-negative and non-motile. They tend to be shorter and thicker when compared to others in the Enterobacteriaceae family. More than eight species and subspecies of the genus *Klebsiella* spp. were recorded in Bergey’s Manual such as: *K. p. subsp. pneumonia, K. p. subsp. ozaenae*, *K. p. subsp. rhinoscleromatis, K. mobilis, K. oxytoca, K. planticola, K. terrigena, K. ornithinolytica*, and *K. granulomatis*[5]. *Klebsiella* can cause different types of healthcare-associated infections, including pneumonia, bloodstream infections, wound or surgical site infections, and meningitis [6]. By using genetic information can discover epidemiologic links between clinical and non-clinical isolates [7].
Klebsiella pneumoniae, a member of the family Enterobacteriaceae, is a rod-shaped, Gram-negative, lactose-fermenting bacillus with a prominent capsule. Typical K. pneumoniae is an opportunistic pathogen that is widely found in the mouth, skin and intestines, as well as in hospital settings and medical devices.

The aim of this study was to identify Klebsiella spp. isolated from different sources by using 16S gene sequencing and to detect the relationship between clinical and environmental samples.

**Materials and Methods**

**Specimens**

In this study, 50 Klebsiella spp isolates were analysed, 25 of which were from leafy vegetables (veg.) collected from local market (Figure 1), while the other 25 bacterial isolates were collected from clinical samples from the Al-Kenday Hospital in Baghdad. Sample collection was carried out from May 2018 to August 2018 (Table-1).

![Figure 1- Shape of Apium graveolens](image.png)

**Isolation and Identification**

All samples in the first isolates were cultured on MacConkey agar and Eosin methylene blue agar plates. The biochemical tests were performed by using vitek-2 system (bioMérieux) in Al-Kenday Hospital laboratory for identification of the isolates.

**Extraction of total DNA**

Ten samples were selected (Table-1).

**Table 1- Selected samples.**

<table>
<thead>
<tr>
<th>No.</th>
<th>Source</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Env.</td>
<td>Apium graveolens</td>
</tr>
<tr>
<td>2</td>
<td>Env.</td>
<td>Apium graveolens</td>
</tr>
<tr>
<td>3</td>
<td>Env.</td>
<td>Apium graveolens</td>
</tr>
<tr>
<td>4</td>
<td>Env.</td>
<td>Apium graveolens</td>
</tr>
<tr>
<td>5</td>
<td>Env.</td>
<td>Apium graveolens</td>
</tr>
<tr>
<td>6</td>
<td>Clin.</td>
<td>Burn swab</td>
</tr>
<tr>
<td>7</td>
<td>Clin.</td>
<td>Urine</td>
</tr>
<tr>
<td>8</td>
<td>Clin.</td>
<td>Urine</td>
</tr>
<tr>
<td>9</td>
<td>Clin.</td>
<td>Burn swab</td>
</tr>
<tr>
<td>10</td>
<td>Clin.</td>
<td>Burn swab</td>
</tr>
</tbody>
</table>
Environment (Env.), Clinical (Clin.)
Total DNA of bacterial isolates was extracted by using the G- spin DNA Extraction Kit (iNiRON biotechnology/Korea, cat.no. 17045).

PCR Amplification and gene sequences
The 16S rRNA gene was amplified using the universal primers 27F (5'-AGAGTTTGATCCTGGCTCAG- 3') and 1492R (5' - GGTACCTTGTAGACTT- 3') (Table-2) [8], and then25μl from the PCR reaction, along with 50μl of primers for this gene, were sent to Macrogene company, (USA) to determine the DNA sequencing.

Table 2- The specific primers for the 16s RNA gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Tm (°C)</th>
<th>GC (%)</th>
<th>Product size</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>5'- AGAGTTTGATCCTGGCTCAG-3*</td>
<td>54.3</td>
<td>50.0</td>
<td>1250 base pair</td>
<td>Millerset al 2013</td>
</tr>
<tr>
<td>Reverse</td>
<td>5'- GGTACCTTGTAGACTT-3*</td>
<td>49.4</td>
<td>42.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Phylogenetic tree design
A specific comprehensive bacterial tree was constructed, in which the observed bacterial variants were compared with their neighboring homologous sequences using the NCBI-BLASTn server [9]. Then, the BLAST results of the observed variants were combined and aligned using the Bayesian neighbor -joining tools [10]. A full inclusive tree, including the 10 observed variants and the 164 neighbor sequences, was made and visualized as a polar cladogram using the Figure Tree tool (http://tree.bio.ed.ac.uk/software/figtree/). The bacterial sequences of each classified phylogenetic species-group in the comprehensive tree were colored appropriately.

Results and Discussion
Klebsiella spp. are commonly found in the environment (soil, water, plant surfaces...etc) and human flora. Klebsiella pneumoniae is a leading cause of hospital-acquired infections [11]. The difficulties in Klebsiella taxonomy are demonstrative and common to other bacterial genera [12]. Comparison of a K. pneumoniae strain collected from the ground with a control strain after spaceflight showed genome variations and phenotypic changes and explained the genomic basis of drug resistance [13].

During our study period, a total of 50 Klebsiella spp. isolates from various clinical samples and leafy vegetable (Apium graveolens) were processed and three Klebsiella spp. species were identified, namely Klebsiella pneumoniae pneumoniae (Kpp), Klebsiella pneumoniae ozaenae (Kpo) and Klebsiella oxytoca (Ko)(Table-3). A previous study by Biradar and Roopa [14] recorded from a total of 100 clinical samples, Klebsiella pneumoniae was the most common species, followed by Klebsiella oxytoca.

Table 3- The number of isolated Klebsiella species from different clinical specimens and the leafy vegetable Apium graveolens

<table>
<thead>
<tr>
<th>No.</th>
<th>Source</th>
<th>Vitek-2</th>
<th>Sample</th>
<th>PCR result</th>
<th>GenBank code</th>
<th>Sample code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Env.</td>
<td>Kpp</td>
<td>Apium graveolens</td>
<td>+</td>
<td>MK371792.1</td>
<td>ITYH1</td>
</tr>
<tr>
<td>2</td>
<td>Env.</td>
<td>Kpp</td>
<td>Apium graveolens</td>
<td>+</td>
<td>MK373026.1</td>
<td>ITYH2</td>
</tr>
<tr>
<td>3</td>
<td>Env.</td>
<td>Kpp</td>
<td>Apium graveolens</td>
<td>+</td>
<td>MK373029.1</td>
<td>ITYH3</td>
</tr>
<tr>
<td>4</td>
<td>Env.</td>
<td>Kpp</td>
<td>Apium graveolens</td>
<td>+</td>
<td>MK373024.1</td>
<td>ITYH4</td>
</tr>
<tr>
<td>5</td>
<td>Env.</td>
<td>Kpp</td>
<td>Apium graveolens</td>
<td>+</td>
<td>MK373023.1</td>
<td>ITYH5</td>
</tr>
<tr>
<td>6</td>
<td>Clin.</td>
<td>Kpp</td>
<td>Burn swab</td>
<td>+</td>
<td>MK373028.1</td>
<td>ITYH6</td>
</tr>
</tbody>
</table>
7 Clin. Kpp Urine + MK373027.1 ITYH7
8 Clin. Kpp Urine + MK373030.1 ITYH8
9 Clin. Ko Burn swab + MK373025.1 ITYH9
10 Clin. Kpo Burn swab + MK373031.1 ITYH10

Figure-2 shows the positive results for the identification of *Klebsiella* spp. by using the 16S *rRNA* gene universal primers 27F and 1492R.

![PCR product](image)

**Figure 2**- PCR product with the band size of 1250 bp. The product was electrophoresed on a 2% agarose gel at 5 volt/cm2 with 1X TBE buffer for 1:30 hours. N: DNA ladder (100).

Several organisms were incorporated within the present phylogenetic tree, including *Klebsiella sp*, *Klebsiella pneumoniae*, *Klebsiella variicola*, *Enterobacter cloacae*, *Enterobacter aerogenes*, *Salmonella enterica*, *Escherichia coli*, and *Enterobacteriaceae* (Figure-3). The currently constructed comprehensive tree indicated the comparison of *Klebsiella* spp. with the observed 174 scanned variants. All studied variants occupied several positions within the tree. However, almost all the targeted sequences were exactly positioned in an extremely close position to *Klebsiella pneumoniae*, indicating a tendency of the targeted sequences to belong to this species. The presented sequences were found to belong to the *Klebsiella* genus. Notably, one of the targeted sequences, MK373024.1, exhibited a close correlation with *E. coli* sequences. This observed close association could not be eliminated, as *Escherichia* and *Klebsiella* organisms belong to the same family, Enterobacteriaceae. Furthermore, these organisms cohabitate with each other in several locations [15]. This notion is applicable for the other species that have close sequence similarity with the targeted *Klebsiella* sequences, such as *Salmonella* and *Enterobacter*, which were found to share high DNA sequence homology with *Klebsiella* species. Nevertheless, the current phylogenetic tree provided conclusive evidence regarding the confirmed identity of the observed variants. Despite the unique characterization of the S-variants that originated from the presence of several substitution mutations, no deviation from *Klebsiella* was revealed (Figure-3). In the presently constructed comprehensive phylogenetic tree, almost all the observed sequence-related species belong to several *Klebsiella* related species. However, these results provide a further indication of the identity of the sequences observed in the studied *Klebsiella* isolates. Thus, this cladogram-based comprehensive tree provided a tool by which the high ability of such selected PCR genetic fragments efficiently identify *Klebsiella* isolates. Moreover, our phylogenetic analysis had a high detection specificity with regard to *rRNA*-based phylogenetic protocols. Accordingly, the currently observed PCR-sequencing protocols, as well as phylogenetic tools, have indicated the relatively close similarity amongst all analysed *Klebsiella* isolates.

In conclusion, by relying on DNA sequencing followed by phylogenetic tree strategies, all ten studied strains were found to belong to *Klebsiella* isolates. On the other hand, the intra-species genetic distance between the local isolates of the *Klebsiella* spp. was higher up to 4.0 when compared to other isolates of the Enterobacteriaceae genus. Universal primers may be mainly shared among most Enterobacteriaceae family members.
Figure 3- The comprehensive phylogenetic tree of the observed *Klebsiella* local isolates. The black arrows refer to the targeted variants, while the other colors refer to the high relative DNA sequences. All the mentioned numbers refer to the GenBank accession no. of each species. The number “4.0” at the bottom of the tree refers to the degree of the scale range among the comprehensive tree categorized organisms.

Biochemical test (vitek-2) results determined the genus, species and subspecies levels for all isolates in this study. On the other hand, the use of the universal primer could only show the genus and species (Table-4).

We consider the development of biochemical tests, such as API 20E and the vitek-2 (60 biochemical and enzymatic reactions test), as well as the characteristic culture and structure shape, as more precise for the identification of species, subspecies and variance than using universal primers or certain primers. A previous study by Podschn and Ullmann[16] recorded that biotyping (a
biochemical test, such as the API 20E system) of Klebsiella spp. is not very suitable as an epidemiological tool and molecular typing method, as its application to the genus Klebsiella, is still in its infancy. However, another study revealed that a suggested selected biochemical test (a test panel consisting of 18 biochemical tests) was able to find one or more positive test results differentiating any species from its closest relative [17]. Siri et al. [18] revealed that a biochemical test (API 20E), and gapA-specific PCR tests could identify Klebsiella spp. from a water sample, and that a biochemical test could identify Klebsiella spp. from a plant (pomegranate) sample [19].

Table 4 - Species classification of isolates from the genus Klebsiella by different taxonomic systems.

<table>
<thead>
<tr>
<th>No.</th>
<th>GenBank code</th>
<th>vitek-2</th>
<th>16S RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MK371792.1</td>
<td>Kpp</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td>2</td>
<td>MK373026.1</td>
<td>Kpp</td>
<td>Klebsiella sp, Klebsiella pneumoniae</td>
</tr>
<tr>
<td>3</td>
<td>MK373029.1</td>
<td>Kpp</td>
<td>Klebsiella sp, Klebsiella pneumoniae</td>
</tr>
<tr>
<td>4</td>
<td>MK373024.1</td>
<td>Kpp</td>
<td>Klebsiella variicola, Klebsiella pneumoniae</td>
</tr>
<tr>
<td>5</td>
<td>MK373023.1</td>
<td>Kpp</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td>6</td>
<td>MK373028.1</td>
<td>Kpp</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td>7</td>
<td>MK373027.1</td>
<td>Kpp</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td>8</td>
<td>MK373030.1</td>
<td>Kpp</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td>9</td>
<td>MK373025.1</td>
<td>Ko</td>
<td>Klebsiella pneumoniae, Klebsiella sp</td>
</tr>
<tr>
<td>10</td>
<td>MK373031.1</td>
<td>Kpo</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td></td>
<td>Other shared species</td>
<td></td>
<td>Enterobacter cloacae, Enterobacter aerogenes, Salmonella enterica, Escherichia coli</td>
</tr>
</tbody>
</table>

The second tree in Figure-4 shows that the degree of the scale range among the comprehensive tree-categorized organisms was 0.9. The root node ancestor was divided into three nodes A, B and C with the isolates MK371792.1 and MK373023.1 from environmental sample showing highly identical gene sequences. The A node was divided into 7 nodes (14 branches) for 8 isolates. Furthermore, the branch lengths between each node had the value of 1. This result suggests that 8 of the 10 clinical and environmental isolates had the same ancestor. On the other hand, the intra-species genetic distance between the individual isolates of the Klebsiella spp. was higher (up to 0.9).
Figure 4-A neighbour joining phylogenetic tree of the observed *Klebsiella* local isolates. All the mentioned numbers refer to the GenBank accession no. of each referred species. The number “0.9” at the bottom of the tree refers to the degree of the scale range among the comprehensive tree categorized organisms. The tree is designed using the regular tree layout option of the Figure Tree software.
Based on our data, using the universal primers 27F and 1492R for the identification, detection, comparison, relationships and source of isolates can be considered. Many researchers have stated that other methods, in addition to primers, are needed to precisely identify species. On the other hand, at least three primers (universal primer, species-specific primer, virulence gene primer) were used for detecting clinical and environmental isolates. Examples include the identification of *K. pneumoniae* from chicks by using a specific gene (GyrA) arvivalence genes such as the mucoviscosity gene MagA and the iron uptake system gene Kfu [20]. Another example is the use of the rDNA16SrRNA gene with restriction enzymes (e.g., BanII for identification of the subspecies *K. pneumoniae*). Identification of the three *K. pneumoniae* phylogenetic groups can be obtained by combining gyrA PCR-RFLP, parC PCR and adonitol fermentation [22]. The universal 16S rRNA bacterial primers 27Fand 1392R can be used to identify isolates at the genus-level, with a concordance rate of 96% and a species-level concordance rate of 87.5% [23]. A multiplex PCR consisting of capsular types K1, K2, K5, K54, K20 and K57, which are related to pathogenicity, and rmpA and wcaG as virulence factors, in addition to the 16S–23S RNA, could be used [24]. PCR detection of the rmpA gene (polysaccharide capsule) showed only 87.5% similarity among *Klebsiella* isolated from wound and burn infections [25]. The molecular detection of some capsular polysaccharide genes (cps) for *K. pneumoniae* revealed a positive concordance rate between 33.3% and 66.6% [26]. Unknown bacteria isolated from clinical and environmental samples can be used for identification by the universal primers 27F and 1492R [27-31].

Our results confirms those obtained by Senthilraj et al. [32]. It can be concluded that 16S rRNA sequence-based identification reduces the time by circumventing biochemical tests and increases specificity and accuracy. In addition, clinical and environmental samples can be identified and compared, and the source of microbes can be determined.

**Conclusion**

All ten studied strains were found to belong to the genus of *Klebsiella* by using the 16S rRNA gene universal primers 27F and 1492R for sequencing. These primers were also used to demonstrate the relationship between clinical and environmental isolates. The phylogenetic tree described that the root node’s ancestor was divided into three nodes (A, B and C), using Bootstrap 4.0.

**Acknowledgments**

The authors are thankful to the administrator of Al-Kenday Hospital in Baghdad for providing the opportunity to collect the samples at their facilities.

**References**


